·人工智能与未来社会（二十四）·

自动驾驶因何“无人”

——对人工智能体“无人化”的伦理批判

田海平[1]　葛中传[2]

【内容摘要】对自动驾驶技术“无人化”目标的伦理批判，需超越决策困境、责任鸿沟等表层难题，深入其背后人工智能体取代人类主体性的哲学危机。以亚里士多德“四因说”为框架剖析可见，“无人化”趋势根植于技术资本借算法理性与数据拜物教，对人类行动自由与德性实践空间的系统性褫夺。应对这一危机，关键在于重申“人为自己立法”的根本原则，构建以“为他者而行动”为指向的德性伦理空间，从而引导自动驾驶技术摆脱“无人化”迷思，回归以人为本的价值轨道。

【关 键 词】自动驾驶　无人化　人工智能体　技术伦理

【作　　者】1 田海平，北京师范大学价值与文化研究中心研究员、哲学学院教授；
2 葛中传，北京师范大学价值与文化研究中心、哲学学院博士研究生。（北京 100875）

【基金项目】教育部人文社会科学重点研究基地（北京师范大学价值与文化研究中心）重大项目“现代科学技术发展的价值问题研究”（25JJD720007）

自动驾驶技术正在重塑人类的出行方式。当资本通过“无人驾驶”的话语将“无人”塑造为目标时，一个更为根本的伦理问题却被这一技术光环所遮蔽：自动驾驶因何“无人”？这一追问并非意在否认技术带来的便利，而是要求我们穿透“工具善”的表象，对潜藏于其下的“无人化”逻辑进行一场前提性的批判。当前学界的讨论多聚焦“无人化”所引发的具体伦理困境，如算法在“电车难题”中的决策悖论、事故后的“责任鸿沟”以及对社会结构的长远影响。然而，这些困境仅是表象，其本质根源于人工智能体作为一种“无主体之主体”，在动力、目的、形式和质料层面系统性地替代乃至褫夺了人之主体性。技术资本恰恰利用了对“无人”的追求，将一种看似中立、高效的算法理性，转变为侵蚀人类行动自由与德性实践空间的隐性权力。

因此，有必要超越对具体困境的被动回应，从伦理学的根本处发起一场批判性考察。本文旨

在揭示，“无人化”并非技术发展的必然归宿，而是特定价值选择与权力运作的结果。通过借用亚里士多德的“四因说”，本文将剖析驱动“无人化”的内在逻辑与外部动力，阐明其如何通过承诺“绝对安全”与“极致效率”，最终将人类世界导向一种被数据拜物教和技术理性所“集置”的生存危机。通过这种根源性的批判，尝试为自动驾驶技术探寻一条属人的、向善的未来路径。

自动驾驶“无人化”的伦理困境

追溯起来看，“无人驾驶”的提法始自“自动驾驶分级”。1965 年，人工智能先驱约翰·麦卡锡在他的论文《电脑控制汽车》中首次提出了“自动司机”（automatic chauffeur）的概念——计算机通过摄像头采集数据模拟人类视觉，以实现对汽车的驾驶。这一设想为自动驾驶汽车提供了设计蓝图。在此后的半个多世纪，自动驾驶始终是技术领域的热门话题。2021 年，“国际汽车工程师协会”（SAE International）发布了“自动驾驶汽车分级标准”。该标准依据智能化程度将“自动驾驶汽车”分为五级，分别为：部分驾驶辅助（L1）、组合驾驶辅助（L2）、有条件自动驾驶（L3）、高度自动驾驶（L4）、完全自动驾驶（L5）。[①]当今，符合“组合驾驶辅助”（L2）标准的汽车已成为新型汽车的主流，而各国也相继开始“有条件自动驾驶”（L3）汽车试点运行工作。[②]

有人据此宣称，我们已经抵达了“无人驾驶”的前夜——L5 级“完全自动驾驶”就是“无人驾驶”，即通过人工智能加持的机器学习，使自动驾驶汽车在最大程度上实现“合规范”的驾驶，而此时的人类驾驶者无须做出决策，自动驾驶汽车上的人只是“乘客”而非“驾驶员”。显然，自动驾驶汽车并非一开始就等同于“无人驾驶汽车”，而是因其“方向盘后面没有人”而被预设为自动驾驶之“目标”，从而被称为“无人车”或“无人驾驶汽车”。[③]尽管目前通用的自动驾驶技术还没有达到能够脱离人类驾驶者而自主运行的程度，但现实中的诸多技术尝试都在向着真正的“无人驾驶”迈进。例如，2021 年开始试运行的“萝卜快跑”无人出租车，截至目前，其“自动驾驶里程超过 2.4 亿公里，其中全无人里程超 1.4 亿公里”。[④]美国车企“特斯拉”也于 2025 年 6 月在其家用车型上实现全球首例在公共高速路上实现车内无人的自动驾驶。[⑤]可以说，“无人驾驶”是未来高度智能化交通的重要组成部分，而先行开展针对其伦理风险的反思和批判，是一种必不可少的前瞻性审视。

自动驾驶的“无人化”面临决策、规则和后果三重伦理困境。首先，自动驾驶“无人化”面临决策困境。以无人车为自动驾驶“智能化行动”的目的善，隐含着无人车作为人工智能体（行动者）试图取代人的“行动权”的预设。于是，交通工具似乎成为行动的主体，而人则仿佛变成了工具的辅助。诸多讨论自动驾驶决策问题的研究，都倾向于将“决策困境”理解为“无人驾驶汽车通过机器学习后如何‘模仿’人做出正确决策”。例如，瓦拉赫和艾伦就借用“电车难题”来描述这一困境。[⑥]但究其根本，无人驾驶技术的机器学习来自无数人类驾驶者的行动事例，而它在危机时刻的逻辑判定可能只依靠算法设计者写下的代码。[⑦]因而，无人驾驶算法的决策在根本上还是人的决策。对于无人驾驶的“决策困境”的讨论，学界主要分为“功利主义”和“义务论”两派。以戈戈尔为代表的功利主义一派认为，个人选择可能会导致囚徒困境，因此应当设置强制性的伦理原则以保证交通安全的最大化和对人伤害的最小化。[⑧]古柯–维拉等义务论一派则认为：功利主义设定的利益比较难以实现，应当依照正当性原则为无人驾驶汽车设定诸多义务规则。[⑨]具体到应用阶段，则有功利主义算法、罗尔斯算法、制动力学算法、伦理旋钮等不同道德算法。[⑩]

但如同“电车难题”揭示的那样，人的所有行动在道德层面没有绝对的正确或错误。自动驾驶“决策困境”的实质不是“自动驾驶汽车的行为决策是否‘正确’”，而是“自动驾驶汽车的行为决策是否是人‘想要’的”。例如，为了拯救生命，人类司机可能会选择冒险超速和闯红灯，在同样情况下无人车能否超越既定规则？对于自动驾驶算法的设计者而言，无论基于何种考虑都应当将使用者的安全放在第一位，但是人类驾驶者的行动则因人而异。[11]个体的诸多行动，尤其是出于自身德性品质的行动，是难以被理性算法复刻的，强而行之会造成对其他不依据德性行动的人的不公正。更进一步可能造成的是算法对控制权的争夺：2019 年 3 月 10 日，埃塞俄比亚航班坠毁事故震惊全球，据事后调查，事故源自波音 737-MAX 飞机的自动驾驶系统在错误情况下同驾驶员抢夺操作权从而导致飞机失控。[12]算法所做出的“不符合人类驾驶者自身要求的决策”同它做出的“完全错误的决策”实际上是相似的，尽管前者未必会造成难以接受的严重后果，但从长期的人类整体道德进步的视角来看，其所造成的后果可能更加严重。

其次，自动驾驶“无人化”面临责任鸿沟。以往技术背后的“他者”并不完全参与到技术主体行动中，但无人驾驶技术背后的他者却在某种程度上成为了行动主体。无人驾驶技术背后的“他者”（或“客体”）可能会取代使用者的主体地位。这导致无人驾驶在道德上面临“责任鸿沟”：无人驾驶的责任主体是车辆使用者还是无人驾驶车辆本身？抑或是算法背后的控制者？马蒂亚斯认为，由于人工智能算法强大的机器学习能力，以往的归责机制将无法适用，责任将失去主体。[13]国内有学者认为“责任鸿沟”根本上来自人工智能体未获得责任主体地位；[14]或者，无人驾驶汽车中的人类驾驶者也应当负有一定责任。[15]显然，无人驾驶汽车遭遇过渡时期的技术、法律和伦理层面的归责难题。例如，当无人驾驶汽车遵守安全驾驶规范，但人类驾驶员不遵行安全驾驶规则时；或者，当人类驾驶员超速行驶，而无人驾驶汽车偶尔跟进超速才安全时；又或者，当无人驾驶汽车自动调整速度以迎合人类驾驶员的危险驾驶行动时，超速行驶的无人驾驶汽车会收到交通处罚单吗？惩戒一条算法、一辆汽车是毫无意义的行为。或许，可以像田纳西州法律那样将无人驾驶汽车的开发者和制造者作为“真正驾驶自动驾驶汽车的人”；[16]但同样，人工智能的算法黑箱也可以为其提供庇护空间，算法的错误完全可以被解释为系统不可预知的漏洞。更严重的问题在于，假使自动驾驶算法在面临抉择危机时突然将控制权交还给人类驾驶者，人类驾驶者无法在极短时间内避险，人类驾驶者反倒成为算法规避责任的“替罪羊”。例如，2025 年发生的“3·29 铜陵小米 SU7 爆燃事故”中，智能辅助驾驶系统（NOA）从发现道路异常情况到提醒驾驶员接管并退出系统，中间仅有 1 秒的时间。[17]反之，如果自动驾驶算法排除人进行瞬时决策行动的可能，人类驾驶者只能为无人驾驶汽车设定起点和终点以供后者行动。此时的人类与其说是驾驶者，倒不如说只能作为乘客——技术背后的他者（资本）完全地占据了技术使用者的实践过程，占据了主体的自我意识。这种被技术强加的“外在主体性”使得主体的全部行动都是他者主导下的行动，主体失去了自身的自由而无法自控，自然也因此失去了承担责任的能力。那么问题就再次回到了“处罚一辆车是否有意义”的层面上了。

最后，自动驾驶“无人化”面临后果难题。驾驶将通过改变人的出行方式影响人的生存方式。威廉斯指出：“应当把技术理解为环境而不是客体。”[18]技术环境蕴含双重内涵：其一，技术因其为环境而是一种自在；其二，技术本身同人的生存息息相关，技术环境是一种人为环境，人借由技术创造了自身的生存境遇。无人驾驶的本质是构建了一种新的技术伦理关系，其核心是人同世界的交往实践。无人驾驶的“无人化”，不仅是行动者层面上的无人，更是人之生活世界的无人。这种后果难题，被称为科林格里奇困境：当初期自动驾驶“无人化”的后果不是很明显时，人们

并不致力于寻求改变；但当因何“无人”问题凸显且演变成对“无人化”的深度质疑时，“无人化”技术就构成了“生存集置”的基础性建构，此时想寻求改变就变得非常困难。[19]安东尼·汤森颇具前瞻性地分析了无人驾驶对于社会生产、公共政策、城市结构等人类世界诸多要素可能造成的整体性、不可逆的社会和伦理影响。[20]未来，无人驾驶汽车会展开一种独特的技术环境而覆盖整个生活世界；它不同于火车只能在铁路上运行，无人驾驶汽车会占据道路交通的绝大部分，而为了信息收集处理高效率和运行安全性，无人驾驶汽车要求智慧道路系统的建设和改造。[21]最终的结果或许是，庞大规模的智慧道路出于效率和安全的考虑会限制人类驾驶员驾驶的汽车进入。

被控制的驾驶：自动驾驶“无人化”的本质

“无人”，通常对应着“autonomous”（自主的）和“driverless”（无人的 / 没有驾驶员的）两种不同表达。人们在技术领域习惯使用“autonomous”；而在人文领域，“driverless”的用法则更为广泛。[22]这种微妙差异，彰显了人们对“作为人工智能体”[23]的无人驾驶的一种人文忧虑。尽管技术专家对机器主体地位或机器自主性似乎充满了信心，但对“人工智能体”可否或应否做到真正“自主”的顾虑则不能完全被消弭。忧思的焦点在于：如果人工智能体可以是具有自主性的“类主体”，那人之人类性又将何去何从？

因此，自动驾驶因何“无人”的追问，根源于“driverless”之“无主体之主体性”的特性。“无主体性”是说，无人驾驶汽车不存在人类驾驶者意义上的主体性；那么，真正控制汽车的“主体”或“主体性”又是何者？是什么取代了“人–主体”在进行自动驾驶？质而言之，无人驾驶技术的“无人化”根源于“人工智能体”所本有的“无主体之主体”的特性。有学者将其最终归入“人工智能体是否具有主体性”的问题上。[24]有观点认为，主体性同人工智能的自主程度相关联，强人工智能可能会拥有主体地位；[25]也有观点认为，可以赋予人工智能体技术性人格。[26]亦有反对意见认为，无论智能程度如何，任何人工智能体都只能作为工具而非主体。[27]笔者认为，以无人驾驶汽车为代表的人工智能体从来不构成任何主体——它是一种“类主体”。“其主体性，从根本上看，只是‘随附于人’的主体性，它从‘拟人化’的或者‘类人化’的关联性结构中获取某种意义的‘主体性形态’。”[28]它仅在讨论部分道德问题时短暂性地具有道德主体地位，但这种道德主体地位不是它的本质属性，而是被其背后的“人”附加的，真正的技术主体有且只有人自己。

因此，“无主体之主体”是作为人工智能体的“本质”而存在的，它一方面在现实行动中为人工智能体提供了“类主体”的道德地位，另一方面则导致了“无人化”的后果。“无人化”并不指人工智能体现实地消灭人类主体或“人–主体”，也不是指在人工智能体的应用中拒斥“人–主体”的参与，而是指人工智能体对“作为驾驶行为主体”的功能性取代（代理）。这种取代或代理，在表面上是人工智能体企图以自主性方式成为“自–主体”，从而将原本的人类主体从技术之使用中排除，但实质上是技术背后“非人–主体”对技术使用者“人–主体”的取代。

由“可受控制的驾驶”内含的四个面向[29]可对自动驾驶因何“无人”作进一步追问，即“为何‘驾驶’最终将要走向‘无人化’”？这“四个面向”包括动力因、目的因、形式因、质料因，它们分别代表了自动驾驶“无人化”趋势背后的社会动力、价值目的、形式结果和物质因素。

（一）自动驾驶“无人化”的动力因

“无人化”似乎首先来自某种社会需求的推动，如同工业革命时期社会生产的需要催生出了

汽车、火车等现代交通工具，“无人驾驶”似乎也来自人类对于“工具善”的追求。自人类发明轮子伊始，“驾驶”行为在带来便利的同时便产生了不容忽视的“不可控”。绕路、延迟、拥堵、事故以及对驾驶者的专业要求等，这些难以控制之处始终是技术要解决的目标。随着人工智能算法的不断迭代更新，通过大数据模型的运算来控制车道流量以缓解交通拥堵，通过整体性的信息收集和处理来规划最合适的路线以高效完成运输目的，通过更加智能理性的算法来控制车辆驾驶以避免人类驾驶员的非专业、非理性行为的自动驾驶技术试图逐步取代人类完成驾驶行为，由此便导致了对于驾驶“无人化”的追求。此为自动驾驶导致“无人化”的动力因——但是，这种因“工具善”而“无人”或“无人化”的动力学根据却并不构成解释“无人化”的真正理由。

从动力因看能动性或主体性，以无人驾驶为代表的“人工智能体”对于“无人化”的追求，在本质上不只是一种“物性”关联，而更是一种“人性”的关联。换言之，无人驾驶之所以“无人”，根本在于人工智能体背后的“自-主体”作为“他者”似乎正在逐步取消人类主体性地位，最极端的情况会导致“人-主体”沦为人工智能体“自-主体”的附庸。从动力结构看，无人驾驶“无人化”的动力因根据，隐含着某种“他者”作为“自-主体”借助人工智能体对“人-主体”的控制。这种控制分为两方面：一方面，人工智能体以自身的技术能力来控制人之实践行动；另一方面，人工智能体背后的技术资本和权力结构通过对算法的掌握来控制人之生活世界，从而在本质上完成“无人化”。

（二）自动驾驶“无人化”的目的因

无人驾驶对人之主体性的取代是对人之“能”和人之“卓越”（德性）的取代。主体行动来自主体自身之“能”的品质。在此维度，主体之“能”指向人之“能-实现”的德性善，它既包括“行动之能”（能动性），即在合理情况下以合理方式采取合理行动，还包括“责任之能”，即承担自身行动后果的能力。从行动逻辑看，行为的发生始于主体的动机，而行动之所以能够顺利发生，不仅在于行动主体“能控物”，更在于行动主体“能自控”，二者皆以行动主体的“目的”之所向为行动根据。

“无法自控”的情况表明：主体行动不由目的因支配，此即“德性之能”的丧失，表现为“失控”。由于“控”之主体既是责任主体，又是“责任之能”所从出者，它便是人为自身行动承担责任的“目的因”。作为技术形式的无人驾驶，根本上是人自身的主体之“能”的品质的一种延伸，即通过算法，人工智能总结出规范精准的驾驶行为模型。这一模型本身是人类驾驶者出色且负责任行动的一种凝结，它是人“驾驶之能”的对象化体现，是类人的主体之能。但是，也应该看到，随着人工智能技术的“具身化”，人的“身体本身”正逐渐从行动中“抽身而去”，无人驾驶的“驾驶之能”正与“人-主体”相脱离，而展现为一种“自-主体”之功能。对于无人驾驶汽车，人只需要简单几个动作，身体根本不需要参与到驾驶行动之中以展现“人-主体”的自身之“能-控”；反之，人在无人驾驶汽车中实际上是处于“不能-控”之状态：驾驶状态的好与坏，突发事件的判断与处理，因交给了人工智能体而同人的操作基本上无关。

更为严重的是，无人驾驶技术对“人-主体”的替代，造成了“因德性善”而“无人化”的技术统治后果。从人的整体性生存看，作为人之生存根本方式之一的交通、出行、上路、驾驶等将人排除在外，行动本身的生命性及其实际性已然消弭。不可否认，如同最初是人之“能”的延伸一样，无人驾驶的目的在于呼应“让生活更美好”，是人“愿好”“愿德性”的行动结果。但当无人驾驶完全取代人而成为行动之主体时，卓越的德性就成为算法的结果，关乎人的好生活不再

是一种生存性实践，而是一种事实性结果。但真正的好生活“在于实现活动”，“而实现活动显然是生成的，而不是像拥有财产那样地具有的”。[30]“无人化”驾驶技术在目的因层面所展现的已然不是人之德性，也难称“算法的德性”，它的一切结果都只是一种必然性事实。既然如此，人的根本性生存、实践还有卓越的可能吗？[31]

（三）自动驾驶的“形式因”

无人驾驶隐含因取消“人之驾驶行为”而展现自由的可能。但这种在“形式”上因自由而“无人化”的技术进步，实则强化了人对技术的依赖及“自由的丧失”。这是形式层面的悖论。由于人自身之“能”同人之“能在”密切联系，而人之“能在”是人之主体性自由的基本规定。无人驾驶通过对人之主体性的取代完成了对人之“能在”的取代。对人而言，世界之可能便转变为世界之必然：假使我想要在某时到达某地，我就必然能在某时到达某地。这既是技术承诺（我们因技术“无人化”而“自由”），也是技术之集置或“牢笼”——偶然性的排除使得交通走向了绝对的控制论。原本应当由人控制的无人驾驶技术反过来开始控制人，作为主体的人成为物的附庸，人之“能在”及其自由生命本质，反而“因自由之故”被技术之“能”所驾驭。自动驾驶在“形式因”层面展现为一种“自由之悖论”：“自由”最终有可能会演变为“控制”或“奴役”。

人类创造算法并开发自动驾驶技术以实现对车辆的控制。在这一过程中，自动驾驶技术无疑是人的造物或“奴隶”，是人的自由生命之“能在”对象化自身的产物。就算“自-主体”的自动驾驶技术有着所谓的“主体性”或“自由”，那也不过是人之主体性或自主性的延伸。问题在于，无人驾驶技术“僭越”了人的主体性，通过对整个驾驶过程的控制，算法将作为人之本质力量对象化到无人驾驶汽车之中，从而改变车辆的独立性。换言之，在无人驾驶的过程中，真正发挥自身能力、“拥有”车辆的是技术和技术背后的存在而非人类车主。因而，不必等到科幻作品中的极端设想成真，即便人工智能体的自我意识尚未觉醒，其对人的奴役也在悄然发生。“无人化”的本质隐蔽着技术对人的剥夺。“因自由”而展现的无人驾驶隐蔽着“全面技术控制”或技术利维坦式的“奴役”：一方面，在完全自动驾驶技术下人几乎无法进行任何自主行动，包括驾驶路线、车速等在内的全部驾驶过程都在自动驾驶技术的掌控之下，而人只能设定起点和目的地——与其说是人的设置促成了自动驾驶汽车的行动，倒不如说人成为自动驾驶汽车的启动程序；另一方面，一旦这种交通方式成了压倒性驾驶行动，无人驾驶的“自-主体”失控，不仅会影响个人生活，甚至会使交通陷入瘫痪。[32]

无人驾驶揭示了一种现实：人工智能体“无人化”面临权力的辩证法。以无人驾驶技术为代表的人工智能体的“无人化”，隐含数字资本谋求控制人的生活世界的行动。正如尤瓦尔·赫拉利所说：“目前最耐人寻味的新兴宗教正是‘数据主义’，它崇拜的既不是神也不是人，而是数据。”[33]“数据宗教”的本质是“数据拜物教”。它坚信数智时代的唯一通路是基于大数据算法的人工智能技术及其应用。人的一切问题似乎在这条通路上都能迎刃而解。作为其衍生，无人驾驶有导向这种“数据拜物教”的风险。无人驾驶的“无人化”，一旦转化为数字资本控制下的无人化，其中隐蔽的“数据拜物教”就会演化为资本权力对人的奴役。

（四）自动驾驶的“质料因”

从生存论视角看，无人驾驶作为“数据拜物教”的特例性力量，属于一种技术权力。追溯起来，这种技术权力是基于人工智能算法对人的能力的超越。人类面对无人驾驶技术背后的强大力量和

算法黑箱，不自觉地“神化”自己所创造的技术。自动驾驶的“质料因”不仅限于“使驾驶变得便利”，更在于其足以改变世界的对物质力量的“神化”。虽然作为个人消费品的自动驾驶汽车并没有脱离“商品”的形式，它实际地以其技术的“物的性质”掩盖了人本身“劳动的社会性质”。当自动驾驶汽车宣称最终会取代人类驾驶者时，实际上是技术背后的资本权力在“理性属性”“绝对权威”方面宣称取代人。一方面，资本为自动驾驶技术赋予“理性属性”。由于作为人工智能体的无人驾驶技术所基于的是数理性的逻辑算法，它的逻辑演绎和归纳既是对人类行为的机械模仿，又摒除了所有非理性因素，资本由此宣称自动驾驶技术能够实现更加安全和高效的驾驶行动；反之，人类驾驶者的驾驶行动就“显得”危险且不可控。另一方面，资本为无人驾驶技术树立了“绝对权威”。基于驾驶算法的相对稳定性，一套合理性、合规范的驾驶模式就会被自然而然地导出，这种“绝对”驾驶令人信服，掌握算法的人（资本）也就掌握了绝对的权力。通过上述两方面的分析可见，资本借助无人驾驶技术实现了“控制”的目的。一是无人驾驶技术要求对道路系统的智能化，从而对整个交通系统产生限制，原有的普通道路将被改造为无人驾驶汽车的专用道路，交通工具和道路深度绑定，人的“交通自由”被进一步压缩。二是无人驾驶主导的智慧道路系统会产生新的资本权力，交通运输垄断兴起。基于无人驾驶汽车运营的新的资本寡头可能会出现，控制社会中的大部分物质流通和社会交往。三是在资本权力主导下，无人驾驶技术会改造城市生活空间。以人为本的城市让位于无人驾驶汽车的高效运行，公共资源被无人驾驶汽车所挤占。四是人在世界中的绝大多数行动都成为算法背后资本控制下的既定行动，高度控制下的社会阶层可能会进一步固化，社会交往实践被压缩。人不再是“想去哪里就去哪里”，而是“只能去算法想让你去的地方”。基于上述“理性且绝对”的自动驾驶模型，整个生活世界可能都将为无人驾驶背后作为他者的资本权力所掌控，主体的整个生活世界都被他者所占据和控制，这就是无人驾驶技术“无人化”的本质。

“无人化”并非是指世界上不再有人，而是人的主体性精神被机器理性及其背后的资本权力所取代和掌控。一方面，“驾驶”不再需要专业性技术，只需一个想法就可以“想去哪儿就去哪儿”，这在表面上看是人拥有了强大的交通能力，是人之能力的巨大提升；但这种“完全的具身性”使得身体“抽身而去”，技术之“能”反而成为人的本质性能力。这是一种更完全的异化，它以一种“无痛苦”“无纷扰”的方式取代了人的主体性，将主体完全消解。另一方面，以往的资本权力主导下的商品拜物教的神秘本质在于隐藏了商品之为劳动产品背后的社会生产关系，而“无人驾驶”则将这种神秘性蔓延到符号化的规训之中，借助资本权力所赋予的力量建立了一套完整的社会文化规训体系，以隐秘的方式进入人的道德生活和本质性的生存活动中。

“回到人本身”：无人驾驶的伦理治理

对于人类个体而言，无人驾驶导致的“无人化”未必构成一种生存性伦理危机。往往只有在自动驾驶汽车危及个体生命时，这种“无人化”后果才会被察觉。但是，对于当下世界而言，“完全自动驾驶”就如同洛夫克拉夫特所创造的“克苏鲁”一样，不断在每个人的耳边和梦境中低语着，使整个世界被引向“无人化”的未来。这是一条通往救赎的“光明之路”，还是一条通往晦暗的“幽灵之路”？问题的关键在于，我们如何在以无人驾驶为代表的人工智能体“无人化”结构中“回到人本身”——这基于一个朴素的价值承诺：无人驾驶技术不是不要人，而是要回到人。

无人驾驶的“无人化”揭示技术作为“自–主体”介入人之生活乃至改变人之主体性的可能。而“自–主体”作为技术、资本、权力等的代理（agent），在人工智能体的行动中如何与“人–主体”相“联结”？这既是技术上一种全新的共在空间构建，更是伦理层面一种全新的道德世界之展现。就自我意识的运动而言，“自–主体”作为“人–主体”本质力量对象化（或异化）的产物，是“人–主体”类生命本质的“宿命”，而“人–主体”的自由又必定是在自身对象化本质中的自我坚持。因此，无人驾驶伦理治理的最高原则，必然优先强调回归“人–主体”的“德性之善”，以应对无人驾驶“无人化”的“决策困境”“责任鸿沟”和“后果难题”。

一方面，需要重审“技术目的”的德性展开，以便将“因何无人”的理由只限定在构建人的好生活的德性框架下。无人驾驶技术因其关涉人之本质性交往而更切近人之“好生活”的德性追求。着眼于无人驾驶技术“让世界更好”的品质之善，就必须将“人–主体”的“德性之善”置于最高地位。这种善不是基于功利主义的计算或某种规则信条的约束，从根本上，它基于“好的人”所建立的“好的算法”。当然，这并不意味着要让每一个程序开发者都成为道德模范，而是旨在建立一个基于技术并合于人之关怀的德性空间。

另一方面，需要重新界定由“无人驾驶”所开放的“空间联结”的伦理意义，以便让无人驾驶技术作为“空间联结”的新型纽带为世界提供“向善”的新可能——我们称之为“因美德而行动”的可能性，即借助无人驾驶，人与世界可以实现彼此通达无碍。从生存视角看，“我们从来都不是单独存在”，世界“在场且澄明”地对人展开为一种共在；主体和他者都处在一种澄明的共在空间中，其中的每个人“按其作为此在本身存在这样一种存在的方式，它是以在世的方式‘在’世界中的，而同时它又在这个世界中以在世界之内的方式来照面”。[34]这种“空间联结”的澄明，由无人驾驶的技术展现为“主体（无论是‘人–主体’还是‘自–主体’）‘因美德而行动’”。“因美德而行动”意味着自我与他者的“理解性的联结”。“理解”是在人的生活世界中去把握其自身存在可能性的能力，而自我与他者的“理解性联结”便是一种相互性的、以彼此为目的的对话，“是在问和答、给予和取得、相互争论和达成一致的过程中实现那样一种意义沟通”。[35]往深里说，自动驾驶汽车所要展开的“理解”的“技术环境”在本质上就是“沟通”的实现。“可被理解的”就是“可以沟通的”。“沟通”既是人类驾驶者同无人驾驶汽车之间的沟通，也是人类驾驶者同自动驾驶技术背后的人的沟通，即人与人的沟通，更是人类驾驶者同自身所处环境的沟通，即人同自我的沟通。基于此，一种应对以无人驾驶为代表的人工智能体“无人化”的伦理治理策略也就得到了进一步“勾勒”。[36]它包括如下三个建议。

第一，将“德性之善”置于无人驾驶技术的绝对优先地位。当前的诸多算法构建，其核心价值都在于形式过程的高效和安全，但“人是‘交通’的动物”，人不是无人驾驶技术要素或构成部分，而是相反，无人驾驶是人之能力和德性的一部分，无人驾驶技术的价值不在于“对人的价值”、而在于“在人之中”作为人之品质的价值。因此，构建一种面向和体现人之主体品质卓越的无人驾驶算法是首要的。它包括：其一，无人驾驶技术的最终目的是人的整体性发展，而非资本的积累；其二，“自主原则”是无人驾驶技术的最基本原则，必须首先保证人在自动驾驶过程中的自主性；其三，“公正”是无人驾驶的主要价值，自动驾驶技术应当保证和促进每个个体在使用该技术时的平等权利和整个社会资源的公正分配。

第二，将无人驾驶技术的算法植入“为了他者而行动”的伦理。“他者”不仅是无人驾驶技术的使用者和开发者，更是整个社会共同体中的所有成员。“为他者而行动”强调：其一，无人

驾驶技术必须尊重每一个人类个体，尤其要尊重其使用者的行动权，通过“踩刹车”和“紧急救援通道”[37]等功能将控制权交还到人类驾驶者手中；其二，无人驾驶技术必须构建“理解的技术环境”以逐渐消除算法“黑箱”带来的负面影响，通过生成性、对话式的人工智能大模型，实现对不同驾驶者的个性化学习，从而达到理解行动、理解责任、理解目的；其三，无人驾驶技术须致力于构建同人类整体的伙伴关系，以实现人工智能与人类社会的协同发展与进化。

第三，无人驾驶技术应当展开为一个人类得以“诗意地栖居”的伦理空间。从现实性上来看，它要求：其一，无人驾驶技术应当始终以人类整体的发展为己任，不仅包括物质流动效率的提高，更包括人类交往水平和道德水平的提高以及社会文明的进步，这将通过无人驾驶技术对生活空间的整体性改造来实现；其二，无人驾驶技术最终应当实现人与世界的“彼此通达”，而非数字控制下对人的压迫和剥削，这不仅需要算法开发者的道德良心，更需要整个社会公权力对此进行前瞻性和预备性的立法规定。在某种意义上，这种“诗意地栖居”的伦理空间之实现需要长久的发展才能达成，但从内在逻辑上，为无人驾驶技术赋予“德性”本质就必然会导向“为了他者”的行动，有助于实现人“诗意地栖居”。在这一过程中必然存在着困难和试错，这不仅需要技术水平的提高以增强硬件的运算处理能力，更需要整个社会民主和文明水平的提高。

基于此，我们认为，对自动驾驶“无人化”风险的防范性治理要遵循六条原则。一是人本原则。以无人驾驶汽车为代表的人工智能体及其背后的制造者、经营者和使用者应当始终以人类群体的发展进步和人类能力的整体性增强为价值目标。二是透明原则。人工智能体所依赖的算法应当具有可理解性，依照其自身算法而进行的行为应当可以被解释；每个人工智能体都应当建立相应的“关系清单”，清单内应当包括所有的投资者、开发者和使用者；人工智能体应当开放给使用者相应权限，使得人类使用者可以训练自己的人工智能体。三是安全原则。以自动驾驶汽车为代表的人工智能体应当首先保证人类的生命安全，保证在充分正确理解使用者的合理意图及其后果的前提下如预期行动。四是公正原则。作为相互性原则，公正原则强调：一方面，以无人驾驶汽车为代表的人工智能体的使用应当保障所有人类交通参与者的基本权利，避免存在偏见和歧视，避免区域交通的不平衡和对生活空间的侵占；另一方面，公正原则要求人类不应当干涉人工智能体的正常合理运行。[38]五是隐私保护原则。人工智能体应当保护使用者的个人数据和隐私权，在未经使用者同意的情况下不得为任何个人或机构收集、调取和提供用户隐私信息。在自动驾驶汽车上这些隐私信息不但包括起点和终点、路线、途经点、偏好地区，更包括驾驶习惯、汽车传感器收集的影像和语音等。六是可控原则。开发者应从技术层面保证人工智能体始终为人类群体（而非某一个体或团体）所掌控；社会层面则应当完善相应的法律、设立相应的机构，对以自动驾驶汽车为代表的人工智能体进行统一的有效监管。

上述原则的落实，最终指向一种前瞻性的“伦理锚定”治理范式。这意味着，在算法开发阶段，就应建立“德性影响评估”机制，审视其对社会正义、个体自主与道德实践空间的长期影响；在立法监管层面，须明确“人类最终仲裁权”，确保在任何情况下，算法都不能剥夺人类对关键决策的知情、选择与否决的权利；在社会文化层面，则应鼓励对技术替代性的公共辩论，警惕资本与技术合谋下的“理性迷思”，培育一种既能拥抱创新又能捍卫人之尊严的技术文化。

自动驾驶的未来，不应是“无人”的寂静世界，而应是“为人”的共融空间。将伦理批判转化为治理智慧，方能驾驭技术浪潮，使技术成为通往“美好生活”的桥梁，而非悬置于人类主体性之上的“达摩克利斯之剑”。在这条道路上，重拾“人为自己立法”的勇气，构建“为他者而

行动”的伦理共同体，是人类对自身命运的积极筹划。在《神秘博士》中有这样一集：在未来，全部人类都被困在车上，在“栓塞”的“高速公路”上一个人终其一生只能行进几十公里，车窗外是致命的废弃物和污染物，而在目不可视的浓雾之后，是如章鱼般的怪物正在用自己的触手控制着每一台车辆，控制着车辆中每一个人的命运。以无人驾驶为代表的人工智能体的“无人化”，本质上是以“无人”的方式实现对人类世界的控制。无人驾驶的危机是人之自由生存的危机，更是世界不断“失控”“疏离”的危机，面对这种“恶兆”般的预言，必须前瞻性地“回到人自身”，将人重新置于价值的高地。对于即将到来的“无人时代”，需要一种忧虑的远见以“警醒”世界：漠视人工智能体的“无人”本质，会使世界导向被支配、被奴役的“无人”状态。

注释：

① 参见“Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles,” SAE International, 2021-04-30, https://www.sae.org/standards/content/j3016_202104/。

② 德国于2017年率先修订了《道路交通法》，增补了自动驾驶的相关法条，美国在同年也通过了自动驾驶汽车法案。我国于2018年开始推行车联网产业标准体系建设（参见《国家车联网产业标准体系建设指南（智能网联汽车）》《四部委关于开展智能网联汽车准入和上路通行试点工作的通知》）。

③ 牛津大学乌希罗实践伦理学中心研究员大卫·埃德蒙兹（David Edmonds）在其编著的《未来道德：来自新科技的挑战》一书中主张用“无人车”指称“自动驾驶汽车”，因为“方向盘后面没有人”。

④ 参见《AmiGo！萝卜快跑落地瑞士》，2025-10-31，https://www.apollogo.com/ch/news/16062。

⑤ 参见《特斯拉Model Y首次实现全自动驾驶交付！》，2025-06-30，https://auto.cnr.cn/hy/20250630/t20250630_527235630.shtml。

⑥ 温德尔·瓦拉赫、科林·艾伦：《道德机器：如何让机器人明辨是非》，王小红等译，北京：北京大学出版社，2017年，第10页。

⑦ 例如，如果算法作者为自动驾驶系统输入了功利主义的行动逻辑，那么自动驾驶汽车在面临“电车难题”时就必然会选择“最大收益”或“最小损失”的行动方案。

⑧ Jan Gogoll and Julian F. Müller, “ Autonomous Cars: In Favor of a Mandatory Ethics Setting,” *Science and Engineering Ethics*, vol.23, no.3, 2017, pp. 681-700.

⑨ Ivó Coca-Vila, “Self-Driving Cars in Dilemmatic Situations: An Approach Based on the Theory of Justification in Criminal Law,” *Criminal Law and Philosophy*, vol.12, 2018, pp. 59-82.

⑩ 参见隋婷婷、郭晓：《自动驾驶电车难题的伦理算法研究》，《自然辩证法通讯》2020年第10期。张学义、王晓雪：《“伦理旋钮”：破解无人驾驶算法困境的密钥？》，《中国人民大学学报》2023年第2期。

⑪ 例如，当一个幼童突然跑到车道上、紧急避让可能会使驾驶者付出生命代价时，部分人还是会出于本能而避让，尽管这可能危及自己的生命。

⑫ 参见《埃塞航空坠机事故最终调查报告发布 确认为系统故障导致》，2022-12-24, https://content-static.cctvnews.cctv.com/snow-book/index.html?item_id=12164718515432650801&toc_style_id=feeds_default&share_to=qq&track_id=96603376-1774-4ae8-8af6-18b00904ae54。

⑬ Andreas Matthias, “ The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata,” *Ethics and Information Technology*, vol.6, no.3, 2004, pp. 175-183.

⑭ 王天恩：《人工智能应用“责任鸿沟”的造世伦理跨越——以自动驾驶汽车为典型案例》，《哲学分析》2022年第1期。

⑮ 白惠仁：《自动驾驶汽车的“道德责任”困境》，《大连理工大学学报》（社会科学版）2019第4期。

⑯ 美国田纳西州的法条将“自动驾驶系统”（ADS）本身定义为“驾驶员”，并且在2017年的州法典修正案中将“人”的定义扩展到了“法人或从事自动驾驶系统（研究、经营或工作）的人”。

⑰ 参见《小米SU7碰撞事故再敲智驾警钟：车企应加强智驾风险提醒，消费者更需时刻保持警惕》，2025-04-01，https://ishare.ifeng.com/c/s/8iCTLHNZaQb?spss=np&channelId=&aman=1aRe45I7a5M6aaTa5b1fb6ka74d09fX1b00579y11e。

⑱ Rosalind Williams, *Notes on the Underground: An Essay on Technology, Society, and the Imagination*, Cambridge, Massachusetts: The MIT Press, 2008, p.127.

⑲ “科林格里奇困境”是指一个行动包含着长远且不可预见的社会后果，这一后果在初期并不明显，一旦显现便难以改变。参见David Collingridge, *The Social Control of Technology*, London: Frances Pinter（Publishers）Ltd., 1980, p. 11。

⑳ 参见安东尼·汤森：《无人驾驶——从想象到现实》，沈

瑜译，北京：中信出版集团，2023年。

㉑ 例如，浙江正在建设的“杭绍甬”高速公路不仅集成了无线充电功能，同时通过搭建大数据云控平台实现车路协同的综合感知体系，通过车联网为“无人自动驾驶”提供辅助。

㉒ 参见Zibo Jin, Daochu Li and Jinwu Xiang, “Robot Pilot: A New Autonomous System toward Flying Manned Aerial Vehicles,” *Engineering*, vol. 27, no. 8, 2023, pp. 242-253; Shu-ping Chen, Guang-ming Xiong, Hui-yan Chen, et al., “MPC-based Path Tracking with PID Speed Control for High-Speed Autonomous Vehicles Considering Time-Optimal Travel,” *Journal of Central South University*, vol. 27, no. 12, 2020, pp. 3702-3720。在这些技术研究文章中，均使用“autonomous”为译。而在谢惠媛（《民用无人驾驶技术的伦理反思——以无人驾驶汽车为例》，《自然辩证法研究》2017年第9期）和张学义、王晓雪（《“伦理旋钮”：破解无人驾驶算法困境的密钥？》，《中国人民大学学报》2023年第2期）等的文章中均使用“driverless”。上引大卫·埃德蒙兹在《未来道德：来自新科技的挑战》一书中使用“无人驾驶”。

㉓“人工智能体”不同于“人工智能”。前者是指基于“人工智能”技术所开发的诸多个体性的产品，诸如无人驾驶汽车、无人轰炸机、人工智能大模型等；后者则是在总体性的意义上，既包括技术也包括相关产品。

㉔ 刘伟兵：《人工智能体是主体吗？“无人化”背后的总体工人》，《学习与实践》2025年第3期。

㉕ 朱凌珂：《赋予强人工智能法律主体地位的路径与限度》，《广东社会科学》2021年第5期。

㉖ 王春梅、冯源：《技术性人格：人工智能主体资格的私法构设》，《华东政法大学学报》2021年第5期。

㉗ 程承坪：《人工智能：工具或主体？——兼论人工智能奇点》，《上海师范大学学报》（哲学社会科学版）2021年第6期。

㉘ 田海平：《人工智能“类人化”的伦理限度》，《东南大学学报》（哲学社会科学版）2025年第3期。

㉙ 此处及下文中的“四个面向”所采用的“动力因”“目的因”“形式因”“质料因”的说法是对亚里士多德形而上学“四因说”的化用，借此以说明驾驶行为如何从“有人控制的驾驶”到“无人驾驶”的转变，因而是在一种道德行动生成论的角度而非原本自然哲学本体论的角度来使用这些说法。

㉚ 亚里士多德：《尼各马可伦理学》，廖申白译，北京：商务印书馆，2003年，第279页。

㉛ 或许在自动驾驶大行其道的时刻，由人完全驾驶汽车的行为就从“交通行为”变成了“竞技”或“娱乐”，人或许可以在赛事的竞争中展现自身的“卓越”，但这在根本上已经和作为人之生存根本的“交通”相脱离了。

㉜ 显见且具有说服力的类比是：在智能手机诞生之前，人尚可以脱离手机生活；而时至当下，没有智能手机几乎任何活动都会受限。不可否认的是，智能手机为人带来了更多自由的可能，但它对人隐性的奴役和控制不容小觑。

㉝ 尤瓦尔·赫拉利：《未来简史：从智人到智神》，林俊宏译，北京：中信出版集团，2017年，第333页。

㉞ 马丁·海德格尔：《存在与时间》，陈嘉映、王庆节译，北京：生活·读书·新知三联书店，2014年，第137页。

㉟ 汉斯–格奥尔格·伽达默尔：《诠释学Ⅰ：真理与方法》，洪汉鼎译，北京：商务印书馆，2011年，第520页。

㊱ 下文所探讨的伦理治理策略涉及“人工智能体的伦理选择的方法论难题”，即“如何将伦理准则植入人工智能体当中？”目前学界的探讨主要分为“自上而下”和“自下而上”两种方式。“自上而下”即在开发端预先设置伦理禁令和信条；“自下而上”则是依靠机器学习实现伦理原则的理解和强化。应当说两种方式的优劣都不言而喻，目前技术领域中采用“由下而上”方法的试验取得了相对较好的成果。笔者认为，应当采取两者结合的方式，既要设定明确的伦理准则，也要通过长期的数据积累和学习来优化算法。而对于伦理准则的设置，既要将规则和义务置于重要地位，更要明确德性算法的重要性。

㊲ 对于“踩刹车”和“紧急救援通道”等功能的论述详见岳缙：《基于自动驾驶“视见”的行动伦理方案》，《江苏社会科学》2024年第4期。

㊳ 新加坡曾于2017年颁布法律，将“干扰自动驾驶汽车运行试验”的行为定义为违法行为。这是一种“保护机器人免受我们侵害的法律”。参见陈西文：《我们，机器人？——人工智能监管及其法律局限》，游传满、费秀艳译，北京：北京大学出版社，2024年，第221页。

编辑　张　蕾

公　示

根据《上海市新闻出版局关于开展2025年度新闻记者证核验工作的通知》要求，《探索与争鸣》编辑部近日开展了2025年度新闻记者证核验工作，拟通过年度核验人员为：李梅、杜运泉、张蕾。特此公示。

监督举报电话：021-53060418

《探索与争鸣》编辑部

2026年1月20日