理论前沿

自动驾驶汽车交通肇事的刑法归责

魏　超

内容提要：自动驾驶汽车的事故原因包括存在制造缺陷与存在设计缺陷两种，在因后者造成的事故中，制造商违反了注意义务，创设出了法不容许的风险，具有刑事违法性。自动驾驶汽车的使用者及制造商与法益损害均存在因果关系，基于信赖原则，使用者对“违反交通法规”不具有预见可能性，无须承担刑事责任。成立过失犯不要求行为人认识到具体的因果流程与致害行为，故只要制造商违反注意义务，创设出法不容许的风险，便对法益损害具有预见可能性，算法黑箱不能阻却其刑事责任。算法歧视有违法治国家平等保护之基本理念，在并未完全避免的情况下，原则上不应允许自动驾驶汽车上路。

关键词：自动驾驶汽车　容许的风险　注意义务　交通肇事　算法黑箱

魏超，苏州大学王健法学院讲师。

一　自动驾驶汽车交通肇事的归责难题

随着人工智能技术的飞速发展，自动驾驶汽车也逐渐从遥不可及的梦想转变成了现实。2024 年 6 月，工业和信息化部、公安部、住房城乡建设部、交通运输部等四部门联合公布《进入智能网联汽车准入和上路通行试点联合体基本信息》，确定 9 个联合体开展 L3、L4 级别智能网联汽车准入和上路通行试点。有公司已在我国武汉、重庆、北京、深圳等地大规模开启了 L4 级别全无人自动驾驶出行服务。但是，人工智能并非能够未卜先知的神明，尽管自动驾驶汽车拥有更强大的信息收集与处理能力，但交通事故仍基于各种原因——如传感器受污染、被遮蔽而无法识别障碍物、距离计算错误撞击前车甚至种种未知原因——频繁发生，由此也引起了民众对于自动驾驶汽车安全性能的广泛关注和担忧。而作为保护法益最后也是最为重要的一道防线，刑法也必须未雨绸缪，有所准备。自动驾驶汽车的交通肇事行为是否具有刑事违法性？其应当在何种情况下、由何人承担刑事责任？内置程序中算法黑箱的出现能否阻却制造商的预见可能性？对于这些我国社会和法治语境中极受关注、极为紧迫、亟需我们面对和解决的问题，学界却或存在巨大争议，或尚

未进行深入研究。基于此,本文以 L5 级别自动驾驶汽车交通肇事为现实面向,以过失犯归责困局为问题导向,盘梳整合既有教义学理论,寻求解决问题之路径,以期促进我国人工智能法律之发展,同时对司法实务有所裨益。

二　自动驾驶汽车交通肇事的刑法归责边界

虽然自动驾驶汽车造成交通事故的种类繁多,但并非均由汽车程序问题导致,故并非所有种类的事故都存在新的理论问题,为了让文章的主线更加清晰,有必要挑选出其中最具代表性的情形,以作为讨论前提。

(一)自动驾驶汽车交通肇事之讨论前提

第一,只包括 L5 等级自动驾驶汽车因自身过错造成交通事故之情形。[1] 不言而喻,事故的发生必然是由一方的过错所致,若主要由被害人一方所致,则根据我国相关司法解释,自动驾驶汽车一方当然无须承担刑事责任,此情形不在刑法的规制范围之内;若是由较低等级自动驾驶汽车的使用者所致,则在符合构成要件的情况下按传统理论直接认定成立交通肇事罪即可,也不存在任何争议;若为汽车程序问题所致,则其他等级的自动驾驶汽车均不若 L5 级别的汽车具有代表性。

第二,使用人不存在违规使用之情况。虽然自动驾驶汽车在绝大多数情况下均能够正常行驶,但其硬件仍然受到现实科技的限制,其运行也必然会受到一定外界环境的影响。例如,在面对浓雾天气、冰雪路面等恶劣环境时,可能并不适合开启自动驾驶模式。故若使用人在已经被告知不宜使用自动驾驶模式的情况下强行开启该模式,进而导致车辆因浓雾难以识别前方路人或车胎打滑发生交通事故,则理当由使用人承担责任。[2] 又如,《道路交通安全法》第 21 条规定,驾驶人驾驶机动车上道路行驶前,应当对机动车的安全技术性能进行认真检查;不得驾驶安全设施不全或者机件不符合技术标准等具有安全隐患的机动车。自动驾驶汽车也是如此,虽然使用人无需实际操控自动驾驶汽车,但仍负有维护和保养义务,如果汽车发出维修警示,而使用人置之不理,则因此类故障,如刹车老化失灵、摄像头有污渍而识别不清等引发的交通事故,理应由使用人承担责任。[3]

第三,汽车电脑并未出现故障,也未产生自主意识,完全是按照预设程序行驶。就前者而言,自动驾驶系统的复杂性使其不可避免地会存在一定漏洞,进而导致程序崩溃、卡顿、死机、车辆不受控制甚至被黑客入侵等现象,此类情形是自动驾驶汽车发展过程中必然遇到的问题。是以,若制造商、程序员等已经尽量避免可能发生的危险,但汽车仍然出现了现有技术无法解决的问题或无法预料的故障并造成人员伤亡,则其可因为缺乏过失而不成立犯罪,[4]若是汽车被黑客入侵引发交通事故,相应的责任当然应当由入侵者承

〔1〕 参见《汽车驾驶自动化分级》(GB/T 40429-2021)。

〔2〕 参见郑志峰:《自动驾驶汽车的交通事故侵权责任》,《法学》2018 年第 4 期,第 26 页。

〔3〕 See Ujjayini Bose, The Black Box Solution to Autonomous Liability, 92 *Washington University Law Review* 1325, 1338 (2014).

〔4〕 Vgl. Ethik-Kommission, Automatisiertes und Vernetztes Fahren Eingesetzt durch den Bundesminister für Verkehr und digitale Infrastruktur, Bericht Juni 2017, S. 11.

担。就后者而言,则涉及有自我意识的电脑能否成为责任主体的问题,笔者支持否定论,认为此类电脑也不具有责任能力,因而难以成为归责主体。[5] 总之,违反交通规范造成事故的撞击行为必须是由制造商设定的自动驾驶汽车内置程序而非其他因素导致的。基于上述前提,下文将从以下基于真实案件改编而来的案件为出发点展开讨论。

案例 1:A 驾驶一辆 L5 级别自动驾驶汽车在高速公路上撞到一位过马路的行人,并致后者死亡。事后查明,事故原因在于后者所穿的白色衣物与明亮的天空非常巧合地一致,导致自动驾驶车辆的人工智能无法有效识别。

案例 2:B 驾驶一辆 L5 级别自动驾驶汽车在高速公路上撞到一位过马路的行人,并致后者死亡。事后查明,事故原因在于死者肤色较深,而自动驾驶车辆未能有效识别其肤色,误将其当成了深色路面。

在这样两个典型的自动驾驶汽车交通肇事的案件中,首先应当探究,其撞死行人的行为是否实现了法不容许的风险,进而符合交通肇事罪的构成要件?

(二)刑法应介入"设计缺陷"引发之事故

根据学界的归纳,自动驾驶汽车发生交通事故的肇因不外乎两种,其一是制造缺陷,即由于科技与材料的限制,即便生产者已尽到所有可能的注意,产品仍然会出现偏离其预期设计的情况,如物理材料强度不足、车辆摄像头、传感器、雷达等感知系统功能障碍等;其二是设计缺陷,即生产者对产品导致损害的可预见风险本可以通过采纳合理的替代设计而减少或避免,但因为疏忽没有采纳该替代设计而使产品具有不合理危险,自动驾驶系统所配置的算法和程序即为其例。[6] 不言而喻,程序员在编写算法程序或者开发驾驶系统时,只能基于当时的科学技术尽可能地规避风险,而不可能超越时代科技设计出无所不能的系统,即便产品已经做了大规模的测试,生产商和审批机构也无法预见到其产品与各类交通参与者以及所有可想象的交通状况、环境条件、能见度和天气条件之间,会发生何种相互作用。因此,如果生产商能够证明其在现有的科学和技术水平之下,已经尽了自己的能力注意和避免损害结果的发生,但受车辆自身性能所限,仍然存在超出了制造商、程序员等无法预见、无可避免的危险,这种对法益残存的风险也就是我们为了换取生活便利等所不得不容忍的所谓"容许的风险"。[7] 就此问题,我国已经于 2021 年 2 月发布了《国家车联网产业标准体系建设指南(智能交通相关)》(下称"《建设指南》"),对车联网应用和产业发展的标准体系作出了详尽的规定,是以,若生产商能够证明其研发和制造行为完全符合相关的硬性生产标准,则因车辆制造缺陷造成的事故便能够因属于仅实现了"法律所容许的风险"而排除刑事责任。[8] 与之相对,在因设计缺陷致害的交通肇事中,制造

〔5〕 参见刘艳红:《人工智能法学研究的反智化批判》,《东方法学》2019 年第 5 期,第 124 页;王钢:《人工智能刑事责任主体否定论——基于规范与语义的考察》,《苏州大学学报(法学版)》2022 年第 4 期。

〔6〕 See Restat 3d of Torts: Products Liability, § 2 (a) (b);二者的详细区别及具体区分,参见王乐兵:《自动驾驶汽车的缺陷及其产品责任》,《清华法学》2020 年第 2 期,第 100-102 页。

〔7〕 参见王莹:《法律如何可能?——自动驾驶技术风险场景之法律透视》,《法制与社会发展》2019 年第 6 期,第 111 页;Valerius, Sorgfaltspflichten beim autonomen Fahren, in: Eric Hilgendorf (Hrsg.) Autonome Systeme und neue Mobilität: Ausgewählte Beiträge zur 3. und 4. Würzburger Tagung zum Technikrecht, 2017, S. 10。

〔8〕 参见龙敏:《自动驾驶交通肇事刑事责任的认定与分配》,《华东政法大学学报》2018 年第 6 期,第 79 页。

商完全可以通过采用合理的替代设计减少或避免产品导致的损害,但其却在具有预见可能性的情况下,因为疏忽大意没有采纳该替代设计,从而违反了注意义务,使得产品制造了“法不容许的风险”,在造成法益损害结果之时,仍然具有刑事违法性。

有部分学者认为此种情形下能够排除构成要件该当性。有学者指出,既然我们的现状是法律允许驾驶行为,而自动驾驶技术的推广显然能将这种法律容许的风险以及这种风险所带来的伤害降到最低程度,那么法律就不是可以允许这种技术,而是应该允许这种技术。〔9〕 不难看出,论者的主要理由在于,自主车辆将减少人为失误造成的车祸死亡和受伤人数,其广泛运用也必将给社会带来巨大的利益,且显著优于传统的人工驾驶,〔10〕因而能够满足概括性的利益衡量原则要求,属于容许的风险,故不能将其所造成的损害归责于使用者、生产者或设计者。另有部分学者认为,在人工智能初期阶段没有明确的产业内标准,制造商的行为规则尚未确立,因此刑法暂时不宜介入。〔11〕 但是,以上观点存在值得商榷之处。

首先,自动驾驶汽车的车祸率较低只能够推导出其更具有安全性,并不能够得出其发生车祸的行为不符合构成要件之结论,此二者一个是自动驾驶汽车与传统汽车的安全性孰高孰低的问题,另一个则是自动驾驶汽车致害是否符合交通肇事罪构成要件的问题,并不处于同一个层面。换言之,自动驾驶汽车车祸率低与其行为成立犯罪之间并非“非此即彼”的互斥关系,而是可能同时并存。如果我们要论证此种情形中自动驾驶汽车并不成立犯罪,应当直接证明其不符合犯罪的成立条件,而不应避重就轻,证明其车祸率较低,因为后者与其是否成立犯罪并无关联。刑法上判断某一行为是否成立犯罪,并不是以其带来了多少利益、拯救了多少人的性命或避免了多少人的生命遭受损害为依据,而是以其是否实施了符合构成要件的行为并造成了法益损害为标准,否则所有数量不对等的对生命紧急避险便失去了讨论的意义。

其次,依照论者的逻辑,更会得出在任何情况下,自动驾驶汽车交通肇事均不成立犯罪之结论——因为无论如何,自动驾驶汽车的车祸率都会比人工驾驶的更低,故无论其因为何种原因造成了事故,都应当属于容许的风险。问题的关键仍然在于行为人是否违反了注意义务,以自动驾驶汽车可能避免更多法益遭受侵害为由,并不能够得出其致人死亡的行为不符合交通肇事罪构成要件之结论。

最后,为了应对自动驾驶技术的出现,我国已经颁布了多项部门规章,对自动驾驶汽车的相关问题从技术标准到法律风险均做出了明确的规定。例如在《建设指南》的第三部分中,相关部门便对车联网(智能交通相关)确立了72项标准项目,包括55项国家标准与17项行业标准,因此有些论者“行为规则尚未确立”的论述已然不符合实际。退一步而言,即便暂时没有行业内的规范,自动驾驶汽车的程序也应当符合一般行为的注意规

〔9〕 参见骆意中:《法理学如何应对自动驾驶的根本性挑战?》,《华东政法大学学报》2020年第6期,第56页。

〔10〕 See Jeffrey K. Gurney, Sue My Car Not Me: Products Liability and Accidents Involving Autonomous Vehicles, 2 *University of Illinois Journal of Law, Technology & Policy* 247, 250-251 (2013).

〔11〕 参见储陈城:《人工智能时代刑法归责的走向——以过失的规则间隙为中心的讨论》,《东方法学》2018年第3期,第36-37页。

范,故在车辆行驶过程中因不符合相应注意义务而造成人身伤亡事故的,纵然因缺少相应的交通运输法规而不符合交通肇事罪的构成要件,却仍然可以通过过失致人重伤罪、过失致人死亡罪来认定其刑事责任。〔12〕 就此问题,我国工业和信息化部、公安部、交通运输部已经于 2021 年 7 月联合发布了《智能网联汽车道路测试与示范应用管理规范(试行)》,其中第 34 条便明确规定,在道路测试、示范应用期间发生交通事故……构成犯罪的,依法追究当事人的刑事责任。

综上,在因符合行业标准的制造缺陷致害的交通肇事中,相关人员并未违反注意义务,不符合过失犯的构成要件,不具有刑事违法性;而在因设计缺陷引发交通事故之时,自动驾驶汽车已然创设了“法不容许的风险”,刑法理应介入其中,认为此时能够以“利益衡量”或“不具有行业规则”而排除构成要件该当性之观点,不但缺乏教义学根据,更与我国既有的规定相违背,因而不宜采纳。

三 自动驾驶汽车交通肇事的刑事责任主体

因“设计缺陷”造成的交通事故符合交通肇事罪的构成要件,故随后的问题在于,在此情况下,乘客与制造商何者应当承担刑事责任?〔13〕

(一)制造商与使用者均与损害存在因果关系

与传统交通事故有别,在全自动驾驶汽车中,车辆在行驶过程中是由预设程序所控制的,所谓的驾驶员只是打开了汽车按钮并输入指令而已,其身份已经无限趋向于纯粹的乘客。正是由于出现了此种转变,使得学界对于自动驾驶汽车交通肇事中的责任主体产生了一定争议。部分学者认为,此时的使用者不应成为责任主体,因为该事故并不是由使用者的行为导致的,汽车的驾驶行为完全由汽车自身控制,事故是由汽车自身在收集到实时数据的基础上进行识别、判断而做出的驾驶行为导致的,使用者的行为与最终的法律效果之间并不存在因果联系。〔14〕 与之相对,少数学者认为,使用者在此情况下仍然需要承担刑事责任。如英格兰德(Armin Engländer)教授在自动汽车紧急避险的情形中便认为,从因果关系的角度考察,实现的法益侵害结果,都产生于使用者发动汽车的行为,故应全部归因于使用者的启动自动驾驶汽车之行为。〔15〕

但是,上述将结果只归因于一方的观点均失之偏颇,因为制造商与使用者对自动驾驶汽车均实施了一定程度的支配行为,故二者与损害结果均存在因果关系。首先,从自然的因果关系方面分析,对于制造商而言,正是其商品中的设计缺陷使得法益损害结果发生,

〔12〕 参见杨宁:《刑法介入自动驾驶技术的路径及其展开》,《中国应用法学》2019 年第 4 期,第 116-117 页。

〔13〕 需要说明的是,因为自动驾驶汽车的生产过程较传统汽车更为复杂,往往涉及多个部门之间的相互配合,即便能够查明因果关系,其中何人来承担责任,也往往涉及企业合规、责任分配、刑事政策等诸多因素,非本文所能解决,故下文仅讨论此种责任在乘客与制造商之间如何分配。

〔14〕 参见皮勇:《论自动驾驶汽车生产者的刑事责任》,《比较法研究》2022 年第 1 期,第 57 页;王德政:《人工智能时代的刑法关切:自动驾驶汽车造成的犯罪及其认定》,《重庆大学学报(社会科学版)》2020 年第 3 期,第 134 页。

〔15〕 Vgl. Engländer, Das selbstfahrende Kraftfahrzeug und die Bewältigung dilemmatischer Situationen, ZIS 2016, 608, 611 f.

其生产制造行为与最终结果损害间存在因果关系并无争议;但是,这并不代表使用者与交通事故的发生不存在因果关系。因为制造商只是将产品生产出来,真正将其投入使用的是按下车辆启动开关的使用者,如果缺少使用者的乘车行为,事故同样不可能发生。由此可见,在自动驾驶汽车交通肇事的案件中,最终的法益损害结果其实是由制造商设置的程序与使用者启动汽车的行为协力完成的,后者的介入并不会阻断前者与结果之间的因果关系,二者处于一种重叠的因果关系之中。[16] 其次,就行为与结果发生的影响力而言,更无理由认为"此时自动驾驶汽车完全受到系统的支配"。诚然,在行驶途中,自动驾驶汽车何时转弯、何时鸣笛、是否刹车等均是由程序所决定,故从表面上看驾驶行为是完全由系统决定的,但是,上述操作并不是由系统单方面随意作出的,而是系统结合特定的外部条件方才启动的,而此类外部条件,如时间、地点、路线等,完全是由使用者决定的。故使用者虽然没有实际地操控汽车,却已经通过决定其他外在条件对汽车的行驶产生了影响,进而与程序系统共同决定了具体的驾驶行为。是以,此时自动驾驶汽车并不是"完全受到系统的支配",而是在使用者的决定与制造商的系统的共同支配下——无论前者是否知晓或承认,都不会影响这种作用力的存在——造成了法益损害结果,其二者也均与结果间存在因果关系。[17] 当然,此种论述并不代表二者已然成立犯罪,因为根据责任主义,唯有当行为人对法益损害至少存在过失的前提下,其才需要承担刑事责任,故接下来的问题在于,使用者与制造商,何者可能成立犯罪?

(二)使用者原则上无须承担刑事责任

从因果关系的类型看,制造商及使用者的行为与自动驾驶汽车交通肇事法益损害结果之间属于一种"累积性竞合"的关系,即虽然每个人的过失行为单独不含有结果发生的危险,但与他人的过失行为相结合,便能够产生结果发生的危险。[18] 部分学者认为,过失竞合理论应当是作为适用于协同作业中的角色分担之信赖原则的适用情形,属于限定过失犯成立的法理,因为此时证明各个行为人的预见可能性或肯定因果关系的存在等较为困难。[19] 依此逻辑,在肯定了行为人与结果具有因果关系的前提下,只要能够准确判断各人对结果是否具有预见可能性,即可判定其是否成立过失犯。

根据我国《刑法》第133条的规定,成立交通肇事罪,行为人不仅要造成他人的法益损害,而且必须违反了相应的交通规范。但无论法律规定还是现实情况,制造商的承诺必然是"本车的程序完全符合现行的法律规范",基于信赖原则,只要制造商在市场销售该产品,并承诺其产品将遵守交通法规且没有做特殊说明,使用者便完全有理由相信制造商所设计的程序在任何情况下都会遵守交通规范。[20] 详言之,使用者在购买自动驾驶车辆时,制造商对他们的承诺不仅仅是"此车能把你送到目的地",更是"此车能够在遵守交通

〔16〕 Vgl. Bock, Strafrecht Allgemeiner Teil, 2. Aufl. 2021, § 5 Rn. 118.

〔17〕 参见彭文华:《自动驾驶汽车犯罪的归责与归因》,《东方法学》2024年第1期,第113页。

〔18〕 参见李世阳著:《共同过失犯罪研究》,浙江大学出版社2018年版,第104页。

〔19〕 参见[日]高桥则夫著:《刑法总论》,李世阳译,中国政法大学出版社2020年版,第215-216页。

〔20〕 See Clint W. Westbrook, The Google Made Me Do It: The Complexity of Criminal Liability in the Age of Autonomous Vehicles, 97 *Michigan State Law Review* 97,127 (2017).

规范的前提下把你安全地送到目的地”。在科技日益发达的时代,知识库系统会为自动驾驶车辆储备最新、最完备的交通运输管理法律法规,加上本文的前提条件是人工智能并未产生自主意识也并未出现故障,其必然会忠实履行内置的程序指令,不会违背法律规定。且如今自动驾驶系统所具备的能力不仅包括简单识别红绿灯、自动巡航等,而是完成更为复杂的如检测物体、避免其他车辆撞击、紧急制动等操作,故无论是从法律的覆盖范围抑或实际的操作可能性层面,使用者都能够充分信赖制造商作出的“遵守交通规则”之承诺是真实有效的。在此情况下,使用者对于自动驾驶车辆违反交通规则并造成交通事故的行为并不具有预见可能性,因而难以成立交通肇事罪。当然,此种结论并非绝对。根据《最高人民法院关于审理交通肇事刑事案件具体应用法律若干问题的解释》第 7 条的规定,成立交通肇事罪并不需要行为人亲自实施驾驶行为,只要其对实际的驾驶者具有一定的支配力,且对其实施违章行为具有预见可能性即可。因此,在“使用者已经知晓车辆可能违反交通运输管理法规,却仍然驱车上路并因此造成交通事故”的情形中,使用者不但对车辆违章具有预见可能性,违反了注意义务,而且开启了驾驶系统,“指使”车辆上路行驶,故仍然能够成立交通肇事罪。

综上,在制造商做出系统会遵守交通法规承诺的前提下,使用者完全能够信赖系统会做出正确的决策,故仅需在传统的过错责任框架下追究其违反注意义务或错误操作自动驾驶系统的过错责任,原则上不应再额外增加其在自动驾驶系统监控方面的注意义务。〔21〕

(三)新过失论语境下制造商刑事责任之证立

最后需要讨论的是,自动驾驶汽车的制造商是否需要为其设计缺陷引发的事故承担交通肇事等过失犯罪的刑事责任。对此问题,学界存在肯定说与否定说之争。肯定说认为,一个能够预见其行为可能危害为刑法所保护的利益的人,就有义务避免这一行为。〔22〕作为自动驾驶汽车的制造商,其当然具有先期的、前瞻性的结果避免义务,如果其在制造、生产阶段就违反了这种注意义务,则在事故发生时可以成立相应的过失犯罪。〔23〕 与之相对,大部分学者支持否定说,但理由却各不相同。如有学者认为,注意义务是指行为人对某一具体行为所可能导致发生的具体结果的预见和避免义务,而非对抽象的某一类行为可能导致什么样结果的预见和避免义务,如果根据抽象判断来认定过失,将导致任何可能造成危害结果的行为都是过失行为,无论行为与结果之间的联系多么间接,行为人对具体结果能否预见,都可能得到肯定过失成立的结果,有违责任主义。〔24〕 还有学者认为,由于算法黑箱的存在,我们并不知晓自动驾驶汽车的运作流程或原理,因而难以对系统背后的车辆制造者是否履行了结果回避义务进行判断,〔25〕也就无法判断制造商是否成立犯罪。

〔21〕 参见刘艳红:《自动驾驶的风险类型与法律规制》,《国家检察官学院学报》2024 年第 1 期,第 128-129 页。

〔22〕 Vgl. Duttge, in: Münchener Kommentar StGB, 3. Aufl. 2017, § 15 Rn. 121.

〔23〕 参见彭文华:《自动驾驶车辆犯罪的注意义务》,《政治与法律》2018 年第 5 期,第 90 页。

〔24〕 参见周铭川:《论自动驾驶汽车交通肇事的刑事责任》,《上海交通大学学报(哲学社会科学版)》2019 年第 1 期,第 41-42 页;江溯:《自动驾驶汽车对法律的挑战》,《中国法律评论》2018 年第 2 期,第 186 页。

〔25〕 参见王霖:《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,第 98 页。

理由在于,人工智能的算法决策才是导致其作出错误判断的关键,而人工智能所做出的行为是其自主性的结果,具有不可预测性,并非设计者可以控制的,也不体现设计者的意图,它不可被解释。既然如此,将责任归于人工智能的设计者,显属不当。〔26〕

本文赞成肯定说,认为在自动驾驶汽车因设计缺陷造成交通事故的情况下,制造商应当承担相应的刑事责任。首先,即便存在算法黑箱,我们仍然能够判断出制造商是否履行了结果回避义务。因为算法黑箱针对的主体是用户而非制造商,是用户因为算法黑箱无法获悉车辆如何行动,而非制造商无法通过算法黑箱控制车辆行动。〔27〕 诚然,自动驾驶汽车在具体情形中如何行动都是系统根据彼时外界环境瞬间做出的决策,但此种决策绝非随机或不可预测,而是系统根据制造商预先设定的程序结合外界状况所产生,其必然存在解释、验证的余地。〔28〕 部分学者认为,由于人工智能算法涉及的数据量超出了人类的运算能力,对数据的特征提取具有很强的随机性,加之部分机器学习算法模型的数据处理过程隐蔽,故算法黑箱是无可避免的,〔29〕此时仍然让制造商承担相应的责任未免有强人所难之嫌疑。但事实上,随着近年来可解释人工智能技术(explainable artificial intelligence)的迅速发展,算法黑箱的技术障碍已被逐渐攻克。纵使部分算法仍然不具有内在透明性,也能够通过各种各样的事后透明方法得以弥补,算法透明在技术上已经完全能够实现。〔30〕 因此,只要该系统没有产生自主意识,汽车就仍然是按照既有的程序行动,即便存在所谓的算法黑箱,专业的鉴定机构也完全能够通过分析代码明晰自动驾驶汽车的运作原理,进而判断出制造商是否存在设计缺陷,有关论者的观点其实是建立在对算法黑箱概念的误解之上,因而并不妥当。

其次,即便如有论者所言,算法黑箱的复杂性让制造商无法完全知晓自动驾驶系统如何做出决策,也并不妨碍我们能够对他们进行归责。第一,“算法黑箱的复杂性”其实是涉及制造商是否违反了注意义务以及其对结果是否具有预见可能性的问题。一方面,就注意义务而言,只有在算法具有可解释性的前提下,人类才可以掌控人工智能技术,这是人工智能技术得以规范发展并持续造福于人类社会的必要条件,故制造商具有对自动驾驶汽车相关程序的解释义务。目前世界人工智能发展的趋势不再仅仅侧重于人工智能的功能发挥,同时也更加关注人工智能的可解释性水平,以此提高自动驾驶汽车的安全性和问责性,并为评估系统的公平性提供证据。2021 年 9 月,国家新一代人工智能治理专业委员会发布了《新一代人工智能伦理规范》,其中第 12 条明确规定,要增强安全透明,在算法设计、实现、应用等环节,提升透明性、可解释性、可理解性、可靠性、可控性,增强人工智能系统的韧性、自适应性和抗干扰能力,逐步实现可验证、可审核、可监督、可追溯、可预测、可信赖。2022 年 3 月实施的《互联网信息服务算法推荐管理规定》第 16 条也明确指

〔26〕 参见刘艳红:《人工智能的可解释性与 AI 的法律责任问题研究》,《法制与社会发展》2022 年第 1 期,第 87 页。

〔27〕 参见金梦:《立法伦理与算法正义——算法主体行为的法律规制》,《政法论坛》2021 年第 1 期,第 31 页。

〔28〕 参见卢有学、窦泽正:《论刑法如何对自动驾驶进行规制——以交通肇事罪为视角》,《学术交流》2018 年第 4 期,第 75 页。

〔29〕 参见陈景辉:《算法的法律性质:言论、商业秘密还是正当程序?》,《比较法研究》2020 年第 2 期,第 131 页。

〔30〕 参见安晋城:《算法透明层次论》,《法学研究》2023 年第 2 期,第 58-60 页。

出,算法推荐服务提供者应当以显著方式告知用户其提供算法推荐服务的情况,并以适当方式公示算法推荐服务的基本原理、目的意图和主要运行机制等。前述规范与生命法益相结合,使得自动驾驶汽车的算法理应具有最高的安全性,而唯有了解算法做出决策的原因及其界限,才能够明晰算法为什么会做出相关的决策,进而在一定程度上知晓其在特定条件下会做出怎样的决策,这样的算法模型才能够被认为是安全可靠的。〔31〕 一旦出现连最初算法设计者也不能予以有效化解的算法黑箱之情形,则此种黑暗算法必将成为强人工智能时代产生潜在风险的根源。〔32〕 具体至自动驾驶汽车中,如果制造商无法清楚地阐释自动驾驶汽车的运行原理,无异于给所有道路交通参与者的命运遮上一层“无知之幕”,将民众的生命交由系统在幕后随机决定。是以,纵使算法黑箱的存在让制造商在事实上难以完全知晓自动驾驶汽车如何做出决策,也并不代表其在规范上不需要知晓自动驾驶汽车如何做出决策,故从规范上判断,制造商应当对系统决策负有解释与告知义务,〔33〕即根据代码阐明汽车在各种情况下会做出的相应抉择及其原因,否则不但有违其解释义务,更将因无法知晓系统将如何行动而有违注意义务。〔34〕

另一方面,即便制造商并不知晓自动驾驶系统如何做出决策,也能够认为其对损害结果的发生具有预见可能性。如今的主流观点认为,因果关系虽然是故意的认识内容,但行为人并不需要认识到所有的因果经过,而是只要对因果经过的基本部分有认识就足够了。因此,在行为人能够认识到构成要件行为制造的危险可能现实化为结果的情况下,因果流程出现偏差并不会影响故意及既遂的认定。〔35〕 根据举重以明轻原理,我们不可能对过失犯罪提出比故意犯罪更高的要求,故在过失犯罪中,行为人也同样无须认识到具体的因果流程,只要其违反了注意义务,且对因果历程具有预见可能性,即能够肯定其具有过失。具体至本文的语境中,制造商是否预想到最终造成损害的方式并不重要,因为对于避免结果发生而言,重要的不是结果以何种方式发生,而是他必须履行注意义务,明确汽车的运行原理,将汽车掌控在手中。〔36〕 作为汽车的制造商,其当然能够预见自动驾驶汽车行驶过程中可能发生的各种状况,故在此情况下,制造商必然会对事故的发生具有预见可能性。至于论者所言“自动驾驶汽车并非设计者可以控制的,也不体现设计者的意图”之观点,也显然值得商榷。诚然,事故的发生看似是由自动驾驶汽车任性、自主决定的,而不是由于错误编程或制造商的过失所造成的,但这种表象是误导性的;实际上,危害可能正是源于编程中的疏忽,而非不能预见的意外。〔37〕 此种疏忽其实已经体现出制造商对于他人

〔31〕 参见沈寓实、徐亭、李雨航主编:《人工智能:伦理与安全》,清华大学出版社 2021 年版,第 216 页;Bartosz Brożek, Michał Furman, Marek Jakubiec & Bartłomiej Kucharzyk, The Black Box Problem Revisited, Real and Imaginary Challenges for Automated Legal Decision Making, 32 *Artificial Intelligence and Law* 427, 433-434 (2024)。

〔32〕 参见姚万勤:《客观归责理论与自动驾驶交通肇事刑事责任的归属》,《大连理工大学学报(社会科学版)》2023 年第 6 期,第 96 页。

〔33〕 参见万方:《算法告知义务在知情权体系中的适用》,《政法论坛》2021 年第 6 期,第 86-88 页。

〔34〕 参见张恩典:《超越算法知情权:算法解释权理论模式的反思与建构》,《东南法学》2022 年第 1 期,第 15 页。

〔35〕 参见周光权著:《刑法总论》(第四版),中国人民大学出版社 2021 年版,第 181-182 页;[日]山口厚著:《刑法总论》(第 3 版),付立庆译,中国人民大学出版社 2018 年版,第 228 页。

〔36〕 类似的分析过程,参见 Kindhäuser, Strafrecht Allgemeiner Teil, 11. Aufl. 2024, § 27 Rn. 45。

〔37〕 See Sabine Gless, Emily Silverman & Thomas Weigend, If Robots Cause Harm, Who is to Blame? Self-Driving Cars and Criminal Liability, 19 *New Criminal Law Review* 412, 432 (2016).

法益的漠视,完全可能成立相应的过失犯罪。而且,正因为此种行为并非由制造商控制,其才会仅成立过失犯罪,若某一行为当真是制造商有意为之,其理应成立相应的故意犯罪。

综上,算法的不可解释性在赋予制造商相应的注意义务的同时,也恰恰代表着其对交通事故具有预见可能性,而制造商对此置若罔闻,放任内置算法黑箱汽车上路之行为,已然违反了相应的结果回避义务,缺少对具体因果关系的预见并不能够阻却制造商的责任。若能够以存在算法黑箱为由逃避责任,则可能会出现行为人使用的软件越智能、技术越复杂,反而越能够逃避法律制裁之局面,从而在高新科技致害中出现"责任真空"之景象,人为地产生法律上的处罚漏洞。当然,此种观点并不意味着只要存在算法黑箱的自动驾驶汽车发生事故,制造商便一律需要承担刑事责任,因为如前所述,即便一款产品在投放市场时符合当前的科技水平,也不能完全排除它将成为"不可预见的危险源",受科学技术的制约,算法黑箱永远难以根除,故只要制造商能够证明其研发和制造行为完全符合相关生产标准,即便该车辆因难以或无法解释的算法黑箱而导致发生事故,其也能够因仅实现了"法律所容许的风险"而排除刑事责任。[38]

第二,算法黑箱的复杂性同样涉及过失犯中对于结果预见可能性程度之争议,本文对此持"抽象说"的立场,即针对还未实际发生的结果或实行行为,无需行为人认识到其具体的发生过程,只要"认识到结果可能发生"或者"预见到结果可能发生"即可。故在过失犯罪中,只要行为人认识到其行为已然创设出法不容许的风险,且最终造成的法益损害与其创设出的风险之间存在因果关系,便能够将结果归责于行为人,并不要求其认识到究竟是哪一个具体行为造成的法益损害。[39] 一方面,绝对具体的预见可能性是无法实现的,对于预见可能性的对象总要进行一定的抽象化。因为所谓"预见",都是指行为人当下对将来可能发生事情的推测,要求绝对正确地预见到自己的行为会导致什么样的结果,以及如何导致结果发生是不合理的。[40] 另一方面,虽然根据程度不同,学界将预见的抽象程度分为能够预见到"具体结果以及因果关系基本(重要)部分"的"具体预见可能性说"与仅具有不安感即可的抽象的"不安感说",[41]且前者占据了通说的地位。但由于将处罚范围限定在近似于存在预见或明显有可能预见的场合,就有可能使得很多过失案件无法得到处罚,因此各学者在具体的运用中,往往对预见对象做抽象的、缓和的解读,使得其在实质上接近抽象的预见可能性的立场。例如,西田典之教授一方面认为"预见可能性,还必须是以特定的构成要件结果为对象的具体的预见可能性,自己的行为也许会造成某种结果这种不特定的、抽象的预见可能性并不够";另一方面又在"住客躺在床上抽烟引起火灾后,由于酒店内未安装自动喷淋装置与隔火帘,造成火势蔓延致32人死亡"的案件中指出,酒店代表只需要对发生火灾之时的结果具有预见可能性,而不需要对起火原因本身也存在预见可能性。[42] 可见其实际上已然对预见可能性的内容做了缓和的理解,即在行

〔38〕 Vgl. Bachmann, Prozedurale Entlastung von Herstellern „smarter" Produkte im Strafrecht?, ZStrR 2022, 77, 90.

〔39〕 Vgl. Frisch, Strafrecht Examenswissen, Examenstraining, 2022, § 2 Rn. 153.

〔40〕 参见吕英杰:《论责任过失——以预见可能性为中心》,《法律科学(西北政法大学学报)》2016年第3期,第88页。

〔41〕 参见[日]前田雅英著:《刑法总论讲义》(第6版),曾文科译,北京大学出版社2017年版,第195页。

〔42〕 参见[日]西田典之著:《日本刑法总论》(第2版),王昭武、刘明祥译,法律出版社2013年版,第238、248页。

为人创设法所不容许的风险的情况下,认识到法益损害的结果即可实现归责,而无须认识到其究竟是在何种具体的情况下造成损害。〔43〕 有学者在传统汽车肇事案件中亦明确指出:只要汽车的制造商知道在恶劣的天气里汽车刹车会失灵却仍然推销汽车,那么他就违反了注意义务。〔44〕 由此可知,违反注意义务并不需要在具体情境中加以判断,只要行为人知道自己的行为创设了法不容许的风险便已经足够。具体至自动驾驶汽车中,同样应当对事故发生时的场景进行一定的抽象,留下基本要素,结合自动驾驶汽车决策的基本原则,根据事故场景是否罕见,判断制造商和使用者是否具有对危害结果的、相对具体的预见可能性。至于论者认为此举可能扩大处罚范围的担忧,其实也大可不必,因为自动驾驶汽车与传统汽车的事故在外观上并无不同,故根据目前的技术,完全能够类型性地判定其预见可能性。〔45〕

最后,采取抽象判断来认定过失,也并不会出现“将任何可能导致危害结果的行为均认定为过失,因而有违责任主义”之现象。否定论者之所以坚持认为“因为结果是通过一番因果经过才发生的,所以为了承认结果的预见可能性是具体的、超过了漠然的不安,因果经过的预见可能性通常是必要的”,〔46〕这其实是担心肯定说会持旧过失论的观点,将结果预见义务作为交通肇事罪的本质,一旦发生交通事故就肯定违法性,进而导致制造商在此类情形中均成立过失犯罪,从而过分扩大处罚范围。〔47〕 但如今的通说新过失论认为,过失的本体在于违反了结果回避义务,即对于社会生活中一般要求的结果回避义务即基准行为的懈怠。只要并未逸脱基准行为,就并无该当于构成要件且违法的行为,〔48〕故只要行为人履行了结果回避义务,便并不会成立犯罪。且就交通肇事而言,立法者已经将现实生活中容易发生事故的情形类型化为相应的义务,并规定在行政法规中,以提醒相关主体加以履行,制造商完全能够知晓其某种设计缺陷可能违背行政法规,进而推测出车辆可能在何种情况下造成何种损害,故其在违背此类规范时并非仅仅会产生极为抽象的“恐惧感”,而是完全能够认识到危险发生的可能性,即论者所谓的“因果经过的预见可能性”。因此,即便自动驾驶汽车造成了交通事故,只要其遵守了所有的交通规则,其相应的行为便是合乎谨慎的,制造商也就能够通过“容许的风险”出罪,并不需要为该结果负责,也就不会出现论者所担忧的其承担结果责任或者严格责任之景象。

综上,过失犯罪中预见可能性的内容只限于行为的注意义务违反性及行为所蕴含的不容许风险,危险的现实化流程及实际发生的具体结果则不属于预见的范围。〔49〕 制造商在生产复杂精密的自动驾驶汽车的过程中必然有确保其安全的义务,无法预见具体的因果流程或法益损害均不能减轻其责任,故当其能够履行却并未履行该义务,以致于给

〔43〕 同样观点,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,第 318 页。

〔44〕 See Sabine Gless, Emily Silverman & Thomas Weigend, If Robots Cause Harm, Who is to Blame? Self-Driving Cars and Criminal Liability, 19 *New Criminal Law Review* 412, 427 (2016).

〔45〕 参见[日]山口厚著:《刑法总论》(第 3 版),付立庆译,中国人民大学出版社 2018 年版,第 253 页。

〔46〕 [日]佐伯仁志著:《刑法总论的思之道·乐之道》,于佳佳译,中国政法大学出版社 2017 年版,第 253 页。

〔47〕 Vgl. Gleß/Weigend, Intelligente Agenten und das Strafrecht, ZStW 2014, 561, 584.

〔48〕 参见周光权著:《刑法总论》(第四版),中国人民大学出版社 2021 年版,第 163 页。

〔49〕 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,第 322 页。

法益造成损害之时,便满足过失犯罪成立的条件,应当对事故中的危害结果承担刑事责任。〔50〕

根据三阶层犯罪理论,只要某一行为符合犯罪构成且缺乏违法阻却事由与免责事由,当事人就应当承担刑事责任,但部分学者却认为,在自动驾驶汽车发生交通事故之时,仍应当"法外施仁",不对其判处刑罚。如我国有学者指出,如果让制造商承担研发后果责任,会造成人工智能这一高新技术遭到变相的压制,因为技术企业及其技术人员不可能愿意仅因涉及系统研发就承担所有责任,从而降低其参与系统研发的勇气和热情,大大阻碍人工智能的发展。〔51〕是以,为了促进自动驾驶汽车发展应用,有学者主张,刑法应持谦抑立场,以民事责任分配如侵权赔偿和强制保险机制作为其风险承担的主要方式。〔52〕但此种观点也有待商榷。

第一,按照有关论者的逻辑,只要某一行为可能降低相关从业人员的热情或阻碍行业发展时,便不应当成立犯罪,若将此观点推而广之,则所有行业的从业者均可以在因过失导致重大损害时以"处罚会阻碍社会发展"等理由逃避刑事处罚,如此行事不但会让法律沦为具文,令人难以接受,同时也缺乏合理的出罪依据,即论者并未说明,此做法是在哪一个阶层排除犯罪。而从前述的论证来看,制造商其实完全符合交通肇事罪的构成要件,且并没有法定的出罪依据,因此,论者的观点其实是以与案情无关的"社会影响"决定某一行为是否成立犯罪,有违《刑法》第3条前半段的明文规定。

第二,论者均是以价值判断作为论证理由,并不具有说服力。一方面,此种观点的前提在于,即使导致一定的牺牲也应当让步于科技进步。但我们同样可以认为,科技终究是为了让民众过上更好的生活,故必须在保护法益的前提下提升科技水平,进而得出与论者相悖的结论。而且从目前我国的立法、司法以及学术研究现状来看,理论与实务界似乎均更加偏向于重视刑法的预防性导向、强调刑法应当阻止危险。〔53〕就此而言,论者的前提似乎与我国目前的整体趋势相悖。另一方面,所谓"阻碍人工智能的发展"同样只是其作出的一种价值判断,在不同的思维模式下,完全可以得出相反的结论。〔54〕因此,在已经符合构成要件的情况下,理应根据刑法的基本原理让制造商承担刑事责任,而不是以其他虚无缥缈的政策甚至价值判断使其出罪,这种方案一方面更符合自动驾驶技术,尤其是高级别自动驾驶技术的设定与宗旨,另一方面有利于督促自动驾驶系统开发者与运营者履行技术安全方面的谨慎义务并鼓励其不断进行技术革新。〔55〕

综上,刑法是刑事政策不可逾越的屏障,〔56〕刑法规定的内容不能通过刑事政策或民

〔50〕 参见蔡仙:《自动驾驶中过失犯归责体系的展开》,《比较法研究》2023年第4期,第77-78页。

〔51〕 参见侯郭垒:《自动驾驶汽车风险的立法规制研究》,《法学论坛》2018年第5期,第158页。

〔52〕 参见付玉明:《自动驾驶汽车事故的刑事归责与教义展开》,《法学》2020年第9期,第140页。

〔53〕 参见张明楷:《论被允许的危险的法理》,《中国社会科学》2012年第11期,第113-114页。

〔54〕 类似观点,参见姚瑶:《人工智能时代过失犯理论的挑战与应对——以自动驾驶汽车交通肇事为例》,《浙江社会科学》2022年第12期,第90页。

〔55〕 参见王莹:《法律如何可能?——自动驾驶技术风险场景之法律透视》,《法制与社会发展》2019年第6期,第110页。

〔56〕 Vgl. v. Liszt, Strafrechtliche Aufsätze und Vorträge, Band 2 1892 bis 1904, 1905, S. 80.

事协议的方式任意创设或解除,自动驾驶汽车制造商的刑事责任也不能以刑事政策或民事责任来免除,[57]只要其违反注意义务,创设出法不容许的风险,便已然对法益损害具有预见可能性,理应承担相应的刑事责任。

四 自动驾驶汽车交通肇事归责理论之实践运用

在明确上述前提后,便可以对前文列举的两个案例逐一分析。从案情中可知,两起事故发生的原因分别在于"死者所穿的白色衣物与天空颜色一致而导致人工智能无法识别"和"车辆未能有效识别死者的肤色,误将其当成了深色路面",故应当除去事故发生的具体地点、死者的身份等与案件发生无关的因素,将场景抽象为"系统未能分辨行人与外界环境",再分析制造商对于事故原因是否具有过失。

(一)自动驾驶汽车设计缺陷的刑法归责

在案例 1 中,根据新过失论,要成立过失犯,必须对结果同时具有预见可能性与避免可能性。就本案中的预见可能性而言,我国有学者认为,汽车的传感器未能将二者区分是自动驾驶车辆本身潜在或固有的危险,本案中的危险是非常偶然地发生,是制造商、程序员等无法预见的,不应要求他们担负起注意义务而承担过失责任。[58] 但此种观点显然值得商榷。诚然,行人穿着与天空颜色相同的衣物在日常生活中可能并不常见,却也绝非难以预料,尤其是在如今多元文化的背景下,年轻人越来越潮流,衣服的色彩也越来越多样,衣着与环境极为相似之状况早已并不罕见,制造商理应注意到此类情形并在系统中设置相应程序对二者加以区分,从而避免事故的发生,故若其并未预见到"行人可能穿着与天空相同颜色的衣物"或者已经预见到却并未在系统中设置相应程序,又或者其虽然预见到,但没有安装符合国家标准的程序,都会有违结果预见/避免义务。就结果避免可能性而言,如前所述,在如今的技术限制下,自动驾驶汽车不可能如人眼般精准地识别出路上所有的行人,故若制造商能够证明,即便设置了相应程序,在相同条件下仍然难以避免碰撞结果发生,则本案便能够以不具有结果避免可能性为由出罪;与之相对,若现有的技术完全可以避免或者减少此类案件的发生,则制造商并未安装程序的行为与损害结果间仍然具有结果避免可能性,其违背义务并造成路人死亡的行为也理应成立犯罪。

在本案例的原型"Joshua Brown 案"发生后,该汽车公司专门从相关的车上收集了录像、雷达记录和声纳传感器数据,以确定出现了何种问题。利用这些数据,该汽车公司升级了自动驾驶软件,现在每辆汽车都使用改进的算法来更好地检测到运动物体。[59] 至为明显的是,任何技术的研发都绝非朝夕之功,若当真遭遇了技术瓶颈,该汽车公司不可能在如此之短的时间内取得技术突破,因此,此事故的发生并非技术所限的制造缺陷,而是因为制造商事先并未预见到此类事件,没有在软件中设置相应的识别程序所致,且这一系

〔57〕 参见王军明:《自动驾驶汽车的刑事法律适用》,《吉林大学社会科学学报》2019 年第 4 期,第 80 页。

〔58〕 参见彭文华:《自动驾驶车辆犯罪的注意义务》,《政治与法律》2018 年第 5 期,第 95 页。

〔59〕 See Jesse Krompier, Safety First: The Case for Mandatory Data Sharing as a Federal Safety Standard for Self-Driving Cars, 2 *University of Illinois Journal of Law, Technology & Policy* 439, 441 (2017).

统瑕疵是完全可以在现有的技术内通过增强测试而被发现并修复的,[60]是一种典型的设计缺陷,理应承担相应的刑事责任。

(二)自动驾驶汽车算法歧视的刑法应对

在案例2中,虽然将肤色较深的行人误认为深色马路可能确实让人难以意料,但此类行为仍然具有成立犯罪的可能性,因为本案是由制造商的算法歧视所造成的,而此种歧视是其在制造过程中可以预见且应当予以避免的。2018年,美国有学者利用人脸识别系统对来自非洲与北欧的上千名国会议员照片进行了人脸识别。结果显示,该系统对白人男性的识别准确率高达99.2%,但其错误率会随着肤色的加深而逐渐提高,尤其是对黑人女性的错误识别率甚至达到了惊人的34.7%。[61] 在论文发表后,Google公司不但没有对此种程序进行修补,反而对作者进行了内部审查,并迫使其辞职。由此可见,制造商完全能够预见到甚至已经知晓其程序中包含着对于深色皮肤行人的算法歧视。在案例2的原型案件发生后,相关研究人员对行人的图像数据进行采集并使用菲茨帕特里克皮肤分型(Fitzpatrick skin type)进行分析后发现,如今的人工智能车辆系统普遍存在算法歧视的问题,深色皮肤者[62]的检测率要比浅色皮肤者低5%,[63]这便意味着深色皮肤的民众在路上行走时,需要承担更多的风险。因此,制造商在开发此种程序时,应尽力避免算法歧视,在发现程序中包含算法歧视后,也必须极力对其进行修正。故其在能够预见甚至明知存在歧视的情况下却因疏忽大意而未加处理,或者故意对此种歧视置若罔闻,最终导致惨剧发生的,理应认定为没有履行注意义务而成立犯罪。

部分学者可能认为,在制造商认识到存在算法歧视且已经尽力减少,但仍然无法完全避免歧视的情况下,则应当将此种算法歧视视为一种容许的风险。但是,若放任内置此种程序的自动驾驶汽车自由上路,其实意味着法律在事实上默认了此种算法歧视的存在,此种默认将至少产生两个重大问题:其一,政府为何能够对部分肤色的人种提供较少保护,其背后的依据何在?此种在他人并未违反任何规范的情况下,仅仅因为肤色问题就遭到"人为歧视"的做法,非但不具有合理性,更可能有违宪法规定的平等原则。其二,若法律因为技术限制而放任内置此种程序的汽车上路,其实也昭示着立法者在此类算法歧视的问题上放弃了对处于不利地位之民众的法律保护。然而,法律具有行为指引的功能,其应当"引导人们做出内容正确的意愿",[64]它以行为人为拘束对象,尤其是在干涉他人法益的情况下,法律应介入其中,对其进行调整与规范。[65] 更何况自动驾驶汽车算法歧视中所涉及的他人利益,不仅仅是价格、财产等身外之物,而是直接关系到具有最高价值的生命法益,如果法律在此类案件中撤回了行为指引,任由内置算法歧视的自动驾驶汽车上

〔60〕 参见杨宁:《刑法介入自动驾驶技术的路径及其展开》,《中国应用法学》2019年第4期,第116-117页。

〔61〕 See Joy Buolamwini & Timnit Gebru, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, 81 *Proceedings of Machine Learning Research* 1, 1 (2018).

〔62〕 指菲茨帕特里克肤色指数在4-6范围内的人,并不需要达到黑人一般的程度,我国民众指数即在此范围之内。

〔63〕 See Benjamin Wilson, Judy Hoffman & Jamie Morgenstern, Predictive Inequity in Object Detection, [21 Feb 2019], https://arxiv.org/abs/1902.11097,最近访问时间[2024-03-18]。

〔64〕 zu Dohna, Die Rechtswidrigkeit als allgemeingültiges Merkmal im Tatbestande strafbarer Handlungen, 1905, S. 150.

〔65〕 Vgl. Rehbinder, Rechtssoziologie, 8. Aufl., 2014, Rn. 100.

路,无异于默认了其对于肤色较深之人提供较少保护的合法性。然则,国家一方面声称对生命实行比任何其他法益都更为严格的保护,另一方面却对部分并未做出任何违法举动的行人提供较少保护,二者显然有自相矛盾之嫌。〔66〕 综上,允许携带算法歧视的自动驾驶汽车上路之行为,虽然可能会带来交通上的便利,却会导致国家对民众生命的保护“区别对待”,让部分民众沦为科技发展的“弃子”,不但会使得民众完全丧失对法律的信心,更有违法治国家的基本原则,因而不宜施行。在算法歧视无法完全通过算法自身加以避免的情况下,最为妥当的方法当然是不容许此类车辆上路,并在日后立法中明确规定相关内容,否则便始终难以避免“对不同肤色国民生命权差别保护”之诟病;若有朝一日,此问题仍然难以解决,而自动汽车的上路已经势不可挡,也应当以其他方式竭力避免不均衡现象的发生,不得放任此种算法歧视发生,例如可以增设一道外置程序,在系统难以识别或存在疑虑时,先减速鸣笛,若对方无动于衷且系统仍然难以识别,则触发停车程序,并提醒乘客下车检查。唯有如此,才符合法治国家对公民生命权平等保护的基本原则。

五 结 语

未来已来,人工智能技术的快速发展使越来越多的风险暴露在公众的视野中,作为其最大的应用场景之一,自动驾驶汽车的发展在备受关注的同时,也对传统的法律体系构成了巨大的挑战。刑法作为保护法益最重要也是最后的一道防线,自然面临着最为严峻的考验,如何防范和应对自动驾驶技术带来的风险已然成为未来刑法最为重要命题之一。自动化技术一方面让驾驶汽车更加安全简便,另一方面也使得其在交通肇事时的责任认定变得更加复杂。在这一过程中,刑法当然不能成为技术进步的障碍,但更不能成为技术风险的帮凶,在保护技术创新与保护社会利益之间,刑法要做出合理的平衡。刑法可以“前瞻”以严密法网,但不能“前站”,以免侵蚀民事、行政法律规范调整的社会关系范围。〔67〕 因此,若自动驾驶汽车的制造商能够证明其事故的发生是由制造缺陷引起的,其便能够因为不具有结果避免义务而出罪;相反,若事故是由于设计缺陷造成的,则其仍然会因为违反了注意义务而成立相应犯罪。而在刑法之外的制度层面,相关部门也应当尽快出台具体可行的技术规范、行业标准等规章制度,以明确各方主体应当遵循的基本准则、能够享有的各项权利以及需要履行的基本义务,为解决事故纠纷、判断法律责任提供基本的技术依据和行业标准。唯此,才能够促进自动驾驶技术及其产业健康、稳定地可持续发展。

[本文为作者主持的 2024 年度四川省犯罪防控研究中心重点项目“行政犯时代违法性认识错误出罪路径研究”(FZFK24-03)的研究成果。]

〔66〕 参见陈璇著:《紧急权:体系构建与基本原理》,北京大学出版社 2021 年版,第 188 页。

〔67〕 参见侯帅:《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》,《中国法律评论》2019 年第 4 期,第 105 页。

Criminal Liability for Traffic Accidents Caused by Autonomous Vehicles

[**Abstract**] The causes of autonomous vehicle accidents include manufacturing defects and design defects. In accidents resulting from manufacturing defects, the deservedness of constituent elements of a crime can be eliminated by "allowable risks" on condition that the relevant production standards are met. In contrast, in accidents resulting from design defects, the manufacturer violates the duty of care and creates unallowable risks and therefore should be held criminally responsible. The driving behavior of an autonomous car is jointly determined by the driving system and the user through the built-in program and the choice of driving route. The relationship between the two is that of "cumulative competition and cooperation". Therefore, both the user and the manufacturer have a causal relationship with the damage to legal interests in a traffic accident. Based on the principle of "trust", the user has no possibility of foreseeing a self-driving car's behavior of "violating traffic regulations" and causing a traffic accident and, according to the doctrine of accountability, needs not bear criminal responsibility for the damage results in principle. However, if a user knows that a self-driving car violates transportation management regulations but still rides it on the road, thus causing a traffic accident, he should be held criminally responsible for the traffic accident. According to the relevant provisions on artificial intelligence ethics, the manufacturer has the obligation to reasonably explain the operation principle of its vehicles. If it cannot clearly explain the operation principle of the autonomous vehicle due to the "algorithmic black box", it violates the duty of care by making it impossible to determine the behavior of the system. The establishment of a negligent offense does not require the actor to recognize the specific causal process and harmful behavior, as long as it is known that his behavior has created an unacceptable risk. Therefore, as long as a manufacturer violates the duty of care and creates a risk not allowed by law, it should be regarded as being able to predict the legal interest damage caused by the vehicle and bear the corresponding criminal responsibility, and it can not be exonerated on the ground that it is unable to foresee the specific harmful behavior because of the existence of the "algorithmic black box". For criminal policy, criminal law is an insurmountable barrier and the content of criminal law provisions cannot be arbitrarily created or removed by means of criminal policy or civil agreement. The belief that imposing criminal liability on the manufacturer for traffic accidents caused by autonomous vehicles will "suppress technology" and "hinder the development of science and technology" is a pure value judgment that lacks a reasonable theoretical basis and therefore cannot be the basis for decriminalization. "Algorithmic discrimination" will turn some people into "abandoned children" of science and technology development, which will not only cause the public to completely lose confidence in law but also violate the basic concept of equal protection by law. In principle, self-driving cars should not be allowed on the road before the relevant issues are resolved.

(责任编辑:贾　元)