



大公国际：“人工智能+”何以向善？

文/刘颖

摘要

2025年8月26日，国务院发布《关于深入实施“人工智能+”行动的意见》这一重大部署，提出我国在2027年、2030年和2035年各个阶段在人工智能发展领域所要实现的目标。当前，在我国将人工智能与经济社会各行业各领域广泛深度融合的关键阶段，如何引导“人工智能+”向善也逐步纳入政策与相关法律法规考虑的范畴。本文将从目标向善、技术向善、治理向善、应用向善等方面进行论述，确保智能技术的发展始终与人类社会的正向价值、我国经济高质量发展的目标对齐。

正文

当前，人工智能正以迅猛之势加速演进，我国在这一赛道上不仅稳稳占据技术发展的前沿梯队，更拥有不可替代的核心优势——丰富的应用场景、超大规模市场空间以及消费者迫切的需求，共同构成了AI发展的中国土壤。国务院日前发布《关于深入实施“人工智能+”行动的意见》，提出我国在2027年、2030年和2035年各个阶段在人工智能发展领域所要实现的目标，此举也旨在充分释放我国的应用场景与市场优势，以AI赋能经济社会发展，是立足当下、着眼长远的关键战略部署。基于此，引导“人工智能+”向善的重要性也逐步凸显，想要牢牢把握AI与产业深度融合的历史机遇，有效应对前所未有的风险挑战，就必须在推进“人工智能+”行动的过程中，精准锚定技术与应用的正确方向，着力构建一套能够保障智能向善落地的良好制度环境。

一、目标向善：以“责任感”定调发展底色

发展人工智能的核心目标，是将其打造为赋能人类社会进步的智能工具与创新引擎，最终服务于高质量发展与人类共同福祉，而非追求技术本身的极致迭代。

从经济维度看，AI旨在驱动产业变革，通过自动化升级传统制造业、以智能算法优化服务业效率、催生新兴业态，破解生产要素瓶颈，构建更高效的经济体系。

从民生维度讲，“人工智能+”的目标是改善生活品质，让“人工智能+医疗”辅助精准诊断、“人工智能+教育”实现个性化教学、“人工智能+养老”提供贴心照护，用技术填补资源鸿沟，提升公共服务的可及性与质量。

从社会与全球维度，AI致力于现代化治理与应对共同挑战，如通过智慧城市系统优化公共管理，借助大数据预测极端天气、疾病传播等风险，为全球可持续发展提供解决方案。

“人工智能+”的目标向善始终内嵌“责任感”的底色，确保其发展不偏离增进民生福祉、促进社会公平的本质方向。

二、技术向善：以“设计向善”筑牢底层根基

技术是人工智能向善的起点，若技术设计本身存在缺陷或偏见，再美好的目标也会落空。这一层面的核心是将“向善”嵌入技术研发的全流程。

首先，算法不能成为“黑箱”。需通过技术手段消除训练数据中的历史偏见，如性别、种族、地域歧视等，同时推动可解释性 AI（XAI）发展，让人工智能的决策逻辑（如信贷审批、招聘、司法辅助等）可追溯、可理解，避免因算法不透明导致的不公或误判。

其次，数据是人工智能的燃料，数据伦理直接决定其底色。需严格遵循“合法、正当、必要”原则，保障用户数据隐私，杜绝未经授权的数据采集、滥用或泄露，同时禁止使用涉及人身权利的敏感数据。

最后，人工智能系统需要既能抵御恶意攻击，也能在出现异常时及时被干预或关停，避免因技术失控引发安全风险。

三、治理向善：以“规则向善”构建约束框架

技术的自律需要制度兜底。“向善”不能仅依赖企业或开发者的自觉，必须通过刚性规则与柔性引导形成治理合力。

首先，在法律法规的刚性约束方面，需明确 AI 研发、应用的“红线”与“底线”，对高风险“人工智能+”场景（如医疗诊断、自动驾驶、公共安全）设立强制准入标准，对 AI 滥用行为需明确法律责任与处罚机制。例如，欧盟《人工智能法案》将 AI 分为“不可接受风险”“高风险”“有限风险”“低风险”四类，对“高风险 AI”要求事前合规评估、事后持续监控，体现了分类治理的向善逻辑。同时，需要进一步明确人工智能出现失职时的责任主体，当 AI 出现失误时，是开发者技术缺陷、使用者操作不当、还是平台方监管缺失所导致的？责任主体是谁？这些都是法律法规需要进一步考量的内容，避免出现责任真空。

其次，在行业标准的柔性引导方面，由政府、学界、企业、公众共同制定跨行业的 AI 伦理标准，引导市场主体主动遵循，例如要求科技企业披露 AI 产品的伦理影响评估报告，或加入行业组织并接受行业监督。

最后，在多方协同的治理生态方面，需进一步构建多元治理体系，政府负责制定规则，企业履行主体责任，学界提供技术与伦理支撑，公众参与标准制定与问题反馈，避免单一主体主导导致的治理失衡。

四、应用向善：以“落地向善”实现价值闭环

人工智能的价值最终通过应用体现，“向善”必须落地为对具体场景的正向影响，而非虚浮的概念，核心是适配场景需求、规避应用风险。

在人工智能进入特定场景前，需先评估其潜在的社会影响，例如，“人工智能+医疗”需通过临床试验验证准确性，避免因误诊延误治疗；“人工智能+教育”需聚焦个性化辅导，而非“替代教师”，能够推动缓解教育资源的不均衡。

此外，“人工智能+”不能进一步加深数字鸿沟。需推动 AI 技术向弱势群体、欠发达地区倾斜，如为视障人士开发 AI 导盲工具、为农村地区提供 AI 农业技术指导，避免仅服务于少数群体，确保技术红利的公平分配。

四、结论

“人工智能+”的向善之路从来不是单一维度的孤军奋进，目标向善、技术向善、治理向善和应用向善，四者环环相扣、协同共生。目标是指南和初心，始终致力于增进人民福祉、赋能高质量发展与实现社会主义现代化；技术是基石，是“人工智能+”向善的起点，缺乏可解释、可管控的技术，治理与应用也便成为空中楼阁；治理是护航，没有刚性约束与柔性引导，技术优势可能异化为风险源头；应用是归宿，唯有在医疗、教育等场景中促进公平、提升效能、实现普惠，向善才不是空洞口号。

报告声明

本报告分析及建议所依据的信息均来源于公开资料，本公司对这些信息的准确性和完整性不作任何保证，也不保证所依据的信息和建议不会发生任何变化。我们已力求报告内容的客观、公正，但文中的观点、结论和建议仅供参考，不构成任何投资建议。投资者依据本报告提供的信息进行证券投资所造成的一切后果，本公司概不负责。

本报告版权仅为本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。如引用、刊发，需注明出处为大公国际，且不得对本报告进行有悖原意的引用、删节和修改。