

自动驾驶汽车的五大伦理问题与原则的思考

李 芳¹, 裴 彧², 徐 峰¹, 高 芳¹, 贾晓峰¹

(1. 中国科学技术信息研究所, 北京 100038;

2. 广西中医药大学马克思主义学院, 南宁 530200)

摘 要: 自动驾驶代替人类驾驶员进行驾驶决策以及驾驶操作, 需要从机器代替人类劳动的整个劳动过程重新审视相关的伦理问题。在实践中, 突破机器决策的伦理“两难困境”, 应以人人平等的伦理共识为底线, 以民意调查为具体的选择标准。为尽力规避自动驾驶系统威胁人类决策的自主性, 宜确立承运人无免责权的默认原则。为了增进社会公平, 需要遵循统一规划和向最不利者倾斜的原则。为有效划归自动驾驶事故责任, 应遵循奥康剃刀原则及责任承担的有效和实用原则。为了在“开放数据获取便利”和“保护隐私”之间实现平衡, 应遵循政府管控数据输入原则。

关键词: 自动驾驶; 伦理; 人工智能

中图分类号: G321 **文献标识码:** A **DOI:** 10.3772/j.issn.1009-8623.2020.10.006

自动驾驶技术在本质上是通过智能机器的自动运行, 代人类之劳。由于将劳动转让给了机器, 因此整个劳动过程及劳动结果相关的伦理问题都需要重新审视。劳动过程可能一帆风顺也可能面临困难。自动驾驶过程是智能机器代替人类驾驶员做出驾驶决策, 并自动执行智能决策系统发出的驾驶指令的过程, 涉及的伦理问题主要包括两类, 一类是在面临人类存在伦理争议决策时, 机器决策的合伦理性过程中面临的决策“两难困境”问题; 一类是在不存在伦理争议的决策中, 人类让渡决策权时的决策自主问题。从劳动结果来看, 劳动的结果可能带来各种收益也可能引发意外或损失, 自动驾驶可能节约时间和经济成本, 带来环境效益, 也可能因交通事故而带来损失, 因此相关的伦理问题就分为成果丰硕时的利益分配问题和发生事故后的责任分配问题。除了以上

四个问题, 现代社会的劳动离不开分工合作, 自动驾驶也不例外, 它是一种由人类提供行驶数据, 再由机器对数据进行分析, 从而规划出最优的社会交通方案的劳动, 在此过程中人类提供了自己的行驶数据, 它是一种隐私数据, 因此自动驾驶还涉及隐私保护问题。

从 2015 年《中国制造 2025》将智能网联汽车列入未来 10 年国家智能制造发展的重点领域以来, 自动驾驶汽车技术研发与路测进入快车道。2020 年 4 月 19 日, 百度宣布 Apollo Robotaxi 服务 Dutaxi 已正式在长沙梅溪湖和洋湖区开放运营, 自动驾驶逐步走进人们的生活、工作场景中。然而, 近年来自动驾驶的快速发展以及伴随的若干交通事故, 引发了各界对自动驾驶的伦理问题的高度关注, 与之相应的“决策两难”困境、人类决策自主、事故责任、社会公平以及隐私等伦理问题,

第一作者简介: 李芳, 女 (1986—), 助理研究员, 主要研究方向为人工智能政策、数字经济与治理。

项目来源: 中国科学技术信息研究所创新研究基金面上项目“自动驾驶汽车领域人工智能伦理风险研究”(MS2020-04); 科技部科技创新 2030-“新一代人工智能”重大项目“重点领域人工智能治理挑战及对策研究”(2019AAA0105300)之子课题“重点领域人工智能伦理风险及对策研究”(课题编号: 2020AAA0105301)。

收稿日期: 2020-08-29

已经成为自动驾驶大规模商业化过程中不可回避的焦点问题。本文基于现有研究，从伦理学角度进行分析，尝试对上述问题做出一些回答。

1 道德决策的“两难困境”与伦理原则

在自动驾驶的伦理问题研究中，机器道德决策的“两难困境”是伦理学界争论的一个经典议题。人们关于自动驾驶汽车面临不可避免伤害的情形中，如何选择伤害目标的道德决策始终存在争议。然而，自动驾驶汽车不断走进人类生存环境的现实，决定了我们需要从争议中探讨一些可供参考的共识性原则，以增强社会对机器的道德决策的信任。

1.1 有关机器道德决策的“两难困境”的讨论

“电车难题”是“两难困境”的典型例子，是由哲学家 Foot^[1]于1967年提出的著名思想实验，基本设定是：驾驶员可以通过变轨使一辆失控的电车撞向1个人或者不变轨使电车撞向5个人。为讨论自动驾驶汽车对碰撞目标的选择困境，后期学者们又改造出“隧道难题”^[2]等变体。在经典的电车难题中主要的考虑因素是人数，但是现实生活中要考虑的因素则十分复杂，比如在老人和年轻人之间怎么选择的问题，以及在人与动物之间选择的问题^[3]。这些问题一旦选择不慎，误用了道德共识，就会导致自动驾驶的伦理问题。Hauser等^[4]针对“电车难题”，对来自120个国家的5000名志愿者进行了网络调查，结果显示，89%的调查对象认为牺牲1个人而挽救5个人是可以允许的。麻省理工大学也对“电车难题”开展了一项上百万人的调查，测算在多达10个变量（人的性别、年龄等各种变量）的情况下人们会做出什么选择^[5]。这类调查研究的缺点在于在有些地方违反了几乎所有的伦理学流派的底线，比如绝大多数伦理学都要求“人人平等”，如若把假想行人的性别、年龄等考虑进去，得到的结论很可能是“撞老人不撞年轻人”或者相反，而麻省理工大学的调查甚至把其他物种（如小狗）也考虑进去了，一旦调查结果是，虽然人们不至于选择为了救1条狗而撞人，但有可能选择为了救10条狗而撞人，这是一个让大多数伦理学家难以

接受的结果。

总结来看，在这一类思想实验下，人工智能面临着决策两难困境，无论哪一种选择，都不可避免地招致对机器决策的伦理质疑。这一矛盾的焦点指向功利主义和义务论两种哲学伦理观念导致的迥异的道德选择标准，功利主义者（以边沁为代表）多倡导最大幸福原则，倾向于“两害相权取其轻”；而义务论者（以康德为代表），多秉持“普遍律原则”，倾向于“不能排斥他人的幸福”。两种标准的对立，导致自动驾驶伦理规范制定的困难。功利主义忽视了个体争议，并且伤害可能无法计算^[6]。当无人驾驶汽车要撞到一个没戴头盔的电动车驾驶人时，而旁边有一个戴着头盔的电动车驾驶人，此时戴头盔的驾驶人很可能成为碰撞的目标^[7]。这显失公平，并且可能引发不遵守秩序等更深层次的道德风险。

功利论和道义论的伦理价值观的碰撞也反映到了伦理规范的实践中，Gerdes等^[8]提出了无人驾驶汽车的三定律：无人驾驶汽车必须不能碰撞行人与自行车；无人驾驶汽车不应碰撞其他汽车，除非与第一条定理冲突；无人驾驶汽车不能碰撞其他物体，除非与第一条与第二条定理冲突。和鸿鹏^[6]认为这三个定律为指导无人驾驶的程序设定提供了一些启发，但其过度强调行为正当性的缺陷也显而易见。除了无人驾驶汽车的三定律以外，英国交通部和德国自动驾驶伦理委员会也分别发布了自动驾驶汽车的8条原则^[9]和20条规则^[10]。陈晓平等^[11]认为德国自动驾驶汽车伦理指南的部分条款是恰当的，但并非没有缺陷，德国的规则误用了功利主义和道义论的原理，导致把事故责任归于人类驾驶员。这些原则还遗漏了一些问题没有解决，比如说为了防止事故发生，是否可以让行人携带信号装置，提醒自动驾驶汽车自己的位置？如果有人买不起信号装置，是否意味着他们就更有可能会被汽车伤害^[9]？这些问题潜在的两种答案都有各自的伦理理由，需要仔细比较各种伦理理由才能得出公正的解决方案。

机器的道德决策还涉及一个现实问题，就是由谁来监督自动驾驶汽车的系统，确保它搭载了正确的道德模块？按照英国交通部的原则，要由

自动驾驶汽车的制造商组成的联合体来进行监督,那么与自动驾驶直接相关的乘客在其中似乎就没有发言权了,这显然是不对的,不过普通乘客不具备相关的技术和伦理知识,也很难进行有效监督。这时候是否需要政府为公民承担起监督的工作呢?既然每家医院都应该依法设立自己的伦理委员会,那么汽车制造商是否应该效仿,成立自己的伦理委员会呢^[8]?这些问题如果处理不好,就有可能造成道德模块的搭载工作无人监管的伦理风险。

以上道德两难问题,双方各执一词,难以立判高下,余露^[12]提出一个新颖的解决途径,把决定权交给乘客,由乘客来决定汽车遵循哪种伦理规则。不过必须指出,除了乘客以外,最好把其他车上的人员以及路上的行人都考虑在内,并且必须坚持康德的普遍律原则,即一个人为自动驾驶汽车做出的伦理选择,不能因为下次他从车内乘客变成车外行人了就发生改变,以免出现麻省理工的调查所显示出的尴尬情况,大多数人都认为应该让自动驾驶汽车上的人牺牲自己去保护其他人,但是一旦设想自己也在车上,就不愿意自我牺牲了。这样的选择是违反普遍律原则的。

1.2 解决“两难困境”的两项原则

对于电车难题,我们不得不承认这样一个事实:一方面伦理学界内部对这个问题没有统一的答案,另一方面伦理学界外部的讨论尚未充分吸收伦理学界有益的营养。具体说来,在伦理学界内部,这个伦理问题不是找不到答案,而是有着互相冲突的答案,这些答案来自延续了两百多年的不同伦理流派,主要以康德的义务论和边沁的功利主义为代表,至今聚讼不休,这意味着我们不能指望在未来一百年内找到一个伦理学界公认的答案,但是自动驾驶汽车很有可能在未来几十年内就出现在世界各地,因此我们不能指望伦理学家及时讨论出一个结论。在伦理学界外部,如上文所述,一些民意调查的结论是有违基本的伦理学共识的。

基于以上现实,本文认为有关自动驾驶汽车“两难困境”的伦理解决方案应遵循以下两项原则:

一是以伦理学共识为底线,不突破人人平等的伦理底线。此处以举例的方式对人人平等原则稍作解释,这条原则怎样运用在电车难题上呢?比如说,如果出现电车难题,左前方是一辆名贵豪车,撞上去的经济损失很大,而右边是一辆普通汽车,撞上去的经济损失很小,根据人人平等的基本伦理规则,我们不希望自动驾驶汽车在做选择时因为车的经济价值而采取不同的态度,而是一视同仁。这条原则还适用于社会名流和普通人、老人和青年人等电车难题中考量的变量。在这一原则基础上,可进一步给出具体的信息收集建议,比如说不用、不应收集路上行人和车上乘客的具体信息,只要知道在某个方位有行人就够了。

二是在满足“一”的前提下,以民意调查为具体的选择标准。也就是说在电车难题中做选择,只要不违背“一”,就按照“二”去执行。以上两条原则既凸显了底线思维,又体现了灵活的态度,兼容并包了专家共识和民意,可以说是中国中庸调和思想的具体应用。

2 决策自主问题与伦理原则

人类一旦接受了自动驾驶的辅助驾驶功能,就不可避免地对智能驾驶系统产生依赖,并且这种依赖会随着驾驶系统智能化水平的提升而不断增强,形成诸如“消费上升容易下来难”一样的棘轮效应。Millar^[2]指出,自动化道德决策可能在使用户完全脱离道德决策循环(Ethical Decision-making Loop)的风险,进而可能对用户的道德自主构成直接威胁^[13]。表面上看这是一个人与机器之间的权利让渡问题,且不说人类愿意让渡权利的程度缺乏广泛共识;从更深层次来看,这种权利的让渡发生在自动驾驶用户与自动驾驶开发商、移动互联网服务商、导航定位服务商等主体之间,由于商业利益驱动,自动驾驶用户的权利让渡可能并非自愿、也不知情,因而值得进一步探究,并从保护消费者权益出发,设定一定的原则。

2.1 自动驾驶决策自主问题发生的三种情境

自动驾驶的自主问题发生在三个情境中,一是驾驶危险出现之前,二是驾驶出现危险的时候,

三是没有驾驶危险的时候。在驾驶危险出现之前，如果自动驾驶汽车的系统是联网的，那么就存在被黑客攻击的危险，可能会被黑客控制汽车进行恐怖活动^[9]；但是如果不联网，自动驾驶就不能与公路上的其他自动驾驶汽车以及天眼等其他智能设备互动，形成更大的交通安全网络^[14]。再考虑一个特殊情况，假设我要驾驶汽车去火灾现场救我的亲人，可是自动驾驶系统判断出此行凶多吉少，那么它是否有权利拒绝带我去到火灾现场呢？第二种情境是当驾驶出现危险的时候，自动驾驶系统应将驾驶权交给人类驾驶员，但问题是所谓的危险时刻、自动驾驶无法自动解决问题的时候，究竟指的是什么情况？在移交驾驶权的过程中是车速不变还是减慢，还是干脆停下来？最后，这些问题是交给政府决定，还是由汽车厂商决定，还是由汽车使用者决定^[9]？第三种情境是没有驾驶危险的情形，同样有可能出现自主性受影响的情况。假设我们为了让两辆自动驾驶汽车不会相撞，最安全的办法就是让它们彼此联网，车与车互相联网以后不仅不会相撞，而且还可以由中心计算机设计出最佳行车路线，确保把堵车的问题减少到最轻的程度。但是这就意味着让一些本来打算走 A 路线（也是本来的最短路线）的汽车去走 B 路线，这样才能实现车辆分流。这样一来，就在一定程度上损害了乘客的自主性，因为走远路不是他们的选择，而是自动驾驶系统替他们做出的选择。更糟糕的情况是，自动驾驶汽车为了让乘客看更多的广告而带他们走了一条较远的路^[15]。除此之外，当危险不存在的时候并不是没有其他风险。还有人指出，如果为了避免危险，自动驾驶汽车被设计得过度小心翼翼，那么就存在缓慢行驶导致塞车的风险，甚至会有人知道既然自动驾驶汽车无论如何不会去碰撞其他汽车，于是故意绕到它前面挑衅，以此寻求刺激，这是过分的谨小慎微反而带来的安全风险。

2.2 维护人类决策自主性的原则建议

为尽力规避自动驾驶系统威胁人类决策的自主性，自动驾驶治理方案宜遵循以下原则性建议：确立承运人无免责权的默认原则，也就是说如果出现法律没有规定、承运人也没有在合同中明确

指出的情形，默认承运人没有免责权。

这个原则是考虑到自动驾驶汽车的制造商有更多的背景知识和时间为意外情况做准备，而乘客难以在意外发生的转瞬之间做好准备，因此将在意料之外情形下发生的相关责任划归承运人。

3 自动驾驶引发的公平问题及解决原则

自动驾驶与人工智能在其他领域的应用一样，会涉及公平的问题，而且由于人工智能是一项通用技术，理论上可以对任何领域产生影响，所以它涉及的公平问题非同小可。人工智能引发的公平问题如果处理得不好，会导致社会出现系统性的分配不公，阻碍共同富裕的实现；如果处理得好，会让中国继经济发展的奇迹之后创造新的奇迹，在分配领域同样成为世界的标杆。自动驾驶将会对社会造成巨大影响，进而引发公平问题，而社会主义中国比资本主义的西方更有可能创造出减少自动驾驶危害公平的解决方案。

3.1 自动驾驶对社会的影响及其引发的公平问题

一般人往往低估自动驾驶对社会的影响，误以为它只不过是为社会节约了职业司机的工资，为业余司机节约了时间。为了估算这方面的收益，我们假设一个私家车司机一旦改乘自动驾驶汽车，所节约的时间的价值等于自己不开车、请网约车代劳的价格。这样算来每辆自动驾驶汽车每年节约的费用只相当于约 2 421 元人民币。

根据 2019 年发布的《中国网约车新就业形态发展报告》，网约车司机的月收入为 2 522 元，假设一名网约车司机每周工作 6 天，每天 10 小时，平均车速 40 公里/小时，那么这名司机每年行驶 12.5 万公里（按一年 52 周计算），为自己挣得 3.026 4 万元，而一名私家车司机每年行驶公里数约为 1 万公里，因此可以估计他每年为自己开车相当于节约了 2 421 元，换句话说，每辆自动驾驶汽车每年节约的价值只相当于约 2 421 元人民币。如果一辆汽车使用 10 年，那么自动驾驶汽车的价格不能高于同类普通汽车价格 2.4 万元，否则消费者还不如购买普通汽车。截至 2017 年 3 月底，我国汽车保有量为 2 亿辆^[16]，就算全部替换成自动驾驶，每年也只能创造 4 842 亿元的潜在价值。

抛开高技术对人的诱惑, 单纯的经济收益似乎并不高, 这就是为什么一般人往往会低估自动驾驶的社会影响的原因所在。实际上这个价值还是无法全部转换成国内生产总值(GDP)的。虽然自动驾驶代替人类驾驶以后, 人类司机因此节省了驾驶汽车的时间, 但是在此时间内他们仍然身处车内, 只不过不需要驾驶而已, 而身处车内的这段时间一般人不会用来工作, 因此节约的时间也就不会转换成工作所创造的GDP。

事实上, 自动驾驶的影响力超出了窄窄的一条公路, 辐射到整个城市。以美国的数据为例, 自动驾驶的首要优点是安全, 美国94%的严重撞车事故是由于人为的原因造成的, 汽车事故在2018年造成了美国36560人的死亡, 而自动驾驶有望大幅减少汽车事故率, 据估计, 自动驾驶一旦普及, 将每年在美国减少3万人的死亡^[17]; 自动驾驶的第二大优点是节省社会损失, 美国2010年由于交通事故致死所造成的经济损失高达2420亿美元, 交通事故致伤所造成的经济损失也有576亿美元, 这些经济损失同样有望在自动驾驶普及以后迎来锐减; 自动驾驶还能减少时间的浪费, 2014年美国人由于交通延误所耽误的时间高达69亿小时, 平均每人每年由于交通的原因耽误21.7小时, 可以预计, 随着自动驾驶的普及, 这个数字会有所缓解。能够节约如此多的时间, 是因为自动驾驶不仅节约了人类司机驾驶汽车的时间, 而且通过优化交通状况, 减少堵车, 节约了车内所有乘客的时间。自动驾驶还能让人类的出行更加方便, 美国65岁以上人口有4900万, 残疾人口有5300万, 自动驾驶汽车的普及将帮助这些人获得更大的行动自由^[18]。

除了上述为人乐道的好处以外, 更加重大的影响还是自动驾驶汽车对城市本身的影响。无论是美国还是中国的自动驾驶汽车都难以与普通汽车在同一条路上行驶, 一是因为自动驾驶需要专门的公路设施, 比如磁力装置^[19-21], 以便提高自动驾驶汽车的安全性^[22], 二是因为自动驾驶汽车相互之间可以通过联网实现互通位置信息^[23], 大幅提高安全性, 安全性提高后, 就可以相应地提高行驶速度, 车速预计高达193公里/小时^[24],

显然这样的时速不适合与普通时速的人类驾驶汽车在同一条路上行驶。既然自动驾驶汽车需要专门的车道, 这就意味着已建设成型的城市不适合大规模投放自动驾驶汽车, 要想大规模投放, 必须兴建新城区。换句话说, 自动驾驶汽车的应用很有可能意味着旧城区的没落和新城区的兴起。

不仅新城区的兴起源于自动驾驶, 而且新城区的具体设计也与自动驾驶息息相关^[25]。可以想象, 当新城区刚刚兴建的时候, 会有一些自动驾驶汽车在其中奔驰, 而相关的行驶数据将被收集起来用于城区规划。居民喜欢驾车购物的商业区, 喜欢驾车观光的景点, 上下班高峰容易出现堵车的交通枢纽……这些都可以通过自动驾驶汽车的行驶所创造的大数据计算而得。自动驾驶不仅可能导致新城区的兴建, 而且还可能参与指导新城区的建设方案。

既然自动驾驶产生的数据价值巨大, 那么如果这些价值所带来的利润全都归于自动驾驶汽车公司, 就会造成同样巨大的不公平。这不仅仅是垄断市场给消费者带来的不公平的交易价格、对同行企业带来的不公平的竞争环境, 而且会将垄断延伸至其他行业, 导致更大范围的不公平。首先, 人工智能属于数据时代的产品, 数据时代的产品的一个共同特征是它们的价值主要在于软件而不是硬件, 而软件的特征在于它是可以几乎零成本地无限生产的, 所以一家公司一旦取得某一领域的领头羊地位, 就很容易让消费者全都以很低的价格购买其产品, 从而将同行业的竞争对手挤出竞争行列; 其次, 人工智能作为通用技术, 又很容易与其他行业产生交集, 比如说上文设想的场景, 自动驾驶汽车与新城区的兴建与规划都产生联系, 这就导致在某一行业取得垄断地位的自动驾驶企业有可能将垄断地位扩张至其他行业, 比如说自动驾驶汽车公司为城市建设提供的建议很可能比传统的城市规划设计公司、房地产行业提供的建议更有竞争优势, 进而把垄断地位从汽车行业扩展至设计业和房地产行业。这种利用人工智能、互联网等通用技术实现跨行业生态垄断的现象已经在亚马逊、谷歌、美团等大型互联网平台企业出现。比如某平台在为用户提供集饮食、

娱乐等于一体的休闲信息服务过程中,在本行业取得了近乎垄断的地位,于是可以以命令的形式要求与其有业务关系的商家使用其提供的共享充电宝,一旦这一商业模式取得胜利,那么该平台将从店铺评分领域扩张至共享充电宝领域,使这两个领域的竞争市场退化成垄断市场。

3.2 解决自动驾驶公平问题的原则

3.2.1 统一规划的原则

集人工智能、信息通信等数字技术于一体的自动驾驶企业很容易在本行业内部和跨行业领域获取垄断地位,造成极大的不公平。这种不公平风险不仅涉及汽车行业的企业、从业者和消费者,甚至还会扩展到公路施工行业、城市建设行业等周边领域,是一个需要综合统筹的问题。职业司机需要适当的再就业培训,从而获得公平的就业机会;消费者需要信息透明度,从而确保公平地获得自己的驾驶数据所创造的价值;交通行业、建设行业需要公平的竞争机会,以免出现垄断。这样一个综合性的问题需要一个强有力的中央领导,居中协调,调度各方,才能解决。假设将这一综合性的公平问题交给自动驾驶汽车企业去解决,不仅缺乏必要的手段,而且阻止跨行业垄断有损此类企业的利益,因此它们缺乏维持公平的动力。

3.2.2 向最不利者倾斜的原则

罗尔斯^[26]认为正义的社会的发展应该让社会中的最小受益者的发展速度不低于社会平均发展速度,也就是说不能让处于社会最底层的人在发展的过程中“掉队”。将罗尔斯的这一原则迁移到对自动驾驶利益相关群体的关注上,可以说公平的自动驾驶行业发展应该是这样的一种发展,在它的利益相关群体中,原来受益最少的群体的发展速度不低于自动驾驶行业相关群体的平均发展速度。比如说自动驾驶普及以后受到冲击最大的群体可能是司机,那么根据向最不利者倾斜的原则,司机行业就有理由得到较多的政策倾斜,通过协助司机实现再就业,确保最小受益者不在发展中“掉队”。

4 事故归责问题与伦理原则

与人们高度关注自动驾驶汽车的安全相对应,

自动驾驶汽车的事故归责问题也始终处于舆论的热点。自动驾驶事故的归责不仅涉及复杂的技术认定问题,也涉及伦理、法律界对自动驾驶系统的主体地位认定的争议,然而,从促进人类出行安全和相关利益主体开发和利用自动驾驶的积极性的目标出发,仍然存在一些可供普遍遵守的原则。

4.1 有关自动驾驶事故责任的讨论

“事故归责”的相关研究主要围绕自动驾驶系统是否具有道德主体地位、是否以及如何担责问题展开。罗尔斯^[26]认为人工智能系统不能作为独立、自主的道德主体,因为计算机系统只是社会技术系统的组成部分,它们并没有精神状态,即使他们可以被理解为有精神状态,它们也没有行动的意图^[27]。国内学者张力等^[28]人也持有类似观点,认为现阶段高度自动驾驶汽车在法律上的定位仍旧是机器,也就是说其本身是无责任能力的。驾驶员要承担一定的注意义务,但产品出厂与事故发生之间的因果关系是一个模糊地带,生产者的责任有待商榷。与之相反,在自动驾驶的状态下,驾驶员的属性与传统驾驶状态下已不尽相同,在事故发生时进行决策的是自动驾驶系统,而非驾驶员^[29]。Sheriff^[30]支持把责任归于自动驾驶系统的设计者,他引用美国空军在驾驶中必须遵循的紧急情况规章,认为乘客在驾驶过程中难以快速作出反应,而自动驾驶汽车的设计师却可以事先考虑到所有应该考虑到的情况^[31]。卢有学等^[32]和冯洁语^[33]认为假如自动驾驶汽车本身满足了除主观罪过(过失)以外交通肇事罪成立所需的全部构成要件,则其是否应受到刑事处罚的关键便在于它能否诞生自主意识,当自动驾驶真正达到完全自动化水平时,就已实现了其自主意识。和鸿鹏^[6]认为无人驾驶汽车具有远离真实情境形成决策的特点,因而必须直面伦理困境并承担相应责任。杨立新^[31]认为自动驾驶机动车发生交通事故时,主要由自动驾驶机动车的设计者、制造者和销售者承担产品责任,只有自动驾驶机动车的使用人因错误操作而使自动驾驶机动车发生交通事故造成损害的特殊情况除外。

事实上,自动驾驶技术所处的阶段不同,

需要人类驾驶员介入的程度也不同,因而其责任划分也存在差异。国际汽车工程师学会(SAE International)将自动驾驶技术分为0~6级,L1和L2级别的功能非常有限,需要人类驾驶员负责大部分驾驶,实际上只是一种“辅助驾驶系统”;L3级别大部分操作由车辆自动完成;L4~L5级别为高度自动驾驶级别,由车辆完成所有驾驶操作,人类驾驶员无须保持注意力,但限定道路和环境条件。2018年百度AI开发者大会上宣布了全球首款L4量产自动驾驶巴士“阿波龙”正式下线,但也仅限在厂区、园区等限定场景使用。从目前的技术水平来看,现阶段量产车大部分处在L2到L3过渡的阶段,L4~L5级别的自动驾驶量产和大规模商业化应用还有待时日。对于L1至L3等级辅助性驾驶阶段涉及的交通违法和事故,除汽车及系统质量原因之外,仍应由驾驶人、所有人或者管理人作为车辆安全监管人员承担法律责任;L4至L5等级完全自动驾驶阶段涉及的交通违法处理、交通事故认定,特别是行政责任、刑事责任、记分等责任承担主体等问题,一定程度上突破了我国现有法律责任框架^[34]。

4.2 事故责任的划归原则

事故归责涉及不同的社会文化背景、法律体系,需要在借鉴国外经验的基础上,针对中国的伦理观念和立法情况,研究适合中国的伦理规范和法律法规。本文认为要解决自动驾驶的规则问题,应该遵循以下一些原则:

(1) 奥康剃刀原则。如无必要,勿增实体,在没有绝对必要的情况下不增加法律责任人的种类。具体说就是如果没有必要,就不应把自动驾驶汽车或任何机器人纳入责任人的范畴。

(2) 责任承担的有效与实用原则。归责是为了让责任人承担赔偿责任,并提示其同类人群尽力降低风险,比如说,如果将惩罚施加在自动驾驶汽车上,汽车没有情感,不会害怕惩罚,那这样并不能让汽车更加专注驾驶,并不能起到降低风险的实际效用,因此这种归责是无效的;再比如说,如果将自动驾驶汽车的技术检验的责任交给车上乘客,那么不具备专业知识的乘客根本无法进行检验,那么这样的规则就是不符合实用原则的。

以上两条原则既考虑到了责任承担主体的有效性,有利于促进相关主体不断改进设计、精进技术、谨慎驾驶;也不会降低各主体开发和使用权自动驾驶的积极性;同时这两条原则还具有顶层设计感,没有对细节指手画脚,但是却指明了方向。

5 自动驾驶的隐私困扰与伦理原则

由于自动驾驶的智能决策依赖于大数据,不可避免地带来了隐私困扰,为了驾驶安全,应该多公开公司相关数据,但是为了保护版权、商业秘密,似乎又应该对这种公开予以限制;同理,为了用户的交通安全,应该多共享用户的相关数据,但为了用户的隐私权,似乎又应该尽量少收集用户的数据。可见隐私问题并不是一个独立的问题,它牵涉到其他同样敏感的问题。进一步我们要问,公开多少数据,向谁公开数据?应该如何设定数据的隐私保护原则?

5.1 自动驾驶的隐私困扰

学界对有关用户数据隐私的研究集中在隐私泄露与安全风险方面,Glancy^[35]、Grunwald^[36]等认为,当人工智能取代驾驶员时,每一辆自动驾驶汽车的信息均是大数据的一部分,个人隐私权利也变得非常重要。倘若这些数据丢失、被盗或遭遇黑客入侵,将极有可能被恶意泄露或利用。鉴于自动驾驶所依赖的大数据技术,Rannenber^[37]强调了行驶数据的收集和可能出现的滥用问题,认为数据的网络化将改变传统驾驶中的“自由”概念,使用自动驾驶汽车受限于数据而存在不自由的危险,在相关立法中应注意自动驾驶汽车消费者的隐私保护。钟纯^[38]认为,尽管无人驾驶是推进人类进步的一项新技术,但它在一定程度上带来泄漏产品使用者隐私以及模糊产权归属等问题。人们不禁要问,自动驾驶汽车可以收集哪些信息?谁有权获取这些信息?这些信息在保存多久之后应该删除?在电车难题中,人们会讨论涉事人的胖瘦,除此之外一个人的年龄和性别也会影响事故的严重程度,这都会为自动驾驶汽车的安全选择提供参考,那么是否要允许自动驾驶系统收集这方面的数据呢?还是说出于隐私权的考

虑，只允许收集路人的位置数据呢？

还有一个令中国用户苦恼不堪的问题，自动驾驶系统收集的这些数据，是否会以“为了更好地服务用户”的名义，共享给其他智能产品，从而其实也就让用户的数据落到了其他人的手上呢？我们经常办了一张银行卡、填了一张单子以后很快就收到卖房广告，这意味着银行把我们的数据卖给了房地产开发公司等营利机构，类似的末端行为是否会在自动驾驶汽车上重演呢？

5.2 保护自动驾驶数据隐私的原则

目前的数据隐私保护规定多以数据相关的商业主体为主要规制对象，然而，商业主体对隐私保护的初衷是合规经营和获取商业利益，其合规或保护程度从根本上取决于数据的商业价值变动，很难保证持久性和公平性。因此，从充分保障各数据主体的隐私权利出发，自动驾驶数据隐私保护应当遵循如下原则：

政府管控数据输入原则。抓输出不如抓输入，用户的数据一旦落入自动驾驶汽车公司手里，就很难保证公司不滥用用户数据，因为我们不可能紧盯这些公司的每一分每一秒每一个员工的所作所为，所以不可能确保没有一个员工滥用数据，把数据卖给其他公司，但是我们可以先在数据落入公司手中之前就由政府掌控的数据过滤器将数据过滤一遍，变成只对自动驾驶有用的数据，而相关的用户姓名、身份、体型等信息一律过滤掉，这样汽车公司即使拿到了这些数据，其他公司也不会对此感兴趣了。该原则与企业经济利益相冲突，一定程度上较难获得企业理解，其贯彻落实需要社会具备强有力的政治领导核心，这不失为与中国治理实践相适应的一种选择。

6 结论

本文讨论了自动驾驶存在的五类伦理风险，包括两类与汽车事故有关的问题：疯狂电车思想实验中的决策问题和事故后的责任归属问题；两类与用户有关的问题：用户在行车过程中的自主性问题和用户在行车中产生的数据的隐私权问题；以及与全社会有关的问题：自动驾驶所创造的价值公平分配问题。其中隐私问题是一个重要的

实践问题，而不是一个复杂的理论问题，因为人人都同意保护消费者的隐私，而其他四个问题全部都是两难的理论问题，并且共有一个深层的伦理关切：权利 - 责任的分配。诚然，疯狂电车的思想实验看起来是在追问应该选择直行还是开入岔道，但如果承认这个问题已经由思想家争论了数百年，我们不可能短期内给出所有人都同意的解决方案，那么这个问题的解决方案就不是直接给出答案，而是回答什么人可以参与回答，具体而言，应该是由伦理学家给出底线，然后设计一套进行伦理决策的全民投票机制；事故的责任归属问题显然属于权利 - 责任的分配问题；用户的自主性需要通过赋予用户自主权来实现，自主权是权利的一种，当然享有这一权利的人也要承担相应的责任，所以它同样是权利 - 责任的分配问题；自动驾驶创造的价值的分配需要根据相关人员的相关权利的多少来进行，因为根据权利多少进行的分配才是公平的分配。由此可见上述四类问题本质上都是权利 - 责任的分配问题。

权利 - 责任的分配问题之所以成为自动驾驶的伦理风险共有的核心问题，或者说自动驾驶相关的权利 - 责任的分配之所以尤其难以解决，是因为自动驾驶作为一种深度学习类型的人工智能技术，它可以部分代替人类的脑力劳动，但同时人们对它的一些特殊表现，包括具有创造力的表现和造成意外损失的表现，却又很难查明问题的根源。可以代替人类脑力劳动，这使得相关劳动的权利和责任的分配与以往不完全一样，人们会考虑是否应该赋予人工智能一部分权利和责任，毕竟它承担了一部分工作。难以查明问题的根源，又使得相关事故的归责与相关价值的分配难以明确，毕竟问题源自哪个人 / 人工智能都不知道，又怎么能够确定责任 / 权利属于哪个主体呢？由此可见，只有回答了人工智能时代的权利 - 责任分配问题，才能从根本上解决自动驾驶所产生的疯狂电车的决策问题、事故责任归属问题和用户自主性的问题，这也是后续需要研究的内容。■

参考文献：

- [1] Foot P. The problem of abortion and the doctrine of the

- double effect[J]. Oxford Review, 1967(5): 4-6.
- [2] Millar J. An ethics evaluation tool for automating ethical decision-making in robots and self-driving cars[J]. Applied Artificial Intelligence, 2016, 30(8): 787-809.
- [3] Joe M. How will self-driving cars make life or death decisions?[EB/OL]. [2020-05-16]. <https://www.weforum.org/agenda/2016/08/the-ethics-of-self-driving-cars-what-would-you-do/>.
- [4] Hauser M, Cushman F, Young L, et al. A dissociation between moral judgments and justifications[J]. Mind & language, 2007, 22(1): 1-21.
- [5] Noothigattu R, Gaikwad S S, Awad E, et al. A Voting-Based System for Ethical Decision Making[EB/OL]. (2017-09-20) [2020-05-16]. <https://arxiv.org/abs/1709.06692>
- [6] 和鸿鹏. 无人驾驶汽车的伦理困境、成因及对策分析[J]. 自然辩证法研究, 2017(11): 60.
- [7] Goodall N. Ethical decision making during automated vehicle crashes[J]. Transportation Research Record: Journal of the Transportation Research Board, 2014(2424): 58-65.
- [8] Gerdes J C, Thornton S M. Implementable ethics for autonomous vehicles[C]. Maurer M, et al. Gerdes Barbara Lenz Autonomous Driving. Berlin Heidelberg: Springer, 2016: 87-102.
- [9] Holstein T, Dodigcrnkovic G, Pelliccione P, et al. Ethical and Social Aspects of Self-Driving Cars[EB/OL]. (2017-09-20) [2020-05-16]. <http://arxiv.org/abs/1802.04103>.
- [10] Luetge C. The German Ethics Code for Automated and Connected Driving[J]. Philosophy & Technology, 2017, 30(4): 547-558.
- [11] 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, 20(3): 1-7.
- [12] 余露. 自动驾驶汽车事故算法及其限度[J]. 自然辩证法通讯, 2019, 41(1): 20.
- [13] Millar, J. Technology as moral proxy: Autonomy and paternalism by design[J]. IEEE Technology and Society Magazine, 2015, 34(2): 47-55.
- [14] Cunneen M, Mullins M, Murphy F, et al. Autonomous vehicles and avoiding the trolley (dilemma): Vehicle perception, classification, and the challenges of framing decision ethics[J]. Cybernetics & Systems, 2019(4): 1-22.
- [15] 尤瓦尔·赫拉利. 未来简史[M]. 北京中信出版社, 2017: 309-313.
- [16] 王茜. 一季度全国机动车保有量突破3亿[EB/OL]. [2020-06-30]. <https://www.mps.gov.cn/n2254098/n4904352/c5682353/content.html>.
- [17] Fleetwood J. Public health, ethics, and autonomous vehicles[J]. American Journal of Public Health, 2017, 107(4): 532-537.
- [18] National Highway Traffic Safety Administration. Automated vehicles for safety[EB/OL]. [2020-06-30]. <https://www.nhtsa.gov/technology-innovation/automated-vehicles>.
- [19] 宋卫平. 美国加利福尼亚州建自动化车辆专用公路[J]. 轻型汽车技术, 2003(02): 22.
- [20] Autonomous Vehicles. Self-driving vehicles enacted legislation[EB/OL]. [2020-06-30]. <https://www.ncsl.org/research/transportation/autonomous-vehicles-self-driving-vehicles-enacted-legislation.aspx>.
- [21] Juice. 重磅! 自动驾驶公路国标出炉, 道路数字化大幕开启! [EB/OL]. (2020-06-30) [2020-07-30]. <http://www.cheyun.com/content/34624>.
- [22] Watzenig D, Horn M. Automated Driving[M]. Berlin: Springer International Publishing, 2017: 17-27.
- [23] Federal highway administration. Cooperative driving automation[EB/OL]. [2020-06-30]. <https://highways.dot.gov/research/operations/Cooperative-Driving-Automation>.
- [24] 若风. 美国加州将设置自动驾驶专用道路限速近200Km/h [EB/OL]. [2020-06-30]. <http://news.mydrivers.com/1/646/646131.htm>.
- [25] 王鹏, 雷诚. 自动驾驶汽车对城市发展的影响及规划应对[J]. 规划师, 2019, 35(8): 81-86.
- [26] 约翰·罗尔斯. 正义论[M]. 何怀宏, 何包钢, 廖申白译. 北京: 中国社会科学出版社, 1988: 292.
- [27] Johnson D G. Computer systems: Moral Entities but not Moral Agents[M]. The Netherlands: Kluwer Academic Publishers, 2006: 195.

- [28] 张力,李倩.高度自动驾驶汽车交通侵权责任构造分析[J].浙江社会学,2018(8):35-43,156.
- [29] 冯洁语.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.
- [30] Sheriff K D. Applying the sudden emergency doctrine to autonomous vehicles: The case of the speeding truck and the polar bear[EB/OL]. [2020-08-04]. <https://ssrn.com/abstract=2807597>.
- [31] 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):75-88,169.
- [32] 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.
- [33] 冯洁语.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):147.
- [34] 焦芳.“自动驾驶”出交通事故谁担责?[EB/OL]. [2020-07-05]. https://www.sohu.com/a/308147308_100018177.
- [35] Glancy D J. Privacy in Autonomous Vehicles[J]. Santa Clara Law Review, 2012(4): 1 171-1 239.
- [36] Grunwald A. Societal Risk Constellations for Autonomous Driving. Analysis, Historical Context and Assessment[C]. Maurer, M., Gerdes, J. C., Lenz, B., Winner, H. Autonomous Driving: Technical, Legal and Social Aspects. New York: Springer Publishing Company, Incorporated, 2016: 651.
- [37] Rannenber K. Opportunities and Risks Associated with Collecting and Making Usable Additional Data[C]. Maurer, M., Gerdes, J. C., Lenz, B., Winner, H. Autonomous Driving: Technical, Legal and Social Aspects. New York: Springer Publishing Company, Incorporated, 2016: 497-508.
- [38] 钟纯.人工智能视域下的伦理挑战与应对——以无人驾驶作为分析对象[J].唐都学刊,2019(3):65-71.

Thoughts on the Five Ethical Issues and Principles of Autonomous Vehicles

LI Fang¹, PEI Yu², XU Feng¹, GAO Fang¹, JIA Xiao-feng¹

(1. Institute of Scientific and Technical Information of China, Beijing 100038;

2. School of Marxism, Guangxi University of Chinese Medicine, Nanning 530200)

Abstract: Autonomous driving replaces human drivers in driving decision-making and driving operations, requiring a re-examination of related ethical issues from the entire labor process. In practice, to break through the ethical dilemma of machine decision-making, the ethical consensus that everyone is equal should be the bottom line, and public opinion surveys should be the specific selection criteria. In order to try to avoid the autonomy of autopilot systems threatening human decision-making autonomy, it is advisable to establish the default principle that the carrier has no immunity. In order to improve social equity, it is necessary to follow the principles of unified planning and preference for the most disadvantaged. In order to effectively assign responsibility for autonomous driving accidents, Aokang's razor principle and the principle of effectiveness practicality in taking responsibility should be followed. In order to achieve a balance between "convenience of data access" and "privacy protection", the principle of government control of data input should be followed.

Key words: autonomous vehicles; ethics; artificial intelligence