

• 科学技术与社会 •

# 自动驾驶无需高阶道德算法吗?

——与李大山等学者商榷

Can Autonomous Vehicles Function Without High-Order Moral Algorithms?

A Critical Discussion with Li Dashan and Colleagues

罗煜中 / LUO Yuzhong 梅其君 / MEI Qijun

(贵州大学哲学学院, 贵州贵阳, 550025)  
(School of Philosophy, Guizhou University, Guiyang, Guizhou, 550025)

**摘要:** 区分自动驾驶的一阶道德算法与高阶道德算法没有意义, 因为有关符合法律规范、满足用户需求、提高安全性能的算法不能算是真正的道德算法。道德算法对于减少自动驾驶的事故损失、明确交通事故的责任主体、增强社会对自动驾驶信任度、促进自动驾驶的普及都是必要的。自动驾驶道德算法应由政府统一规定行业标准, 即采取强制的伦理设定, 而制定何种内容的争议, 则可以通过广泛而深入的讨论予以消除。自动驾驶越“智能”, 交通肇事的归责就越前置化, 不管有没有道德算法, 自动驾驶的责任都是“先定”的, 但“责任先定”并不意味着算法供应商承担“原罪”, 也不必然导致责任归属过度决定。

**关键词:** 自动驾驶 道德算法 伦理规范

**Abstract:** It is futile to differentiate between first-order and higher-order moral algorithms in autonomous vehicles, as algorithms related to adhering to legal norms, fulfilling user needs, and enhancing safety performance cannot be deemed true moral algorithms. Moral algorithms are essential for minimizing accident losses in autonomous vehicles, clarifying accountability in traffic incidents, boosting public confidence in autonomous vehicles, and facilitating their widespread adoption. Ethical algorithms for autonomous driving should be uniformly regulated by the government to establish industry standards, that is, adopting mandatory ethical settings. Controversies over the specific content of such standards, however, can be resolved through extensive and in-depth discussions. As autonomous vehicles become increasingly “intelligent”, the attribution of risk for traffic accidents is preemptively established; regardless of the presence of moral algorithms, the responsibility for autonomous vehicles is “pre-determined”. However, this “pre-determined responsibility” does not imply that algorithm suppliers bear “original sin”, nor does it necessarily lead to an excessive determination of responsibility attribution.

**Key Words:** Autonomous vehicles; Moral algorithms; Ethical norms

中图分类号: B82-057; N031 DOI: 10.15994/j.1000-0763.2026.02.010 CSTR: 32281.14.jdn.2026.02.010

收稿日期: 2025年5月14日

作者简介: 罗煜中(1995-)男, 贵州贵阳人, 贵州大学哲学学院博士研究生, 研究方向为技术哲学。Email: 380629197@qq.com

梅其君(1972-)男, 湖南汉寿人, 贵州大学哲学学院教授, 研究方向为技术哲学与技术人类学。Email: 2320500418@qq.com

随着自动驾驶的快速发展,自动驾驶的伦理问题,尤其是自动驾驶的道德算法问题成为学术界讨论的热点。大多数学者认为应给自动驾驶植入道德算法,争论的焦点是应植入何种算法。但也有学者提出不同的看法,如李大山认为自动驾驶无需高阶道德算法。李大山在“自动驾驶无需高阶道德算法”(以下简称“李文”)一文中首先对一阶道德算法和高阶道德算法进行了区分,然后对给自动驾驶植入高阶道德算法“既不必要也不可行”进行了分析,最后提出避免道德两难的关键是改进自动驾驶技术的观点。<sup>[1]</sup>与此类似,苏令银认为,“人们不需要也不可能把伦理准则嵌入人工智能机器”,<sup>[2]</sup>隋婷婷等学者也提出“以技术手段解决伦理问题”。<sup>[3]</sup>针对这种“以技术手段解决伦理问题”的主张,我们认为有必要予以反驳。由于李文进行了“充分”的论证,具有代表性,因此,本文主要针对李文的论述进行反驳。需要指出的是,如非特别说明,本文所说的自动驾驶是指L4-L5级别的自动驾驶,即无人驾驶。因为在高级别的自动驾驶迅猛发展的当今,讨论低级别的自动驾驶意义不大,而且也只有无人驾驶把传统的驾驶者变成了乘客,从而改变了交通事故发生后的责任归属。

## 一、区分一阶道德算法和高阶道德算法是否有意义

李文首先对一阶道德算法和高阶道德算法进行了区分。所谓一阶道德算法是指实现“如识别红绿灯、礼让行人、礼让救护车、不得占用应急车道等”功能而“负载特定价值取向”的算法,而所谓高阶道德算法是指“能够重新定义行动者的算法”。<sup>[1]</sup>但这种区分并不成功。李文所谓的一阶道德算法虽然也涉及减少事故和损失的伦理因素,但与道德决策基本上没有关系,不属于道德算法的范畴,或者说不是学术界所讨论的道德算法。

首先,李文所谓的一阶道德算法虽然也与道德相关,但基本上都属于法律规范或市场的要求。李文所说的“识别红绿灯、礼让行人、

礼让救护车、不得占用应急车道”,一般的道路交通安全法都会对此作出明确的要求,驾驶者必须严格遵守。虽然实现这些功能“负载特定价值取向”,但与复杂的道德决策并没有关系。这主要是法律规范层面的要求,而不是伦理道德层面的考量。没有这一基本能力,自动驾驶汽车就无法合法上路,更不用说获得消费者的信任。真正的道德决策往往涉及复杂的情境和价值判断,如在紧急情况下如何权衡保护乘客与行人之间的利益,这就超出了一般道路交通安全法范畴。李文所说的要求车辆“以最短时间将用户送至目的地”,满足“过程平稳”“内部空间宽敞”“内饰高档”“车载应用丰富易用”等,<sup>[1]</sup>这些都是旨在提高效率和用户满意度,满足市场的需求,与道德决策无关,实际上也不属于道德算法的范畴。

其次,李文试图以丽贝卡·德文纳尔(Rebecca Davnall)的“刹车直行优先”理论为例来说明一阶道德算法与高阶道德算法的区别。“刹车直行优先”有两个理由:一是在制动力学层面减少事故概率;二是“自动驾驶系统缺乏必要的信息来判断急转弯后带来的损失一定小于直行”。<sup>[1]</sup>李文认为,这两个理由都是基于安全的考虑,因而该理论具有一阶道德算法的特征。但李文又说,“德文纳尔的理论仍然有不少高阶道德算法的痕迹”,如“是否应该增加与前方车辆发生事故的概率来降低与尾随车辆发生事故的概率”。<sup>[1]</sup>

那么,按照李文的对一阶道德算法和高阶道德算法的区分,德文纳尔的“刹车直行优先”理论究竟属于一阶道德算法,还是属于高阶道德算法呢?可见,李文对一阶道德算法和高阶道德算法的区分并不成功。李文又列举了芥川等人的轮胎扰动控制算法,“该算法致力于消除轮胎的外部噪音,抑制轮胎与路面产生的滑移与微震,最终达到改善汽车转向时轮胎稳定性的目的。”<sup>[1]</sup>该算法同样是聚焦于自动驾驶的安全性能,是一种纯技术的改进。将这样的算法归为道德算法有什么意义呢?一种算法之所以被称为“道德算法”,不仅因为它关涉伦理道德,符合伦理的要求,更重要的是它要么

作为人类道德决策的助手，能够帮助人们作出符合伦理要求的决策，要么作为独立的道德主体，能够分辨善恶，自主地作出道德决策。

把符合法律规范、满足用户需求、提高安全性能所作的算法上的改进称为道德算法，是对道德算法这一概念的误用。学术界讨论的道德算法一般都是指李文所说的高阶道德算法，也只有在高阶道德算法问题上才有争议，而所谓的一阶道德算法，不论是追求安全，还是效率和舒适，基本上都没有什么争议。可见，李文对一阶道德算法与高阶道德算法的区分并没有意义。

为避免歧义，下文继续沿用李文的高阶道德算法这一说法，但下文所说的高阶道德算法就是学术界所讨论的道德算法，亦即，下文不再对道德算法进行一阶和高阶的区分。

## 二、高阶道德算法是否必要

李文认为，高阶道德算法是不必要的。李文列举了支持高阶道德算法的三个常见理由，然后逐一进行了反驳，但李文反驳的理由并不充分。

李文列举支持高阶道德算法的第一个理由是，“高阶道德算法虽然不能避免但能降低损失”。<sup>[1]</sup>李文这样反驳：“产业界可以研发更敏锐的感知系统，通过算法优化获得更节约资源的计算能力，通过性能更强大的芯片提升数据处理能力，从而降低自动驾驶发生事故的概率”，“间接体现了最小值的最大化、帕累托最优等原则”。<sup>[1]</sup>不可否认，开发更敏锐的感知系统和提升数据处理能力，固然可以减少事故发生，降低损失，但技术开发具有不确定性，开发时间长短难以预测。即使开发出了更敏锐的感知系统，也只能降低事故发生的概率，而不可能完全避免事故的发生，尤其是在自动驾驶完全取代人类驾驶之前，必然有一个自动驾驶和人类驾驶并行的过渡阶段，这一过渡阶段可能是自动驾驶交通事故的高发期。即使将来自动驾驶完全取代了人类驾驶，有些事故仍然不可避免，如机械故障、制动失灵或路人、动

物等不可预测的行为。李文将制动系统失灵作为自动驾驶道德两难产生的前提也是片面的。制动系统失灵只是自动驾驶道德两难产生的原因之一。

在实际驾驶过程中，造成自动驾驶道德两难的情况非常复杂，如自动驾驶车辆迎面驶来一辆失控的重型卡车。在不可避免发生事故的情况下，自动驾驶车辆仍然需要在多个不理想的选择中作出决策，单纯依赖技术优化而不考虑道德决策的系统可能会导致难以预料甚至不公正的后果。如果自动驾驶车辆因不预先考虑在多个不理想的选择中作出决策——这种不作为也会被认为是一种事先确定的选择——而造成了不公正的后果，其存在的合法性就会受到质疑。道德算法能够为自动驾驶系统提供一个在紧急情况下进行公平决策的框架，确保在无法完全避免损失时作出的决策符合社会的伦理标准。忽视道德决策可能会导致系统在关键时刻缺乏必要的伦理指引，从而无法真正实现对公众安全的保障。

李文列举支持高阶道德算法的第二个理由是，“高阶道德算法使得损失以一种人类可接受的方式发生，照顾了人类情绪”。<sup>[1]</sup>对于第二个理由，李文是这样反驳的：如果没有道德算法，自动驾驶面临道德两难时只能随机选择，但人类驾驶面临道德两难也是随机选择，而且有些道德算法也并没有带来确定性，因此，给自动驾驶植入道德算法就没有必要。不可否认，人类驾驶员面临突发的道德两难时往往来不及思考，所做出的行为因个体化和情绪化而具有不确定性，但人类驾驶员的这种随机性是可以理解的，而且人类驾驶员要对他的选择进行负责。如果自动驾驶也像人类驾驶员那样随机选择，至少不符合人们对它的期待。自动驾驶代替人类驾驶的一个最重要的理由就是可以减少事故，增加确定性，而不仅仅是减少人类的劳动。

李文以罗尔斯式算法为例来说明某些道德算法并没有带来确定性，也没有说服力。李文所列举的是非常极端的情况——“两个车辆行为的最小值相等且没有其他可比较的效用”。<sup>[1]</sup>



在这种情况下,系统只能作随机选择。但这种极端的情况非常少,而且通过在实际应用中不断反馈,道德算法可以不断优化和调整,进一步减少这种需要随机选择的情况。即使是在这种极端情况下,系统作随机选择也是可以理解和接受的。

李文列举支持高阶道德算法的第三个理由是,高阶道德算法“有助于界定交通事故的责任主体”。<sup>[1]</sup>对于第三个理由,李文从两个方面进行了反驳:一是近六年(截止2022年)报道的自动驾驶交通事故的责任归属都比较清楚,因此无需高阶道德算法;二是发达国家对自动驾驶的事故责任归属问题进行了探索,“暂未看到要求植入高阶道德算法的法律法规实践”,因此,“缺乏高阶道德算法会导致责任归属危机过度忧虑了”。<sup>[1]</sup>

关于第一个方面,李文反驳的依据是近六年报道的自动驾驶交通事故,但近六年的民用自动驾驶尚处于L2-L3阶段,是还需要人类驾驶员的低级阶段的自动驾驶。李文列举的一类交通事故就是由于驾驶员未严格按照操作手册进行操作而导致的。一般道德算法所说的自动驾驶都是指L4-L5级别的自动驾驶,即无人驾驶。低级阶段的自动驾驶最终都会被无人驾驶所淘汰,而且淘汰的速度远超出公众的预期。李文以低级阶段的自动驾驶的交通事故数据为依据来反驳高级阶段的自动驾驶的需要显然不合逻辑。L2-L3阶段的自动驾驶数据虽然对L4-L5级别的自动驾驶有一定的参考价值,但L4-L5级别的自动驾驶车辆已完全接管驾驶任务,无需人类驾驶员的介入,这与L2-L3阶段的自动驾驶有本质的区别。如果不事先考虑可能发生的交通事故的责任归属,一旦发生排除路人或其他车辆违规的交通事故,如何明确事故责任呢?

关于第二个方面,李文显然忽视了法律的回应性和滞后性。科学技术的快速发展导致许多新问题产生,但法律不可能很快作出相应的调整 and 应对,因为法律需要保持相对的稳定性,以确保其权威性和可预测性,而且法律的制定需要稳定的社会预期和足够从中提炼刚性的法

律规则的伦理资源,<sup>[4]</sup>其过程通常较为繁琐和缓慢。道德算法涉及深层次的伦理问题,平衡不同的道德原则需要政策制定者、技术开发者、伦理学家和公众的广泛讨论和研究。只有在形成广泛的社会共识后,才能制定相应的法律法规。而且,法律制定还需考虑国际合作与协调。自动驾驶技术和相关道德算法问题是全球性挑战,不同国家在道德观念、技术水平和法律体系上的差异,需要国际间的沟通与合作。这一过程也是复杂而缓慢。因此,李文所说的没有看到“高阶道德算法的法律法规实践”也就很正常了。相反,如果没有经过充分的酝酿,过早地制定法律规定很可能会阻碍技术的发展。

事实上,目前很多国家都在尝试将道德算法写入法律。德国2021年制定了《自动驾驶法》。该法对L4级自动驾驶明确规定,“在无法避免的损害中,面对不同法益的损害,应考虑到不同法益的意义,并以对人的生命的保护为最优先”;“在对人类生命的危害无法避免时,不得根据个人特征预先规定其他的权重”。<sup>[5]</sup>虽然很多条款还有待进一步明确,但开启了道德算法的法律法规实践。

在法学界关于自动驾驶讨论中,很多学者也主张构建道德算法。郑玉双认为,“解决自动驾驶的伦理和法律挑战需要重建算法正义观,并构建出符合算法正义观的自动驾驶道德算法”,并进一步提出,“基于算法正义理念,最大化最小值算法是克服人机协作之技术困境并解决碰撞难题的较佳方案。”<sup>[6]</sup>储陈城从法理上详细地论证了自动汽车程序设计中解决“电车难题”的刑法正当性,即自动驾驶的程序设计如果满足“被允许的危险的构成要件”,“可以直接在构成要件该当性阶段被排除归责”。<sup>[7]</sup>这说明,预设道德算法是法律实践之所需,而非“过度忧虑”。

上面对李文反驳高阶道德算法的必要性的理由进行了逐一反驳,事实上,除了李文所列举的支持高阶道德算法的三个常见理由之外,支持高阶道德算法还有一个重要的理由,即增强社会对自动驾驶的信任度,促进自动驾驶的普及。道德算法不仅仅是技术上的提升,也是

一种社会契约的体现。当公众看到自动驾驶汽车能够在复杂的道德困境中作出符合社会期望的决策,他们对这种新技术的接受度和信任度将极大提升。社会信任度的增强不仅有助于技术的推广和普及,也能够减少因不信任而导致的社会成本和阻力。

如果认为道德算法不必要并简单地予以放弃,当遇到不可避免的碰撞时,自动驾驶只能作随机选择。这种缺乏明确方向的决策方式不仅可能导致不道德行为的发生,乃至触碰法律的所维护的底线,而且会因其丧失对行为的指引功能和激励作用而导致个人权利空间被压缩。<sup>[8]</sup>“机器的自由度越大,它就越需要道德标准”。<sup>[9]</sup>正如白惠仁指出,“对于当下自动驾驶汽车所面临道德困境的可能解决方式是为其预设‘道德算法’。”<sup>[10]</sup>只有通过建立明确的道德框架,才能确保在复杂的道德决策中,自动驾驶能够作出符合社会伦理标准的选择,从而不仅提升技术的合理性与透明度,也为技术的未来发展奠定更加稳固的社会基础。

### 三、高阶道德算法是否可行

李文不仅认为高阶道德算法“不必要”,而且认为它“不可行”。李文列举了植入高阶道德算法不可行的四点理由:“一是由谁制定有争议;二是制定何种内容有争议;三是导致算法供应商负有‘原罪’;四是导致责任归属过度决定。”<sup>[1]</sup>对李文的这四点理由,下面逐一剖析。

首先,关于由谁制定的争议问题,在自动驾驶道德算法的讨论中,道德算法由谁制定确实存在两种观点。一种观点认为,道德算法由供应商提供,让消费者选择,即所谓个人化的伦理设定(personal ethics setting)。另一种观点则认为,道德算法应由政府统一规定行业标准,即强制的伦理设定(mandatory ethics setting)。但是,由谁制定的争议并非不能消除。经过广泛而深入的讨论,目前学术界逐渐形成共识,即自动驾驶道德算法只能是强制的伦理设定。因为个人化的伦理设定,不仅会使得自动驾驶的行动模式缺乏一致的协调性,从而降

低整个行业的效率和安全价值,<sup>[11]</sup>而且会造成更多的道德问题,给个体带来沉重的负担,最终必然导致“囚徒困境”。<sup>[12]</sup>政府作为公共利益的代理人和决策者,能客观公正地引领目前存在的法律伦理困境的解决,促进法律和伦理的协同。<sup>[13]</sup>虽然强制的伦理设定限制了理性个体的选择,也可能造成偏见和歧视,但可以通过适度调整,如加入情境归责等办法进行弥补。<sup>[14]</sup>

其次,关于制定何种内容的争议问题,在自动驾驶的具体算法内容上,代表性的算法有利己主义算法、功利主义算法、罗尔斯式算法、机器学习和数据驱动算法等。各种算法既有其合理性,也有不足,自然也就有争议。争论的焦点主要集中在:(1)在不可避免的碰撞事故中,当生命冲突处于乘客和车外人员之间时,是否需要将乘客置于优先获得保护的地位,即是否遵循乘客优先的原则?(2)在不可避免的碰撞事故中,当生命冲突处于车外人员之间或乘客与车外人员之间时,是遵循旨在保全多数人生命的多数原则,还是遵循听天由命的随机数原则?上述争议并非“无解”。我们认为,只要确立几条基本的原则以及这些原则之间的优先顺序,这些争议可以消除。按优先顺序,这些需要确立的基本原则依次是:(1)生命权平等原则;(2)遵守交通法规优先原则;(3)最小损害原则;(4)不改变路线原则。其中,生命权平等原则是最基本的原则,优先于其它原则,即在坚持生命权平等的原则下适用遵守交通法规优先原则。同理,如果都遵守交通法规,则适用最小损害原则;如果损害相等,则适用不改变路线原则。下面我们用这些基本原则对上述争论的焦点进行分析。

关于第一个争论,主张乘客优先原则的理由主要有两点:一是“市场论证”,即只有乘客优先,才能促进自动驾驶的发展,而自动驾驶的发展能减少交通事故,减少交通事故蕴含着保护生命的道德价值;二是“接管论证”,由于自动驾驶把驾驶员变成了乘客,汽车制造商在接管驾驶活动的同时,也接管了乘客的生命健康权,因此有保护乘客生命健康权的义务。<sup>[15]</sup>



乘客优先原则的两个论证都存在缺陷。优先保护乘客不仅可能会因受其他交通参与者的抵制而阻碍自动驾驶的发展,也会因违背了技术发展的初衷而妨碍造福全社会目标的实现,甚至造成意料之外的恶性竞争。<sup>[16]</sup>“接管论证”也与搭车原则相矛盾,“会导致车主乘客以其他乘客的生命作为保障自己安全的工具”。<sup>[17]</sup>乘客优先原则意味着,当出现生命冲突时,保护乘客是以牺牲无辜的车外人员为代价的,或者说是潜在地把无辜的车外人员当作工具。这不符合生命权平等原则,违背了公平正义的要求,无论是市场论证,还是接管论证,都不足以为此辩护。人类驾驶员在紧急情况下,如躲避突然冲来的失控车辆,会本能地猛打方向盘而因此撞到行人,一般可以适用紧急避险而阻却法律责任,但这与作为一般行为准则的事先设计乘客优先的编程有着本质的区别,因为设计程序不存在危险的迫切性,因而无法通过紧急避险来免责。

关于第二个争论,随机数原则主张通过随机数生成器来决定保全哪些生命,其理由是生命拥有绝对最高价值,不能用数量进行衡量。<sup>[15]</sup>选择随机数原则是一种不严肃、不负责任的行为,违背了人类理性的要求,也不符合一般公众的道德直觉。多数原则,即最小损害原则是在伤害不可避免的情况下,尽可能挽救多数人的生命,实现更大的善。它没有义务论者所批评的“制造伤害”(doing harm)的主观意图,也没有造成更坏的结果。最小损害原则以生命权平等原则为前提,并不违反“人是目的”原则。最小损害原则也是以遵守交通法规优先原则为前提的,如果只能在撞向不遵守交通法规的多数人和撞向遵守交通法规的少数人之间选择,则选择撞向不遵守交通法规的多数人。因为遵守交通法规优先原则,不仅是对无过错的一方优先保护,符合公平原则,而且是对守法行为的一种激励。以生命权平等原则和遵守交通法规优先原则为前提的最小损害原则实际上是一种以义务论进行调试的规则功利主义,其对普适规则的追求,本质上与义务论相同。

再次,关于高阶道德算法是否导致算法供

应商负有“原罪”和责任归属过度决定的问题,李文认为,高阶道德算法必然导致“责任先定”,从而导致责任归属“过度决定”。在法学界,关于人工智能在交通事故中扮演的角色也有“主体肯定说”与“主体否定说”之争,“主体肯定说”认为人工智能可以承担刑事责任,而“主体否定说”则认为人工智能不能承担刑事责任。如果按照“主体肯定说”的观点,人工智能可以承担刑事责任,不存在“责任先定”。如果按照“主体否定说”的观点,交通事故的责任由生产者和使用者承担,其中使用者主要承担不进行日常检查和风险制止的责任,<sup>[18]</sup>其它责任则由生产者承担。自动驾驶越“智能”,使用者的责任就越小,而生产者的责任就越大。生产者生产自动驾驶车辆的过程包括但不限于算法设置,而生产者的生产风险控制义务是先期的、前瞻性的。<sup>[19]</sup>质言之,“责任先定”是自动驾驶区别于人类驾驶的一个特征,自动驾驶越“智能”,交通事故的风险归责就越被前置化,不管有没有道德算法,责任都是“先定”的。

“责任先定”并不意味着算法供应商承担“原罪”,算法供应商虽然提供算法,但他们并不是“算法决定者”。“碰撞算法的设计是程序员将制度化的社会道德共识或偏好转化为代码和决策机制”,而包括算法供应商在内的生产者“并不是决定算法之道德立场和法律责任形态的适格主体”。<sup>[6]</sup>这就是说,道德算法的设定一定是包括生产者、使用者、管理者乃至路人等利益相关的多方主体反复讨论、协商的结果,是体现正义的透明的社会决策。如果道德算法是由公民投票通过或者是在广泛民意的基础上由职能部门决定,那么算法供应商这个“原罪”也就自然消解了,也就不存在所谓的责任归属“过度决定”。

## 余 论

“以技术手段解决伦理问题”的观点隐含的逻辑前提是:技术是价值中立的。当代技术哲学研究表明,技术并非价值中立,而是负荷

价值、内嵌道德的。自动驾驶的算法不是价值中立的,算法本身存在偏见,而且偏见无法完全消除。算法设计者的主观选择和价值判断不可能绝对中立,“算法设计者本身的价值观念也可能被带入学习模型构建和程序代码编译的过程。”<sup>[19]</sup>

例如,自动驾驶总要考虑一些无法避免的事故,也就必然要考虑碰撞的决策模型,设计碰撞算法,而碰撞算法涉及碰撞中的每个个体,它如何对待这些个体?它对待这些个体的方式是否体现正义?质言之,碰撞的决策模型必然预设着特定的伦理和法律立场。这就超出了纯粹的技术范畴,正如王天恩指出,“智能算法发展到一定程度,算法本身就会逐渐拥有伦理属性”。<sup>[20]</sup>而且,算法与其支持公司的利益是紧密关联的,在某种程度上,算法是利益垄断的产物,也是利益分配的工具。

“以技术手段解决伦理问题”的主张不是积极地解决问题,而是回避问题,而自动驾驶的快速发展已不允许回避问题。直面问题不妨从讨论现实责任入手,因为自动驾驶对现行的相关法律制度造成了冲击,“自动驾驶汽车导致事故后,对具体的自动驾驶风险仅作概括性地描述,并不符合汽车运行的实际情况,而后续的自动驾驶汽车的法律规制体系存在整体性的缺失,将会导致对自动驾驶汽车的监管陷入无序化的尴尬境地。”<sup>[21]</sup>既然自动驾驶不可能完全避免交通事故,而交通事故发生后的责任又必须追究,那么构建自动驾驶法律责任体系就势在必行。自动驾驶技术的核心在于高度智能化的系统,需要具备复杂的感知、决策和执行能力,以应对各种路况和突发情况,这些都依赖于复杂的算法设计。

由于算法在自动驾驶中的核心地位,构建自动驾驶法律责任体系就不可避免地要探讨自动驾驶的算法正义问题。这是因为,“自动驾驶算法致损的责任应由谁来承担,以及在碰撞困境中谁应该承受损失等问题,都是正义理念要解决的”。<sup>[6]</sup>而探讨自动驾驶的算法正义问题,也就不可避免地要涉及道德算法。“如何设定体现‘善’的程序,归根结底取决于人自

身的价值取向,不能寄望于技术的完善。”<sup>[22]</sup>道德算法有助于界定自动驾驶交通事故的责任归属和形成明确的法律框架。通过在算法中植入道德决策模型,道德算法可以更好地保障在紧急情况下车辆的决策透明性,从而为法律责任的分配提供明确的依据。至于应用何种道德算法则不仅需要从理论上予以融贯的阐明,而且需要利益相关主体进行商谈,达成共识并形成契约,体现分配正义。基于道德算法所作出的选择是在不得已的情况下所作出的优先性选择,它不可避免地留有道德遗憾,面临道德质疑。<sup>[22]</sup>面对技术特别是人工智能的飞速发展,人类的道德实践不得不作出调整。人工智能极大地增强了人类改造社会的能力,但人类也要为这种能力的应用承担更大的责任,而不是听之任之。

#### [参考文献]

- [1] 李大山. 自动驾驶无需高阶道德算法[J]. 自然辩证法通讯, 2023, 45(6): 10-17.
- [2] 苏令银. 能将伦理准则嵌入人工智能机器吗?——以无人驾驶汽车为例[J]. 理论探索, 2018, (3): 38-45.
- [3] 隋婷婷、郭晓. 自动驾驶电车难题的伦理算法研究[J]. 自然辩证法通讯, 2020, 42(10): 85-90.
- [4] 郑戈. 电车难题与自动驾驶系统算法设计中的伦理考量[J]. 浙江社会科学, 2022, (12): 37-47.
- [5] 张韬略、钱榕. 迈入无人驾驶时代的德国道路交通法——德国《自动驾驶法》的探索与启示[J]. 德国研究, 2022, 27(1): 85-101.
- [6] 郑玉双. 自动驾驶的算法正义与法律责任体系[J]. 法制与社会发展, 2022, 28(4): 145-161.
- [7] 储陈城. 自动驾驶汽车设计中解决“电车难题”的刑法正当性[J]. 环球法律评论, 2018, 40(3): 82-99.
- [8] 于涵. 自动驾驶汽车的道德算法设计——一种风险分配立场[J]. 天府新论, 2021, (4): 101-108.
- [9] Wallach, W., Allen, C. *Moral Machines: Teaching Robots Right from Wrong*[M]. Oxford: Oxford University Press, 2009, 23.
- [10] 白惠仁. 自动驾驶汽车的“道德算法”困境[J]. 科学学研究, 2019, 37(1): 18-24.
- [11] Belay, N. 'Robot Ethics and Self-Driving Cars: How Ethical Determinations in Software Will Require a New Legal Framework'[J]. *Journal of the Legal Profession*, 2015, 40(1): 119-130.

- [12] Gogoll, J., Müller, J. 'Autonomous Cars: In Favor of a Mandatory Ethics Setting' [J]. *Science and Engineering Ethics*, 2017, 23(3): 681-700.
- [13] 李飞. 无人驾驶碰撞算法的伦理立场与法律治理 [J]. 法制与社会发展, 2019, 25 ( 5 ) : 167-187.
- [14] 孙保学. 自动驾驶汽车事故的道德算法由谁来决定 [J]. 伦理学研究, 2018, ( 2 ) : 97-101.
- [15] 姚沅辰. 乘客优先与随机数原则: 自动驾驶情形下的生命冲突的破解逻辑 [J]. 法制与社会发展, 2024, 30 ( 3 ) : 205-224.
- [16] 郑志峰. 自动驾驶汽车“电车难题”的伦理困境与立法因应 [J]. 法学, 2024, ( 11 ) : 107-123.
- [17] 陈景辉. 自动驾驶与乘客优先 [J]. 华东政法大学学报, 2020, 23 ( 6 ) : 6-19.
- [18] 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 [J]. 学术交流, 2018, ( 4 ) : 73-80.
- [19] 陈禹衡. 算法优化语境下自动驾驶汽车交通肇事的困境纾解 [J]. 苏州大学学报 ( 法学版 ), 2021, 8 ( 3 ) : 60-73.
- [20] 王天恩. 智能算法规则规律一体化的伦理意蕴 [J]. 学术界, 2020, ( 7 ) : 42-50.
- [21] 刘艳红. 自动驾驶的风险类型与法律规制 [J]. 国家检察官学院学报, 2024, 32 ( 1 ) : 114-130.
- [22] 谢惠媛. 民用无人驾驶技术的伦理反思——以无人驾驶汽车为例 [J]. 自然辩证法研究, 2017, 33 ( 9 ) : 39-43.
- [责任编辑 李斌]