自动驾驶汽车事故伦理分析及应对策略

褚灵强 计算机学院 212050175

**摘要：**无人驾驶汽车在改变人类出行方式的同时，也给人类社会带来了巨大的挑战。本文介绍了当前自动驾驶汽车事故中存在的道德困境，指出了道德困境产生的原因，以及在事故中各方可能存在的责任。同时，就缓解当前存在的道德困境提出了相关建议，如通过构造具体责任人来取代自动驾驶这一抽象责任人，以及通过法律明确各方责任界限，完善补偿机制等方法。

**关键词：**人工智能；工程伦理；自动驾驶汽车

# 1、引言

随着新一代人工智能算法和产品的快速开发和部署，人工智能在各个领域发挥着越来越重要的作用，并对现代社会结构产生了重大影响。神经网络模型和算法已被广泛应用于各种决策场景，如刑事审判、交通控制、金融贷款和医疗诊断等等。

腾讯研究院院长司晓介绍，全球人工智能创业企业已达2000多家，仅腾讯公司就有四大实验室正在开展人工智能相关研究。事实上，有些人工智能商业场景迟迟不能落地，不是被技术“绊”住了，而是伦理研究“拖了后腿”。事实上，我们每个人每天都会做出各种的道德决策。当一个司机选择踩刹车来避免撞上一个乱穿马路的行人时，他就做出了将风险从行人身上转移到车内人身上的道德决策。但自动驾驶中有一个著名的伦理困境，即当它无可避免要撞人时，是撞向人多的一边还是人少的一边。这是一个典型的伦理困境，它表明了伦理在技术发展中的重要性。

同时，随着基于人工智能的自动决策系统得到不断的推广和应用，也在安全性和公平性等方面带来潜在的风险，并且已经造成了几例惨剧。特别是在自动驾驶系统的安全性方面，美国亚利桑那州一名女性在过马路时被一辆Uber无人驾驶SUV撞倒并最终死亡，这是全球首起自动驾驶汽车撞死人类的事故，引发了广泛关注和担忧。自动驾驶汽车是目前民众接触最多的人工智能应用之一，且由于它的实在性，其所带来的风险、道德问题也更加显著。

# 2、自动驾驶汽车的道德困境

人工智能的责任伦理问题是现在极为突出的一个道德困境。在世界范围内，自动驾驶技术在实际测试、使用中由于操作失误、设计缺陷等出现了多起伤亡事故。2016年，美国佛罗里达州某驾驶人员在开启自动驾驶情况下与一辆拖拉机拖车相撞后不幸死亡。2016年，中国发生国内首起自动驾驶伤亡事件，特拉斯车主在京港澳高速河北邯郸段直接撞击道路清洁车，所驾驶车辆当场严重损坏，驾驶人员当场死亡，随后家属向特拉斯公司发起索赔，经过一年多取证，特拉斯公司终于承认发生事故时汽车处于自动驾驶状态，自此后特拉斯公司将自动驾驶翻译为自动辅助驾驶。

在机器人伤人事件中，法律责任的界定是非常困难的。显然人类不可能要求机器人本身负责，因为它是不具备人类意识的人工制造物，我们无法起诉机器人。因此，责任将在使用者、销售者与制造者之间划定，因现阶段的人工智能并非早期机械化机器，现代人工智能拥有学习能力，能自主运算、自主运行，那么事故发生时难以判断出错阶段，导致安全事故责任难以厘清。

根据传统责任理论，某人应对其行为负责意味着将某种能力归属此人，并且，其行为是其运用此种能力的结果。自亚里士多德以来，使得归责成为可能的能力被普遍认为是对行动的控制和觉知。然而，行动者对其不能控制的事情，比如天要下雨，则无需负责。行动者能够控制行动通常意味着他能够做不同于他实际所做的行动。比如，一个行动者在路口选择了往右走，如果他能控制自身行动，那他当时也可以选择往左走。这种形式的控制体现了常识意义上的“自由意志”。一些传统理论的确认为自由意志是责任归属的必要条件：因为你的行动是你自由选择的结果，所以你应该对它负责。

传统责任理论的另一个原则是，责任的归属以行动者关于自己行动的自我知识为条件。如果一个人根本不知道自己在做什么，比如睡着的时候，他是不用对其行为负责的。如果一个人知道自己在做什么，但他还是选择那样做，那么他应该对自己所做的事情负责。更进一步的要求是，当一个人被问到为什么那样做时，他有能力为自己的行动提供基于他信念和意图的解释。

但传统责任理论无法清晰地告诉我们，在一场因自动驾驶汽车导致的事故中，究竟应该是由谁来承担责任。核心原因在于，对于自动驾驶汽车来说，它没有通常意义上的操作人员，也没有一个可完整定义的操作模式，这是由于人工智能的黑盒状态导致的。自动驾驶汽车的研究人员并不是清晰地给定了驾驶规则，而是由人工智能无法完全预测的自动做出决定。并且，人工智能具有自主学习的能力，研究人员无法完全预知自动驾驶汽车的行为。此外，如果是需要驾驶员介入，而驾驶员没有及时介入的情况下，我们可以认为责任人是驾驶员。但实际情况是，当紧急情况发生时，驾驶员往往没有足够的时间介入驾驶。因此，按照传统责任理论，我们无法将责任归结于驾驶人或研究人员。这导致一个困境：当我们想追责时，却找不到合适的承担责任的主体，从而产生了责任空缺。

# 3、自动驾驶汽车伦理问题可能的责任人

## 3.1 驾驶人存在的责任问题

对于驾驶人来说，责任问题分为两种情况，即目前的自动辅助驾驶和完全自动驾驶，区别在于驾驶员是否能够对驾驶过程进行控制。对于完全自动驾驶来说，驾驶人在驾驶过程中无法或者很难及时介入驾驶，因此在事故中驾驶人无需承担主要责任。但由于其选择了自动驾驶，驾驶人应该可以认识到自动驾驶可能带来的风险，那么其自主选择了自动驾驶，便需要为事故承担一定责任，但这种责任应该是次要的。对于自动辅助驾驶来说，在驾驶人可以干预汽车运行的情况下，驾驶人自然有义务注意道路和交通，并在必要时进行干预，以避免发生事故。然而，事故通常并不容易预见，尤其是在驾驶人有明显的疲劳、分心的情况下，如果要求驾驶人必须时刻关注路面情况，那么自动驾驶也就失去了它的意义。所以，除非驾驶人在一开始就做出了错误的决定，比如在闹市区选择过大的车速等，否则很难对其进行追责。

## 3.2 自动驾驶研究人员的责任问题

对于自动驾驶研究人员来说，在软件没有恶意编程、系统没有安全漏洞的情况下，是很难被追责的。当自动驾驶系统存在明显漏洞，如特斯拉动能回收系统在湿滑路面情况下，会导致用户刹车踏板变硬的情况下，自动驾驶研究人员应该被追究责任，因为这个是一个清晰的可选择的逻辑问题，是由于研究人员在编程过程中对情况考虑不全面，系统控制逻辑混乱所导致的。但如果是自动驾驶系统的驾驶决策出现问题，研究人员是很难对其负责的，因为驾驶决策并没有清晰的逻辑选择，是由人工智能在黑盒状态下自主进行的，研究人员很难预测，更无法对其负责。

# 4、自动驾驶汽车伦理问题可能的解决方法

## 4.1 自动驾驶汽车安全公司

在责任空缺或者说只存在抽象责任人的情况下，一个自然而然的方式是引入一个具体化的责任人。比如，可以成立自动驾驶汽车安全公司，由其开发一个标准化的安全包，对自动驾驶过程的安全问题负责。其安全包的原则应该是尽可能明确、清晰的，以减少对抽象责任人的引入。同时还应包含一个监测模块，负责记录自动驾驶系统接收到的所有指令、升级情况以及关键运行数据，并定期对智能系统进行测试。当自动驾驶安全公司确定所有行动者都无重大过失时，惩罚落实并终止于自动驾驶安全公司；当自动驾驶发现某个人类行动者有重大过失时，人类行动者就应该接受惩罚。一般来说，惩罚的目的既是为了让责任人意识到自己的错误，也是为了让所有社会成员知道任何挑战法律规范的行为都要付出代价，也可以促进该类事件的改善。引入专门的自动驾驶汽车安全公司，当没有明确人类责任人时，将其作为责任主体，通过对其的惩罚，可以使其认识到自己在安全包上存在的漏洞，并且对安全包进行改进，很好的实现了惩罚的目的。

此外，为了保证自动驾驶汽车安全公司可以对自动驾驶导致的事故负责，自动驾驶汽车安全公司应该至少由制造商、拥有者、使用者等成员组成，且成员应当由法律保证强制加入。即，制造商在生产自动驾驶汽车前、拥有者在购买自动驾驶汽车前、使用者在使用自动驾驶汽车前，必须加入自动驾驶汽车安全公司；而自动驾驶汽车在销售前也必须在一家自动驾驶汽车安全公司注册。

当然，自动驾驶汽车安全公司可以由政府出面组建，强制要求所有相关公司加入。并且，自动驾驶汽车的广泛普及，实际上是由全体公民做出的决定，由此引发的事故由代表全体公民的政府进行负责也是合乎伦理道德的。

## 4.2 完善的法律与补偿机制

在无法找到明确责任人的情况下，对事故进行处理的另一个方式是尽量弥补事故所形成的伤害，即对自动驾驶汽车的伤害采取补偿。补偿机制应该尽可能地消除受害人所受的损害，并且采取保险或其他集体化减轻风险的方案。从这个角度来看，受害者可以随时获得足够且合理的补偿，而不需要诉诸于法律机制，法院也不需要扩大现有的责任模式，以确保受害者能得到补救。

因为在自动驾驶汽车正常运行的情况下，事故的发生可以被视为一种意外，那么只需提供一定的补偿就可以解决这一问题。正如电刚刚出现的时候，有一个伦理学家站出来说：电力不可能普及的，因为人们在使用电的过程中有可能被电死，但是在电杀死人的过程中，电力作为一项技术无法承担责任。但是100年后的今天，我们并没有遇到因电导致的伦理问题。原因在于，我们为各方指定了完善的法律法规，无论是制造者还是使用者，只要在合法框架下行动就无需承担责任。当制造商制作出不符合规范的电器时，他需要承担责任；当使用者知悉用电规则却没有遵守规则时，需要承担责任。当各方都没有过错，但仍然电死人时，普遍被认为是不幸的意外。因此，只要明确各方的责任与义务，并提供完善的补偿机制，可以在一定程度上缓解自动驾驶汽车的道德困境。当然，这会需要一定的时间，来完善法律法规，以及让民众认可符合规范下的自动驾驶汽车的安全的。

# 5、结语

在由自动驾驶汽车引发的事故中，责任人的界定是十分困难的，很有可能各方都不存在过错，我们无法找到明确的责任人对这场事故负责，而自动驾驶汽车自身作为机器还无法跨越技术和概念框架，不能作为道德责任的归责对象，导致道德困境的产生。本文对事故各方进行了可能的责任认定，并且给出了当各方无明确责任情况下可能的解决方法：一方面政府需要完善法律法规，明确各方的责任边界，并建立完善的补偿机制；另一方面，可以引入一个最终责任人，自动驾驶汽车安全公司，由它负责监管自动驾驶汽车中的所有安全问题，并对其安全负责。

我相信，随着自动驾驶技术的发展和法律法规的完善，以及民众对自动驾驶汽车的信任，自动驾驶汽车事故的道德困境必将和百年前电的道德困境一样逐渐消失。

# 6、参考文献

[1] 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.

[2] 吴戈. 人工智能发展带来的问题及其伦理思考[J]. 中州学刊, 2021, 43(03): 93-95.

[3] 李园园. 社会规范下公众对自动驾驶汽车道德判断及其对行为偏好影响研究[D]. 浙江工业大学, 2020.