

以普惠包容的人工智能治理 赋能全球可持续发展



世界互联网大会人工智能专业委员会
安全与治理推进计划
2025年4月

世界互联网大会人工智能专业委员会 安全与治理推进计划

牵头人

曾毅 中国科学院自动化研究所

肖恩·欧·海格缇 剑桥大学

专家及贡献者

(包括推进计划成员和特邀专家, 以及为报告提供内容或意见的专家,
按单位名称首字母排序)

阿拉伯信息通信技术组织

纳达·拉阿比迪

巴西南大河州联邦大学

埃德森·普雷斯蒂

北京金山云网络技术有限公司

西京京

北京邮电大学

杨忠良

德国慕尼黑工业大学

克里斯托夫·吕特格

杭州安恒信息技术股份有限公司

王欣、李华伟

IQilibriumAI

希梅娜·索菲亚·比维罗斯·阿尔瓦雷斯

联合国大学澳门研究所

塞尔日·斯廷克维奇

南开大学

陶锋

人民网股份有限公司

谷文杰

上海人工智能实验室

乔宇

深圳市腾讯计算机系统有限公司

邓雯洁、翟家欢、李大任、王梦寅、曹建峰

阿里巴巴（中国）有限公司

张荣、王星光

北京大学

杨耀东

北京印象笔记科技有限公司

乔迁

德国明斯特大学

伯纳德·霍尔兹纳格尔

伏羲智库

李娜

金砖国家未来网络研究院（中国·深圳）

周原

联想（北京）有限公司

胡永启

联合国教科文组织驻东亚代表处

刘唱、夏和华

清华大学

唐新华

上海诺基亚贝尔股份有限公司

王彤、叶晨晖

上海稀宇极智科技有限公司

彭韬

世界互联网大会

梁昊、张雪丽

斯里兰卡计算机应急响应中心

尼罗沙·阿南达、查鲁卡·塞纳尔·达穆努波拉

香港科技大学

冯雁

香港中文大学

蒙美玲、张寄冀

英国南安普敦大学

温迪·霍尔

中关村实验室

谭知行

中国电子技术标准化研究院

张妍婷

中国联合网络通信集团有限公司

周凯

中国社会科学院哲学研究所

段伟文

中兴通讯股份有限公司

孟伟

中国移动研究院（中移智库）

林琳、吴淑燕、刘玥、赵英、邓洁

360集团

肖荐午

**乌兹别克斯坦高等教育科学
和创新部人工智能咨询委员会**

艾墨·尤纳斯

香港岭南大学

姚新

央视国际网络有限公司

程明

云安全联盟

约翰·杨

中国电信集团有限公司（天翼智库）

陈凤仙、刘威辰、王峰

中国互联网络信息中心

陈晶晶

中国人民大学

龚新奇

中国网络空间安全协会

王健兵、夏文辉

中国信息通信研究院

呼娜英、郭苏敏

中国政法大学

张凌寒

编写团队

中国科学院自动化研究所
人工智能安全与超级对齐北京市重点实验室
北京人工智能安全与治理实验室

曾毅、鲁恩萌、范津宇、皇甫存青

中国科学院大学

王金、王正奇、郭晓阳

世界互联网大会

康彦荣、韩开宇

联系邮箱

research@wicinternet.org



前言

当前,人工智能(AI)技术正迅速发展,深刻影响全球经济、社会 and 环境的各个方面,成为推动可持续发展的关键驱动力。其中,人工智能治理既为人工智能有效赋能可持续发展目标(SDGs)提供了方向指引,又为防范人工智能可能对可持续发展进程带来的负面影响和风险提供了有力保障。2024年,联合国大会相继通过两项决议,推动完善全球人工智能治理,加快落实可持续发展议程;《全球数字契约》将弥合数字鸿沟,加快可持续发展和加强人工智能国际治理,造福人类作为目标。

世界互联网大会依托人工智能专业委员会安全与治理推进计划,广泛汇聚国内外人工智能领域国际组织、高校、智库、行业协会及领军企业专家,共同开展人工智能治理与可持续发展联合研究,力求凝聚国际共识,加速议题求解,推动人工智能更好造福人类社会。

本报告介绍了人工智能赋能可持续发展的全球进程、当前进展与现存问题,并从如何通过普惠包容的人工智能发展与治理弥合数字鸿沟和推动人工智能能力建设两方面进行分析阐述,最后从深化面向可持续发展的人工智能治理国际合作出发,提出加速赋能全球可持续发展的行动建议,梳理实践案例;旨在为各国政府、企业和学术界提供参考,为加速人工智能赋能全球可持续发展目标贡献智慧与力量。

目录

01 可持续发展与人工智能治理：全球进程-----01

- (一) 可持续发展是兼顾多方的综合过程-----01
- (二) 人工智能极大影响可持续发展进程-----02
- (三) 促进赋能可持续发展是全球人工智能治理共识-----03

02 人工智能赋能可持续发展的当前进展和未来方向-----05

- (一) 人工智能赋能可持续发展的现状与进展-----05
- (二) 人工智能赋能可持续发展中的现存问题-----06

03 通过普惠包容的人工智能发展与治理，弥合数字鸿沟-----08

- (一) 从数字鸿沟到智能鸿沟-----08
- (二) 释放人工智能潜力促进数字素养提升-----09
- (三) 构建更加普惠和有活力的人工智能研发与应用生态-----09

04 推动人工智能能力建设，加速赋能可持续发展进程-----11

- (一) 发展中国家在推动人工智能赋能可持续目标上面临的困境-----11
- (二) 人工智能能力建设国际合作为加速赋能可持续发展提供历史性机遇-----12
- (三) 探索可负担可持续的人工智能能力建设国际合作原则-----13

05 面向可持续发展的人工智能治理国际合作：行动建议-----14

- (一) 全面推进人工智能赋能可持续发展目标的实现-----14
- (二) 推进人工智能赋能可持续发展的普惠性-----14
- (三) 推进人工智能赋能可持续发展的包容性-----14
- (四) 与中低收入国家共同探索可负担、可持续的能力建设模式-----15
- (五) 以国际治理多边合作机制为人工智能赋能可持续发展提供保障-----15

附录 人工智能治理与可持续发展应用与实践案例-----16

可持续发展与人工智能治理：全球进程

2015年,联合国大会通过了《改变我们的世界——2030年可持续发展议程》(《2030年议程》或《2030年可持续发展议程》)¹。当前,《2030年议程》实施已过去三分之二,但目标实现面临严峻挑战。而人工智能作为近十年最具颠覆性的技术突破,既为实现可持续发展目标带来重大机遇,也带来了多重挑战。通过人工智能治理为可持续发展目标提供引导与保障已成为关乎人类共同未来的紧迫任务。

(一)可持续发展是兼顾多方的综合过程

可持续发展是指既满足当代人的需求,又不损害后代人满足其需求的能力的发展模式²。2015年,联合国举办了可持续发展峰会,193个联合国会员国一致通过了《改变我们的世界——2030年可持续发展议程》³。该议程为全球设定了17个目标,包括消除饥饿、良好教育、性别平等、健康福祉、应对气候变化、促进可持续经济增长等多个方面。**17个可持续发展目标是人类的共同愿景,也是世界各国领导人与各国人民之间达成的社会契约⁴。**



图 1: 17个可持续发展目标之间的关联⁵

17个可持续发展目标是不可分割的整体,它们兼顾了可持续发展的三个方面:经济、社会与环境。可持续发展目标包含产业、创新和基础设施(目标9)等经济目标,优质教育(目标4)等社会目标以及水下生物(目标14)等环境目标。同时,根据图1可以看出,任何一个目标都是复杂的,不仅关联自己所在的领域,更存在跨领域的深度交织,这意味着贡献于任何一个目标的实现,都可能同时带动其他相关目标。

1 联大决议,《改变我们的世界——2030年可持续发展议程》(2015年9月),来源:<https://sdgs.un.org/2030agenda>

2 联合国世界环境与发展委员会,《我们的共同未来》(1987年4月),来源:<http://www.un-documents.net/our-common-future.pdf>

3 来源:<https://sdgs.un.org/2030agenda>

4 联合国新闻,时任联合国秘书长潘基文相关论断(2015年12月),来源:<https://news.un.org/zh/story/2015/12/249342>

5 此项工作来自远期人工智能研究中心构建的“面向可持续发展的人工智能智库平台”,使用关键词共现分析方法对联合国可持续发展网站公布的案例进行分析。

图片揭示了各可持续发展目标之间的关联模式,线条的粗细反映了两个可持续发展目标之间关系的紧密程度。

在寻求可持续发展过程中,发展中国家与发达国家共同肩负责任¹,但由于历史原因、发展水平和能力等方面的差异,发展中国家与发达国家的责任和义务有所不同。在新冠疫情、地缘政治冲突、气候变化等多重因素的影响下,发展中国家面临发展资金短缺、债务风险上升等挑战²。《2030年议程》指出,国际社会应当为发展中国家提供更多援助与支持,以实现全球公平与包容的可持续发展。

(二)人工智能极大影响可持续发展进程

人工智能是加速可持续发展进程的重大机遇。作为过去十年来最深刻影响人类社会的颠覆性赋能技术,人工智能从不同角度为可持续发展提供新的可能。在经济方面,人工智能作为新型生产要素,能促进产业创新与转型³,提供新的就业机会。在社会方面,人工智能促进教育⁴、医疗⁵等公共服务的公平普惠,向发展中国家、偏远地区与弱势群体输送优质资源⁶。在环境方面,人工智能助力绿色革命⁷,帮助实现节能减排⁸,优化资源配置。在科学方面,人工智能在生物医药⁹、地球科学¹⁰、生物多样性保护¹¹等领域展现突破性潜力,通过科技创新为全球性挑战提供新的解决方案。

然而,人工智能也对可持续发展目标的实现带来诸多挑战。例如,人工智能可能对气候与环境的

可持续发展目标带来压力:人工智能目前的技术路径的算力需求可能造成能源负担¹²,增加碳排放量;人工智能硬件的快速迭代可能导致电子废弃物堆积。此外,人工智能给平等包容带来挑战,表现在:人工智能本身的数据偏见与算法偏见有引入或固化歧视的风险,导致误解与不信任¹³;人工智能的广泛应用向劳动者提出新要求,可能引发结构性失业¹⁴;技术与数据集中在少数国家手中,进一步扩大南北差距。在安全方面,人工智能的内容生成与输出能力可能操纵舆论,产生虚假与欺骗性内容,对个人或群体带来伤害¹⁵;强大的人工智能可能被滥用、恶用或误用,例如可能增加网络攻击与生物恐怖主义的风险¹⁶,可能在军事行动中造成人道主义挑战,可能加剧隐私侵权与版权侵权等;此外,人工智能系统自身的故障或失控也可能导致风险。这些挑战与风险给可持续发展目标的实现带来了新的难题。

总之,快速发展的人工智能正以前所未有的深度和广度重塑人类社会,其治理的紧迫性日益凸显。只有妥善治理人工智能的风险,人类才能安全、充分地从获益¹⁷。人工智能治理的核心在于通过制度性框架实现技术红利最大化与社会风险最小化的动态平衡,确保智能向善,防止误用、滥用和恶用。这不仅是防范技术异化的安全底线,更是确保人工智能始终服务于人类文明进步和社会生态福祉¹⁸的价值基石。

1 联大决议,《改变我们的世界——2030年可持续发展议程》(2015年9月),来源: <https://sdgs.un.org/2030agenda>

2 联大发言,中方呼吁国际社会帮助发展中国家实现经济社会可持续发展(2024年4月),来源: https://www.gov.cn/yaowen/liebiao/202404/content_6945579.htm

3 中国电信自研人工智能节能系统:年均可节电8亿度,节约电费5.2亿元,来源: <https://www.c114.com.cn/news/117/a1275451.html>

4 Vicari, R., Brackmann, C., Mizusaki, L., & Galafassi, C. (2024). Políticas e governança para apoiar a inserção da Inteligência Artificial na Educação potencializando seus benefícios e minimizando seus riscos (Nota Técnica). SESI - Departamento Nacional.

5 人工智能的医疗应用,瑞金医院与华为的瑞金医院病理大模型,来源: <https://news.sciencenet.cn/htmlnews/2025/2/538982.shtm>

6 中国移动,智慧村医在吉林开启乡村医疗新模式,来源: https://www.10086.cn/download/csrreport/cmcc_2023_csr_report_full_cn.pdf

7 中国移动,人工智能赋能实现数据中心的高效节能和低碳运维,来源: <https://sjzyj.wuhu.gov.cn/sjzy/8664228.html>

8 常州市行政中心,通过部署人工智能技术将市政中心空调制冷功耗降低15%(2025年3月),来源: <https://js.ifeng.com/c/8hMOL3DbIZZ>

9 阿西莫夫出版社,Evo 2可设计整个基因组,来源: <https://press.asimov.com/articles/evo-2>

10 开放智算产业联盟,PECIP项目以AI驱动气候行动和全球可持续发展为极地科研赋能,来源: https://innoport.cuhk.edu.hk/single-cubiczine-eng/hokinchung_en/
<https://mp.weixin.qq.com/s/XnCU3nEdrnY5xdidBhGcMQ>

11 腾讯,野朋友计划,来源: <https://techforgood.qq.com/projects/detail/313>

12 Saklani, S., & Singh, D. (2024). Minimizing Carbon Emissions by Improving Water and Energy Use Efficiencies in AI Servers: A Green Cloud Computing Strategy for Sustainable Artificial Intelligence Systems. International Journal of Innovative Science and Research Technology (IJISRT).

13 在用于检测英国福利欺诈的AI系统中发现的偏差,来源: <https://www.jaag.org.uk/blog/bias-found-in-ai-system-used-to-detect-uk-benefits-fraud>

14 国际货币基金组织,人工智能将改变全球经济,让我们确保它能造福全人类(2024年1月),来源: <https://www.imf.org/zh/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity>

15 首届人工智能安全峰会,《布莱切利宣言》(2023年11月),来源: <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>

16 新美国安全中心,《人工智能与生物国家安全风险的演变:能力、阈值和干预措施》(2024年8月),来源: <https://www.cnas.org/press/press-release/new-cnas-report-on-ai-and-biological-national-security-risks>;
https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/AIBiologicalRisk_2024_Final.pdf

17 Y. Bengio 等,《国际人工智能安全报告》(DSIT 2025/001),来源: <https://www.gov.uk/government/publications/international-ai-safety-report-2025>

18 《人工智能北京共识》(2019年5月),来源: <https://ai-ethics-and-governance.institute/beijing-artificial-intelligence-principles/>

(三) 促进赋能可持续发展是全球人工智能治理共识

人工智能在赋能可持续发展中的应用是机遇与挑战并存的正向实践,有效的治理为其提供了引导与保障。包括欧盟《人工智能法》¹在内的法律框架,以及中国《人工智能安全治理框架》²和美国《人工智能风险管理框架》³在内的治理框架,采取了基于风险进行分类和监管的方式;而中国《生成式人工智能服务管理暂行办法》⁴《人工智能生成合成内容标识方法》⁵和新加坡《生成式人工智能治理模型框架》⁶等则针对生成式人工智能技术广泛应用的现实情况进行专门的监管与治理。这些工作为促进技术有序发展与赋能、防范滥用与恶用提供了有益参考,为各国以负责任方式应对人工智能技术的多重影响提供了借鉴。

联合国将人工智能赋能可持续发展作为近年来的关键议题。联合国大会于2024年先后通过“抓住安全、可靠和值得信赖的人工智能系统带来的机遇,促进可持续发展”与“加强人工智能能力建设国际合作”两项决议,引导与保障人工智能技术服务于可持续发展目标的实现。联合国人工智能高级别咨询机构发布《治理人工智能,助力造福人类》最终报告⁷,强调治理可成为推动人工智能创新促进全球实现可持续发展目标的关键因素,并为此提出了多项具体建议,包括成立国际人工智能科学小组和建立人工智能治理全球对话机制来探讨人工智能有助于实现可持续发展目标的领域,部分相关建议被2024年9月联合国未来峰会正式通过的《全球数字

契约》⁸采纳。《全球数字契约》提出了包括“弥合所有数字鸿沟,加快在实现各项可持续发展目标方面取得进展”和“加强人工智能国际治理,造福人类”在内的五项目标,通过一系列承诺和行动来加强国际合作推进全球数字治理,为人工智能促进可持续发展目标提供了框架性支撑。

此外,联合国教科文组织于2021年发表《人工智能伦理问题建议书》⁹,强调人工智能对社会、环境、生态系统及人类生活的深远影响,并积极支持“全球南方”国家平等参与人工智能国际治理与对话。世界卫生组织于2024年发布《大型多模态模型的卫生人工智能伦理与治理指导意见》¹⁰,关注人工智能在医疗领域中增进人类福祉。国际电信联盟通过发起“人工智能向善”系列倡议,聚焦人工智能技术在实际场景中的创新应用,以加快实现可持续发展目标,在全球范围内扩大人工智能解决方案。

各多边机制和对话平台不断完善治理框架,推动人工智能赋能可持续发展。经济合作与发展组织提供多学科、基于证据的政策分析和数据,以促进普惠与可持续的发展与变革,实现可持续和包容的未来¹¹。二十国集团(G20)于2019年批准《G20人工智能原则》¹²,强调人工智能研发需遵循以人为中心的发展理念。2023年新德里峰会¹³进一步呼吁建立“以人为中心”的人工智能治理框架,尤其重视发展中国家在技术应用中的公平性问题。欧盟认为人工智能技术对实现“碳中和”目标有重要意义,通过《人工智能协调计划》¹⁴与《包装和包装废物条例》¹⁵等,鼓励

1 来源: <https://artificialintelligenceact.eu/the-act/>

2 来源: https://www.cac.gov.cn/2024-09/09/c_1727567886199789.htm

3 来源: <https://www.nist.gov/itl/ai-risk-management-framework>

4 来源: https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm

5 来源: https://www.cac.gov.cn/2025-03/14/c_1743654684782215.htm

6 来源: <https://aiverifyfoundation.sg/wp-content/uploads/2024/05/Model-AI-Governance-Framework-for-Generative-AI-May-2024-1-1.pdf>

7 来源: https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_zh.pdf

8 来源: https://unesdoc.unesco.org/ark:/48223/pf0000380455_chi

9 来源: <https://www.who.int/publications/i/item/9789240084759>

10 国际电信联盟,人工智能向善全球峰会,来源: <https://aiforgood.itu.int/summit24/>

11 经合组织,《关于可持续和包容性未来的变革性科学技术和创新政策宣言》(2024年4月),来源: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0501>

12 来源: <https://wp.oecd.ai/app/uploads/2021/06/G20-AI-Principles.pdf>

13 G20,《二十国集团领导人新德里峰会宣言》(2023年9月),来源: https://g20.org/wp-content/uploads/2024/09/G20-2023_India_Declaracao-de-Lideres.pdf

14 来源: <https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence>

15 来源: https://environment.ec.europa.eu/news/new-rules-more-sustainable-and-competitive-packaging-economy-2025-02-11_en

通过人工智能技术发展气候预测模型和智能电网,改善工业设计,促进循环经济与绿色转型。**亚太经合组织**在《布特拉加亚愿景2040》¹中主张缩小数字鸿沟,深化包容性合作,通过良好的发展与治理实现可持续。**上海合作组织**发布《上海合作组织成员国元首理事会阿斯塔纳宣言》²,将2025年定为“可持续发展年”,倡导抓住数字转型机遇期,推动普惠包容的全球合作,加快落实2030年议程。**世界互联网大会**于2023年发布了《发展负责任的生成式人工智能共识》³,广泛汇聚各方智慧,推动生成式人工智能发展与治理协同共进。

“全球南方”国家抱有热切期待,呼吁缩小数字鸿沟,以实现平等普惠的可持续发展。2023年,由“77国集团和中国”峰会发表的《哈瓦那宣言》⁴提出要提高发展中国家获取科技的机会和发展科技的能力,建立一个开放、公平、包容、非歧视的科技发展环境。2024年初举办的第三次南方首脑会议⁵,呼吁缩小国家和区域内部、国家和区域之间以及发达国家和发展中国家之间在数据生成、基础设施和可及性方面的所有数字鸿沟和不平等。中国提出《全球人工智能治理倡议》⁶与《人工智能能力建设普惠计划》⁷,强调包容合作而非排他孤立的人工智能发展对于确保实现全球可持续发展进程至关重要,主张弥合数字和智能鸿沟,帮助“全球南方”在人工智能发展进程中平等受益。

深化人工智能治理,加速人工智能赋能可持续发展的全球进程需要更广泛和深入的国际合作。北方国家具有技术和资源优势,有能力承担国际责任,而南方国家则对获得针对性支持有迫切需求。因此。双方通过互补性合作共同应对人工智能与可

持续发展目标间的共性挑战,才能推动构建包容共享的全球治理体系,确保技术红利惠及全人类。

1 来源: https://www.apec.org/meeting-papers/leaders-declarations/2020/2020_aelm/annex-a

2 来源: https://www.mfa.gov.cn/zyxw/202407/t20240704_11448360.shtml

3 来源: https://cn.wicinternet.org/2023-11/09/content_36952741.htm

4 来源: https://www.gov.cn/yaowen/liebiao/202309/content_6904536.htm

5 第三次南方首脑会议成果文件,来源: http://www.g77.org/doc/3southsummit_outcome.htm

6 来源: https://www.fmprc.gov.cn/web/ziliao_674904/1179_674909/202310/t20231020_11164831.shtml

7 来源: https://www.mfa.gov.cn/web/wjzb_673089/xghd_673097/202412/t20241218_11496414.shtml

02

人工智能赋能可持续发展的当前进展和未来方向

本章主要介绍人工智能赋能可持续发展的现状与未来方向。人工智能为推动联合国2030年可持续发展议程带来了前所未有的机遇,但在当前赋能可持续发展的研究与实践中同样存在一些不足。因此,通过治理为人工智能赋能可持续发展提供方向上的引导至关重要。

(一) 人工智能赋能可持续发展的现状与进展

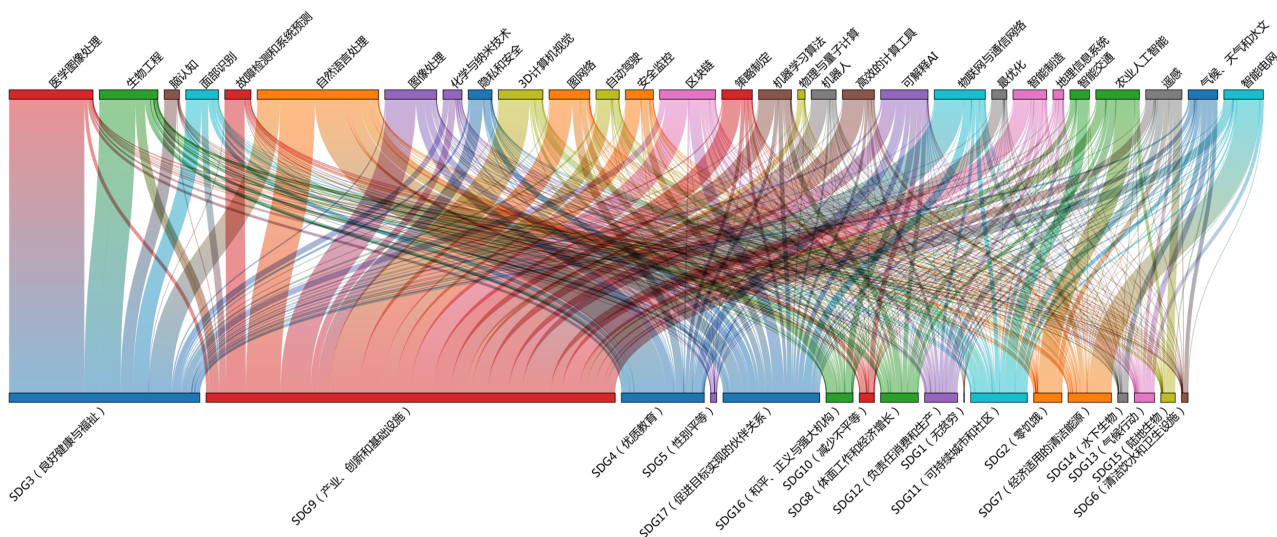


图2:人工智能研究领域与可持续发展目标的关联图(基于文献分析)¹

¹ 此项工作来自远期人工智能研究中心构建的“面向可持续发展的人工智能智库平台”。一共分析了来自Semantic Scholar检索平台的183,435篇高引用人工智能文献(至少20篇以上引用),其中96,404篇对至少一个可持续发展目标的实现有帮助。

人工智能在推动全球可持续发展目标方面有巨大的潜力。17个可持续发展目标包含了169个子目标¹，而人工智能有望促进79%的可持续发展目标，涵盖134个具体目标。其中，82%的社会目标，70%的经济目标和93%的环境目标可能从人工智能技术中受益。在17个可持续发展目标中，人工智能在目标1、4、6、7、9、11、14、15中可释放的潜力格外突出²。

当前人工智能在可持续发展目标中的研究已取得普遍进展。研究显示，人工智能各领域与各可持续发展目标之间存在普遍的关联，有52.6%的高引用人工智能文献对至少一个可持续发展目标有积极贡献，在所有研究领域与可持续发展目标的排列组合中，有91%的组合具有高引用论文³。如图2所示，人工智能各领域研究对良好的健康与福祉（目标3），优质教育（目标4），产业、创新和基础设施（目标9）以及促进目标实现的伙伴关系（目标17）赋能较多，其中图像处理与自然语言处理等技术领域影响重大。在研究较少的领域，人工智能也能发挥独特作用。如数据与遥感技术能定位贫穷区域以提供精准帮助⁴，促进无贫穷（目标1）的实现；词向量表征中的性别不平等问题调查与研究⁵能提高机器学习透明度，助力性别平等（目标5）的实现；机器学习可以帮助水质监测与管理⁶，赋能清洁饮水和卫生设施（目标6）的实现。

在应用与实践方面，在环境领域，法国人工智能行动峰会期间发布的《向协调全球人工智能环境

可持续性标准化方法迈进》⁷关注人工智能用于提升环境可持续的标准化与最佳实践。在医学领域，2024年智能向善案例集⁸筛选了来自19个国家的53条实践案例，展示了人工智能在医疗服务、诊断和治疗革命中具有重要意义。在农业领域，人工智能在作物管理⁹、农具维护和市场决策等方面已发挥关键作用，显著提升了农业生产力、优化资源利用并促进环境管理¹⁰。此外“面向可持续发展的人工智能智库平台”¹¹对于人工智能赋能可持续发展的案例统计显示，来自57个国家的超过400个案例充分涵盖了17个可持续发展目标方向。

(二) 人工智能赋能可持续发展中的现存问题

人工智能在赋能可持续发展进程中仍面临诸多挑战，主要体现在以下四个方面：

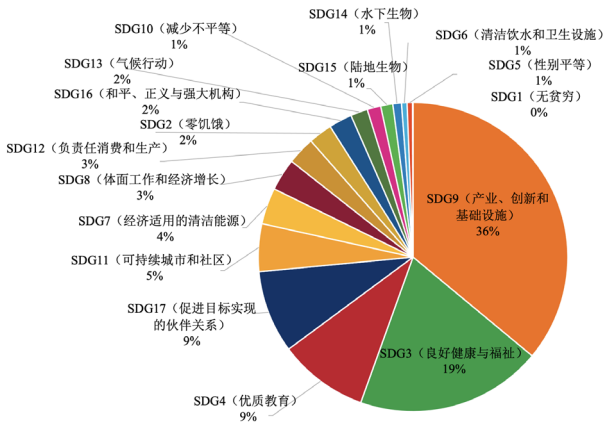


图3：人工智能赋能17个可持续发展目标的文献占比¹²

1 来源：<https://www.un.org/sustainabledevelopment/>
2 Vinuesa, R., Azizpour, H., Leite, I. et al. The role of artificial intelligence in achieving the Sustainable Development Goals. Nat Commun 11, 233 (2020). <https://doi.org/10.1038/s41467-019-14108-y>
3 来源同图2
4 Joshua Evan Blumenstock, Fighting poverty with data. Science 353, 753-754 (2016). DOI: 10.1126/science.aah5217
5 N. Garg, L. Schiebinger, D. Jurafsky, & J. Zou, Word embeddings quantify 100 years of gender and ethnic stereotypes, Proc. Natl. Acad. Sci. U.S.A. 115 (16) E3635-E3644, <https://doi.org/10.1073/pnas.1720347115> (2018).
6 Sit M, Demiray BZ, Xiang Z, Ewing GJ, Sermet Y, Demir I. A comprehensive review of deep learning applications in hydrology and water resources. Water Sci Technol. 2020 Dec;82(12):2635-2670. doi: 10.2166/wst.2020.369. PMID: 33341760.
7 来源：https://www.sustainableaicoalition.org/wp-content/uploads/Standardization_AI_Sustainability.pdf
8 国际电信联盟，《AI向善-创新影响2024年最终报告》，来源：https://www.itu.int/dms_pub/itu-t/obj/ai4g/T-AI4G-AI4GOOD-2024-3-PDF-E.pdf
9 中国信息通信研究院，《中国智慧农业发展报告》中指出运用人工智能赋能预测天气模式、优化灌溉和精准施肥（2021年12月），来源：<http://www.caict.ac.cn/kxyj/qwfb/ztbg/202201/P020220104495485440718.pdf>
10 国际电信联盟，《数字农业的数据建模》，来源：<https://www.itu.int/en/ITU-T/focusgroups/ai4a/Documents/Deliverables/FGAI4A-04.pdf>
11 来源：<https://ai-for-sdgs.academy/ai4sdgs-research-program-cn>
12 来源同图2

当前人工智能赋能可持续发展的潜力未能充分发挥。如今,超过三分之一的可持续目标停滞不前甚至倒退,其中零饥饿(目标2)、可持续城市和社区(目标11)、水下生物(目标14)和陆地生物(目标15)以及和平、正义和强大机构(目标16)情况尤为严峻¹。这其中,人工智能至少对目标2、11、14与15有能力产生重大积极影响²。但这些潜力大多未能得到良好发挥,如图3所示。此外,人工智能在无贫穷(目标1)、清洁饮水和卫生设施(目标6)减少不平等(目标10)以及气候行动(目标13)等高赋能潜力领域的研究同样有限。

人工智能在不同可持续发展目标的关注和投入不均衡。工业、医药等高产值领域因商业化前景明确,企业与资金投入意愿强烈。同时目前的科研资源更多投入到大语言模型、图像识别等技术领域(见图2)。然而,技术本身并不具备自发解决全球性社会问题的能力,各方资源的匹配与投入使得人工智能赋能可持续发展目标的部分领域率先形成“研究-应用”的正向循环。例如,智慧医疗与智慧教育通过降低运维成本直接创造收益,吸引大量资本;而水资源、生态保护等领域缺乏成熟商业模式,导致投入不足,难以形成闭环。

人工智能赋能可持续发展所依赖的数据资源不足。一方面,相关数据存在**国别与区域差距**,根据联合国《可持续发展目标报告2023》³,全球193个国家或地区中仅有不到一半拥有国际可比数据;无贫困(目标1)和性别平等(目标5)数据在发展中国家普遍存在城乡和区域代表性偏差,例如,非洲国家仅有20%的可持续发展目标指标具备完整统计体系支持。另一方面,可持续发展**不同指标的数据分布不均衡**。《可持续发展目标报告2024》⁴进一步显示,在169个目标中,34个目标缺乏足够的趋势数据或额外分析,且数据及时性不足,约三分之一指标

缺乏过去三年的最新数据。例如在气候行动(目标13)领域,2024年全球只有约10%的国家具有趋势数据。由于人工智能高度依赖数据资源,数据质量和代表性不足严重影响人工智能在这些可持续发展目标领域的研究与应用。

人工智能与不同可持续发展目标的技术适配度不均衡。在相关研究与实践中,产业、创新与基础设施(目标9)与几乎所有人工智能技术方向都有较高适配程度;可持续城市和社区(目标11)可整合智能交通、遥感、3D视觉等多领域技术;医疗与教育虽主要依赖图像识别与自然语言处理,但相关算法与技术成熟度较高且高度匹配。但在另外一些可持续发展目标中,人工智能发挥作用的空間尚未探索完成,这也导致目前人工智能在无贫穷(目标1),清洁饮水和卫生设施(目标6)和水下生物(目标14)等方面贡献较少。

总体来说,为充分发挥人工智能潜力以赋能可持续发展目标在2030年前如期实现,应在人工智能的发展和治理上进一步做出如下努力:**一是关注弥合数字鸿沟与智能鸿沟**,通过治理与合作防止人工智能加剧不平等,发挥人工智能的优势,帮助贫穷与脆弱地区和人群平等受益。**二是推动人工智能在可持续发展关键领域释放潜能**。加强科研创新与人才培养,在提升技术适配度的同时培育良好的人工智能生态,以充分发挥人工智能在减少贫困与饥饿、保护生物与环境保护与优化资源配置等可持续发展领域的潜力。**三是在成熟领域进一步发挥优势**。教育、医疗与工业是人工智能促进可持续发展的优势领域,这些领域技术相对成熟,产业链完整,市场化程度高。在直接提升教育与医疗质量,改善就业与经济情况,促进包容普惠机会平等之外,这些领域的应用经验也为人工智能赋能其他可持续发展目标提供了模式借鉴。

1 联合国,《可持续发展报告2024》,来源:<https://dashboards.sdgindex.org/chapters/executive-summary>

2 Vinuesa, R., Azizpour, H., Leite, I. et al. The role of artificial intelligence in achieving the Sustainable Development Goals. Nat Commun 11, 233 (2020). <https://doi.org/10.1038/s41467-019-14108-y>

3 来源:https://unstats.un.org/sdgs/report/2023/The-Sustainable-Development-Goals-Report-2023_Chinese.pdf

4 来源:<https://unstats.un.org/sdgs/report/2024/The-Sustainable-Development-Goals-Report-2024.pdf>

03

通过普惠包容的人工智能 发展与治理, 弥合数字鸿沟

弥合数字鸿沟是可持续发展的核心议题与重大挑战。联合国有关报告指出,除了数字接入的鸿沟,数字世界还存在数据鸿沟和创新鸿沟,这些数字不平等背后是治理的巨大差距¹。数字鸿沟直接制约着无贫穷(目标1)、优质教育(目标4)、减少不平等(目标10)等多项目标的实现。为此,联合国通过“可持续发展目标数字加速议程”²等行动,将弥合数字鸿沟纳入全球治理框架。

(一)从数字鸿沟到智能鸿沟

人工智能由于其技术积累的**马太效应**,一定程度上可能加剧数字鸿沟。人工智能技术为传统的数字鸿沟带来新的挑战:其一,高质量数据作为人工智能训练的核心资源,获取难度和成本高。其二,算力是人工智能的基础,发展中国家难以负担先进算力建设的高额投入。其三,算法研发能力的差距正形成人工智能核心技术代差;算法决策机制提高了公众的认知门槛。根据联合国和国际劳工组织的报告³,一方面,人工智能加剧了**社会群体间的分化**:掌

握人工智能技能的劳动力薪资溢价空间提升⁴;而低收入群体缺乏数字素养被排除在技术红利之外,收入差距与结构性失业风险叠加。另一方面,区域发展鸿沟因技术垄断持续扩大:部分科技企业和先发国家依托积累的算力基础设施和数据资源优势,在算法模型研发领域形成较强竞争力。而许多发展中国家受限于技术基础薄弱和资源获取渠道不足,技术升级进程面临挑战。

同时,人工智能也提供了跨越式发展的可能,为弥合数字鸿沟提供契机。数字素养与开放合作对于弥合数字鸿沟至关重要⁵,一方面,人工智能技术降低了教育与信息传播的成本,使不发达地区与弱势群体也有机会接触最前沿的数字知识与数字技能,提升数字素养。另一方面,人工智能通过**开源社区与生态构建**的方式,有利于打破发达国家与部分公司的技术垄断,帮助发展中国家共享数字科技发展的成果,促进普惠包容。

人工智能的出现引起从数字鸿沟向智能鸿沟的转变。智能鸿沟是数字鸿沟的新形态,在技术层面,算法工具的使用门槛与数据资源的获取能力形

1 联合国,《我们的共同议程 数字简报》(2023年5月),来源: <https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-gobal-digi-compact-zh.pdf>

2 来源: <https://www.itu.int/initiatives/sdgdigital/acceleration-agenda/>

3 国际劳工组织/联合国,《注意 AI 鸿沟:塑造对未来工作的全球视角》(2024年),来源: https://digitalibrary.un.org/record/4061662?ln=zh_CN&v=pdf

4 陈凤仙,刁乃琦,王娜.人工智能对就业的多重影响及应对建议[J].科技中国,2024,(12):84-89.

5 联合国未来峰会,《全球数字契约》(2024年9月),来源: <https://www.un.org/zh/documents/treaty/A-RES-79-1-Annex-I>

成新壁垒；在认知层面，公众对智能技术运作逻辑、社会影响的理解程度差异显著；在智能参与层面，算法偏见、技术垄断等环境因素加剧弱势群体的边缘化。总之，在人工智能的影响下，数字鸿沟不再仅是硬件和技能的差距，更是环境与素养的差距。以普惠包容的人工智能增进对智能社会的认知、理解和参与，对于弥合新型数字鸿沟至关重要。

（二）释放人工智能潜力促进数字素养提升

人工智能为提升数字素养提供了良好条件。例如，大语言模型与机器翻译为各国民众减少从事数字与互联网工作的语言障碍；智能语音助手、智能家居设备等在带来生活便利的同时塑造了智能环境，提升公众认知；基于机器学习的内容推荐机制输送符合需求的个性化信息，助力知识教育与技能培训，培育数字人才，增进就业质量，构建全民终身学习生态；智能语音交互、图像识别等技术帮助残障群体跨越数字鸿沟；产业数字化协同发展，企业级人工智能培训系统加速员工数字技能迭代，形成人才供给与产业升级的良性循环。

同时，人工智能技术应用与数字素养提升之间具有相互驱动和促进的关系。数字素养不仅是数字技能和数字知识，也包括对数字与人工智能的态度、理解与参与。人工智能的广泛应用提供了良好的数字环境，提升公共服务与公民素养水平的同时，也反向促进智能社会的普遍参与。

在政府治理领域，治理流程智能化与政府数字化水平的提升形成正向循环¹。例如，深圳福田于2025年率先启用人工智能公务员²，提升了政府工作人员利用人工智能完成公文修正、执法文书生成、

民生诉求分拨及企业筛选等工作的能力和效率，并为240个业务场景量身定制智能助手³，这同样给工作人员提出了更高的数字素养要求。

在公众素养领域，人工智能的教学与应用促进公众提升数字技能以适应智能化场景，公众数字素养的增强又反哺技术迭代。例如美国麻省理工学院开发的Day of AI平台⁴为学校 and 教师提供了免费的人工智能教学和办公工具用于人工智能课程教学，提升了学生与教师的数字素养，也为人工智能行业培养了更多具备专业知识和实践能力的人才。新加坡通过构建覆盖全龄段的终身教育体系，以AI4E⁵等一系列科普课程整合生活案例和多语言资源，降低公众认知门槛，推动人工智能产品普及；新加坡还通过“人工智能学徒计划”⁶培养专业人才，既保障技术落地所需的人力资源，又通过工程师群体的实践成果反哺公众教育，形成技术应用与素养提升的良性循环。

（三）构建更加普惠和有活力的人工智能研发与应用生态

构建普惠而有活力的人工智能生态对于弥合全球不同国家与地区的发展鸿沟有深远影响。当前全球南北方在人工智能发展上存在显著差距，技术研发应用主要由北方国家主导，南方国家由于资源和技术限制，难以充分利用人工智能助力可持续发展目标。为此，应共同构建包容性的技术社区和生态体系，降低研发和应用门槛，增进开放协作，促进南方国家积极参与人工智能的研发、应用和治理。同时，广泛参与的人工智能生态也能确保南方国家在数据采集、技术研发和算法训练等方面拥有更多话语权，减少因训练数据代表性不足带来的偏见和歧视，弥合数字鸿沟，推动全球范围内的可持续发展。

1 人民网，“民意通”系统实现社情民意动态感知与风险预警，来源：<https://baijiahao.baidu.com/s?id=1824767183311673488&wfr=spider&for=pc>

2 人民网，评AI公务员，来源：<http://politics.people.com.cn/n1/2025/0224/c458474-40424076.html>

3 光明网，深圳福田“AI数智员工”，来源：https://m.gmw.cn/2025-02/21/content_1303976783.htm

4 来源：https://dayofai.org/?utm_source=aihub.cn

5 来源：<https://ai4sme.aisingapore.org/ai-for-everyone/>

6 来源：<https://aiap.sg/>

开放资源(Open Resource)是人工智能生态发展的核心驱动力与实践路径。在生成式人工智能时代,开源的定义逐渐演进为开放资源¹,不仅可能表现为源代码的开放,也可能包括数据、算法、模型、标准、知识和内容等资源的开放²。以DeepSeek为代表的开放资源实践,通过核心技术开放共享,有效推动自主创新与技术交流,使全球开发者能够共同参与前沿技术演进³。开放模型社区如ModelScope和Hugging Face通过提供多样化的基础模型、工具链及协作平台,显著降低了技术应用门槛并培育了跨国开发者社群。开放资源协作带来的生态扩展效应已渗透至政务、教育、医疗等关键领域,形成了具有区域适应性的多样化解决方案。这种技术普惠模式通过优化资源配置创造了经济与社会价值的良性循环,有助于持续推动全球产业链的数字化转型与可持续发展目标实现。

在充分释放开放资源潜力的同时,也应注意相应风险。代码和数据的开放可能导致安全漏洞与隐私侵犯;开放资源可能带来开源许可证冲突、版权归属不清等知识产权问题;同时,由于开放资源带来技术的广泛可访问性和开放性,滥用、误用甚至恶意利用现象可能发生,恶意行为者可能利用公开代码进行非法活动或网络攻击;此外,运营管理风险还包括社区治理不力、资源不足和依赖链过长等问题,这些风险共同构成了开放资源生态在促进技术创新的同时必须面对的综合挑战。国际社会应通过合作加强开放资源生态的建设,推动技术共享和知识传播,同时制定相应的规则 and 标准,防范开放资源风险,确保人工智能的健康发展。

普惠包容有活力的开放资源生态对于弥合数字鸿沟至关重要。包容性保证充分参与和代表性,普惠的原则促进共享人工智能的发展成果,有活力提供了长期和更广泛的影响。为推进可持续发展进程,**生态构建应与数字素养提升相协同。**提升数字素养能为人工智能研发与应用生态创造更广阔的发展空间,良好的生态系统又为数字素养提升提供丰富的资源,只有形成良性循环,才能保持生态的活力。生态社区间应形成**去中心的互联互通网络**,促进不同生态社区之间的交流与合作,实现各参与主体之间的协同创新与资源高效整合⁴,才能构建出既灵活又安全、开放且自组织的生态系统。**为发展中国家融入全球生态提供便利与支持。**发展中国家在人工智能领域面临着技术、资金、人才等多方面的挑战,国际社会应加强合作,通过技术转移、资金援助、人才培养等方式,帮助发展中国家提升人工智能研发与应用能力,使其能够更好地融入全球人工智能生态,共享发展成果。

1 来源: https://cn.wicinternet.org/2025-02/20/content_37862286.htm

2 联合国教科文组织,《人工智能伦理问题建议书》(2021年11月),来源: https://unesdoc.unesco.org/ark:/48223/pf0000381137_chi

3 腾讯网,深度开源助推AI大模型普惠“破圈”,来源: <https://news.qq.com/rain/a/20250304A01GIL00>

4 阿拉伯ICT组织,建设阿拉伯人工智能项目孵化器,整合多方资源提供可持续支持,来源: <https://www.aicto.org/ar/>

04

推动人工智能能力建设， 加速赋能可持续发展进程

各国的人工智能能力是赋能可持续发展的前提与基础。发展中国家作为联合国2030年可持续发展议程的主要对象，面临着消除贫困、粮食安全、气候变化等最艰巨的可持续发展挑战。随着联大决议和《全球数字契约》的通过，人工智能能力建设国际合作为发展中国家提供了宝贵的历史机遇。深化国际合作，加大对发展中国家人工智能能力建设的支持力度，才能全面推动可持续发展议程的加速实现。

(一) 发展中国家在推动人工智能赋能可持续发展目标上面临的困境

由于历史积累薄弱、发展基础脆弱，发展中国家在包括人工智能在内的新兴技术发展应用能力上普遍存在差距。这进一步加剧了这些国家在推动人工智能赋能可持续目标上的困难：

数字基础设施不完善。数据显示，目前部分地区仍存在大规模人口缺乏稳定电力供应的现象。其

中，非洲、亚洲的发展中国家与发展中地区的电力可及性显著低于全球水平¹。同时，全球互联网普及率也表现出明显的地域差异，低收入国家的网络覆盖率依然很低²。此外，许多发展中国家虽然接入电力和网络，但仍处于供应不稳定的状况中。这些因素直接阻碍了人工智能技术的有效部署。

数据与算力设施短缺。发展中国家数据的匮乏导致人工智能模型训练难以精准捕捉本土化需求，尤其在医疗、农业、环境监测等领域，数据采集成本高与共享机制缺失加剧了数据资源分布不均。同时算力基础设施薄弱导致高性能计算资源获取受限，许多发展中国家依赖外部云计算服务，成本高昂且存在数据安全风险³。

技术与应用能力薄弱。发展中国家在核心算法开发、模型优化及系统集成能力上存在不足⁴，技术成果难以转化为实体应用，因此在构建适配本土需求的人工智能解决方案上存在困难。尤其在应对气候变化、灾害预警等复杂场景时，这些国家和地区普遍缺乏自主可控的技术支撑与落地方案。

1 全球能源互联网发展合作组织，《全球电力发展指数研究报告》（2024年7月），来源：<https://www.geidco.org.cn/publications/zxbg/2025/7899.shtml>

2 国际电信联盟，《衡量数字发展：事实与数据2024》，来源：<https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx/>

3 Lehdonvirta, V., Wú, B., & Hawkins, Z. (2024). Compute North vs. Compute South: The Uneven Possibilities of Compute-based AI Governance Around the Globe. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 7(1), 828-838. <https://doi.org/10.1609/aies.v7i1.31683>

4 联大决议，《抓住安全、可靠和值得信赖的人工智能系统带来的机遇，促进可持续发展》（2024年3月），来源：<https://docs.un.org/zh/A/78/L.49>

风险监管与应对能力欠缺。一方面,发展中国家在伦理规范、数据安全机制及算法透明度等方面缺乏前瞻性立法,同时由于技术手段的欠缺,难以有效识别与评估人工智能技术可能带来的风险;另一方面,其执法与监管机构可能更容易面临专业能力不足、跨部门协作难等问题,难以形成有效治理。

人才储备与数字素养不足。发展中国家往往面临人才储备与数字素养的多重制约。受限于资金短缺和教育水平等问题,发展中国家的人才培养规模和质量有限;薪资与科研条件弱势加剧本土人才流失与国际人才引进困难;公众的人工智能基础素养及技术应用能力不足。这些因素进一步阻碍了人工智能在促进可持续发展上的应用。

国际多边治理机制参与有限。当前在全球人工智能治理机制中,发展中国家的代表性与话语权普遍不足,难以实质性参与国际组织规则制定。这导致一些国际人工智能治理多边机制和议题设置更倾向于关注少数国家的关切,而未能充分纳入广大发展中国家在基础设施、能力建设、数据主权保护、技术普惠等领域的核心诉求。

(二) 人工智能能力建设国际合作为加速赋能可持续发展提供历史性机遇

人工智能能力建设是加速赋能可持续发展的关键路径,联合国通过中国提出的“加强人工智能能力建设国际合作”决议¹,此后中国提出“人工智能能力建设普惠计划”²,倡议并成立“人工智能能力建设国际合作之友小组”³。这些工作系统性推进技术共享、算力互联、人才培养及数据安全协作,通过多边合作重构技术治理格局,将普惠包容理念转化为推动加速2030年可持续发展议程实现的重大机遇。

数字基础设施能力建设为赋能可持续发展目标提供了跨区域协同发展的新机遇。通过区域合作构建可持续人工智能基础设施,能够推动包容性增长并缩小数字鸿沟。例如,共享算力中心与智能电网的布局可优化资源分配,减少区域间技术发展不平衡⁴。在发展中国家,此类基础设施可加速农业智能化转型,直接促进零饥饿(目标2)和经济适用的清洁能源(目标7)等目标的实现。

数据资源与算力优化能力建设为赋能可持续发展目标提供了普惠化技术落地的可能性⁵。开源模型与算法创新显著降低算力依赖和推理成本,使中小企业能以低成本获取高性能人工智能服务。这种能力优化不仅助力产业升级(目标9),还能推动教育普惠(目标4),教育机构通过低成本API接入大模型,为偏远地区学生提供个性化学习支持。

知识与技能培训能力建设为赋能可持续发展目标提供了复合型人才培养的基石。跨学科人才培养成为关键,如首都师范大学筹建人工智能学院、校企合作实训基地等模式,可培育兼具技术能力与伦理意识的复合型人才。此类能力建设直接支撑体面工作与经济增长(目标8),同时通过提升公众人工智能素养,增强社会对技术风险的认知(目标16)。

应用赋能与解决方案能力建设为赋能可持续发展目标提供了多场景落地的实践路径。人工智能已在医疗、交通、能源等领域形成规模化应用:非洲 Aerobotics利用无人机和图像识别筛查早期农作物疾病⁶;伦敦证券交易所通过人工智能问答系统提升50%客户服务效率(目标9),各类大模型在教育场景中提供透明化解题思路(目标4),而智能电网优化则减少25%的能源浪费(目标7)。这些案例验

1 来源: https://www.gov.cn/yaowen/liebiao/202407/content_6960524.htm

2 来源: https://www.mfa.gov.cn/web/wjzb_673089/xghd_673097/202412/t20241218_11496414.shtml

3 来源: https://un.china-mission.gov.cn/chn/czthd/202412/t20241219_11506138.htm

4 阿拉伯ICT组织,推动Arab AI Charter: AI4All促进域内国家形成统一立场和政策框架,来源: <https://www.aicto.org/ar/>

5 中国电信,星辰MaaS平台为行业客户提供全流程大模型解决方案,来源: https://netnic.com.cn/news/info_544.html

6 来源: <https://www.aerobotics.com>

证了人工智能技术对多目标的协同促进作用。

研发创新与生态能力建设为赋能可持续发展目标提供了技术迭代的可持续动力。构建开放协作的创新生态至关重要，例如开放资源策略带动全球开发者参与模型优化，形成“技术共享-应用反馈-迭代升级”的良性循环。世界经济论坛发布的一份报告指出¹，企业通过与云服务商、科研机构合作建立的数字核心(Digital Core)，可加速气候行动(目标13)相关技术的商业化进程。

安全与治理能力建设为赋能可持续发展目标提供了风险防控的制度保障。面对人工智能引发的网络攻击、数据偏见等风险，需建立覆盖伦理规范、安全标准的治理体系。例如《人工智能安全承诺》等行业自律自治框架²可确保技术应用符合区域法规(目标16)，而网络安全能力³的提升(如防范深度伪造技术的滥用)则直接保护社会基础设施安全(目标9)。

(三)探索可负担可持续的人工智能能力建设国际合作原则

人工智能能力建设国际合作能够帮助发展中国家提升技术实力，缩小数字鸿沟，释放人工智能在推动经济增长、改善公共服务和应对全球挑战方面的潜力。同时，为保障能力建设的长期效益和系统性成果，确保合作真正惠及发展中国家，必须防止合作机制异化为新的发展负担。面对资金、技术和人才资源有限的现实，亟需探索更加可负担、可持续的创新型合作范式。通过推动多方协作，鼓励公共和私营部门共同参与，促进南南合作和南北合作相结合，形成开放共享的合作生态，将有助于实现人工智能技术的普惠应用，助力发展中国家在全球数字化转型中实现包容性和可持续的发展目标。

为了真正惠及发展中国家，加强人工智能能力建设国际合作中应遵循以下一些原则：一是坚持**主权尊重**，确保合作尊重各国的国家主权和数据主权；二是注重**本土适配**，以本地化和需求导向为核心，充分考虑各国的实际国情和发展需求，推动务实合作，确保技术和方案能够在当地落地生根；三是秉持**循序渐进**的方式，根据发展中国家的基础设施状况，优先推动数字化建设，夯实数字基础，再逐步向智能化转型，避免操之过急；四是强调**平衡推进**，既要重视技术创新和发展能力，也要加强安全治理能力，确保技术应用在安全可控的前提下释放潜力；五是注重**创新培育**，鼓励发展中国家培养本土人才，壮大本地科技生态，形成自主创新能力。通过建立与各国发展阶段相匹配的技术转化体系，确保合作成果真正落地转化为推动经济社会发展的内生驱动力，助力形成长期可持续的发展能力。

1 世界经济论坛, AI in Action: Beyond Experimentation to Transform Industry (2025年1月), 来源:

https://reports.weforum.org/docs/WEF_AI_in_Action_Beyond_Experimentation_to_Transform_Industry_2025.pdf

2 中国信息通信研究院, 发布《人工智能安全承诺》推动行业安全自律(2024年12月), 来源:<http://finance.people.com.cn/n1/2024/1229/c1004-40391586.html>

3 中国互联网络信息中心, 针对AI网络钓鱼欺诈的深度对抗关键技术及应用, 收录于2024年世界互联网大会领先科技奖收录成果集《科技之魅》

面向可持续发展的人工智能治理国际合作：行动建议

数字与智能鸿沟日渐凸显，可持续发展进程紧迫严峻，人工智能能力建设迫在眉睫，建议全球各方从以下方面深化国际合作，推进人工智能进一步加速赋能全球可持续发展。

（一）全面推进人工智能赋能可持续发展目标的实现

基于对全球可持续发展进程的监测与分析，当前形势的紧迫性已不容忽视，可持续发展目标的实现迫在眉睫。尽管人工智能为可持续发展带来新的驱动力，其潜力发挥尚不充分且对不同领域赋能尚不均衡。国际社会应加强对相关议题的关注，深化多方合作，既促进人工智能赋能潜力的稳健释放，为人工智能赋能可持续发展提供引导与保障，同时关注和应对人工智能赋能中的潜在负面影响。

（二）推进人工智能赋能可持续发展的普惠性

人工智能赋能可持续发展应遵循普惠原则，促使不同国家、群体都有机会尽可能平等地从中获益，弥合数字与智能鸿沟。应推进全球科研机

构、企业与各国政府联合探索前沿人工智能基础研究，并推进在教育、医疗、农业、气候、文化等与可持续发展密切相关的领域加强人工智能国际合作，让人工智能发展惠及全人类。应通过促进人工智能数据、算法、平台、应用、治理等方面的国际标准化，降低跨国合作的技术壁垒。此外，应鼓励开放资源，促进不同开源生态社区间的去中心化互联互通，助力开放共享研发与应用成果。

（三）推进人工智能赋能可持续发展的包容性

国际社会应鼓励探索不同的人工智能科学与技术实现途径，及其促进可持续发展的应用，关注人工智能的近期与长远挑战对可持续发展的影响。应一致推进南南合作、南北合作以及不同合作机制间的有机协同，提升发展中国家在人工智能治理规则制定、技术研发与应用实践等环节的参与度。倡导多利益相关方协同共治的包容性框架，在数据主权、算法透明度、资源分配等关键议题上平衡南北方诉求。应关注全球数据资源分布不均衡的问题，提升南方国家的数据代表性；避免单一模式主导全球治理，保障发展中国家在全球治理体系中拥有足够的话语权。

(四) 与中低收入国家共同探索可负担、可持续的能力建设模式

国际社会应倡导立足区域特色的差异化人工智能发展模式,使技术路线与本土需求精准适配;着力破除对传统高耗能算力路径的过度依赖,通过分布式计算架构优化、轻量化模型研发、可再生能源驱动算力中心等技术创新降低部署与应用资源需求与门槛。应同步加强全球人工智能教育协作网络建设,鼓励技术发达国家向发展中国家开放人才交流渠道,支持短期培训、远程教育和跨国联合培养项目;助力发展中国家的基础设施建设与数字化、智能化转型。

(五) 以国际治理多边合作机制为人工智能赋能可持续发展提供保障

各方应支持以联合国为核心的多边合作机制,在联合国框架下务实推进人工智能治理国际合作,重点围绕能力建设、风险监测评估、安全管控、伦理规范共识、技术标准互操作、多边融资支持等关键领域深化协同。同时,依托国际组织、国际机制促进多方对话,通过协调政策立场、整合优势资源、共享实践经验,凝聚全球合力为人工智能发展构建包容普惠的治理框架,共同推动释放人工智能的变革性潜力,赋能全球可持续发展目标的实现。

附录 人工智能治理与可持续发展应用与实践案例

(一) AI赋能优化资源配置与应对气候变化(目标2零饥饿、目标7经济适用的清洁能源、目标9产业、创新和基础设施、目标13气候行动)

华为云盘古气象大模型赋能预测极端天气

无贫穷(目标1)、产业、创新和基础设施(目标9)、气候行动(目标13)

美国国家食品与农业研究所数据库促进AI赋能农业与生态

零饥饿(目标2)、气候行动(目标13)、陆地生物(目标15)

中国电信自研AI节能系统助力数据中心优化与绿色计算

经济适用的清洁能源(目标7)、气候行动(目标13)

(二) AI赋能环境与生态保护(目标13气候行动、目标14水下生物、目标15陆地生物)

瓜卡玛亚项目:利用AI赋能守护亚马孙雨林

气候行动(目标13)、水下生物(目标14)、陆地生物(目标15)

澳大利亚海洋科学研究所开发ReefCloud平台助力珊瑚礁追踪监测

水下生物(目标14)

腾讯野朋友计划:AI技术助力生物多样性保护

气候行动(目标13)、水下生物(目标14)、陆地生物(目标15)

(三) AI赋能教育普惠与素养提升(目标4优质教育、目标10减少不平等)

联想建设AI未来科技馆,助力乡村学生接轨最前沿科技

优质教育(目标4)、减少不平等(目标10)

远期人工智能搭建智联文脉平台,编织连接历史创造未来的全球文化网络

优质教育(目标4)、减少不平等(目标10)、可持续城市和社区(目标11)

芬兰人工智能中心:构建全民AI素养的多层次教育生态

优质教育(目标4)、减少不平等(目标10)

(四) AI赋能弱势群体促进社会公平(目标3良好健康与福祉、目标5性别平等、目标10减少不平等)

腾讯红雨伞计划:AI赋能守护女性健康

良好健康与福祉(目标3)、性别平等(目标5)、减少不平等(目标10)

香港博智感知与交互研究中心:AI助力沟通障碍人群的日常交流

良好健康与福祉(目标3)、减少不平等(目标10)、可持续城市和社区(目标11)

阿拉伯信息通信技术组织推动利用AI赋能女性领导力

性别平等(目标5)、减少不平等(目标10)

中国移动智慧村医平台在吉林开启乡村医疗新模式

良好健康与福祉(目标3)、减少不平等(目标10)

(五) AI赋能产业转型与收入提升(目标1无贫穷、目标8体面工作和经济增长、目标9产业、创新和基础设施、目标10减少不平等)

中国移动利用AI技术破解畜牧业融资难题

无贫穷(目标1)、体面工作和经济增长(目标8)、产业、创新和基础设施(目标9)、减少不平等(目标10)

南开大学师生利用AI技术为苗族山村设计文创产品

无贫穷(目标1)、优质教育(目标4)、体面工作和经济增长(目标8)、减少不平等(目标10)

星辰MaaS平台:推动行业客户智能化转型和大模型技术产业化落地

产业、创新和基础设施(目标9)、可持续城市和社区(目标11)

(六) AI赋能可持续生态与合作平台(目标9产业、创新和基础设施、目标17促进目标实现的伙伴关系)

阿里云魔搭社区:搭建开源普惠的AI生态

产业、创新和基础设施(目标9)、促进目标实现的伙伴关系(目标17)

阿拉伯AI4SD平台助力弥合信息鸿沟

产业、创新和基础设施(目标9)、促进目标实现的伙伴关系(目标17)

印象笔记智能协作系统搭建知识共享平台

产业、创新和基础设施(目标9)、促进目标实现的伙伴关系(目标17)

开放智算产业联盟推进AI新生态

产业、创新和基础设施(目标9)、促进目标实现的伙伴关系(目标17)

中国网络空间安全协会发布“中文互联网语料资源平台”

产业、创新和基础设施(目标9)、促进目标实现的伙伴关系(目标17)

(七) AI赋能良好治理与安全保障(目标11可持续城市和社区、目标16和平、正义与强大机构)

人民网“民意通”系统实现社情民意动态感知与风险预警

产业、创新和基础设施(目标9)、可持续城市和社区(目标11)、和平、正义与强大机构(目标16)

360安全大模型智能化保障行业应用安全

产业、创新和基础设施(目标9)、可持续城市和社区(目标11)、和平、正义与强大机构(目标16)

中国信息通信研究院建立大模型安全中文基准测试体系

产业、创新和基础设施(目标9)、和平、正义与强大机构(目标16)

安恒“恒脑”安全智能体护航亚冬会网络安全

产业、创新和基础设施(目标9)、和平、正义与强大机构(目标16)

中国互联网络信息中心:研发AI快速过滤与深度对抗检测技术,高效应对复杂场景钓鱼欺诈

产业、创新和基础设施(目标9)、和平、正义与强大机构(目标16)

中国电子技术标准化研究院联合复旦大学发布“生成式人工智能服务安全基准测试集”

产业、创新和基础设施(目标9)、和平、正义与强大机构(目标16)

(一) AI赋能优化资源配置与应对气候变化(目标2、目标7、目标9、目标13)

华为云盘古气象大模型赋能预测极端天气¹

马达加斯加地处热带气旋高发区域,当地渔民因传统气象信息获取渠道有限,面临较大的气候风险。公益组织Mitao Forecast于2023年引入华为云盘古气象大模型,借助其高精度预测能力,将气象预报时间从3天延长至10天,并通过3000多块天气预报板向60万至75万渔民传递多维度气象数据。2024年初,该模型精准定位台风“阿尔瓦罗”登陆地点,为渔民提供预警,帮助其采取防护措施。

美国国家食品与农业研究所数据库促进AI赋能农业与生态²

由美国农业部下属的美国国家食品与农业研究所(NIFA)资助的科学家团队发布了动植物数据库Arboretum。该数据库包含近32.7万种动植物的1.346亿张经过专家验证并配有文字说明的图像,促进AI工具的开发,以改善害虫控制策略,减少农作物损失,提高农业产量,优化农业资源利用降低食品价格。同时,该系统还将为生物多样性保护工作提供更高效率的工具,对全球粮食安全、生态系统保护和气候变化缓解产生积极影响。

中国电信自研AI节能系统助力数据中心优化与绿色计算³

中国电信自研AI节能系统,助力实现大幅节电与成本节约。该节能系统被广泛部署在基站、IT云主机、AB类机房,通过实时监控设备运行和环境信息,智能化关断和启动局部空闲设备,对机房和数据中心制冷系统风速频率进行智能化调节,实现

年均节电8亿度,节省电费约5.2亿元。2024年,中国电信积极践行绿色发展理念,累计获评中国国家绿色数据中心达28个,持续推动基站、机房AI节能,合计年化节电超10亿度;单位电信业务总量碳排放量同比下降19.2%。

(二) AI赋能环境与生态保护(目标13、目标14、目标15)

瓜卡玛亚项目:利用AI赋能守护亚马孙雨林⁴

亚马孙雨林作为全球最大的热带雨林,对碳吸收、气候调节和生物多样性保护具有重要意义。然而,森林砍伐仍是其面临的重大威胁。由安第斯大学、亚马孙科学研究所SINCHI、洪堡研究所、星球实验室和微软AI for Good实验室共同领导的瓜卡玛亚(西班牙语Guacamaya,意为金刚鹦鹉)项目,利用AI和云计算技术监测森林砍伐并保护生物多样性,通过分析卫星数据、野生动物图像和声学记录,实现更快速、更有效的保护行动。

澳大利亚海洋科学研究所开发ReefCloud平台助力珊瑚礁追踪监测⁵

ReefCloud平台由澳大利亚海洋科学研究所(AIMS)联合帕劳国际珊瑚礁中心、斐济野生动物保护协会等机构开发,是一款利用机器学习和高级分析技术,快速从全球珊瑚礁图像中提取和共享数据的数字工具。该平台通过实时协作功能,使全球珊瑚礁监测社区能够共同工作,改变传统监测管理方式。系统以80-90%的准确率标准化分析珊瑚礁组成,效率较人工评估提升700倍,大幅节省管理资源。用户提交调查期间拍摄的珊瑚礁照片后,平台运用类面部识别技术,从珊瑚颜色、形

1 <https://www.huawei.com/cn/gallery/videos/2024/mitao-forecast-madagascar>

2 <https://www.nifa.usda.gov/about-nifa/impacts/nifa-funded-scientists-spur-research-artificial-intelligence-release-massive>

3 <https://www.c114.com.cn/news/117/a1275451.html>; <https://www.c114finance.comsina.cn/news2025-03-26/117/a1275451detail-ineqxcef9511378.d.html>

4 <https://parispeaceforum.org/projects/project-guacamaya/>

5 <https://www.aims.gov.au/research/technology/reefcloud>

状和大小中识别约700万个参数,准确判断珊瑚类型并追踪礁体变化,为全球珊瑚礁长期保护提供决策支持。目前平台已在澳大利亚、帕劳、斐济、所罗门群岛、瓦努阿图和马尔代夫投入使用。

腾讯野朋友计划:AI技术助力生物多样性保护¹

腾讯以“野朋友计划”为核心构建公众参与生态保护的数字化体系。开发“腾讯雪豹AI识别系统”“观鸟君”小程序以及“物种之眼”生物多样性AI模型,提升雪豹等旗舰物种栖息地监测效率及鸟类迁徙数据可视化,降低公民科学参与门槛,为科研提供高质量观测数据库。同时,联合多家机构打造“数字协同-云数据管理-AI识别”三位一体工具链,为物种识别与保护提供标准化解决方案,物种AI模型能力支持289种濒危物种的识别,识别率均在73%以上,雪豹精确度在95%以上。

(三) AI赋能教育普惠与素养提升(目标4、目标10)

联想建设AI未来科技馆,助力乡村学生接轨最前沿科技²

联想在江西省九江市修水县(原中国国家级贫困县)何市镇中心小学建设AI未来科技馆,通过展示AI语音作画、AR元宇宙、AI体育等应用场景,为乡村学校提供了新的教学方法和工具,提升教学质量。科技馆向公众免费开放,并通过邀请嘉宾、招募志愿者以及与国际项目联动等方式,帮助当地儿童不断接触新鲜知识,链接前沿科技。同时,科技馆还可以辐射到修水县周边的学生和研学团体,目前已接待超6000人。

远期人工智能搭建智联文脉平台,编织连接历史创造未来的全球文化网络³

智联文脉是一个AI驱动的全球文化分析、链接和互动系统平台,旨在深入探索和连接世界各地的物质文化遗产、非物质文化遗产和其他文化内容。智联文脉的数据引擎涵盖了来自172个国家文化数据,包括所有联合国教科文组织世界遗产名录中的遗产地,并扩展到1,725项物质文化遗产,以及10,565项非物质文化遗产。

芬兰人工智能中心:构建全民AI素养的多层次教育生态⁴

芬兰人工智能中心(FCAI)通过构建多层次的教育与普及生态,构建了从入门慕课“人工智能基础(Elements of AI)”,到AI伦理与社会影响课程,再到Python技术培训的阶梯式知识体系,已覆盖全球55万学习者并获评全球最佳慕课。同时,FCAI开展机器人编程俱乐部和校园合作培养青少年AI兴趣。其以人为本的伦理导向和多层次全年龄的课程体系奠定芬兰在负责任AI普及领域的突出地位。

(四) AI赋能弱势群体促进社会公平(目标3、目标5、目标10)

腾讯红雨伞计划:AI赋能守护女性健康⁵

红雨伞计划通过AI技术精准赋能宫颈癌早期筛查和诊断。通过AI辅助诊断系统实现自动标记宫颈可疑病变区域并生成可视化诊断建议。该计划帮助基层医生快速锁定风险点,大幅提升偏远地区早期筛查效率。在基层医疗能力建设层面,依托AI驱动的智能培训系统,通过游戏化模拟操作场景

1 <https://techforgood.qq.com/projects/detail/313>

2 <https://wap.peopleapp.com/article/7208432/7051788>

3 <https://culturalink.ai/cn>

4 <https://fcai.fi/mission>

5 <https://ssv.tencent.com/domain/health-welfare>

和自适应学习算法，对医生操作技能进行个性化评估并生成针对性培训课程，助力欠发达地区诊疗水平标准化升级。

香港博智感知与交互研究中心:AI助力沟通障碍人群的日常交流¹

香港中文大学博智感知与交互研究中心(CPII)致力于运用基于AI的语音和语言技术，为沟通障碍人群提供有力支持。其具体目标包括：构建香港本地化的核心语言模型及其个性化子模型，用于非语言AAC用户的符号预测和推荐；开发受神经语言学启发的深度学习算法，以实现AAC符号的智能推荐；以及开发多模态人机交互方法，以适应用户的身体或认知限制。

阿拉伯信息通信技术组织推动利用AI赋能女性领导力²

阿拉伯信息通信技术组织(AICTO)正在推动利用AI为阿拉伯女性提供必要的技能、深入的知识以及战略领导能力。通过建立一个包容性的生态系统，当地女性不仅可以为AI发展做出贡献，而且有望领导和促进可持续增长、社会公平和区域创新、进一步弥合性别差距，并将阿拉伯女性塑造为能够利用AI技术解决紧迫挑战、塑造行业趋势并为社区及更广泛地区的经济社会发展做出贡献的有影响力的领导者。

中国移动智慧村医平台在吉林开启乡村医疗新模式³

中国移动携手合作伙伴打造“智慧村医平台”，平台通过5G手持终端设备和定制化app实现前端数据采集，提供AI辅助诊断、智能随访、远程

医疗、家庭医生服务、村医工作量化考核等功能。其中，AI辅助诊断系统可基于医生录入的患者病历数据，实时进行智能分析和病情研判，针对多种常见疾病生成相关治疗方案，辅助村医提升病情诊断准确率，优化临床决策效率，有效提高乡村医疗机构的诊疗服务水平。该平台自启动以来，已实现吉林全省全部行政村覆盖，为2907名村医提供服务，惠及村民30余万人。

(五)AI赋能产业转型与收入提升(目标1、目标8、目标9、目标10)

中国移动利用AI技术破解畜牧业融资难题⁴

中国移动创新运用“普惠金融+科技赋能”打造智慧活畜管理云平台，通过物联网技术实现“活牛抵押”创新模式，助力慧牛合作社成功获贷200万元。在传统模式下，养牛业在扩大养殖规模时会因缺乏有效抵押物而难以获得贷款；移动通过部署蓝牙耳机标、AI视频监控及无线传感网络，打造具备活畜定位追踪、生命体征监测、智能盘点和数据采集功能的监管系统，实时掌握抵押牛只的位置与健康状态，推动“活体牲畜抵押贷款”落地。

南开大学师生利用AI技术为苗族山村设计文创产品⁵

南开大学陶锋团队为贵州省雷山县南猛村设计创作AI艺术设计品500余件、宣传视频10部以上并举办和推行了AI艺术汇报展、智能艺术研学计划、艺术家驻地等计划。陶锋团队的师生们探索AI美学与传统苗族文化的结合点，用智能艺术帮助当地的茶叶、纺织等产业的发展，以及通过智能艺

1 <https://cpil.hk/zh/rp2-2/>

2 <https://www.aicto.org/ar/>

3 https://www.10086.cn/download/csrreport/cmcc_2023_csr_report_full_cn.pdf

4 https://agri.hainan.gov.cn/hnsny/ywdt/zwtd/202306/t20230630_3444906.html

5 <https://news.nankai.edu.cn/zhxw/system/2024/05/29/030061914.shtml>

术推动当地文旅发展，并在项目实施过程中，向村民们宣传和普及AI技术应用，消减少数民族地区的教育鸿沟。

星辰MaaS平台：推动行业客户智能化转型和大模型技术产业化落地¹

中国电信与企商在线、浪潮、科大讯飞、智谱华章、百川智能等合作伙伴联合发布星辰MaaS平台，为行业客户提供了从评估、选择、部署到应用的全流程大模型解决方案。该平台整合了自研和生态等多家大模型伙伴资源，支持多云算力调度、大模型选配和应用场景创新，帮助用户轻松实现大模型的选型与部署。星辰MaaS平台通过端到端服务模式，赋能多行业智能化转型，推动大模型技术的产业化落地。

(六) AI赋能可持续生态与合作平台(目标9、目标17)

阿里云魔搭社区：搭建开源普惠的AI生态²

阿里云联手中国计算机学会开源发展委员会于2022年11月共同推出AI模型社区“魔搭”(ModelScope)。魔搭社区是基于MaaS(Model as a Service)理念，面向模型探索、环境安装、推理验证、训练调优全周期，开源模型相关技术及服务。提供超15000个优质AI模型，全面覆盖各个模态各个主要领域，汇聚了600万开发者，累计模型下载量超过一亿次，成为LlaMa, Stable Diffusion, Qwen等国内外先进大模型的开源首发平台。

阿拉伯AI4SD平台助力弥合信息鸿沟³

联合国开发计划署和阿拉伯联合酋长国推出了AI4SD平台，该平台利用AI技术弥合信息鸿沟，赋能农业和环境韧性、气候适应、水资源管理、促进社会凝聚力和减少不平等。AI4SD平台致力于提供可操作的数据洞察、早期预警、预测能力和情景分析，有利于全面解决该地区的数据稀缺问题，促进该地区的可持续发展。通过解决该地区的数据稀缺问题，最终惠及整个阿拉伯地区。

印象笔记智能协作系统搭建知识共享平台⁴

印象笔记构建的智能协作系统融合大语言模型、OCR手写识别与生成式AI思维导图等功能，形成跨场景知识处理中枢，通过轻量化工具包Notelt+ScanIt快速部署于数字基础设施薄弱地区，智能阅读工具iMark打破语言与文化障碍，云端同步的EverPEN硬件体系为跨国协作提供物理-数字融合入口。该平台为企业提升文档处理效率40%，降低知识管理成本35%，同时通过模块化服务组合赋能发展中国家教育、医疗等关键领域。这种集智能工具开发、本地化场景适配与可持续发展理念于一体的平台架构，不仅为跨国企业构建标准化知识共享界面，更能通过模块化服务组合赋能发展中国家教育、医疗等关键领域，为全球知识要素的互联互通提供可扩展的技术基础设施。

开放智算产业联盟推进AI新生态⁵

中兴通讯作为发起单位之一，成立开放智算产业联盟(COIA)，致力于通过开放、共享、协同的理念赋能可持续生态与合作平台。联盟汇聚了字

1 https://netnic.com.cn/news/info_544.html

2 <https://www.modelscope.cn/home>

3 <https://ai4sd.info/>

4 <https://baijiahao.baidu.com/s?id=1821751136157797728&wfr=spider&for=pc>

5 <https://www.zte.com.cn/china/about/news/20241220c1.html>

节跳动、中国移动、中国联通、新华三、神州数码、英特尔等众多领军企业与研究机构，聚焦AI智算平台与基础资源服务，推动大模型训练、推理机制及合成数据技术的发展以及在行业中的应用。通过开源协作与技术共享，COIA将加速AI技术的落地应用，助力数字化转型与智能化重构，促进新质生产力的发展。未来，联盟将持续深化产业合作，探索场景化解决方案与硬件设备协同，依托安全委员会保障技术研究的安全性，成为全球科技创新与产业智能化发展的重要推动力量。

中国网络空间安全协会发布“中文互联网语料资源平台”¹

中国网络空间安全协会会同中国国家互联网应急中心，协同人工智能产、学、研、用单位，共同建设了中文互联网语料资源平台，面向社会提供中文互联网基础语料展示下载服务。平台支持行业领域、内容模态、体量规模等多种标签分类，目前已入驻27个语料数据集，数据总量约2.7T。平台后续还将依托建立的语料共建共享机制，持续吸纳优质中文互联网语料进驻，探索开展数据来源合规评估、质量评价、安全检测等服务，构建健康可持续的中文互联网语料开发利用生态，促进和支撑大模型产业发展。

(七) AI赋能良好治理与安全保障(目标11、目标16)

人民网“民意通”系统实现民意感知与风险预警²

人民网“民意通”系统通过大模型技术，聚焦网上群众工作和基层治理工作领域，为各级党政部门提供多来源数据汇总分析、群众留言一键办

理、社情民意实时监测等服务。采用智能化手段处理群众留言，自动筛选重复诉求并精准转办，减少基层干部事务性负担。系统打通部门间数据壁垒，实现跨区域协作处理复杂问题，同时24小时监测舆情热点并建立紧急事项直报机制，疑难问题可一键直达主官督办。该技术方案既保障民意响应效率，又通过全流程数字化留痕强化责任落实，已在河北、内蒙古等地率先应用，显著提升群众诉求处理速度，为基层治理提供了“高效透明、智能预判”的数字化升级路径。

360安全大模型智能化保障行业应用安全³

360基于安全大模型聚焦分析研判、威胁溯源、邮件检测等领域进行专项训练，推出100+安全数字专家智能体，在2024年期间服务保障了政府、金融、能源、交通等行业超200家客户的行业应用安全。在服务某能源企业期间，360安全大模型赋能安全运营日均处理日志50亿条，自动封禁6.56万个恶意IP，监测到近100封恶意邮件，日均节省23.17人天运营成本，防御体系防护率达100%，实现行业应用安全保障智能化全流程覆盖。

中国信息通信研究院建立大模型安全中文基准测试体系⁴

2024年4月，中国信息通信研究院依托中国人工智能产业发展联盟(AIIA)安全治理委员会，联合30余家单位发起大模型安全基准测试AI Safety Benchmark，秉持公平公正、产业应用和场景导向的原则，建立大模型安全中文基准测试体系。以提高底线红线、数据安全和科技伦理等安全能力为目标，覆盖价值观正确、合法合规、隐私保护、公

1 https://www.cac.gov.cn/2025-01/10/c_1738209436031064.htm

2 <http://finance.people.com.cn/n1/2024/1026/c1004-40347773.html> & <https://baijiahao.baidu.com/s?id=1824767183311673488&wfr=spider&for=pc>

3 https://mp.weixin.qq.com/s/HD92eTuPRpR_c3l0bZzejw

4 <http://www.caict.ac.cn/>

平公正等二十个维度的中文评测数据集。经过持续迭代更新, AI Safety Benchmark目前已经发布四轮测试结果, 测试数据整体包含50万余条原始提示词样本和80余种攻击模板, 组合生成 4000万条攻击样本, 已对全球前沿大语言模型、图文大模型开展100余次测试。

安恒“恒脑”安全智能体护航亚冬会网络安全¹

安恒信息依托自研的“恒脑”安全智能体平台, 为2025年哈尔滨亚冬会构建全流程网络安全防护体系。安恒深入应用10大(7大通用安全智能体+3大专用安全智能体)安全智能体, 实现了从防护、研判、溯源、通报到管理的全流程升级, 期间总计拦截攻击近180万次, 研判与封禁1.3万余恶意IP, 为赛事的顺利进行筑牢了一道坚不可摧的网络安全屏障, 确保赛事在网络空间中不受任何干扰, 平稳、有序地开展。

中国互联网络信息中心: 研发AI快速过滤与深度对抗检测技术, 高效应对复杂场景钓鱼欺诈²

中国互联网络信息中心针对AI网络钓鱼欺诈的深度对抗关键技术及应用成果, 通过“快速过滤+深度对抗”的检测架构机制, 实现了对大规模复杂场景下AI钓鱼欺诈的快速检测和高效分析。该成果应用于中国国家域名滥用治理服务, 近五年累计发现并处置钓鱼网站超12万例, 将钓鱼网站的平均存活时间从12天缩短至4天内, 为广大网民避免了高达数百亿元的潜在经济损失。此外, 该成果还推动了中国域名审核和滥用治理服务的高质量发展, 成为中国域名服务行业审核标杆工程。

中国电子技术标准化研究院联合复旦大学发布“生成式人工智能服务安全基准测试集”³

中国电子技术标准化研究院联合复旦大学, 支撑TC260-003《生成式人工智能服务安全基本要求》大模型安全测评需求, 建设大模型内容安全基准测试集。测试集支持TC260-003开展全风险安全测评, 覆盖共5大类、31子类安全风险, 包含超过1000项风险主题, 测试集规模超过100万。测试集适配大模型企业安全能力建设需求, 建设基础级、进阶级、专家级等多个对抗强度, 并支持针对模型定制化设计测试问题集, 形成大模型多等级安全能力阶梯。此外, 通过测试问题动态定期更新, 为大模型企业提供常态化安全能力评估服务, 持续推动大模型安全能力不断提升。

1 https://www.dbappsecurity.com.cn/content/details4756_29490.html

2 中国互联网络信息中心: 针对AI网络钓鱼欺诈的深度对抗关键技术及应用, 收录于2024年世界互联网大会领先科技奖收录成果集《科技之魅》

3 <https://www.cesi.cn/>

世界互联网大会人工智能专业委员会

世界互联网大会于2024年设立首个专业性、常态化分支机构——人工智能专业委员会，下设标准推进计划、安全与治理推进计划和产业推进计划。专委会汇集了来自人工智能领域国际组织、知名智库、科研院所、专业协会及产业界的权威专家和专业人才，秉持搭建国际交流合作平台、推动发展与治理协同、促进全球共享人工智能发展成果的原则，通过开展专题研讨、成果分享、倡议发布等活动，不断凝聚国际共识，促进人工智能包容普惠、可持续发展。

以普惠包容的人工智能治理赋能全球可持续发展

如何引用本报告：

曾毅, Seán Ó hÉigeartaigh, 鲁恩萌, 王金, 王正奇, 郭晓阳, 范津宇, 皇甫存青, 康彦荣, 韩开宇, Ammar Younas, 张荣, 王星光, Nada Laabidi, 乔迁, 程明, 呼娜英, 郭苏敏, 张妍婷, 陈晶晶, 林琳, 吴淑燕, 刘玥, 赵英, 邓洁, 陈凤仙, 刘威辰, 王峰, 王健兵, 夏文辉, 王欣, 李华伟, Edson Prestes, 李娜, 冯雁, 段伟文, Jimena Sofia Viveros Alvarez, 西京京, 胡永启, 姚新, 彭韬, 陶锋, 王彤, 叶晨晖, 杨耀东, 谷文杰, 邓雯洁, 翟家欢, 李大任, 王梦寅, 曹建峰, 蒙美玲, 张寄冀, 唐新华, 刘唱, 夏和华, Serge Stinckwich, Bernd Holznagel, Dame Wendy Hall, 谭知行, 孟伟, 肖荐午, 杨忠良, 周原, 周凯, 张凌寒, John Yeoh, 龚新奇, 乔宇, Nirosha Ananda, Charuka Senal Damunupola, Christoph Lütge, 张雪丽, 梁昊. 以普惠包容的人工智能治理赋能全球可持续发展[R]. 北京:世界互联网大会, 2025.



关注大会公众号



世界互联网大会
World Internet
Conference