

【法学研究】

自动驾驶汽车设计者刑事过失归责的构造

刘 颖

(四川师范大学 法学院, 四川 成都 610066)

摘 要:随着自动驾驶汽车技术的日趋成熟与自动驾驶汽车的普及,以自动驾驶汽车为代表的新型风险所引发的交通事故不可避免。新型交通事故存在因果关系复杂、责任判断困难、牵涉人员多元等特点,加之算法的“黑箱”属性以及当前《刑法》仅就缺陷产品故意责任进行规定的现实情况,势必会造成设计者过失肇事行为刑事规制的难题。在现有的刑法体系内,修正的旧过失论与新过失论均不能合理解释设计者的过失致害行为,应当从一般过失切入,以修正的危惧感说对自动驾驶汽车因设计者过失肇事规则路径予以论证。

关键词:自动驾驶汽车设计者;过失犯;缺陷产品;修正的危惧感说;刑事责任

中图分类号:DF 6 **文献标识码:**A **DOI:**10.13486/j.issn.2097-4973.2026.03.014

现代技术带来了传统社会不存在的便利,也创造了许多新的、人造的、不确定的危险源。^{[1]16}人工智能技术出现和发展所蕴含的风险其实是新型风险的重要表现形式。^[2]如果将新型风险一味规划进社会成员分担的范畴,那么此时现代技术带来的风险是远大于传统风险的,也就违背了人工智能创造与发展的初衷。就算是在限定主体范围内分配风险,但如何做到公平分配、有效分配也是一个值得思考的问题。

随着智能化时代的来临与人口老龄化的进一步加剧,企业更加热衷于无人化、智能化,意欲以人工智能代替人类工作以提高生产效率,这一现象在汽车驾驶领域愈发突出。“自动驾驶汽车是指外置技术领先的硬件设备,内置自动化、智能化的驾驶系统的新兴汽车。”^①当前自动驾驶汽车属于弱人工智能^[3],根据指示与命令进行的活动始终处于人类所预期技术范围之内,其当然不具有刑事责任能力,其工具属性表露无遗。^[4]随着神经元控制算法的参与愈发深入和算法内

容不断精简,自动驾驶汽车的智能性、自主性、自适应性等特点让交通事故中设计者、监管者等参与者的责任判断难度更上一层楼。而众多可追责主体中,设计者因其对算法的设计与决策,其通过汽车装载的自动驾驶系统在驾驶领域也发挥着重要的控制作用。^[5]算法设计者通过自动驾驶系统对汽车自动行驶、自动监测环境等多种功能存在着决定性控制作用体现其对行驶安全性的控制地位,设计者当然成为责任分配主体。^[6]因此,自动驾驶汽车使用领域的安全性也应考虑设计者这一因素。

在自动驾驶领域,因果关系复杂、涉案人员广泛和危害结果严重等特征不可避免地引发刑事责任归责困境。人工智能技术愈发成熟与自动驾驶汽车的推广,势必会带来两个难题:一是遭到外部侵入而沦为犯罪工具,二是因内部存在缺陷而致人损害。^[7]前者需要承担责任的主体不言自明,而后者在刑法上或许面临着无罪可罚的忧虑。2016 年 1 月 20 日,在国内第一起特斯拉

① 参见《智能网联汽车道路测试与示范应用管理规范(试行)》第 37 条。

收稿日期:2025-10-29

作者简介:刘 颖(2000—),女,四川泸县人,在读硕士研究生,主要从事刑法学研究。E-mail:2778533133@qq.com

自动驾驶致死案中,自动驾驶汽车因为系统尚不完善导致未能成功识别前方车辆致使事故发生。^[8]2021年8月12日,林某在汽车的自动驾驶模式下于福建莆田高速公路上发生车祸而去世。^[9]前述案例均属于自动驾驶汽车之缺陷致人损害,在此前提下,如何厘清众多行为与结果错综复杂的因果关系并进行责任分配是一大难题。本文就自动驾驶汽车的设计者因过失行为导致交通肇事之问题进行探析,以期有效解决自动驾驶汽车设计缺陷中设计者的刑事过失责任分配难题。

一、过失犯理论归责路径检视

(一) 过失犯理论

1. 旧过失论 旧过失论源于古典犯罪论,主张过失之核心在于结果预见可能性,认为故意犯和过失犯罪只在责任阶层存在差异。即,结果一旦发生,行为人的行为就具备危险性。^{[10]228}从而在事后考虑行为人当时是否对结果的发生有预见可能性。此时的预见可能性为“主观的预见可能性”^{[11]190},即行为人因违反了应当保持精神紧张状态的预见义务而对结果之发生缺乏预见。此时,成立过失犯的行为标准并不明确。

针对这点,修正的旧过失论对过失的实行行为进行修正,主张从构成要件该当性与有责性两个阶层对过失犯进行判断^{[12]211},若“具有实质的不被允许的危险行为”^{[13]39}引发了危害结果,并且行为人对于该结果的出现有预见可能性,即应以过失犯论处。该预见可能性虽然为具体预见可能性,但并没有严苛至需要预见到现实具体结果和因果流程详细内容之程度,达到较为具体程度即可,即预见到特定的构成要件结果的发生和引起结果发生的因果流程的基本部分。^{[14]224}

2. 新过失论 不同于旧过失论,新过失论认为过失犯与故意犯二者在不法阶层、违法性阶层与有责性阶层均存在区别。新过失论从行为无价值出发,将重点放在了行为人客观上是否实施了防止危害结果发生的行为,而并非不考虑主观方面。^{[15]129}但是在新过失论中,将过失作为注意义务的违反来把握,必须先有结果预见可能性的前提性存在,进而才可得出结果回避义务这一结论。^{[16]292}因此,新过失论与旧过失论之于具体的

结果预见可能性这一点上是共同的。

3. 新新过失论(危惧感说) 过失论的预见可能性可以分为具体的结果预见可能性与危惧感两种类型,与要求具备具体的结果预见可能性的旧过失论和新过失论相反,新新过失论认为过失之成立不需要结果的预见可能性。该理论主张按照一般人的标准,行为人如果产生了可能会发生什么的恐惧感,却没有采取措施以避免结果发生,即为过失。^{[17]241-242}然而新新过失论标准的模糊性将会带来过失犯处罚范围扩大的效果。

(二) 过失犯理论述评

1. 旧过失论的弊端 由于人工智能技术的迅速发展及进步,带来“波兰尼悖论”^[18]束缚被解除的效果,算法不再受制于人类的表达能力,能够进行无监督学习和强化学习。^[19]相较于监督学习来说,算法设计者对于无监督学习和强化学习两种学习方法的可控度更低,对机器如何分析、比较和组合数据是不得而知的。并且人工智能不断学习与变化的趋势使得我们无法从表面的、现在的输入结果来推断运算结果,其中的因果关系是难以理解的、无法解释的。^[20]算法的这种“黑箱”属性使得根据数据得出的统计意义上的因果关系与现实规范意义上的因果关系截然不同。^[20]这就造成了算法设计者对于自己的行为所引发危害结果之间的因果关系预见不能之后果。因此,事前的预见可能性在人工智能产品引发的社会危害事件中是缺席的。旧过失论对于特定构成要件的结果和引起结果发生的因果流程的基本部分要有预见这一条件,具体到人工智能中,就算不要求预见具体的时间、地点、行为人等要素,但是对于决策在特定情况下是否会对他人造成伤害,会带来何种伤害等有预见可能性要求。但是正如前述的算法的不透明性,使得根据静态的数据预测算法的决策及侵害结果是不现实的,因此,难以实现对因果关系及侵害结果的预见。

旧过失论对过失的判断标准为预见可能性和具备法益侵害性的过失行为。从结果无价值的角度出发,对案件进行判断时往往会出现只看结果而忽略行为的问题,从而落入结果责任的错误逻辑,这无疑限制了容许风险作为正当化事由的效力范围。人工智能可以便利驾驶,也可以引

发事故;可以提高生产效率,也能误伤工人……风险的增加带来刑罚范围的扩张,如果便利生活是以人们承担刑事责任为代价,那各行各业的发展都会停滞不前。过失犯的实行行为,仅限于实质危害性达到一定程度的行为,且造成了不法的结果。^{[21]166}此时,由于过失犯相较于故意犯的特殊性,过失行为的法益侵害性更为模糊,那么在实践中以这一标准判断行为的实质危害性是否达到一定程度是不切实际的。为此,平野龙一教授提出在具有结果发生的客观预见可能性时可以认定为实质的不被允许的危险。^{[22]194}但此种带有唯心主义色彩将二者混为一谈的判断方式更为不妥。除此以外,修正的旧过失论亦以结果无价值为立场,却又通过实行行为的法益侵害性以区分故意犯和过失犯,这一违法性判断方式难言单纯客观。^[23]

2. 新过失论的弊端 新过失论以结果回避义务为核心,尽管结果回避义务此时是第一顺位,但是在第二顺位的结果预见可能性要求以具体预见可能性的形式以限制处罚范围。不同于旧过失论要求高度的具体预见可能性,新过失论并没有就预见可能性的高低提出详细的标准。从学说史发展之角度来看,新过失论相较于旧过失论更进一步,在其基础上添加结果回避义务发展而来,从而对于预见可能性的要求应该与其相当,但新过失论的主张者却往往语焉不详,未就结果预见可能性的程度详细展开。^[24]从而在自动驾驶汽车事故这一特殊命题中,新过失论与旧过失论面临着同一问题:具体的结果预见可能性和算法黑箱之间的冲突难以化解。由于结果避免义务以对不法事实发生预见可能性为前提,那么在行为人对不法事实没有预见可能性时,结果避免义务也就难以得到实施,过失也就当然不存在。^{[25]391}

3. 新新过失论(危惧感说)的修正 正如前述,由于算法的黑箱属性,设计者对算法的决策及侵害结果预见不能,具体的结果预见可能性并不适用于设计者的过失判断。过失犯常被认定为“开放的构成要件”(意指法律仅对部分构成要件要素进行了描述,剩余部分由法官事后进行评价增补)^[26],同时,事前为行为提供指向性作用的行为规范也不可忽视。行为规范的事前判断

是过失犯将结果预见可能性作为要件的原因之一,但如果从“禁止引起侵害”的角度来理解过失的行为规范,那么此时的过失犯与故意犯所违反的都是事后判断的同一规范。^{[11]187}但是在过失犯中行为人并不追求结果的发生,所以“禁止引起侵害”这类行为规范无法为行为人提供事前的指向性作用。此时,构成要件结果不能为过失犯的行为规范提供基础,那么即便在没有结果的具体预见可能性的情况下,过失犯的行为规范也能被违反。

即便对结果的具体发生没有预见,也可判断预见可能性,此观点在判断自动驾驶汽车设计者的过失行为时具备相当的合理性。因此前述旧过失论与新过失论并不适用于对设计者的过失判断,仅在新新过失论(危惧感说)的立场下才能实现对自动驾驶汽车设计者的过失归责。

然而,虽然新新过失论不会面对结果的具体预见问题,但是新新过失论标准的模糊性将会带来过失犯之处罚范围的扩大效果。因此,修正的危惧感说主张,不是对结果之发生有恐惧感就会带来产生结果避免义务,而是将结果修正为法益,弹性理解预见可能性与结果回避义务之间的关联性,即仅在具备法益关联性时,承认结果回避义务的发生。

二、修正的危惧感说设计者过失归责路径展开

在论证修正的危惧感说在设计者过失归责中的当然合理性后,随之而来的就是如何在具体案件中适用该归责路径。本文拟通过如下案例论证修正的危惧感说之具体适用路径。

例如,甲驾驶着由汽车制造公司乙所设计、生产的自动驾驶汽车(第3级)以40 km/h的速度在公路上行驶。在自动驾驶模式下,甲闭目养神时,小孩A突然从路旁人行道冲进公路中间。自动驾驶汽车因为将小孩识别为不明物体,认为其没有危险性,没有执行紧急应对措施,导致小孩A死亡。

在上述案例中,是否因驾驶员甲为自动驾驶汽车的直接控制者,就应由甲一人承担刑事责任?汽车制造公司乙是否过失甚至故意?如若成立过失,应当如何认定汽车制造公司乙的注意

义务呢?

道路情况时刻变幻莫测,但是作为驾驶员应当时刻观察道路情况以获取相关信息并且及时采取合理措施。就算是在第三级别的自动驾驶汽车开启了自动驾驶模式在道路上行进时,注视前方仍然是驾驶员不能免去的最基本的义务。并且由于自动驾驶汽车的特殊性,驾驶员更应当时刻观察道路情况、获取道路情况信息,以做好随时切换成手动模式的准备。因此,在自动驾驶模式下,甲身为驾驶员,有注视前方的义务。^{[27]276}

驾驶员甲承担相应的过失责任并不意味着制造公司乙可以免责。根据信赖原则,行为人因信赖他人将实施恰当的行为而采取了不恰当的行为致使他人权益受损的,不应追究行为人责任。^{[28]389}在该案例中,如果驾驶员甲尽到了注视前方的义务,且基于信任汽车制造公司乙所设计、生产的汽车没有缺陷,能够正常运行,那么自动驾驶汽车因其“正常运行”侵害了他人权益,驾驶员甲便不必承担责任。本案中,自动驾驶汽车将冲进道路中的小孩识别为不明物体,认为其没有危险性,进而没有执行紧急应对措施显然不属于正常运行的范畴。因此,在甲承担相应刑事责任后,汽车制造公司乙并不当然地排除责任。如果自动驾驶汽车的设计者过失造成系统故障、程序错误等致使事故发生,那么其过失责任应当如何被认定?

(一)注意义务之前提:抽象的预见可能性和结果回避可能性

在修正的危惧感说中,结果回避义务被视为过失的本质,因而提出行为具有结果回避可能性是当然的,但是在有结果回避可能性之前,行为人必然有对自己的行为会有什么样的发展之认识的可能。

一方面,就预见对象而言,修正的危惧感说所要求预见的对象不同于修正的旧过失论与新过失论,而是将结果修正为法益。^{[29]275}此时并不必然要求有结果的预见可能性,行为人具备某种危险可能发生的不安感便可。^{[30]81-82}另一方面,就预见程度而言,前两种理论对于结果要求有具体的预见可能性,则必然要具备高度的预见可能性,而后者感觉这种结果可能发生并且有种危惧感即可。

虽然算法的设计者无法预见自动驾驶汽车会做出何种决策,但是其毋庸置疑是可以预见在异常情况下自动驾驶汽车会做出不好的决策的。^{[27]282}在异常情况(在自动驾驶汽车中的异常情况,是指当下只有让人类接手才能正确处理的情况^[24])下,只要一般人能够预见到算法可能会做出不好的决策以致法益受到侵害,并对于此有不能忽视的危惧感,就已经表明设计者有着相应预见可能性。^{[31]146}

此外,在有抽象的预见可能性之前提下,应当有能够采取措施以避免结果发生之可能。此时的结果避免可能性是一种回避措施的履行可能性,即事前的结果回避可能性。^{[13]84}如在前述案例中,小孩A突然从路旁人行道冲进公路中间时,自动驾驶汽车应当执行紧急应对措施,让甲接管汽车。设计者在进行算法设计时,对于汽车可能会对突然冲进道路中的行人造成身体法益甚至生命法益侵害的情况有预见可能性,在有结果避免可能性时,设计者应当对于类似的异常情况采取避免的措施。

(二)注意义务之内容:结果避免义务

藤木英雄教授指出,科学技术的发展给人们带来了一个又一个未知的危险,在这种情况下,暂且不论危险是以什么方式发生的,当危害发生时有着完全无法忽视的恐惧感时,那么在假设结果是可预见的情况下,将预防责任作为注意义务下达命令是适当的。^{[32]241}主流观点认为该理论的抽象预见可能性会导致注意义务之标准降低,从而扩大刑罚范围。^{[10]230}虽然修正的危惧感说的危惧感可以被概括为一种抽象的预见可能性,但主流意见的批判仍然值得商榷。修正的危惧感说并非主张行为人对于结果之发生怀有恐惧感这一条件一旦满足,就一定产生与之相对应的结果避免义务,而是要肯定预见可能性与结果回避义务之间的法益关联性。^{[14]226}在此前提下,不能以大概的有或者无的思路来追问预见可能性,而应从预见可能性与结果避免义务之间的相对关系进行弹性思考。“通过考虑以什么样的结果回避义务为前提来相对地决定。”^{[33]98}即,要求的结果避免义务越高,就需要更高层次的预见可能性,要求的结果避免义务越低,需要的程度的预见可能性就更低。

根据藤木英雄之理论,注意义务之内容限于结果回避义务。^{[13]43} 注意义务是能够避免结果发生者为了避免特定结果而被要求采取特定结果避免措施的负担。^{[34]57} 详言之,因为过失的双重地位,行为人被要求在构成要件阶层违反客观的注意义务,在责任阶层违反主观的注意义务。^{[35]373}

一方面,就客观的结果回避义务而言,从一般人的标准出发的同时还应当考虑到行为人当时所处的情况。依照《人工智能治理原则》所提出的“安全可控”、“共担责任”、“敏捷治理”三项原则,自动驾驶汽车设计者因为其数据输入、训练、检测的特殊职能,在数据处理时,至少应当对数据采取最基本的检测措施以提高自动驾驶汽车的可靠性、可控性。2018 年,在国内第一起特斯拉自动驾驶致死案中,由于设计车辆时所输入的样本存在问题,使得自动驾驶汽车在行驶时未能成功分辨出前方车辆导致事故发生。^[8] 如果设计者尽到使用数据异常检测工具检查这一基本义务,该结果就不会发生的话,那么就应当追究其刑事责任。此外,如果面临可能会产生异常风险的情况,设计者应当采取结果避免措施,具体如告知生产者、销售者、使用者等相关责任主体可能产生的风险和影响等。就前述案例而言,就使用者甲面临异常情况时应负临时接管义务,设计者应及时告知生产者,生产者可以做出相关警示,以便使用者面临特殊情况时能够及时采取相应措施。另一方面,就主观的结果回避义务而言,若在构成要件阶层未满足一般之注意义务,就暂时推定行为人具有违法性,进而进入违法性阶层与有责性阶层的审查。若不存在阻却违法事由或行为人不具有有责性(过失犯有责性的判断包括责任能力、不法意识与客观必要注意义务的履行能力三个方面^{[36]527}),则行为人构成过失犯罪。

三、设计者过失责任之承担

缺陷自动驾驶汽车过失责任之追究,实质上就是缺陷过失产品责任之追究。在没有缺陷过失产品罪之前提下,面对人工智能技术引发的风险,在现有体系内如何追究设计者的过失责任,是一个值得思考的问题。

我国《刑法》对于缺陷产品仅从故意层面进

行定罪处罚,当现实生活中出现过失缺陷产品导致损害后果时,常常在民法或者行政法中寻找法律依据。设计者如果是故意设计缺陷而造成损害自然难逃刑责,若是过失行为造成损害,带来的风险也将突破立法者所允许的界限,难以将其归入被允许的风险范畴。^{[31]186} 与传统产品不同的是,人工智能产品的自主性与不可预测性带来危害后果的广度与深度丝毫不亚于故意的产品犯罪。如果仅从民法或者行政法对该行为进行规制,该犯罪行为势必会脱离法律可控范围。

刑法需要对随社会发展而新出现的犯罪做出反应,但部分学者动辄主张构建新的刑事责任体系以追究责任之主张^[37],这样未在既有体系中进行解释便求之于重构的思路并不值得推崇。构建新的理论体系虽然能够解决问题,但却易使理论体系本身产生新的矛盾。与其如此,不如首先考虑在既有的理论体系下进行解释,以既存的理论体系对新生事物进行规制。在面对诸如此类的缺陷时,解释者应该抛弃“等待”,积极地对该缺陷进行解释。^{[38]219} 对于个别问题,在适当的刑事政策的引领下结合社会实际情况对法律规范进行解释,即可实现既定目标。就现行《刑法》而言,对自动驾驶汽车设计者的过失行为致使交通事故的归责方式主要包括如下几种。

(一)生产伪劣产品罪与生产不符合安全标准的产品罪

自动驾驶汽车由于预先设计和编制程序使得其在投入使用领域后也会持续受到设计者的影响。自动驾驶汽车设计者因其影响领域的变广、影响深度的增强,其应当履行更多注意义务,并在事故发生后承担相应的法律责任。当前,我国《刑法》“破坏社会主义市场经济罪”章第 140 条到 150 条对一般产品与特殊产品涉及的故意犯罪的刑事责任予以规定。但是该章节的生产伪劣产品罪或者生产不符合安全标准的产品罪均不能合理评价自动驾驶汽车设计者之过失致害行为。

第一,受侵害之法益难以被全面、合理、确切评价。^[5] 法益的侵害是实质犯罪概念的核心。^{[36]8} 自动驾驶汽车在公共交通领域引发的交通事故极易使不特定或者多数人的生命法益、身体法益与公众生活的平稳和安宁受到侵害。^{[39]882} 该犯罪

行为更多在公共安全或者公民人身权利、民主权利方面造成法益侵害,而非社会市场经济秩序的破坏。不同的条文是不同程度的法益受到侵害的具象化,即便在同种法益下,不同的犯罪行为也会受到不同种类、不同幅度的刑罚处罚。^[40]以生产伪劣产品罪或者生产不符合安全标准的产品罪评价设计者之过失行为,势必会造成受侵害之法益无法被完整、合理、确切评价的问题。有罪乱罚不仅难以维护社会安定,反而会违背罪刑法定原则与罪责刑相适应原则。

第二,过失不能包容评价为故意。生产伪劣产品罪或者生产不符合安全标准的产品罪要求行为人主观罪过为故意,且带有非法牟利之目的。在故意犯中,如果缺少结果回避可能,那么往往仅能从行为无价值的角度进行考虑,在此前提下,行为人至少不能成立既遂犯。因此,结果回避可能性实质上为故意犯和过失犯之共同要件。^{[14]213}根据结果回避可能性的高低、违法性的高低、责任承担的高低,可以将故意与过失之间的关系定义为位阶关系。^{[28]343}故而,故意可以降低评价为过失,而过失不能升格评价为故意。同时,评价亦不可能是任意的、无限制的。因此,在没有被刑法明文规定“过失生产伪劣产品罪或者过失生产不符合安全标准的产品罪”时,不能以生产伪劣产品罪或者生产不符合安全标准的产品罪对设计者的过失行为进行定罪处罚。

第三,生产伪劣产品罪或者生产不符合安全标准的产品罪之主体无法涵摄设计者。生产伪劣产品罪或者生产不符合安全标准的产品罪之主体仅包括了生产者,生产者与设计者虽然均属产品生产流水线前端,但是生产者无法控制算法设计阶段的风险,二者不能混为一谈。虽然生产者与设计者具有相似性,但是二者属于不同的职能分工,以生产者之名,罚设计者之罪,难言无有罪类推之嫌。

(二)交通肇事罪

虽然生产领域的罪名不能界定与评价自动驾驶汽车生产者的过失行为,但是自动驾驶汽车的产品本质属性与设计者对于使用阶段的控制地位,决定了可以从产品使用领域考虑如何追究设计者的责任。此时,应当从产品使用之具体情形来展开讨论。

在自动驾驶汽车根据既定程序实施了自动驾驶行为,造成交通事故后,以交通肇事罪考虑自动驾驶汽车设计者的责任的归责路径并不明朗。一方面,交通肇事罪的主体对于交通管理规定的违反是出于故意,但就设计者出于疏忽大意或者过于自信的过失设计了程序而言,设计者并非故意违反交通管理规定。^[41]另一方面,从自动驾驶汽车核心系统的设计、检测、制造到销售再到实际使用等多个场景中,虽然设计者在产品使用领域享有控制权,但并没有直接参与驾驶过程。如果以交通肇事罪追究其责任,就会出现驾驶行为主体与责任主体不一致的情况,这一情形明显超出了交通肇事罪主体的涵摄范围。^[42]故而,以此罪追究设计者责任有失合理性。

(三)重大责任事故罪

就重大责任事故罪而言,有学者认为此罪不仅能评价所侵害的法益,还能确切评价行为人的主观罪过。^[5]但是,一方面,自动驾驶汽车的研发、测试过程的风险与该罪的生产安全风险并不一致,不能简单将二者等同,故不能以重大责任事故罪法益概括评价自动驾驶汽车在使用领域所侵害的法益。^[43]另一方面,本罪主体要求为对生产、作业相关的负责人员,与前述两罪类似,虽然自动驾驶汽车设计者虽然在汽车进入使用领域后对其还有控制权,但并不意味着设计者应当履行安全生产义务,且该罪主体难以涵括自动驾驶汽车设计者。若一味适用该罪以规制设计者之过失行为,难免有类推适用之嫌。

(四)过失致人重伤罪与过失致人死亡罪

从上述分析可知,不论是生产伪劣产品罪还是生产不符合安全标准的产品罪、重大责任事故罪抑或是交通肇事罪,均不能评价设计者之过失行为。但对该行为如若不加以抑制,只会引发更严重的危害后果。因此,在《刑法》中特别过失条款无法规制该类危害行为时,我们可以考虑将该行为放在一般过失(过失致人死亡或过失致人重伤罪)的视角下进行思考,构建其归责路径。现有《刑法》关于过失缺陷产品类和人工智能类刑法条文的缺位并不代表只追究此类犯罪的故意行为,对该类过失犯罪仅追究民事或行政责任。

因此,在现有刑法体系内,以过失致人重伤罪与过失致人死亡罪对设计者的过失致害行为

定罪处罚,并以修正的危惧感说对过失致人重伤罪与过失致人死亡罪进行解释,不失为一条合理的归责路径。

人工智能技术在席卷各个领域的时候也带来了风险,在自动驾驶领域,因果关系复杂、涉案人员面广和危害结果严重等特征不可避免地引发自动驾驶汽车的刑事责任归责困境。在人工智能当前不能成为刑事主体的背景下,当自动驾

驶因设计过失造成交通事故,设计者不可避免地成为归责主体。根据我国目前《刑法》,交通领域与产品领域的罪名都不能对其定罪处罚,然而对该犯罪行为不加以抑制只会引发更严重的危害后果。因此,特别条款无法规制该犯罪行为时,我们可以通过修正的危惧感说的解释将该犯罪行为放在从一般过失的视角下思考,以追究设计者的责任。

参考文献:

- [1] 劳东燕. 风险社会中的刑法: 社会转型与刑法理论的变迁[M]. 北京: 北京大学出版社, 2015.
- [2] 叶良芳, 马路瑶. 风险社会视阈下人工智能犯罪的刑法应对[J]. 浙江学刊, 2018(6): 65-72.
- [3] 陈祖瀚. 刑事责任主体不宜涵括人工智能[J]. 铁道警察学院学报, 2019, 29(5): 77-82.
- [4] 刘宪权. 人工智能时代我国刑罚体系重构的法理基础[J]. 法律科学(西北政法大学学报), 2018, 36(4): 47-55.
- [5] 刘宪权. 涉人工智能产品犯罪刑事责任的归属与性质认定[J]. 华东政法大学学报, 2021, 24(1): 50-59.
- [6] 王乐兵. 自动驾驶汽车的缺陷及其产品责任[J]. 清华法学, 2020, 14(2): 93-112.
- [7] 吴汉东. 人工智能时代的制度安排与法律规制[J]. 法律科学(西北政法大学学报), 2017, 35(5): 128-136.
- [8] 央视网. 国内首起“特斯拉”车祸致死案: 确认为“自动驾驶”[EB/OL]. (2018-04-19)[2024-3-25]. <http://news.cctv.com/m/a/index.shtml?id=ARTIZzv9BLsbAmvSi0oF0lhJ180419>.
- [9] 央视网. 蔚来汽车频陷安全风波, L2 级自动驾驶暴露识别缺陷[EB/OL]. (2021-08-17)[2024-7-06]. <https://auto.cctv.com/2021/08/17/ARTI11nLfYdQ3RGjW9CnM9CS210817.shtml>.
- [10] 西田典之. 日本刑法总论[M]. 王昭武, 刘明祥, 译. 2 版. 北京: 法律出版社, 2013.
- [11] 高桥则夫. 刑法总论[M]. 李世阳, 译. 北京: 中国政法大学出版社, 2020.
- [12] 付立庆. 刑法总论[M]. 北京: 法律出版社, 2020.
- [13] 程皓. 注意义务比较研究: 以德日刑法理论和刑事判例为中心[M]. 武汉: 武汉大学出版社, 2009.
- [14] 井田良. 講義刑法学・総論[M]. 2 版. 东京: 有斐阁, 2018.
- [15] 劳东燕. 被允许的危险理论与过失犯的归责[M]//梁根林, 埃里克·希尔根多夫. 违法性论: 共识与分歧: 中德刑法学者的对话(三). 北京: 北京大学出版社, 2020.
- [16] 佐伯仁志. 刑法总论的思之道·乐之道[M]. 于佳佳, 译. 北京: 中国政法大学出版社, 2017.
- [17] 山口厚. 刑法总论[M]. 付立庆, 译. 3 版. 北京: 中国人民大学出版社, 2018.
- [18] 潘斌. 重思人工智能的辩证法: 从人工智能到人类未来[J]. 天津社会科学, 2020(3): 23-29.
- [19] 贾开. 人工智能与算法治理研究[J]. 中国行政管理, 2019(1): 17-22.
- [20] 沈伟伟. 算法透明原则的迷思: 算法规制理论的批判[J]. 环球法律评论, 2019, 41(6): 20-39.
- [21] 刘艳红. 刑法学(上)[M]. 2 版. 北京: 北京大学出版社, 2016.
- [22] 平野龙一. 刑法総論[M]. 东京: 有斐阁, 1972.
- [23] 周光权. 风险升高理论与存疑有利于被告原则: 兼论“赵达文交通肇事案”的定性[J]. 法学, 2018(8): 66-78.
- [24] 刘嘉铮. 人工智能产品算法设计者的犯罪过失判断: 以危惧感说的核心观点为立场[J]. 重庆大学学报(社会科学版), 2024, 30(3): 228-241.
- [25] 黄荣坚. 基础刑法学(上)[M]. 台北: 元照出版有限公司, 2012.
- [26] 蔡桂生. 德国刑法学中构成要件论的演变[J]. 刑事法评论, 2012, 31(2): 1-67.
- [27] 福田雅树, 林秀弥, 成原慧. AI 联结的社会: 人工智能网络化时代的伦理与法律[M]. 宋爱, 译. 北京: 社会科学文献出版社, 2020.
- [28] 张明楷. 刑法学(上)[M]. 6 版. 北京: 法律出版社, 2021.
- [29] 吕英杰. 缺陷产品的刑事责任研究[M]. 北京: 法律出版社, 2019.

- [30]高桥则夫. 规范论和刑法解释论[M]. 戴波, 李世阳, 译. 北京: 中国人民大学出版社, 2011.
- [31]热娜古·阿帕尔. 人工智能的刑事责任研究[M]. 北京: 研究出版社, 2022.
- [32]藤木英雄. 刑法講義総論[M]. 东京: 弘文堂, 1975.
- [33]井田良. 刑法总论的理论构造[M]. 秦一禾, 译. 北京: 中国政法大学出版社, 2021.
- [34]藤木英雄. 公害犯罪[M]. 丛选功, 徐道礼, 孟静宜, 译. 北京: 中国政法大学出版社, 1992.
- [35]林书楷. 刑法总则[M]. 台北: 五南图书出版股份有限公司, 2010.
- [36]王皇玉. 刑法总则[M]. 台北: 新学林出版股份有限公司, 2022.
- [37]皮勇. 论自动驾驶汽车生产者的刑事责任[J]. 比较法研究, 2022(1): 55-70.
- [38]张明楷. 刑法分则的解释原理[M]. 北京: 高等教育出版社, 2024.
- [39]张明楷. 刑法学(下)[M]. 6版. 北京: 法律出版社, 2021.
- [40]赵运锋. 刑法法益的认识定位与功能分析: 兼论法益分析对以刑制罪的影响[J]. 北方法学, 2017, 11(1): 86-98.
- [41]侯帅. 自动驾驶技术背景下道路交通犯罪刑事责任认定新问题研究[J]. 中国应用法学, 2019(4): 92-106.
- [42]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[J]. 河南社会科学, 2019, 27(4): 54-62.
- [43]贾济东, 岳艾洁. 人工智能事故过失犯认定的中国方案: 规范型塑与理论因应[J]. 河北法学, 2023, 41(10): 57-78.

Theoretical construction of criminal negligence imputation for designers of autonomous vehicles

LIU Ying

(College of Law, Sichuan Normal University, Chengdu 610066, China)

Abstract: As autonomous vehicle technology matures and such vehicles gain widespread adoption, traffic accidents arising from the novel risks represented by autonomous vehicles are inevitable. These new types of traffic accidents are characterized by complicated causal relationships, difficulties in liability determination, and involvement of multiple parties. Coupled with the “black box” nature of algorithms and the reality that the current Criminal Law only stipulates liability for intentional conduct concerning defective products, this will inevitably create obstacles to the criminal regulation of traffic accidents caused by designers’ negligent acts. Within the existing criminal law system, neither the revised old negligence theory nor the new negligence theory can reasonably explain harm-causing negligent conduct attributable to designers. It should start from general negligence and adopt the revised risk-awareness theory to demonstrate the regulatory path for traffic accidents of autonomous vehicles caused by designers’ negligence.

Keywords: designers of autonomous vehicles; negligent offender; defective products; revised risk-awareness theory; criminal responsibility

(责任编辑: 周新颜)

引用格式 刘颖. 自动驾驶汽车设计者刑事过失归责的构造[J]. 山东航空学院学报, 2026, 43(3): 106-113.

Liu Ying. Theoretical construction of criminal negligence imputation for designers of autonomous vehicles[J]. Journal of Shandong University of Aeronautics, 2026, 43(3): 106-113.