

推动人工智能包容性发展 实现技术进步与社会价值相协调：

在中国高层发展论坛 2026 年年会

“人工智能治理：合作与挑战”专题研讨会上主题演讲

李 萌

(中国可持续发展研究会, 北京 100038)

中图分类号: TP18

文献标识码: A

文章编号: 1005-0566(2026)04-0001-04

人工智能是当下中国乃至全球范围内成长空间最广阔、发展增速最迅猛的核心领域,更是驱动新质生产力发展的关键引领性力量。党的十八大以来,以习近平同志为核心的党中央准确把握全球科技前沿趋势,对我国人工智能发展进行前瞻谋划和系统部署。2026 年 3 月发布的《中华人民共和国国民经济和社会发展第十五个五年规划纲要》明确要求,推动建立人工智能全生命周期风险管理制度^[1]。作为兼具技术属性与社会属性的新兴领域,人工智能能否真正形成规模化、高质量的新质生产力,绝非单纯取决于技术迭代的先进性与产业发展的成熟度,更核心的是能否为经济社会全局发展注入正向动能,助力破解经济社会结构中的深层次矛盾与问题;能否牢牢把握“智能向善”^①的核心导向,科学驾驭、有效引导技术发展,确保人工智能始终以造福人类为根本宗旨,杜绝其产生损害人类利益、扰乱社会秩序的负面影响。

当前人工智能正经历深刻质变,超级智能体时代的到来,安全问题已然上升为首要议题。随着大模型能力提升、多模态技术突破与智能体自主决策演进,人工智能正在从辅助工具转向自主执行者,加速融入各类关键行业。技术能力边界不断拓展带来了现实和潜在风险的全面升级,系

统失灵、技术滥用与各类衍生风险成为全球人工智能治理面临的三大核心挑战。系统失灵方面,如自动驾驶系统在复杂环境下生成错误指令,机器人遭受恶意攻击导致关键工序中断,智能体因安全机制缺陷、逻辑错误等引发误操作风险等。技术滥用方面,如利用人脸识别技术伪造身份实施诈骗,智能金融系统被入侵造成经济损失,人工智能生成虚假实验结果进行学术造假,利用人工智能炮制软文扰乱市场竞争秩序,利用人工智能伪造法律文书干扰司法公正等。衍生风险方面,如人工智能可能替代标准化岗位加剧结构性失业风险,发达国家与发展中国家的能力差距和智能鸿沟扩大可能加剧全球范围内的不平等。

人工智能治理任务复杂而艰巨:一是随着人工智能能力边界的持续拓展,如何在追求技术效率的同时,确保发展的公平性、包容性与可持续性,成为亟待破解的核心命题;二是随着人工智能应用场景的不断丰富,新型风险持续涌现,如何通过动态调整治理方式实现有效应对,成为急需解决的重要问题;三是随着人工智能产业链条日益复杂,参与主体更加多元,如何构建高效协同的多元治理机制,成为亟待完善的关键环节;四是随着人工智能全球竞争不断加剧,治理规则分化趋势

作者简介:李萌,中国可持续发展研究会理事长,十四届全国政协委员。曾任科学技术部副部长。主要从事科技政策、人工智能战略和治理、产业创新研究。

^① 我国《全球人工智能治理倡议》提出了“以人为本、智能向善”倡议。

显现,如何凝聚全球共识、构建公平包容的国际治理框架,成为亟须突破的重大挑战。

发展人工智能必须走包容性发展之路,兼顾速度与质量、效率与公平、发展与安全,解决好以下 3 个关键问题。一是创新与向善同频,积极应对民众关切的社会议题。超越“不作恶”底线,主动引导人工智能符合伦理规范,鼓励面向公共福祉的智能应用。二是妥善处理技术应用过程中带来的诸多风险挑战,包括帮助劳动者应对“AI 换人”挑战,以及前瞻识别、评估和防控深度伪造、网络安全等风险。三是从源头出发,确保技术本身安全、可靠、可控发展。既要从基础层推动算法透明可解释、数据安全可信,构建技术研发全生命周期监管框架,实现风险早期预警,也要从应用层制定智能体、端侧智能、人形机器人等新技术应用的安全标准等。

因此,人工智能包容性发展的主要目标是:打造兼具速度、温度、绿色、善度的人工智能新质生产力。“速度”就是技术创新,促进经济发展;“温度”就是惠及民生,促进社会公平;“绿色”就是绿色低碳,促进可持续发展;“善度”就是智能向善,促进安全可控。实现包容性发展的具体路径如下。

一、由微观主体增效向宏观经济增长拓展,发展全要素生产率友好型人工智能

在微观经济主体层面,人工智能技术的应用已展现出显著价值,通过优化业务流程、重塑运营模式、压缩生产运营成本、提升资源配置效率,有效拓宽了企业的盈利边界与增长空间。如在物流领域,人工智能赋能供应链,可使成本下降 5.88%,全行业潜在降本规模高达 6.77 万亿元^[2],充分印证了人工智能在微观主体增效中的强大作用。但从经济发展全局来看,当前企业微观经营效率的提升,尚未充分转化为宏观经济的持续增长动能,新一轮人工智能技术创新的红利,对全要素生产率的整体拉动作用仍有待释放。

从技术发展规律来看,重大新技术体系的落地应用,往往存在双重特征:一方面能够大幅降低物质资源消耗,减少无效投入;另一方面其对全要

素生产率的提升存在一定滞后期,难以在短期内快速显现成效。推动人工智能包容性发展,其核心目标就是加快缩短这一滞后期,推动技术优势切实转化为国内生产总值(GDP)增长动能与全要素生产率的全面提升,让不同地区、行业、群体间协同发展,把人工智能发展红利全面转化为全民可感知的收入增长与生活改善,避免“智能技术差距”演变为“社会分配鸿沟”,真正防范马太效应,实现发展成果全民共享。

二、由人力替代型向就业创造型转变,发展就业友好型人工智能

任何新技术的普及应用,都会伴随就业替代与就业创造的双重效应,人工智能作为颠覆性技术,其劳动力替代率远高于传统技术,其发展进程也完美契合工业革命以来新技术影响就业的杰文斯悖论:即技术提升虽会替代部分劳动力岗位,但同时会催生新产业、新需求,最终带动社会就业总量实现增长,且人工智能的就业替代与就业创造并非简单的对称关系,呈现出短期局部替代与长期分布式增长相交织的非对称特征。

当前,人工智能对就业的影响是岗位替代与岗位创造并存。在单点领域的岗位替代效应较为直观,而新型就业岗位创造则具有分散性、滞后性,这种时空不对称性容易引发社会就业焦虑。因为相比于其他技术,社会公众对人工智能的劳动替代感受更为强烈。流程化工作(如整理记录、基础翻译、标准化代码生成等)的人工智能替代效应明显,正在加大初等技能劳动者就业难度。新创造的岗位普遍具有更高的创造性和更复杂的行动性技能要求,劳动力市场正在向技能密集型、创造密集型方向转型。“就业友好”需关注技术应用的公平性,确保技术红利惠及更多劳动者^[3]。要看到人工智能时代创造性思维、独立前瞻判断、复杂实践能力以及优秀价值观更为宝贵,这类岗位始终存在稀缺性,人工智能无法替代。

就业是民生之本,新技术带来的就业影响是社会公众感受最直接、最关切的问题,因此就业治理是人工智能包容性发展中必须优先重视的核心课题。推动就业友好型人工智能发展,既要依托

人工智能技术催生新兴产业、新业态、新模式,培育更多高质量新型就业岗位,也要建立健全过渡性就业保障机制,化解集中性就业替代与分布式就业增长的矛盾,让杰文斯悖论的正向效应充分释放。要构建多主体协同的治理体系,鼓励企业建立新岗位,对应给予税收与政策扶持等激励,着力培育高质量就业供给,推动技能培训体系转型。通过“就业友好型人工智能”、终身技能培训、社会保障政策等,让人工智能发展成为稳定就业、扩大就业的重要推动力,让民众共享人工智能发展红利。

中国高度重视人工智能时代的就业治理,国家层面已经建立了完善的社保制度,形成了失业缓冲与兜底机制。未来在具体工作中要探索短期局部替代与长期全域扩容之间平滑过渡路径,化解集中替代与分散增长的时空差异问题,实现技术进步与就业扩容的良性互动。

三、由高能耗向低能耗转变,发展绿色低碳友好型人工智能

“十五五”是实现碳达峰的关键时期。根据世界银行与国际电信联盟(ITU)2024年联合报告测算,目前信息与通信技术行业的碳排放约占全球总量的1.5%~4%^[4]。行业内常说“人工智能的尽头是能源^[5]”,这一观点精准道出了人工智能对能源支撑的高度依赖。因此,加快推进绿色人工智能迫在眉睫。

当前,人工智能产业能耗居高不下已成为不可忽视的现实问题,尤其是在大模型训练、海量数据运算等环节;与此同时,风电、光伏、水电等新能源产业受电网消纳能力限制,出现弃风、弃光、弃水等资源浪费现象。二者形成鲜明反差。破解这一矛盾,必须将绿色低碳作为人工智能发展的核心导向,制定绿色人工智能发展战略,构建绿色人工智能技术标准与认证体系,建立多部门协调机制展开联合行动,推动人工智能降耗减碳。

要聚焦当下短板,布局低能耗绿色计算体系。一要统筹推进绿色算力基础设施建设,从技术源头优化算法模型、软件架构,在推理、渲染等应用端场景,大幅削减云计算、边缘计算用户的算力消

耗。二要创新算力结构,加快发展芯片集成堆叠、资源池化、网络切片、新型热管理等高效算力技术,提高算力效率。三要探索新计算范式,加大光计算、类脑计算、量子计算等新型计算技术的研发投入,突破传统计算模式的能耗瓶颈,从根本上降低数据计算与传输的能源消耗与碳排放,推动人工智能产业向绿色低碳转型。

在评价体系构建上,要树立循序渐进的绿色发展标准:现阶段加快建立以“能效比”为核心的绿色人工智能衡量体系,聚焦单位能耗的智能产出效率;未来逐步演进为“能效比”与“碳效比”相结合的综合评价体系^[6],其中“碳效比”特指单位碳排放对应的智能涌现能力。尽管当前智能涌现的评价维度多元、各类大模型性能测评标准尚未统一,但随着技术发展与行业探索,未来必将形成共识性、统一化的衡量标准。“能效比”与“碳效比”的边际效益持续提升,既是实现“双碳”目标的关键支撑,更是扭转人工智能高耗能标签、推动人工智能实现可持续发展的关键路径,对经济社会整体绿色低碳转型具有深远意义。

四、由通用人工智能“恐惧焦虑”向人机协同转变,发展人机友好型人工智能

智能社会的到来,是人类社会历经工业革命250余年发展后面临的最深刻变革,也带来了前所未有的挑战。随着头部企业对通用人工智能(AGI)的持续探索,智能体的能力不断迭代升级,引发了社会各界对人类未来命运的担忧与讨论。当前,关于人与人工智能的关系,存在“工具论”“伙伴论”等不同观点,但无论何种定位,人工智能发展必须坚守“以人为本”的底线,这一原则与早期机器人发展第一定律“机器人不得伤害人类”一脉相承,也应是大型模型、智能体时代人工智能发展的根本遵循。真正成熟的文明选择应该是在制度、架构和物理层面,永久拒绝任何形式的“智能主权让渡”^[7]。面对人工智能可能带来的潜在风险与负面影响,我们既不能视而不见、放任自流,也不必盲目恐慌、过度焦虑。人工智能本身并无善恶,其发展走向完全由人类掌控,只要我们精准把握技术发展规律,建立健全风险防控机制,提前

划定技术发展边界,就能有效规避智能“意外”,实现趋利避害。

要始终把智能体发展与人的关系定位在技术层面和工具范畴,要么是牛马,为人劳动干活,要么是猫狗,作为人的陪伴宠物。引导人类与人工智能协同进化、相互适应、共同成长,让人机协同成为社会发展的常态,确保人工智能始终服务于人类福祉。

五、由一国一企获益向全球共享转变,发展全球友好型人工智能

随着人工智能技术能力边界的持续拓展,全球范围内对技术滥用、失控的担忧日益加剧,构建高效、协同的全球人工智能治理体系已成为各国科学界、政府部门的普遍共识。但由于各国技术发展水平、发展理念存在差异,在人工智能治理的态度、思路与路径上仍存在分歧,难以形成统一合力。

中国始终坚持人工智能开源开放、包容普惠的发展道路,以实际行动推动全球友好型人工智能建设:通过效率革命降低经济运行门槛,推动人工智能技术平权,让更多主体能够共享技术红利;通过开源生态建设降低技术研发门槛,让人工智能创新不再局限于少数国家与企业。中国的开源发展路线,不仅为国内产业发展注入活力,更为全球人工智能领域尤其是硅谷等核心创新区域带来了发展福利,进一步凝聚了“人工智能应该是造福全人类的国际公共产品”^[8]的国际共识。包容、普惠的全球人工智能治理体系是实现全球公平发展的基础条件^[9]。人工智能发展没有单方胜出者,全球人工智能大国肩负着共同的责任与使命——让技术始终坚守向善轨道,造福全人类。实现这一目标,需要加强各国发展战略、治理规则的对接与协调,加快构建具有广泛国际共识的全球治理框架与标准规范,形成安全、可控、包容的全球治理体系,逐步将人工智能打造成为更安全、可靠、普惠的基础设施力量。加大对欠发达国家的技术

支持与帮扶,助力其提升人工智能应用水平,缩小国家间的智能发展差距,消弭全球智能鸿沟,在全球范围内实现包容共生与价值共享。

在全球人工智能发展进程中,人类必须确立主体地位、发挥主导作用,要由人类定义人工智能的未来,而非让人工智能定义人类的未来,坚决杜绝人工智能脱离人类掌控。要通过敏捷治理与国际协同合作,真正让人工智能成为惠及全人类的全球公共产品,推动构建人类命运共同体框架下的智能文明新生态。

参考文献:

- [1] 中国政府网. 中华人民共和国国民经济和社会发展第十五个五年规划纲要[EB/OL]. (2026-03-13)[2026-04-20]. https://www.gov.cn/yaowen/liebiao/202603/content_7062633.htm?enable_bottom_share_style=1&hybrid_event_param=HybridEventP.
- [2] 京东集团. 京东率先在行业内发布工业供应链大模型发起“智赋千业 万亿降本”行动[N]. 经济日报,2025-09-25(12).
- [3] 李西顺. 就业友好型发展方式的特征、挑战与对策[J]. 人民论坛·学术前沿,2025(6):60-72.
- [4] SETH A, BALLAN S, GRAY V, et al. Measuring the emissions and energy footprint of the ICT sector: implications for climate action. Washington, D. C: World bank group, 2024[R/OL]. (2024-03-21)[2026-04-20]. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099121223165540890>.
- [5] 孙亚军. AI 能耗困扰并非无解[N]. 经济日报,2024-04-25(4).
- [6] 李萌. 当前人工智能发展和治理中几个热点问题的认识[J]. 中国科学院院刊,2026,41(3):435-441.
- [7] 李国杰. 基于可判定性理论的人工智能系统安全风险分类[J]. 计算机研究与发展,2026,63(3):539-547.
- [8] 习近平. 向 2025 世界智能产业博览会致贺信[N]. 人民日报,2025-09-06(1).
- [9] 薛澜,梁正. 构建包容、普惠的全球人工智能治理体系:现实挑战与对策建议[J]. 人民论坛·学术前沿,2025(9):23-32.