

走向“有意义的人类控制”

——自动驾驶汽车的主体悬置与责任锚定

刘成科^{1,2},葛燕¹

(1.安徽农业大学马克思主义学院,安徽合肥230036;2.安徽农业大学外国语学院,安徽合肥230036)

摘要:自动驾驶作为一个动态发展的技术连续体,在具体应用场景中呈现出从人类主导过渡到人机协同再到机器主导等不同人机关系样态,其自主性与人类控制权之争引发了各种伦理、法律和社会等问题。人机互动关系演变背后折射的是人类控制形式的不断进化,从本质上看,自动驾驶汽车的责任鸿沟问题与不同技术嵌入层级下的人类控制密切相关。源于应对人工智能伦理困境的“有意义的人类控制”,可作为解决自动驾驶汽车责任鸿沟问题的一种全新技术治理理念。借助“追踪-回溯”双向路径设计,实现对多元责任主体的锚定。在“有意义的人类控制”框架下,将自动驾驶置于技术层级与人类控制的关系界面之上,坚持以科学的数据系统为支撑,以责任锚定为导向,以人类控制为中心,通过伦理与法律的整合进路,并结合完善的保险制度,破解自动驾驶主体悬置的困境。未来,应加强自动驾驶的跨学科交叉研究,尤其需要借助科技与人文的常态化对话机制,强化各利益相关方的伦理自觉和法律认知,摒弃简单的技术设计视角,实现“有意义的人类控制”转向,从而进一步推动自动驾驶技术的发展。

关键词:自动驾驶;主体悬置;责任锚定;技术嵌入;有意义的人类控制

中图分类号:N031

文献标识码:A

文章编号:1008-7699(2023)01-0027-08

作为一种颠覆性的技术,人工智能的快速发展正在深刻影响着众多领域。其中,自动驾驶是传统驾驶革命性的飞跃,更是公众关注的焦点,备受期待。与传统汽车相比,自动驾驶汽车能够有效减少事故发生率,使得人类在驾驶汽车的过程中更安全、更轻松。^[1]虽然自动驾驶的优势非常显见,但是经典的“电车难题”一直是无法回避的道德决策困境。如今,我们已不只是局限在思想实验中考量自动驾驶的伦理问题,现实生活中相关案例也不断出现。^①不仅是技术安全隐患的议题,更有事故责任的归属问题。当前,各大车企和互联网公司开始加大对自动驾驶汽车的研发力度,而自动驾驶汽车的伦理解决方案迫切需要贯穿于技术研究的全过程。

对此,学界从主体缺位和过错标准认定困难、从中短期和长期两个技术阶段的责任划分,以及结合自动驾驶的程序原理和级别设定考量刑事归责等不同角度出发,对自动驾驶汽车可能会出现的问题进行了讨论。^[2-4]不言而喻,自动化程度对于准确把握自动驾驶汽车事故的责任分析至关重要,其本质关切是自动驾驶中的人类控制问题。可遗憾的是,学者在人类控制的框架下探讨自动化程度与事故责任归因的研究相对较少,也未能基于智能机器与人类控制的关系界面思考责任主体的锚定问题。^②

收稿日期:2022-08-03

基金项目:安徽省哲学社会科学规划一般项目“人工智能时代翻译技术的伦理治理研究”(AHSKY2020D134)

作者简介:刘成科(1979—),男,安徽合肥人,安徽农业大学马克思主义学院/外国语学院特任教授,哲学博士;葛燕(1999—),女,安徽肥西人,安徽农业大学马克思主义学院硕士研究生。

① 2016年5月,美国佛罗里达州发生了一场因特斯拉电动汽车在自动驾驶模式下导致撞车事故。2018年3月,亚利桑那州更是发生了无人驾驶测试车辆致人死亡的事故。2023年2月,美国国家公路交通安全管理局表示,由于“全自动驾驶”测试软件可能会增加撞车风险,特斯拉将在美国召回36.2万辆汽车,再次引发了各界对自动驾驶汽车安全问题的关注。

② 近年来,在人工智能治理理念上,学界对价值敏感性设计、负责任创新等问题的探讨仍然占据主流地位。

因此,本文拟在分析自动驾驶事故责任主体的基础上,结合自动驾驶汽车的自动化等级,从“有意义的人类控制(Meaningful Human Control, MHC)”^①的视角,^[5]审视人机交互复杂情境下可能事故的责任锚定,尝试为自动驾驶汽车的责任处理提出相应的建议。

一、主体悬置的理论困境与实践探索

从功利视角来看,智能机器要具备主体地位,必须满足目的正当性和手段適切性等条件,符合工具理性的要求。目前,仅仅出于应对事故归责的目的,并不能简单地将作为智能机器的自动驾驶汽车赋予主体地位,使其独立承担相应的责任。

通常,传统汽车发生交通事故时,主体界定一般较为清晰,我国法律对此也有相应的条款可以适用,责任划分并非难事。然而,自动驾驶汽车出现事故后,现有道德和法律框架显然已不能解释和判定责任的归属。那么,究竟谁可以成为自动驾驶汽车的责任主体呢?自动驾驶车辆的驾驶员,还是自动驾驶汽车本身,抑或相关的生产厂商?答案众说纷纭,莫衷一是,出现了谁都似乎有责任而谁都不愿承担责任的“主体悬置”现象,进而产生了一直困扰我们的“责任鸿沟”^②问题。

需要说明的是,此处的“责任主体”是一个更上位的概念,既可以指代“道德主体”,对应的是伦理责任;又可表示“法律主体”,对应于法律责任。根据自动驾驶汽车事故发生的具体情境,再去确定应诉诸道德主体抑或法律主体的阐释框架。

自动驾驶汽车能否作为独立责任主体的问题,学术界争论已久。实质上,这关涉的是机器智能体作为道德主体的要件问题,而具有意识、能够进行思考和判断是成为道德主体的必要条件。可见,道德或伦理主体必须能独立控制自己的意志,并能清楚知晓自身行为,也可以充分预见其可能产生的责任或后果。但是,道德主体也经历着不断的变迁,已经从具有自由意志的人类拓展到一切生命体,进而再由生命体延伸至技术人工物,最后拓展到智能人工物。^[6]据此,自动驾驶汽车作为一种智能人工物,理论上也可以被视为某种“道德主体”。

但是,是否真正需要在实践中赋予自动驾驶汽车独立的道德地位呢?首先,目前人工智能尚未达到预想的强人工智能甚至超人工智能状态,还不足以独立承担责任。有学者基于费希尔道德责任理论,主要从引导性和管制性控制两个方面进行了分析,认为最终还是应该由人来承担相应的责任,而不是机器智能体;^[7]其次,从道德主体的角度来看,即使机器智能体具备道德主体的地位,且具备一定的“自主性”,但需要谨慎的是,“机器自主性不等于道德自主性”,^[8]不能把二者简单等同起来。从某种程度上说,自动驾驶汽车虽具有一定的机器自主性,但终究因为无法承担责任而不能成为道德谴责或法律追偿的对象。当下伦理进行行为举止规范的对象只是自然人,对于人工智能体基本没有约束力。最后,即便人工智能具有道德地位,机器权利和义务的履行方式也无法开展具体有效的监督,进而又会带来一系列新的伦理问题。显然,将自动驾驶汽车视为道德主体并不能解决主体悬置的困境。

此外,从法律承担角度来说,赋予智能机器法律人格以此期望达到限定责任的作用,并不符合正义的要求,因为智能机器人无法响应法律所要求的行止,无法理解和接受法律的调整并且财产对其也是没有意义的。不能将智能机器与法人相类比,因为自然人与法人之间存在着特殊的联结意义,这种关系是无法被忽视的,也是智能机器所无法代替的。因此,从理论上讲,赋予智能机器以法律人格是无法成立的。^[9]倘若只是简单地希望智能机器为所发生的意外事故承担责任,便赋予其法律人格,这是没有必要

^① 目前,国内学界对“有意义的人类控制”理念的探讨尚处于起步阶段,散见于哲学和法学等不同学科领域。已有研究将其直译为“有意义的人类控制”,实际上,该理念意在实现人类对自动驾驶汽车等智能机器的有效控制,进而保证后续相关责任的明晰。所以,笔者认为,该术语的翻译还有待进一步商榷,也可尝试译为“有效人类控制”。为了保持与以往研究术语上的一致性,本文仍然采用“有意义的人类控制”的译法。参见刘永安:《在有意义的人类控制下发展负责任的人工智能》,《自然辩证法通讯》2022年第9期。

^② “责任鸿沟(responsibility gap)”也可译为“责任缺口”。

的,而且在法律上也缺乏法理基础。

有学者指出,如果未来时机成熟,可以尝试赋予智能机器人法律地位,从而让其承担相应的法律责任。^[10]的确,近年来尝试赋予人工智能独立人格的实践也从未中断。^①从现实案例中可以看出,未来自动驾驶汽车被赋予独立的人格似乎也符合历史发展的规律。^[11]不过,一些国家和地区虽然已经赋予人工智能机器所谓的“独立人格”,但是这些实践当前都只是处于尝试摸索阶段,并没有大规模地推行,也没有这些被赋予人格地位的机器人应用和发展的具体图景。

即便如此,在未来很长一段时间内,仍然将由人作为道德主体或法律主体,而不是人工智能本身。换言之,主体(如自动驾驶汽车)将以某种形式隐身于具有自由意志、理性能力的自然人之中。不过,既然最终都将由人类作为伦理或法律主体,在复杂的现实交通事故中,又该由哪些人为自动驾驶汽车负责?如何应对自动驾驶汽车责任鸿沟带来的挑战?

下面基于技术嵌入程度的差异性,即自动驾驶汽车的分级体系,具体讨论自动驾驶汽车事故中的责任锚定问题。

二、从技术嵌入到“有意义的人类控制”

(一)技术嵌入是一个复杂动态系统

目前自动驾驶汽车是一个较为广义的概念,容易给大众带来一定程度的认知偏差。其实,根据人工智能技术的嵌入程度,自动驾驶汽车在实践中的智能化水平差异巨大,有的仅仅属于辅助驾驶的范畴,而有的却已接近真正意义上的无人驾驶。

2021年4月,国际自动机工程师学会(SAE International)和国际标准组织(ISO)联手发布了更新版《SAE汽车驾驶自动化分级(SAE Levels of Driving Automation)》。当前,SAE标准J3016依旧是全球汽车行业公认的自动驾驶系统等级的划分标准,对自动驾驶水平的分类和定义作出详细的解释。^[12]按照J3016的标准,明确了SAE六个不同层级的支撑条件及定义,从零级(L0)无自动驾驶到五级(L5)完全自动驾驶,自动化程度逐步提升,按照道路机动车辆及其运行条件,划分至相应的层级。^[13]

我国市场监管总局(标准委)发布的《汽车驾驶自动化分级》国家标准也于2022年3月1日正式实施。无论是SAE J3016,还是我国的《汽车驾驶自动化分级》,都有一个共同的特点,即依据智能技术嵌入程度进行分级和分类,把自动驾驶视为一个不断发展的动态连续体,且预见了几种层级的自动驾驶汽车在未来很长一段时期内基于不同应用场景或用途,将处于并存共生的状态。

有鉴于此,单纯讨论自动驾驶汽车的主体或者责任主体是没有意义的,必须结合其智能技术嵌入的程度,在具体分级层次下统筹考虑人机关系,方能更为全面呈现潜在的主体,进而完成相关责任的锚定。那么,自动驾驶水平的级别与责任相匹配,人们选择怎样的级别,就必须对事故承担相对应的责任,这是一种借助事故车辆的技术嵌入程度或自动化程度而锚定责任主体的方式。

根据SAE J3016,L0~L5六个等级对应的自动化程度呈递增趋势。当级别处在L0~L2阶段,自动驾驶系统主要充当辅助驾驶的角色,人在整个驾驶行为中仍然处于主导地位,因而此类情形下自动驾驶汽车交通事故的责任分配与传统汽车不应有太大区别。其实,从更严格意义上说,当驾驶员借助辅助驾驶技术不幸发生事故时,我们仍然可以在传统交通事故的判定标准下进行责任的认定。在L3~L5阶段,技术嵌入程度已超出了辅助驾驶的范畴,所以驾驶责任分配应当区别于传统汽车,需要根据具体情况,重新进行责任锚定。

一般来说,责任锚定需满足两个条件:一是责任主体必须具有自我意识并能够独立控制自身行为,

^① 例如,2016年,欧盟提议将特定的权利与义务赋予最先进的自动化机器人,将之定位为“电子人”。2017年,沙特阿拉伯给予机器人索菲亚独特的公民身份。2018年,韩国也出台了《机器人法案(草案)》,明确了机器人的“电子人格”。

“二是存在一个或一些能够被责任人感知的他者与其在道德上相关。”^[14] 据此,与自动驾驶汽车相关的主要责任主体是驾驶员和制造商。其中,制造商又包含设计和生产等各环节,应被视为一个法人的概念,属于集体责任主体。同时,自动驾驶汽车与其使用者的边界关系将因技术嵌入而发生改变,两者主导地位的变化也导致了责任呈现出一种复杂动态性的现象,人类控制的内容和方式也发生了历史性的变革,是从人类主导过渡到人机协同再到机器主导的发展过程。

需要说明的是,由于车辆本身的质量问题引起的责任问题必须排除在外,因为当自动驾驶的汽车有质量问题时,无论处于什么状态下,都毫无疑问由汽车制造商承担相应责任。下面将尝试对不同技术嵌入程度下自动驾驶汽车的责任承担主体进行分类讨论。

(二)基于技术层级的人类控制

L0 阶段,即无自动驾驶阶段,属于传统汽车模式,完全依靠驾驶员自己控制汽车。此时,人类控制就是一般意义上的概念内涵。对于传统模式下的交通事故责任归属,目前的法律法规已经成熟,完全可以应对道路上出现的复杂的意外情况,所以该阶段出现任何交通事故,依照现有的交通法律法规即可,兹不赘述。

L1 阶段,也被称为“驾驶辅助”阶段,重点提供各种辅助驾驶的功能。驾驶员只能在特定的时间和地点内,将汽车某些部分的操作主动权交由自动驾驶汽车系统完成,如自适应巡航、防抱死刹车等,此时系统拥有的自主权只是部分的控制权。虽然驾驶员让渡了一部分对汽车的操作控制权,但并没有改变驾驶员对汽车的绝对控制权,人类控制的角色没有发生根本性变化,所以发生事故时,驾驶员的责任认定也是明确和清晰的。

相较于辅助驾驶来说,自动化程度略高的 L2 阶段,技术嵌入程度从 L1 只有典型系统下的自动化操作,发展到现阶段的部分自动化操作。该阶段自动驾驶项目在数量上有所增加,但是仍然处于驾驶者监督之中。驾驶员既能够授权给自动驾驶汽车,又可以随时收回车辆操作权,所以这一阶段车辆的主要控制者仍然为驾驶员,相应事故的责任主体也理应为驾驶员。与 L1 类似,该阶段的自动驾驶系统还不具有任何成为责任或道德主体的可能,需要在驾驶员的授权下选择和执行相应的操控任务。换句话说,虽然自动驾驶系统在一些任务是控制者,但其根源是驾驶员的主动授权。

需要指出,根据更新版 SAE 自动驾驶等级的标准,L0~L2 级自动驾驶都属于自动驾驶辅助的范畴。此时驾驶员必须处于一种正常驾驶的状态,并随时保持对驾驶环境的观察和判断,从而最大程度地确保驾驶行为的安全。其中,L1 和 L2 级自动驾驶系统仍然属于“驾驶支持系统(Driver Support Systems)”。由此可见,这两个阶段自动驾驶的技术嵌入程度十分有限,仅提供一定的技术支持,协助驾驶员提升自身的驾驶体验和安全。

当技术嵌入水平提高到 L3 后,责任主体则由原先单一的驾驶员变更为驾驶员和制造商,人类控制的形式也发生了变化。在“有条件自动化”模式下,驾驶员需要及时响应自动驾驶系统的请求,进而接管车辆的操控权。驾驶员在接到功能请求时,必须驾驶汽车。在 SAE 标准里,L3 级采用了“回退就绪用户”的概念,与备用驾驶员的角色类似,以及部分自动化回退的可能性。换言之,假若驾驶员未能及时响应系统发出接管车辆的请求而造成事故的,必须承担其应有的责任,这一点是毋庸置疑的。如果系统没有发出请求要求驾驶员接管车辆而发生事故的,那么根据我国《侵权责任法》《产品质量法》,汽车制造商应该对事故造成的损失负责。

L4 阶段的“高度的自动化”是指,在特定情形下,自动驾驶系统负责监控行驶环境和完成驾驶操作。如果驾驶员操作超出了该阶段自动驾驶的特定情况,或者明明知道不属于特定情形而随意启动自动驾驶模式,从而导致事故的发生,此时,相关责任主体十分明确,驾驶员必须为自己的不当行为承担相应的事故责任。^[15] 另外,在驾驶员已经意识到危险或自动驾驶汽车对外部环境判断失误而造成的损失,同 L3 中的部分情况类似,主要应由驾驶员与制造商共同承担责任。如果当时的环境和操作都符合该阶段自动驾

驶汽车的相关要求,那么该状况下自动驾驶汽车就完全不需要任何人为的操作,也就与完全自动驾驶机动车没有区别,^[16]这种情况可以与以下自动驾驶最高阶段合并讨论。

自动化水平达到最高阶段 L5 之后,就是真正意义上的无人驾驶,车辆系统可以自主完成所有驾驶操作,可以在任何条件下驾驶车辆,不再需要驾驶员接管驾驶。这是对现有归责乃至伦理体系最大的挑战,其中涉及用户、系统设计者、制造商、保险公司等多重关系。当汽车处于完全自动化的状态时,驾驶员对此时事故的发生毫无防范能力,因此也不用承担责任。^[17]该模式下的交通事故,驾驶员不仅不应该承担责任,反而应该受到赔偿。显然,此时的事故责任主体已经发生了变化,由原来的驾驶员完全转移至制造商。但是,这并不意味着驾驶员可以完全置身度外,仍然需要积极配合各种后续的调查取证工作。

当然,如果因为黑客等因素造成自动驾驶系统的损坏而导致事故的,制造商可以在一定程度上申请减免责任,同时需要对造成过错的第三方进行追责。那么,如何划分这部分的责任,将在下文提及,但是如何找出攻击自动驾驶汽车程序的黑客,则需要专业技术的支持。

根据 SAE 所提供的自动驾驶水平分类,我们对自动驾驶模式逐一进行了分析,进而尝试判定在不同情形下发生交通事故时相应的责任承担主体。虽然看似只有用户和制造商,是一种简化处理的方式,但是仍然存在责任遗漏的风险。然而,从某种程度上说,责任过于分散意味着无责任可言,也就是本文重点提及的主体悬置现象。

基于技术嵌入的框架讨论自动驾驶汽车的责任主体问题,必须考虑到如下几点:其一,责任主体是复杂动态变化的;其二,自动驾驶汽车不具备道德能动性,无法直接赋予其责任主体地位;最后,责任主体可以是集体范畴,或者相关第三方主体,如保险公司等。

不过,从技术层级的技术治理角度看,只能粗略地开展责任主体的框定,在具体的应用场景中,仍然面临着走入另一个主体悬置的循环之中。当然,关于自动驾驶汽车技术层级的讨论是必要的,因为技术层级总是对应着不同的人类控制关系,尤其当自动驾驶处于更高层级的抽象控制关系时,如何识别这种人类控制,又如何基于人类控制的视角锚定相应的责任主体?当前,“有意义的人类控制”这一全新的治理理念可为自动驾驶汽车的责任锚定提供借鉴。

(三)“有意义的人类控制”

“有意义的人类控制”最早应用于致命性自主武器系统,旨在解决自主武器可能带来的国际法、人道主义等法律、伦理挑战。^①该理念自提出之日起,就受到各方的高度关注。虽然大家对“有意义的人类控制”的具体内涵和实践路径等核心问题存在不同见解,但各界均对理念的前景十分看好。

“这一概念有两个基本构成:‘有意义’与‘控制’。‘有意义’指代人类控制智能机器行为并承担道德责任的基础条件。‘控制’涉及人类主体对智能机器的控制形式的多元化”。^[18]“有意义”作为“控制”的前置修饰语,强调人类控制的有效性。只有“有效”的控制,才是“有意义”的,即能够有助于解决主体悬置的问题,进而实现责任锚定。如果智能机器没有脱离其工具属性,那么就仍然处于人类的有效控制之下。归责或者锚定是否有效,取决于人类能否对智能机器实现有意义的控制,这又与智能系统的自主性密切相关。

显然,“更高水平的系统自主性可以且应该与人类控制和责任相结合”,^[19]人类控制水平应该与自动驾驶的自主性程度同步提高。遗憾的是,现有的一般道德责任和自由意志的哲学讨论中,难以涵盖包括自动驾驶在内的自主性水平较高的智能机器行为。德西奥(Filippo Santoni de Sio)和范登霍温(Jeroen van den Hoven)借鉴了“引导控制(guidance control)”^[20]的概念,将之应用于自主武器系统,以及更一般的自主系统,提供了第一个完整的“人类对自主系统有意义的控制”的哲学解释,^[19]这为解决自动驾驶责任真空

^① 2015年,《特定常规武器公约(Certain Convention Weapons, CCW)》会议便开始关注“有意义的人类控制”这一理念,此后的多次会议也进一步推动和拓展了该理念的发展。

问题奠定了理论基础。

“有意义的人类控制”的理念也逐渐被运用到自动驾驶领域。^[21]诚然,自动驾驶系统与致命性智能武器有很大的区别,将其扩展到自动驾驶领域,需将其与自动驾驶汽车的各类场景结合起来,决不能简单照搬。自动驾驶汽车发展迭代迅速,如何根据不同技术层级更加深入理解“有意义”和“人类控制”的题中之义,这是一个迫切需要回答的问题。

究竟应该如何实现“有意义的人类控制”?毋庸置疑,控制总是与责任捆绑在一起,明晰控制主体和控制形式是有效应对自动驾驶责任鸿沟等问题的关键。可见,由谁来控制?需要明确的是哪些主体参与到自动驾驶的人类控制中;如何控制?确定的是自动驾驶的控制方式问题。对于上述基本问题的思考和回答,必须与人工智能发展,尤其是自动驾驶技术的发展水平相匹配,使“有意义的人类控制”在自动驾驶实践中具备更强的可行性和灵活性。

对此,德西奥和范登霍温提出了一个基于“追踪(tracking)-回溯(tracing)”双重条件的人工智能伦理治理框架,改变了传统单纯技术治理的思路。总体来说,为了实现“有意义的人类控制”,“追踪”条件要求决策系统应该能够体现人类行为的道德理由,这就赋予“人类控制”更加丰富的内涵与外延;与此对应,“回溯”条件则需要系统行为能够还原至系统设计或人机互动等不同阶段中的至少一个人类责任主体,从而避免相关主体逃避自身可能承担的伦理或法律责任。^{[19]5}那么,对应自动驾驶的应用场景,如何满足上述“追踪”“回溯”的条件?自动驾驶过程中有意义的人类控制有哪些?人机互动的每个阶段能否对应至少一个人类主体?这些人类主体能否对自身的伦理或者法律角色有较为清晰的认知?

必须承认,随着智能机器的出现,人类控制形式的确发生了根本性变化,逐渐从显性控制转换成隐性控制,从具象控制演化为泛在控制,甚至出现了多元主体参与的控制链。^①对于自动驾驶来说,直接操作意义上的控制形式也逐渐被抽象意义上的控制所替代。从一定程度上说,作为责任主体的人类对智能机器的控制呈现出日益弱化的态势,而智能机器的自主性却不断地延伸和拓展。正因如此,智能机器自主性与人类控制权之争引发了一系列伦理、法律和社会等问题。

从本质上看,自动驾驶之所以可能会出现责任主体悬置问题,仍然可归因于人类控制的削弱或隐蔽。智能时代,在让渡部分控制权的前提下,人类控制的内涵逐渐从具体与在场操作发展到抽象与远端控制。人类控制与自动驾驶可能导致的风险之间,在时序上难以简单地以因果性来进行推断,在空间上也存在明显的距离效应,加上自动驾驶技术的体系化也极大地增加了人类控制的复杂性。^②

总之,必须从整体上系统把握该理念的丰富内涵,应对自动驾驶汽车带来的责任真空问题。对于自动驾驶汽车而言,未来应将该理念作为伦理治理的切入点,在结合大量案例的基础上,更加清晰界定“有意义的人类控制”的概念内涵和实践路径。

三、走向“有意义的人类控制”的责任锚定

基于上述分析,我们更加明确了不同技术等级下事故责任主体的可能归属问题,这直接关系到自动驾驶汽车未来的发展。有鉴于此,笔者借鉴“有意义的人类控制”理念,并结合我国自动驾驶汽车的实践现实,尝试给出如下建议。

首先,需要制定自动驾驶汽车专门法律法规。法律的强制约束力可为“有意义的人类控制”提供强有力的实施保障。我国现行的交通事故方面的法律并不能完全适用于自动驾驶汽车,亟需在“有意义的人

^① 人工智能技术集成性和交互性等核心特征决定了人类控制形式发生了巨大变革,这也是某种程度上造成主体悬置或责任鸿沟的主要原因之一。

^② 为了应对这一困境,梅卡奇(Giulio Mecacci)和德西奥提出了一种“理由接近度量表(proximal scale of reasons)”,用于诠释系统行为与人类控制之间的关系问题。参见 MECACCI G, DE SIO F S. Meaningful human control as reason-responsiveness: the case of dual-mode vehicles. *Ethics and information technology*, 2020(22): 103-115.

类控制”的框架下加强自动驾驶方面的立法工作,并将人类控制作为关键的切入点,需要严格界定其概念边界。随着自动化程度的增强,传统意义上驾驶员的概念内涵将被消解,逐渐转变成人类控制参与者的角色。因此,自动驾驶技术的开发者、制造者、设计者等都应纳入法律约束的对象,不只是局限在一般意义上的车辆驾驶员。格外需要注意的是,对于L5阶段,即自动驾驶汽车完全处于无人驾驶的状态下,更需要基于人类控制的内容和形式的变化,尽早出台专门针对无人驾驶汽车阶段的相关法律法规。因此,新修订的法律必须要紧紧围绕不同阶段人类控制的差异性,给出详细的解释与规定。

其次,必须践行自动驾驶汽车伦理共识。除了完善相应法律法规之外,还需要坚持“伦理先行”,特别是针对完全自动驾驶阶段,如何破解“电车难题”式的伦理困境,这需要多方主体通过开放、透明的协商机制,达成一种伦理共识。同时,应该广泛借鉴其他国家在自动驾驶伦理建设方面的一些先进做法,如“德国自动驾驶伦理指南”等,^[22]增强自动驾驶汽车相关人类主体对自身伦理和法律角色的充分认知。唯有通过伦理和法律的双重规制,才能更好地保障自动驾驶汽车的良性发展。

再者,推行强制购买自动驾驶汽车专用保险。伴随着系统自主性的持续提升,自动驾驶汽车事故的责任主体也发生了动态变化。虽然产品责任具有很强的调整适应性,但对于自动驾驶这个高风险的新兴产业来说,引入责任保险制度是非常有必要的。^[23]对于责任实践的可行性来说,建立自动驾驶汽车的保险制度具有十分重要的现实意义。具体来说,当自动化水平逐渐提高时,需要强制驾驶员和制造商购买相应的保险。就驾驶员投保而言,如果因驾驶员对自动化的等级判断有误或者操作不当,以及没有及时接管自动驾驶车辆所引发事故而造成的损失,可以在保险责任限额内,由保险公司赔偿。就制造商投保而言,对于黑客的恶意行为、产品的技术难题,或者现有系统中没有预知的不足等所导致的交通事故造成的损失,可由保险公司进行赔偿。因此,借助自动驾驶汽车保险,既可以分担制造商和驾驶员等不同人类控制主体的责任,又能有效保护受害人的利益,达到法律和保险双管齐下的良好效果。

最后,优化自动驾驶全生命周期的数据系统。由于自动驾驶汽车的分级系统,责任锚定更多依赖于驾驶数据的有效存储、获取和分析。未来应该研发更加强大的数据系统,不仅提供自动驾驶汽车的内部数据和外部数据,还应该包括研发、设计和生产等全周期的数据,便于日后开展“追踪-回溯”分析,从而让最终的责任划分更加精准明晰。此外,自动驾驶汽车数据系统还能对道路状况和交通事故进行智能化分析,帮助驾驶员提前预知道路状况,并可及时通知有关部门处理意外发生的交通事故。如果自动驾驶汽车得到普遍使用后,建议将数据信息化,建立全国范围内的大数据信息分析中心。在确保信息安全的情况下,由专业人员监控分析,并为自动驾驶汽车研发提供现实的依据,促进自动驾驶行业快速发展。

从本质上看,自动驾驶汽车的责任鸿沟问题与不同技术嵌入层级下的人类控制密切相关。源于应对人工智能伦理困境的“有意义的人类控制”,可作为解决自动驾驶汽车责任鸿沟问题的一种全新技术治理理念。该理念强调多学科的研究路径,^①“要求人类保证对智能机器的最终控制并承担道德责任,并试图倡导当代新兴技术的社会控制哲学”,^[24]借助“追踪-回溯”双向路径设计,实现对多元责任主体的锚定。

在“有意义的人类控制”框架下,将自动驾驶置于技术层级与人类控制的关系界面之上,坚持以科学的数据系统为支撑,以责任锚定为导向,以人类控制为中心,通过伦理与法律的整合进路,并结合完善的保险制度,破解自动驾驶主体悬置的困境。未来,应该加强自动驾驶的跨学科交叉研究,尤其需要借助科技与人文的常态化对话机制,强化各利益相关方的伦理自觉和法律认知,摒弃简单的技术设计视角,实现“有意义的人类控制”转向,从而进一步推动自动驾驶技术的发展。

^① 该框架本身就是由荷兰代尔夫特理工大学的工程师、哲学家和心理学家等组成的多学科团队牵头的名为“自动驾驶系统的有效人类控制”研究项目,旨在通过哲学、行为心理学和工程学等多学科的联合工作,实现对自动驾驶汽车的有效控制。参见 DE SIO F S, MECACCI G, CALVERT S, et al. Realising meaningful human control over automated driving systems: A multidisciplinary approach. *Minds and machines*, 2022: 1-25.

参考文献:

- [1]李伟,华梦莲.论自动驾驶汽车伦理难题与道德原则自我选择[J].科学学研究,2020(4):588-594+637.
- [2]尤婷,刘健.自动驾驶汽车的交通事故侵权责任研究[J].湘潭大学学报(哲学社会科学版),2021(2):32-36.
- [3]冯浩语.人工智能技术与责任法的变迁——以自动驾驶技术为考察[J].比较法研究,2018(2):143-155.
- [4]付玉明.自动驾驶汽车事故的刑事归责与教义展开[J].法学,2020(9):135-152.
- [5]DE SIO F S, MECACCI G, CALVERT S, et al. Realising meaningful human control over automated driving systems: A multidisciplinary approach[J]. Minds and machines, 2022; 1-25.
- [6]闫坤如.人工智能机器具有道德主体地位吗? [J].自然辩证法研究,2019(5):48.
- [7]顾世春,费希尔道德责任理论视角下自动驾驶汽车使用者道德责任研究[J].自然辩证法研究,2021(2):114.
- [8]黄闪闪.无人驾驶汽车的伦理植入进路研究[J].理论月刊,2018(5):187.
- [9]冯珏.自动驾驶汽车致损的民事侵权责任[J].中国法学,2018(6):116.
- [10]司晓,曹建峰.论人工智能的民事责任:以自动驾驶汽车和智能机器人为切入点[J].法律科学(西北政法大学学报),2017(5):173.
- [11]张继红,肖剑兰.自动驾驶汽车侵权责任问题研究[J].上海大学学报(社会科学版),2019(1):20.
- [12]闰玺池,冀瑜.SAE分级标准视角下的自动驾驶汽车事故责任承担研究[J].标准科学,2019(12):51.
- [13]SAE International.SAE levels of driving automationTM refined for clarity and international audience[EB/OL].[2022-05-12].
<https://www.sae.org/blog/sae-j3016-update>.
- [14]白惠仁.自动驾驶汽车的“道德责任”困境[J].大连理工大学学报(社会科学版),2019(4):16.
- [15]万丹,詹好.自动驾驶汽车责任主体和道德两难问题的哲学分析[J].社会科学战线,2021(11):24-32.
- [16]杨立新.自动驾驶机动车交通事故责任规则设计[J].福建师范大学学报(哲学社会科学版),2019(3):81.
- [17]季若望.智能汽车侵权的类型化研究——以分级比例责任为路径[J].南京大学学报(哲学·人文科学·社会科学),2020(2):125.
- [18]俞鼎.“有意义的人类控制”:一项新的人工智能伦理治理原则[N].中国社会科学报,2022-12-27(06).
- [19]DE SIO F S, VAN DEN HOVEN J. Meaningful human control over autonomous systems: A philosophical account[J]. Frontiers in robotics and AI, 2018, 5.
- [20]FISCHER J M, RAVIZZA M. Responsibility and control: A theory of moral responsibility[M]. Cambridge, UK: Cambridge University Press, 1998.
- [21]MECACCI G, DE SIO F S. Meaningful human control as reason-responsiveness: The case of dual-mode vehicles[J]. Ethics and information technology, 2020(2):103-115.
- [22]陈晓平,翟文静.关于自动驾驶汽车的立法及伦理问题——兼评“德国自动驾驶伦理指南”[J].山东科技大学学报(社会科学版),2018(3):1-7.
- [23]郑志峰.自动驾驶汽车的交通事故侵权责任[J].法学,2018(4):23.
- [24]俞鼎,李正风.智能社会实验:场景创新的责任鸿沟与治理[J/OL].科学学研究.[2022-12-15]. <https://doi.org/10.16192/j.cnki.1003-2053.20221129.001>.

(下转第44页)

Application of Judicial Confirmation Procedure in Ecological Environment Damage Compensation Cases

LIANG Yong, ZHU Ye

(*Environmental Resources Court, Shandong Higher People's Court, Jinan 250101, China*)

Abstract: After the event of ecological environment damage, the agreement of compensation for the damage to the ecological environment reached by the obligee and the obligor through the consultation procedure is the primary means to quickly repair the environment. In practice, it often occurs that the obligor delays or refuses to perform the compensation agreement. It wastes administrative resources and judicial resources when the obligee files a contract lawsuit to demand the obligator to perform. As the civil compensation agreement is reached between equal civil subjects, the obligee can apply to the people's court for judicial confirmation so as to obtain the compulsory execution. It's important to note that the compensation agreement, different from the mediation agreement, is special in the application of judicial confirmation system. The main body of the application should not be limited to the compensation obligee and the compensation obligor, but can be extended to the local government or its departments and agencies which engage in the specific work of emergency response and environmental restoration, or the parent company, affiliated company or the guarantor of the obligor. In case that a compensation agreement applies for judicial confirmation, if there is no third-party mediation organization to preside over it, it may also be under the jurisdiction of the grassroot court where the damage occurs or where the damage affects. The review should be substantive, and the court may organize experts and scholars in the industry to discuss and give opinions on the compensation agreement, and publicize the compensation agreement. Meanwhile, in view of the substantive review of compensation agreement, the time limit for review should not be limited to 30 days as stipulated in the *Civil Procedure Law*.

Key words: ecological environment; consultation procedure; compensation agreement; judicial confirmation

(责任编辑:董兴佩)

(上接第 34 页)

Towards “Meaningful Human Control”: Subject Suspension and Responsibility Anchoring in Self-driving

LIU Chengke^{1,2}, GE Yan¹

(1. *School of Marxism, Anhui Agricultural University, Hefei 230036, China*;

2. *School of Foreign Languages, Anhui Agricultural University, Hefei 230036, China*)

Abstract: Self-driving, as a dynamic technology continuum, presents different man-machine relationships in specific scenarios, such as the transition from human-led mode to man-machine cooperation mode, and then to machine-led mode, which leads to various ethical, legal and social problems. The various human-machine interactions reflect the continuous evolution of human control. In addition, from the perspective of forms of human control, self-driving has also experienced a shift from explicit control to implicit control, from concrete control to ubiquitous control, and then to multi-agents participating in the control chain. Regardless of changing forms of human control, meaningful human control in self-driving is the key to the problem. In view of this, by introducing the levels of technology embedding, this paper analyzes the problems of subject suspension that may exist in the accidents of self-driving cars. Under the framework of “Meaningful Human Control(MHC)”, the boundaries of responsibilities are clarified according to the different human-machine interfaces and forms of human control, and the dilemma of subject suspension in self-driving is resolved through an integrated approach of ethics and law, combined with a sound insurance system. In addition, a full-cycle data system based on “tracking-tracing” should be optimized to help anchor the corresponding responsibilities, which ensures the healthy and rapid development of self-driving technology. In the future, it is expected that the new concept of MHC will be applied to explore more systematically the mechanisms and practical paths to address the responsibility gap in self-driving.

Key words: self-driving; subject suspension; responsibility anchoring; technology embedding; meaningful human control(MHC)

(责任编辑:江 雯)