

新一代人工智能伦理规范

新一代人工智能伦理规范为深入贯彻《新一代人工智能发展规划》，细化落实《新一代人工智能治理原则》，增强全社会的人工智能伦理意识与行为自觉，积极引导负责任的人工智能研发与应用活动，促进人工智能健康发展，制定本规范。

第一章 总则

第一条 本规范旨在将伦理道德融入人工智能全生命周期，促进公平、公正、和谐、安全，避免偏见、歧视、隐私和信息泄露等问题。

第二条 本规范适用于从事人工智能管理、研发、供应、使用等相关活动的自然人、法人和其他相关机构等。（一）管理活动主要指人工智能相关的战略规划、政策法规和技术标准制定实施，资源配置以及监督审查等。（二）研发活动主要指人工智能相关的科学研究、技术开发、产品研制等。（三）供应活动主要指人工智能产品与服务相关的生产、运营、销售等。（四）使用活动主要指人工智能产品与服务相关的采购、消费、操作等。

第三条 人工智能各类活动应遵循以下基本伦理规范。（一）增进人类福祉。坚持以人为本，遵循人类共同价值观，尊重人权和人类根本利益诉求，遵守国家或地区伦理道德。

坚持公共利益优先，促进人机和谐友好，改善民生，增强获得感幸福感，推动经济、社会及生态可持续发展，共建人类命运共同体。（二）促进公平公正。坚持普惠性和包容性，切实保护各相关主体合法权益，推动全社会公平共享人工智能带来的益处，促进社会公平正义和机会均等。在提供人工智能产品和服务时，应充分尊重和帮助弱势群体、特殊群体，并根据需要提供相应替代方案。（三）保护隐私安全。充分尊重个人信息知情、同意等权利，依照合法、正当、必要和诚信原则处理个人信息，保障个人隐私与数据安全，不得损害个人合法数据权益，不得以窃取、篡改、泄露等方式非法收集利用个人信息，不得侵害个人隐私权。（四）确保可控可信。保障人类拥有充分自主决策权，有权选择是否接受人工智能提供的服务，有权随时退出与人工智能的交互，有权随时中止人工智能系统的运行，确保人工智能始终处于人类控制之下。（五）强化责任担当。坚持人类是最终责任主体，明确利益相关者的责任，全面增强责任意识，在人工智能全生命周期各环节自省自律，建立人工智能问责机制，不回避责任审查，不逃避应负责任。（六）提升伦理素养。积极学习和普及人工智能伦理知识，客观认识伦理问题，不低估不夸大伦理风险。主动开展或参与人工智能伦理问题讨论，深入推动人工智能伦理治理实践，提升应对能力。

第四条 人工智能特定活动应遵守的伦理规范包括管理规范、研发规范、供应规范和使用规范。

第二章 管理规范

第五条 推动敏捷治理。尊重人工智能发展规律，充分认识人工智能的潜力与局限，持续优化治理机制和方式，在战略决策、制度建设、资源配置过程中，不脱离实际、不急功近利，有序推动人工智能健康和可持续发展。

第六条 积极实践示范。遵守人工智能相关法规、政策和标准，主动将人工智能伦理道德融入管理全过程，率先成为人工智能伦理治理的实践者和推动者，及时总结推广人工智能治理经验，积极回应社会对人工智能的伦理关切。

第七条 正确行权用权。明确人工智能相关管理活动的职责和权力边界，规范权力运行条件和程序。充分尊重并保障相关主体的隐私、自由、尊严、安全等权利及其他合法权益，禁止权力不当行使对自然人、法人和其他组织合法权益造成侵害。

第八条 加强风险防范。增强底线思维和风险意识，加强人工智能发展的潜在风险研判，及时开展系统的风险监测和评估，建立有效的风险预警机制，提升人工智能伦理风险管控和处置能力。

第九条 促进包容开放。充分重视人工智能各利益相关主体的权益与诉求，鼓励应用多样化的人工智能技术解决经济社会发展实际问题，鼓励跨学科、跨领域、跨地区、跨国界

的交流与合作，推动形成具有广泛共识的人工智能治理框架和标准规范。

第三章 研发规范

第十条 强化自律意识。加强人工智能研发相关活动的自我约束，主动将人工智能伦理道德融入技术研发各环节，自觉开展自我审查，加强自我管理，不从事违背伦理道德的人工智能研发。

第十一条 提升数据质量。在数据收集、存储、使用、加工、传输、提供、公开等环节，严格遵守数据相关法律、标准与规范，提升数据的完整性、及时性、一致性、规范性和准确性等。

第十二条 增强安全透明。在算法设计、实现、应用等环节，提升透明性、可解释性、可理解性、可靠性、可控性，增强人工智能系统的韧性、自适应性和抗干扰能力，逐步实现可验证、可审核、可监督、可追溯、可预测、可信赖。

第十三条 避免偏见歧视。在数据采集和算法开发中，加强伦理审查，充分考虑差异化诉求，避免可能存在的数据与算法偏见，努力实现人工智能系统的普惠性、公平性和非歧视性。

第四章 供应规范

第十四条 尊重市场规则。严格遵守市场准入、竞争、交易等活动的各种规章制度，积极维护市场秩序，营造有利于人工智能发展的市场环境，不得以数据垄断、平台垄断等破坏市场有序竞争，禁止以任何手段侵犯其他主体的知识产权。

第十五条 加强质量管控。强化人工智能产品与服务的质量监测和使用评估，避免因设计和产品缺陷等问题导致的人身安全、财产安全、用户隐私等侵害，不得经营、销售或提供不符合质量标准的产品与服务。

第十六条 保障用户权益。在产品与服务中使用人工智能技术应明确告知用户，应标识人工智能产品与服务的功能与局限，保障用户知情、同意等权利。为用户选择使用或退出人工智能模式提供简便易懂的解决方案，不得为用户平等使用人工智能设置障碍。

第十七条 强化应急保障。研究制定应急机制和损失补偿方案或措施，及时监测人工智能系统，及时响应和处理用户的反馈信息，及时防范系统性故障，随时准备协助相关主体依法依规对人工智能系统进行干预，减少损失，规避风险。

第五章 使用规范

第十八条 提倡善意使用。加强人工智能产品与服务使用前的论证和评估，充分了解人工智能产品与服务带来的益

处，充分考虑各利益相关主体的合法权益，更好促进经济繁荣、社会进步和可持续发展。

第十九条 避免误用滥用。充分了解人工智能产品与服务的适用范围和负面影响，切实尊重相关主体不使用人工智能产品或服务的权利，避免不当使用和滥用人工智能产品与服务，避免非故意造成对他人合法权益的损害。

第二十条 禁止违规恶用。禁止使用不符合法律法规、伦理道德和标准规范的人工智能产品与服务，禁止使用人工智能产品与服务从事不法活动，严禁危害国家安全、公共安全和生产安全，严禁损害社会公共利益等。

第二十一条 及时主动反馈。积极参与人工智能伦理治理实践，对使用人工智能产品与服务过程中发现的技术安全漏洞、政策法规真空、监管滞后等问题，应及时向相关主体反馈，并协助解决。

第二十二条 提高使用能力。积极学习人工智能相关知识，主动掌握人工智能产品与服务的运营、维护、应急处置等各使用环节所需技能，确保人工智能产品与服务安全使用和高效利用。

第六章 组织实施

第二十三条 本规范由国家新一代人工智能治理专业委员会发布，并负责解释和指导实施。

第二十四条 各级管理部门、企业、高校、科研院所、协会学会和其他相关机构可依据本规范，结合实际需求，制订更为具体的伦理规范和相关措施。

第二十五条 本规范自公布之日起施行，并根据经济社会发展需求和人工智能发展情况适时修订。

国家新一代人工智能治理专业委员会

2021 年 9 月 25 日