

附件 2

驾驶自动化技术研发伦理指引

1. 目的

为指导驾驶自动化技术研究开发和应用的规范开展，防范技术研发与产品应用过程中的科技伦理风险，推动该领域健康发展，研究提出驾驶自动化技术研究开发伦理指引。

2. 术语

2.1 驾驶自动化功能（driving automation feature）

驾驶自动化功能是指驾驶自动化系统在特定的设计运行条件内执行部分或全部动态驾驶任务的能力。

2.2 汽车驾驶自动化分级（taxonomy of driving automation for vehicles）

基于驾驶自动化系统能够执行动态驾驶任务的程度，根据在执行动态驾驶任务中的角色分配以及有无设计运行范围限制，将驾驶自动化由低到高分成 0 级至 5 级，分别为应急辅助（emergency assistance）、部分驾驶辅助（partial driver assistance）、组合驾驶辅助（combined driver assistance）、有条件自动驾驶（conditionally automated driving）、高度自动驾驶（highly automated driving）、完全自动驾驶（fully automated driving）。

0 级驾驶自动化（应急辅助）系统：不能持续执行动态

驾驶任务中的车辆横向或纵向运动控制，但具备持续执行动态驾驶任务中的部分目标和事件探测与响应的能力。

1 级驾驶自动化（部分驾驶辅助）系统：在其设计运行条件下持续地执行动态驾驶任务中的车辆横向或纵向运动控制，且具备与所执行的车辆横向或纵向运动控制相适应的部分目标和事件探测与响应的能力。

2 级驾驶自动化（组合驾驶辅助）系统：在其设计运行条件下持续地执行动态驾驶任务中的车辆横向和纵向运动控制，且具备与所执行的车辆横向和纵向运动控制相适应的部分目标和事件探测与响应的能力。

3 级驾驶自动化（有条件自动驾驶）系统：在其设计运行条件下持续地执行全部动态驾驶任务。

4 级驾驶自动化（高度自动驾驶）系统：在其设计运行条件下持续地执行全部动态驾驶任务并自动执行最小风险策略。

5 级驾驶自动化（完全自动驾驶）系统：在任何可行驶条件下持续地执行全部动态驾驶任务并自动执行最小风险策略。

2.3 先进驾驶辅助功能（advanced driver assistance function）

先进驾驶辅助功能是指驾驶自动化系统在特定的设计运行条件下辅助驾驶员执行部分动态驾驶任务的功能。包括“应急辅助”“部分驾驶辅助”和“组合驾驶辅助”功能。

2.4 自动驾驶功能（automated driving function）

自动驾驶功能是指驾驶自动化系统在特定的设计运行条件下代替驾驶员持续自动地执行全部动态驾驶任务的功能。包括“有条件自动驾驶”“高度自动驾驶”和“完全自动驾驶”功能。

3. 基本原则

3.1 以人为本、增进福祉

驾驶自动化技术的研究开发应坚持以人为本的发展思想，充分体现人文关怀，积极引导科技向善，应有利于促进经济发展、社会进步、民生改善，有利于促进交通效率提高和生态环境可持续发展。

3.2 安全优先、尊重生命

驾驶自动化技术的研究开发和应用必须充分考虑安全性，既要尽最大可能保证用户及参与者的生命安全，还应高度重视保护车辆周围人员生命与环境安全。在面对不可避免的交通事故或极端行车环境时，应确保驾驶自动化系统能高度尊重生命，积极寻求有效应对方法最大限度降低对人的伤害。

3.3 公平公正、避免偏见

应确保在驾驶自动化系统的正常运行过程中或在面临紧急情况时，系统决策不受人人类个体属性和个体特征（如年龄、性别、种族、身体特征、社会地位、财富水平或城乡背景等）的影响。应避免单一技术路线或单一技术主体造成的

技术垄断，以防止驾驶自动化系统成为部分群体道德观念的表达载体，导致对其他群体的不公正对待。确保在驾驶自动化系统设计阶段所有相关群体的利益得到平等考量。

3.4 知情保障、信息公开

与驾驶自动化技术相关的算法、模型等内容应当被清晰记录、随时可查。依照相关法律法规、国家标准及其他有关文件在数据记录、存储、读取以及重要数据保护等方面的要求，保证与驾驶自动化功能相关的重要数据得到安全记录、妥善保存，保障利益相关方的知情权。涉及新的、可能具有争议的技术时，应充分讨论，广泛听取利益相关方和公众意见。

4. 一般要求

4.1 合法合规

驾驶自动化技术的研究开发和应用应严格遵守我国法律法规和相关规定，遵守科技伦理基本原则和要求以及行业协会、产业联盟及其他科学共同体达成的专业共识和技术规范。在驾驶自动化技术的国际推广和相关产品的海外应用中，还应遵循国际公认的伦理准则和目标国家的法律规范。

4.2 效益提升

驾驶自动化技术的研究开发和应用应以提升综合效益为导向。应通过技术创新与产业协同推动经济增长，培育智能网联汽车产业生态；应致力于增强公众出行便利性，优化交通服务可及性，积极改善特殊群体出行体验，降低社会交

通参与成本；应有效提升交通系统运行效率，缓解交通拥堵，助推智慧交通基础设施与新型智慧城市建设；应注重提升环境效益，通过应用驾驶自动化技术提高能源利用效率，降低交通运输领域碳排放，助力绿色低碳发展。

4.3 控制风险

驾驶自动化技术的研究开发应设计完善的风险监测与应急机制，确保相关责任主体能够及时发现、干预并控制风险，将潜在威胁降至合理可控的范围内。驾驶自动化系统内置的风险应对策略应当成为事后追责的重要参考依据。

4.4 及时调整

驾驶自动化技术研究开发过程中应科学评估技术成熟度与现实可行性，应充分考虑和吸纳社会公众的意见和建议，保证各利害关系者利益诉求尽可能得到满足，并根据技术进步、市场变化等及时优化调整。

4.5 隐私保护

驾驶自动化技术的研究开发和应用应建立健全数据全生命周期隐私保护机制，严格遵循知情同意、最小必要、去标识化等原则。数据采集须限定于实现驾驶功能所需范围，避免留存可识别特定自然人的原始数据；对行踪轨迹、地理环境等敏感信息应通过技术手段实现实时模糊化处理，建立数据访问权限动态审查机制，对涉密机构等特殊区域的地理信息实施访问阻断。

5. 不同类型驾驶自动化技术的伦理规范要求

5.1 先进驾驶辅助阶段

先进驾驶辅助阶段是指 2 级及以下的驾驶自动化。这一阶段由人类驾驶员与驾驶自动化系统共同完成行车过程中对目标和事件的探测与响应，人类驾驶员完成动态驾驶后援任务，驾驶自动化系统在有限的设计运行范围内发挥辅助作用。为了保证驾驶安全性，当驾驶员¹主动请求驾驶自动化系统退出时，立即解除系统控制权。因此，在先进驾驶辅助阶段，责任主体通常是人类驾驶员。

具有先进驾驶辅助功能汽车的生产制造标准应当符合国家相关法规的基本安全要求。先进驾驶辅助阶段的驾驶自动化系统应能够提供不低于传统有人驾驶汽车所能提供的安全性，并在特定场景中优于人类驾驶员，以体现技术优势。系统设计需降低用户²的学习成本、操作复杂性和经济负担，确保技术的易普及性和可用性。在产品开发中，应充分考虑老年人、残疾人等特殊群体的需求，体现技术应用的公平性和包容性。应向所有用户明确告知驾驶自动化系统的核心功能、使用条件、使用方式、安全应对措施等内容。驾驶自动化系统遵循的技术标准、编程规则和决策逻辑需保持清晰透明，在责任追溯过程中能够被政府等监管部门按需审查。鼓励驾驶自动化系统设计方和生产者主动设置伦理审查机构

¹ 包含传统驾驶员（conventional driver）和远程驾驶员（remote driver）。

² 与驾驶自动化相关的人类角色的统称，包括传统驾驶员、远程驾驶员、乘客、动态驾驶任务后援用户、调度员等主体。

开展伦理评估（审查），通过适当引入外部审计机制等方式提升社会信任。

5.2 有限制阶段的自动驾驶

有限制阶段的自动驾驶包括 3 级和 4 级驾驶自动化。在这一阶段，自动驾驶系统能够持续完成车辆的横向和纵向运动控制，并承担目标与事件的探测与响应。自动驾驶系统需具备识别动态驾驶任务后援用户的接管能力，并在特定条件下及时向动态驾驶任务后援用户发出介入请求，或在设计运行范围内自动执行最小风险策略以保障车辆及周围环境的安全。基于此，有限制阶段的自动驾驶责任主体因具体运行场景而异。当自动驾驶系统作为自动驾驶责任主体时，应当能够根据有关法律法规、技术标准向相关自然人主体或法人主体（包括但不限于自动驾驶系统开发者、自动驾驶系统生产商、自动驾驶汽车权利人、自动驾驶系统用户）清晰追责。

有限制阶段的自动驾驶除了应当遵循前述先进驾驶辅助相关的科技伦理指引以外，需要遵从以下几点要求：

一是在自动驾驶系统的设计、开发阶段，影响自动驾驶系统功能的关键算法、训练数据等内容的重大变更应当被清晰记录，在需要时能够被政府等监管部门及时调取并清晰审查。

二是自动驾驶的决策行为应当可追溯、可审查，以便在发生侵权行为、碰撞行为或其他违法行为时，能够通过追溯算法内容明确相关的自然人责任主体或法人责任主体。

三是自动驾驶系统应按照用户预设的导航习惯、行程和路径偏好等行车策略执行全部动态驾驶任务。若因突发情况必须对行车策略进行变更，系统应通过清晰的方式提前告知用户变更内容和变更原因。

四是自动驾驶算法的设计和实现应符合社会最大公约数的价值倾向，体现广泛的社会认同。

五是自动驾驶决策选择应体现对全体人类的平等尊重，任何基于个体属性的差异化决策表现出的歧视与偏见均属伦理不可接受范围。为此，算法设计者和汽车生产商需在决策机制中消除潜在的偏见，确保技术公平性。

5.3 无限制阶段的自动驾驶

无限制阶段的自动驾驶为 5 级的驾驶自动化。这一阶段由完全自动驾驶系统完成车辆的横向和纵向控制，负责行车过程中对目标和事件的探测与响应，并在紧急情况下采取最小风险策略，以保障车辆及周围环境的安全。5 级自动驾驶不受设计运行范围的限制，能够在任何可行驶条件下正常运行。若存在潜在安全风险，自动驾驶系统可暂缓满足用户退出自动驾驶的请求，并执行最小风险策略以达到最小风险状态。除非特殊情况下有用户主动介入操作，否则无限制阶段自动驾驶的责任主体以自动驾驶系统为主。当自动驾驶系统作为自动驾驶责任主体时，应当能够根据有关法律法规、技术标准向相关自然人主体或法人主体（包括但不限于自动驾驶系统开发者、自动驾驶系统生产商、自动驾驶汽车权利人、

自动驾驶系统用户）清晰追责。

无限制阶段的自动驾驶除了应当遵循前述先进驾驶辅助、有限制阶段的自动驾驶相关的科技伦理指引以外，需要遵从以下几点要求：

一是在完全自动驾驶系统执行全部动态驾驶任务时，车辆所能提供的安全性应不低于拥有驾驶资格、驾驶经验丰富且具备道德敏感性的人类驾驶员在相同情境下所能达到的安全水平³。

二是自动驾驶系统的决策应具有稳定性，表现在一定范围内的时间一致性和空间一致性，即在相同行车情境下，自动驾驶系统不应因时点不同或空间差异而做出相互矛盾的决策。

三是在遵守法律法规和技术标准的前提下，不同厂商、型号、版本的自动驾驶系统在面临相同驾驶情境时，其决策逻辑应具有较高的一致性，以体现技术规范化与社会公平性。

四是在人类驾驶员执行动态驾驶任务时，在充分保护用户隐私并获得合法授权的前提下，允许自动驾驶系统基于即时的行车环境预判相关主体的行动意图。

五是当人类驾驶员可能实施危害社会的行为时，自动驾驶系统应及时向利益相关方发出警示，必要时采取合理方式干预相关操作，以避免重大社会危害或违法行为。

³ 为此，需要依据有关法规和标准，对自动驾驶系统的测试里程、场景覆盖率、决策响应时间、道德图灵测试、道德学习机制、决策社会接受度等进行多维度测试和系统性验证。

6. 科普宣传

技术研发人员和企业应积极开展面向公众的科学普及工作，引导社会公众以理性、科学的态度认识驾驶自动化技术。在发布重大研究成果时，相关主体应坚持客观准确、实事求是，避免片面夸大、隐匿风险，不得编造、传播与驾驶自动化系统实际功能和效果不相符的虚假信息。此外，在研发、生产、推广等过程中，相关主体应向社会公众明确告知并清晰阐释驾驶自动化产品的功能分级、使用条件、使用方式等注意事项，及时澄清和纠正可能导致公众对驾驶自动化技术产生误解的误导性信息，引导并敦促用户合理、正确地使用不同分级的驾驶自动化产品，避免公众认知偏差和产品误用、滥用。

本指引由国家科技伦理委员会人工智能伦理分委员会研究制定，定期评估，适时修订。

国家科技伦理委员会人工智能伦理分委员会

2025 年 6 月

主要参考文件

- [1] 《智能汽车创新发展战略》(2020)
- [2] 《汽车驾驶自动化分级 GB/T 40429-2021》(2021)
- [3] 《中华人民共和国网络安全法》(2016)
- [4] 《中华人民共和国数据安全法》(2021)
- [5] 《中华人民共和国道路交通安全法》(2021)
- [6] 《智能网联汽车准入和上路通行试点实施指南(试行)》(2023)
- [7] 《国家车联网产业标准体系建设指南(智能网联汽车)》(2023)
- [8] 《新一代人工智能伦理规范》(2021)
- [9] 《智能网联汽车 术语和定义 GB/T 44373-2024》(2024)
- [10] 《智能网联汽车 自动驾驶数据记录系统 GB 44497-2024》
(2024)