

国研院专题研究报告



CAICT
中国信通院

中国—东盟人工智能发展与 治理合作：进展观察和推进建议

中国国际问题研究院
中国信息通信研究院
二零二四年十二月

人工智能技术正引领新一轮科技革命和产业变革向纵深发展，有望重塑人类经济、社会生活的各个方面。与此同时，人工智能技术的迅速发展也带来新的安全和伦理挑战，成为全球治理的重要议题。中国与东盟是全面战略伙伴，在数字经济和智能发展领域拥有坚实的合作基础。面对人工智能的技术浪潮，中国与东盟应进一步加强合作，共同分享技术进步带来的发展红利，并尽力规避和减轻技术引发的风险挑战。

中国国际问题研究院和中国信息通信研究院组织专门力量，从全球人工智能发展与治理态势、中国的人工智能发展与治理、东盟的人工智能发展与治理、中国与东盟的人工智能合作等方面展开系统研究，并在此基础上对中国—东盟人工智能合作进行展望，形成此篇研究报告。





CAICT
中国信通院

中国—东盟人工智能发展与治理合作： 进展观察和推进建议

中国国际问题研究院
中国信息通信研究院

二零二四年十二月

目录

01

前言

2

02

全球人工智能发展与治理态势

6

（一）人工智能技术产业发展迅速

6

（二）人工智能技术发展不平衡问题日益突出

10

（三）全球人工智能治理框架正在酝酿

16

03

中国的人工智能发展与治理

24

（一）中国人工智能技术和产业应用蓬勃发展

24

（二）中国加快构筑人工智能治理体系

28

（三）中国积极参与全球人工智能治理体系建设

31

04	东盟的人工智能发展与治理	38
	（一）东盟国家人工智能发展潜力巨大	38
	（二）东盟国家人工智能治理迅速起步	46
05	中国与东盟的人工智能合作	56
	（一）中国与东盟人工智能发展合作基础深厚	56
	（二）中国与东盟人工智能治理合作前景可期	60
06	展望	68
	（一）推进基础设施发展，构筑智能转型基础	68
	（二）加快技术应用，共建创新生态	69
	（三）加强治理对话与合作，筑牢安全和伦理屏障	70
07	附件	76
	附件 1 全球人工智能治理倡议	76
	附件 2 人工智能全球治理上海宣言	79
	附件 3 人工智能能力建设普惠计划	83

01 前言

01 前言

人工智能技术正引领新一轮科技革命和产业变革向纵深发展。人工智能不仅可以赋能各行各业，也能够创造新的产业形态、开辟新的经济增长点，将成为影响国家综合实力和国际竞争力的重要因素。人工智能还为人类解决能源、公共卫生、城市规划、环境保护等重要问题提供新的途径。与此同时，人工智能技术的飞速发展也带来隐私、安全、伦理等多方面挑战，并可能导致国际社会“智能鸿沟”的加深，相关全球治理进程正在推进。中国与东盟是全面战略伙伴，双方地缘相邻、人文相亲、理念相近，在数字经济和智能发展领域已经拥有坚实的合作基础。双方应进一步加强在人工智能发展和治理方面的合作，深化全面战略伙伴关系，携手推进构建更为紧密的中国—东盟命运共同体建设。

02

全球人工智能发展与 治理态势

- 人工智能技术产业发展迅速
- 人工智能技术发展不平衡问题日益突出
- 全球人工智能治理框架正在酝酿

02

全球人工智能发展与治理态势

全球人工智能技术产业迎来爆发式增长，智能时代加速到来，深刻改变人类的生产生活方式。如何深化国际合作、促进人工智能发展红利普惠共享，共同应对技术应用带来的潜在风险，已成为国际社会的共同关切。

（一）人工智能技术产业发展迅速

人工智能自 1956 年达特茅斯会议诞生以来，已历经三次发展浪潮。随着 2022 年 11 月 ChatGPT 上线，生成式大模型又推动人工智能发展进入新阶段。据《全球人工智能产业发展白皮书（2024 年）》显示，2023 年全球人工智能产业规模达 7078 亿美元，同比增长 19.3%。^① 随着产业规模的不断扩张，人工智能的经济社会影响日益扩大。

^① 深圳市人工智能行业协会、深圳市易行网数字科技有限公司：《全球人工智能产业发展白皮书（2024 年度）》，2024 年 4 月，https://www.sz.gov.cn/cn/xxgk/zfxxgj/zwdt/content/post_11531060.html。



1. 人工智能技术发展近十年取得重大突破

人工智能技术发展迄今经历四个阶段（图 1）：由符号主义主导的萌芽探索期和专家系统期，以及由连接主义主导的机器学习期和大模型期。人工智能技术在上世纪 90 年代开始的机器学习期实现突破，开始大规模投入实践。**在图像识别领域**，卷积神经网络（CNN）等人工智能算法得到充分应用，通过卷积层、池化层和全连接层的组合，能够有效地提取图像中的局部特征和全局信息。2015 年，微软亚洲研究院何恺明等四名研究员研发的 ResNet 首次在图像识别领域超越人类。**在语音识别领域**，端到端的深度学习可被用于识别汉语、英语等多个语种。2015 年 12 月，百度研究院宣布其 Deep Speech 系统实现超越人类水平的汉语识别。**在自然语言处理领域**，人工智能被广泛用于文本分类、情感分析、机器翻译等任务。特别是注意力机制（Attention Mechanism）的引入，使得模型能够更好地聚焦于输入序列中的重要部分，进一步提升了处理长文本的能力。谷歌 2017 年研发的 Transformer 模型不仅通过自注意力机制（Self-Attention）在机器翻译任务中取得了显著的成果，还成为推动人工智能进入大模型期的关键基础。

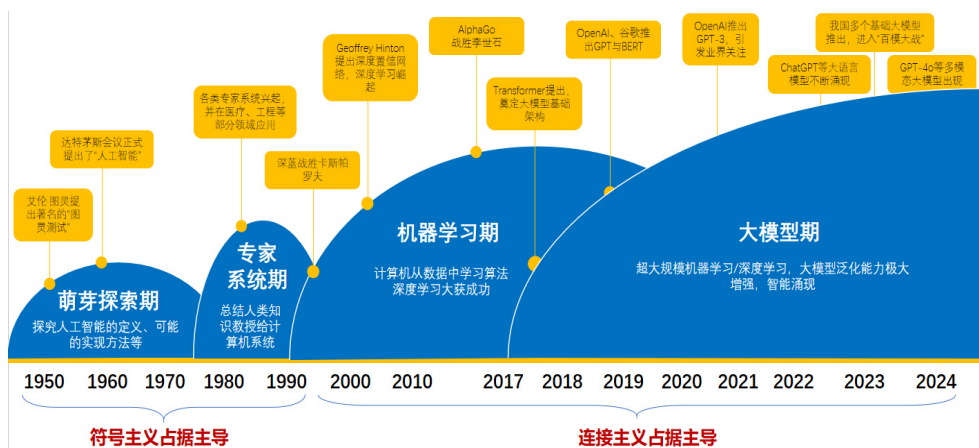


图 1 人工智能发展历程
来源：中国信息通信研究院

2. 大模型技术使人工智能初步具备一定通用性，可应用于诸多领域

随着人工智能技术发展进入大模型期，生成式大模型的崛起引发全球高度关注。这些模型通常具有少则数十亿、多则上万亿个参数，通过大规模的预训练在大量未标注数据上学习通用的特征表示，再通过微调在特定任务上进行优化。相比于传统人工智能，大模型在多个任务和场景中展现出强大的通用性和适应性（图 2）。

首先是规模可扩展性强。在大模型中，参数规模、计算量和训练数据增加可带来模型性能持续提升，而在传统人工智能模型中，规模增大不能带来性能持续提升。**其次是多任务适应性强。**在大模型中，一个模型能同时支持多种任务和多个模态，甚至实现跨模态，而在传统人工智能模型中，一个模型只能胜任单任务和单模态。**最后是能力可塑性强。**大模型通过模型微调、思维链提示等措施即可实现能力进一步增强，传统人工智能模型训练完成后能力无法进一步拓展。

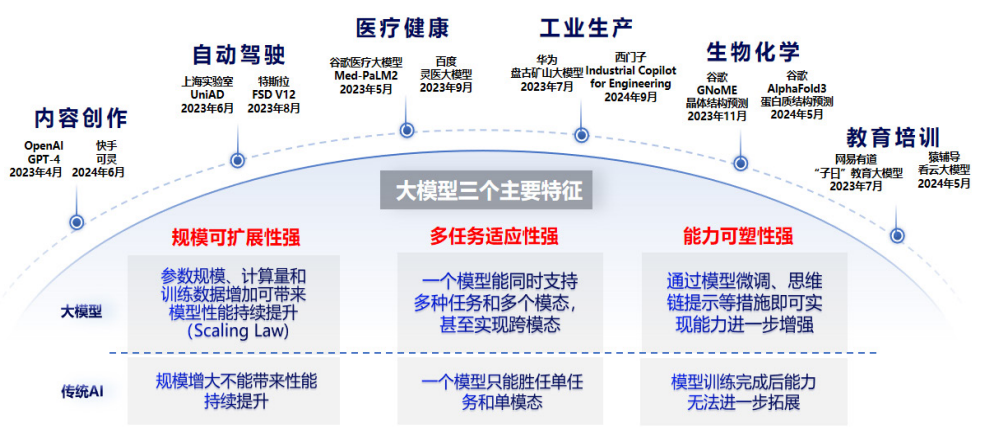


图 2 大模型特征与应用
来源：中国信息通信研究院



3. 人工智能已在众多产业领域获得深度应用

在内容创作领域，许多媒体公司和内容平台使用大模型自动生成新闻稿、文章和创意内容，提高内容生产的效率和质量。目前，在财经新闻、商业插画等领域，人工智能所生成的作品效果已得到广泛认可，可以经过微调甚至直接投入商业应用。**在自动驾驶领域**，智能驾驶系统能够基于摄像头、雷达和超声波传感器等收集数据、提取环境信息，如车辆、行人、交通标志等，并据此实时分析路况并调整驾驶策略，显著提高了驾驶的安全性和舒适性。在自动驾驶技术推动下，新能源汽车行业实现迅猛发展。**在医疗健康领域**，通过学习大量的医疗影像数据，人工智能能够辅助医生进行疾病诊断，提高诊断的准确性和效率。目前人工智能已经在利用肺部 CT 影像检测肺癌、识别皮肤癌、筛查眼科疾病等方面取得了显著成果。此外，人工智能还可以用于患者的个性化治疗方案推荐，通过分析患者的基因组数据、病史和生活习惯等信息，为患者提供更加精准的治疗建议。**在工业生产领域**，人工智能已具备沉淀专家经验、提高生产效率的能力，在替代人从事危险工作、重复劳动方面发挥重要作用，同时在 CAD 研发设计、智能装备等方面已经体现出明显的潜力。如在智慧矿山中，人工智能在矿山智能生产、设备管理、安全作业以及经营决策等领域发挥重要作用。**在生物化学领域**，大模型技术被用于蛋白质结构预测等科研工作。2024 年，谷歌 DeepMind 公司的丹米斯哈萨比斯 (Demis Hassabis) 和约翰乔普 (John M. Jumper) 因共同领导了对 AlphaFold2 的开发，展现了人工智能在攻克长期悬而未决的生物学难题上的巨大潜力而荣获诺贝尔奖。**在教育培训领域**，数字人可以作为虚拟教师，帮助学生进行学习，指导员工完成培训，特别是对于贫困落后地区的儿童而言，人工智能技术可以增强优质教育资源的可得性，帮助有效缩小城乡差距。综上所述，人工智能已经为全球各行业创造了巨大的商业价值，催生出一批新型产业，深刻改变了人类的生产生活方式，对经济、社会产生深远影响。

4. 人工智能将对经济社会产生更为广泛和深入的影响

人工智能技术既有提升效率、促进增长等正面作用，也潜藏着替代就业、扩大数字鸿沟等风险，需加以高度关注。**从经济增长角度看**，生成式大模型作为新的通用目的技术，将推动劳动生产率的提升。高盛于 2023 年 3 月发布报告认为，从历史经验看，人工智能如果能够得到普遍应用，将在 10 年内推动全球 GDP 年均增长 7%，或近 7 万亿美元。^①**从就业影响角度看**，全球约有 18% 的工作可能被 AI 自动化所取代，影响 3 亿工作岗位；但对不同行业的冲击有所差异，相较于发达经济体，新兴经济体能够被自动化替代的工作岗位较少。^②**从社会治理角度看**，一方面大模型可以通过分析海量社会数据，帮助政府部门进行城市规划、交通管理、公共安全保障等工作。另一方面，生成式人工智能也可能因可得性不足、输出结果存在偏见等问题使弱势群体处于更加不利的地位，损害社会公平。

（二）人工智能技术发展不平衡问题日益突出

与人工智能技术快速发展和应用形成鲜明对比的是，大多数国家特别是发展中国家尚未能真正接触、使用人工智能并从中受益。联合国人工智能高级别咨询机构对来自 38 个国家的 120 余名专家的调查发现，多数专家认为人工智能将对较高收入国家更早产生积极的影响，对人工智能助力减少不平等持谨慎态度（图 3）；发展中国家获取基础

^① “The Potentially Large Effects of Artificial Intelligence on Economic Growth”, March 26 2023, <https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html>.

^② 同上。



设施、数据、模型、人才等关键要素的能力远低于发达国家。^①

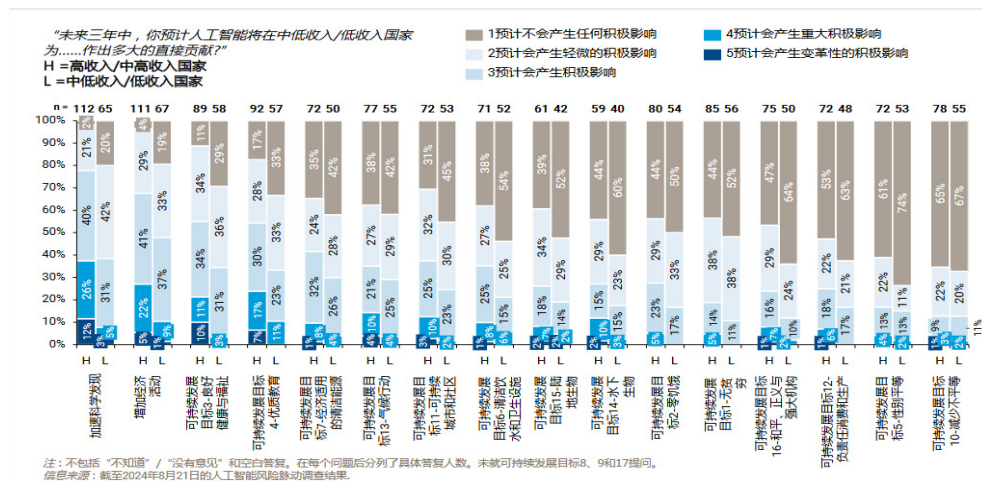


图 3 未来三年人工智能助力国家发展与可持续发展目标
来源：联合国“人工智能高级别咨询机构”最终报告

1. 数字基础设施鸿沟明显

随着大模型、生成式人工智能技术革新，人工智能由之前的点状应用，逐步规模化发展到赋能广泛行业，对基础设施也提出了更高要求。然而发展中国家在网络接入、算力等方面仍有不小差距。一是**网络接入方面**，根据国际电联（以下简称“ITU”）《衡量数字发展：2023年事实与数字》报告，全球互联网连接取得稳步进展，但总体发展仍不均衡，越来越多的低收入国家被抛在后面。^②2022年，全球仍有26亿人无法接入互联网，非洲仅有37%的人口使用互联网，还不足美欧

^① “Final Report: Governing AI for Humanity, September 2024”, <https://www.un.org/ai-advisory-body>.

^② 国际电联：《衡量数字发展：2023年事实与数字》，2023年11月27日，<https://news.un.org/zh/story/2023/11/1124322>。

水平（约 90%）的一半。低收入国家每个固定宽带用户的月平均流量为 161GB，仅占全球平均数据使用量 257GB 的六成。5G 移动网络分布不均衡，高收入国家 89% 人口已被 5G 网络覆盖，而低收入国家几乎没有 5G 服务。网络接入水平低限制了发展中国家对人工智能服务的使用。**二是算力方面**，根据经合组织（以下简称“OECD”）定义，人工智能算力是指支持专门的人工智能工作负载和应用的硬件和软件堆栈，包括数据中心、超级计算机和云计算等。^① 根据市场调研机构 Synergy Research Group 的数据，截至 2023 年底，全球超大规模数据中心主要集中于北美、亚太和欧洲。^② 大模型规模和精度提升，智能化应用场景拓展，都对算力支撑提出更高要求。据华为《智能世界 2030》报告预测，2030 年人工智能算力需求将达到 864 ZFLOPS，算力需求十年百倍的增长将成为常态。^③ 缺乏可负担的算力将限制发展中国家融入全球人工智能产业生态。

2. 数据集资源分布不均

数据是人工智能模型训练的基础，也是决定其性能和专业性的关键因素。目前，大多数主流人工智能大模型都基于英语语料库训练，如 ChatGPT-3 训练语料中，英语占比约 92.6%，英语、法语、德语、

^① “A blueprint for building national compute capacity for AI”, OECD, February 28 2023, <https://www.oecd-ilibrary.org/docserver/876367e3-en.pdf?expires=1728870337&id=id&accname=guest&checksum=891E32CEBE7040A85F291B4406294F95>.

^② “Hyperscale Data Centers Hit the Thousand Mark; Total Capacity is Doubling Every Four Years”, Synergy Research Group, April 17 2024, <https://www.srgresearch.com/articles/hyperscale-data-centers-hit-the-thousand-mark-total-capacity-is-doubling-every-four-years>.

^③ “Intelligent World 2030”, Huawei, <https://www.huawei.com/en/intelligent-world>.



西语等 16 种语言的头部效应明显，合计占比超过 99.24%^①，其他小语种的贡献非常有限。数字经济领先的国家较早开展数字化转型，在生产生活和公共服务各领域积累了大量的有效数据。大多数发展中国家受数字化进程限制，海量信息尚未转化为可用数据，经过标注、清洗等一系列数据处理流程后形成的高质量数据集更为缺乏。此外，发展中国家公共数据开放和整合也面临挑战。世界银行 2022 年数据统计表现指数（Statistical Performance Indicators, SPI）显示，非洲、拉美、亚太区域发展中国家在人口普查数据、行政数据、地理空间数据的开放共享方面表现较为落后。^②

3. 大模型集中度较高

从全球已发布的大模型的分布来看，截至 2023 年 5 月，中国和美国的大模型总数已超过全球大模型总数的 80%（图 4）。^③ 法国、德国、俄罗斯、以色列等国家拥有部分大模型，其他发展中国家基本缺位。大模型全球分布非常集中的部分原因在于高昂的成本。根据斯坦福研究团队测算，2023 年 OpenAI GPT-4 和谷歌 Gemini Ultra 训练成本分别约为 7800 万美元和 1.91 亿美元。^④

值得一提的是，人工智能开源模型为技术扩散提供了机遇，但也面临着较大的安全挑战。开源模式通过开放源代码促进不同开发

^① 姚旭晨：《人工智能的负效应：没有大语言模型的语种未来会消亡》，2024 年 6 月 11 日，https://mp.weixin.qq.com/s?__biz=MjM5NDEyNjE0MQ==&mid=2652685829&idx=1&sn=51114897eec387531d0a95bd752658e3。

^② World Bank Group data, <https://datanalytics.worldbank.org/SPI/>。

^③ 中国科学技术信息研究所：《中国人工智能大模型地图研究报告》，2023 年 5 月，<https://www.istic.ac.cn/html/1/284/338/1506840038330332778.html>。

^④ “The AI Index Report: Measuring Trends in Artificial Intelligence”，Stanford University Human-Centered Artificial Intelligence, <https://aiindex.stanford.edu/report/>。

者的获取、修改和分发，显著降低了个人和企业接触和使用人工智能技术的成本，提高了发展中国家部署本地化大模型的可能性。但与此同时，开源系统也面临着价值观偏差、滥用和易受攻击等一系列安全问题。

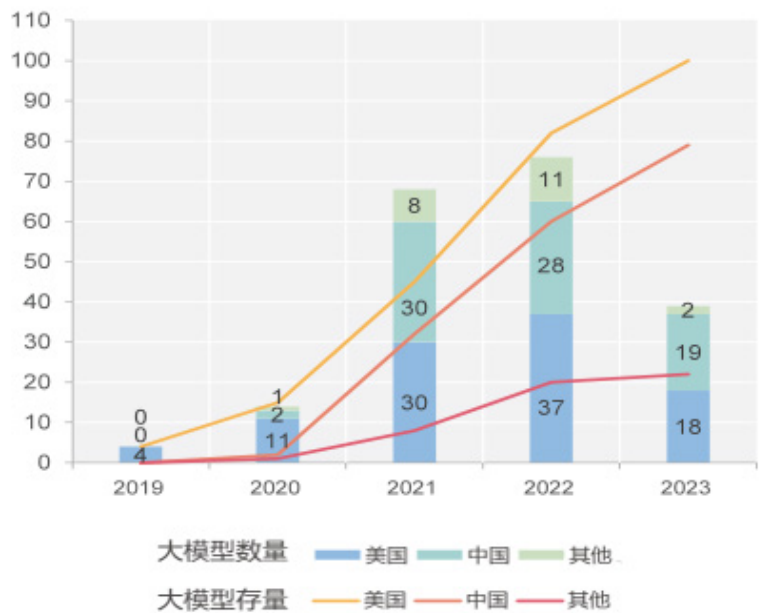


图 4 全球大模型数量

来源：中国科学技术信息研究所，《中国人工智能大模型地图研究报告》

4. 人才差距显著

顶级人才方面，根据清华大学《人工智能发展报告 2011-2020》，入选人工智能 2000 榜单的高层次学者共计 1833 位（去重处理），覆盖全球 37 个国家，主要集中在北美洲、欧洲和东亚地区，南美洲、非洲的高层次学者数量稀少。其中，美、中、德高层次人工



智能学者的数量为全球前三，其余国家数量均低于 100 人（图 5）。^①专业人才方面，发展中国家在物联网、数据科学、自然语言处理、机器学习等人工智能相关专业的教育基础薄弱、科研水平不足。联合国教科文组织对二十国集团成员的统计显示，发展中国家科学、技术、工程和数学（STEM）领域毕业生、信息通信技术毕业生数量明显低于发达国家。南非、土耳其 STEM 毕业生占比仅为 18.71%、15.8%，约是排名第一的德国的半数（35%）（图 6）。^②此外，全球对人工智能人才需求旺盛导致发展中国家的人才外流问题严重。原因之一是发展中国家的人工智能商业模式仍处于早期发展阶段，难以提供充足的就业岗位。



图 5 入选 AI 2000 榜单的高层次学者国家分布

来源：清华-中国工程院知识智能联合研究中心、清华大学人工智能研究院，《人工智能发展报告 2011-2020》

^① 清华-中国工程院知识智能联合研究中心、清华大学人工智能研究院：《人工智能发展报告 2011-2020》，<https://static.aminer.cn/misc/pdf/zpAIreport2020.pdf>。

^② “Toolkit for Artificial Intelligence Readiness and Capacity Assessment”, UNESCO, September 2024, <https://www.gov.br/mcom/pt-br/acao-a-informacao/governanca/governanca-de-tic-1/documentos-g20/p4-g20-dewg-brasil-2024-toolkit-for-ai-readiness-and-capacity-assessment.pdf>。

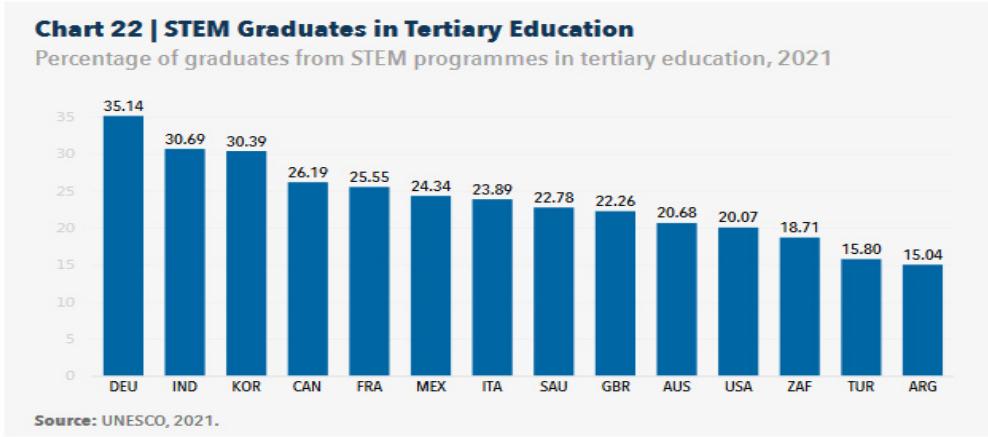


图 6 部分二十国集团成员 2021 年 STEM 毕业生占比
来源：UNESCO, Toolkit for Artificial Intelligence Readiness and Capacity Assessment

（三）全球人工智能治理框架正在酝酿

1. 各方对人工智能风险的认识正在逐步加深

人工智能技术作为战略性、通用性技术，可能引发系统性、全球性风险，受到各方高度关注。**从技术本身看**，人工智能大模型的“幻觉”“涌现”效应导致内容不可信、能力不可控，治理对象的不确定性显著增加。**从技术使用者来看**，人工智能系统能力日趋强大，其使用门槛却日益降低，使用者利用人工智能技术从事各类违反伦理、危害公共安全活动的风险急剧增加，而对人工智能系统使用者的管理却更加困难。**从人机互动前景来看**，劳动替代、创新替代、甚至知识和观念的替代在未来均可能发生。人工智能技术对个人、企业、社会和国家安全均可能造成复杂的影响，随着技术应用领域的扩大，潜在的风险清单仍在不断增加。



图 7 人工智能在不同维度引发风险

来源：中国信息通信研究院

2. 主要经济体基于本国国情采取不同模式推进人工智能治理

欧盟出台全球首部人工智能专门性法律，延续“规制领先、产业跟随”的思路。2024年8月，欧盟《人工智能法》正式生效，采取基于风险的人工智能监管理念，按照应用场景对人工智能系统实行分级监管，对高风险领域设定严格的前置审查程序和合规义务。

美国秉持“以治理促发展”的理念，核心目标是维护和扩大本国人工智能产业的国际领先优势。拜登政府2023年10月发布《安全、可靠、值得信赖的人工智能》行政令。2024年10月，拜登总统发布有关人工智能的国家安全备忘录，将人工智能技术纳入国家安全考虑，确保美国在全球先进人工智能系统开发中的领导地位。

英国采用“促进创新、分散监管”思路，有意引导人工智能安全治理国际议程。英国发布《促进创新的人工智能监管方法》白皮书，要求各行业主管部门根据其领域人工智能实际使用情况评估风险，酌情考虑行业指南、技术标准、工具箱等较轻的干预手段。

中国兼顾发展和安全，审慎监管，营造有利于发展的环境，倡导全球人工智能治理合作，重视弥合智能鸿沟。2021年12月，发布《新一代人工智能伦理规范》，给人工智能全生命周期引入伦理指引。2023年7月，出台全球第一部针对生成式人工智能的监管法规《生成

式人工智能服务管理暂行办法》，以提高模型透明度为主要治理思路，为促进生成式人工智能创新、规范发展释放积极信号。2024年9月，发布《人工智能安全治理框架》，按照风险管理的理念，针对模型内生安全风险和应用安全风险提出相应技术应对和综合防治措施。国际上，2023年10月，习近平主席在第三届“一带一路”国际合作高峰论坛上向全球发出《全球人工智能治理倡议》，倡导各国遵循“以人为本、智能向善”的原则，在人工智能治理中加强信息交流和技术合作，形成具有广泛共识的人工智能治理框架和标准规范，并增强发展中国家在人工智能全球治理中的代表性和发言权。2024年7月，在上海举办的世界人工智能大会暨人工智能全球治理高级别会议发表《人工智能全球治理上海宣言》，提出促进人工智能发展、维护人工智能安全、构建人工智能治理体系等一系列主张。为弥合数字和智能鸿沟，特别是帮助全球南方在人工智能发展进程中平等受益，中国提出“人工智能能力建设普惠计划”，呼吁各方对人工智能能力建设加大投入。

总体来看，全球主要经济体都在探求创新发展和安全保障间的平衡，从伦理和价值观对齐、提升人工智能系统透明度、训练数据治理、知识产权保护、信息内容安全、防范滥用等各角度开展治理实践。在人工智能技术更迭迅速、风险具有较大不确定性的现阶段，欧盟、中国均采取了重事前监管、兼顾事中和事后监管的全生命周期监管模式。但不同于欧盟较为全面的人工智能法案，中国采取“小切口”立法策略，针对特定应用领域提出监管要求，体现了敏捷治理的思路。美国联邦层面立法进展缓慢，大多为政策或指引，侧重行业自律方式。



表 1 主要经济体人工智能相关政策法规

国家	人工智能相关政策法规	时间
欧盟	《人工智能法》	2024 年 5 月
美国	《人工智能权利法案蓝图》	2022 年 10 月
	《人工智能风险管理框架》（AI RMF1.0）	2023 年 1 月
	《安全、可靠、值得信赖的人工智能行政令》	2023 年 10 月
	《人工智能风险管理框架：生成式人工智能情况》	2024 年 6 月
	《建立先进人工智能模型和计算集群开发的报告要求拟议规则》	2024 年 9 月
	《关于推进美国在人工智能领域领导地位、利用人工智能实现国家安全目标及实现安全、可靠和可信赖人工智能的备忘录》	2024 年 10 月
英国	《生成式人工智能框架》	2024 年 1 月
	《促进创新的人工智能监管方法》	2023 年 7 月
中国	《新一代人工智能发展规划》	2017 年 7 月
	《新一代人工智能伦理规范》	2021 年 12 月
	《互联网信息服务算法推荐管理规定》	2021 年 12 月
	《互联网信息服务深度合成管理规定》	2022 年 11 月
	《生成式人工智能服务管理暂行办法》	2023 年 7 月
	《科技伦理审查办法（试行）》	2023 年 9 月
	《全球人工智能治理倡议》	2023 年 10 月
	《人工智能安全治理框架》	2024 年 9 月

来源：中国信息通信研究院

3. 联合国成为全球人工智能治理的重要推动力量

人工智能技术带来的复杂影响不断突破地理疆界，亟需全球性解决方案。联合国作为拥有最广泛会员、建立在对联合国宪章普遍认同

之上、并拥抱全球文明多样性的机构，在协调全球人工智能发展和治理重大问题上具有天然的优势和合法性，为各方共享知识和经验、讨论规范和原则提供了关键平台。

一是联合国大会已通过有关人工智能治理的重要决议。2024 年，联大通过了两份具有里程碑意义的人工智能决议，即《抓住安全、可靠和可信赖人工智能系统带来的机遇，促进可持续发展》和《加强人工智能能力建设的国际合作》，强调弥合人工智能鸿沟、推动实现安全可靠的人工智能系统治理办法及框架，提升发展中国家能力建设国际合作。决议展现了联合国利用人工智能促进可持续发展的积极姿态，也为真正发挥人工智能潜力、实现人工智能包容、普惠和可持续发展，助力实现联合国 2030 年可持续发展议程提供了方向指引。

二是积极酝酿成立协调全球人工智能治理的新机构。联合国秘书长组建的人工智能高级别咨询机构于 2024 年 9 月发布《治理人工智能、助力造福人类》最终报告，提出开展人工智能治理政策对话及组建国际人工智能科学小组、标准交流中心、能力发展网络、全球人工智能基金等多项建议，并提议在秘书处内设立人工智能办公室，强化联合国系统内协调。其中，人工智能治理政策对话和国际人工智能科学小组被纳入《全球数字契约》，已在未来峰会获得成员国授权，有望率先落地。

三是联合国系统内专门机构已开展积极实践。**教科文组织** 193 个成员国共同通过了《人工智能伦理问题建议书》，形成了全球首个人工智能伦理规范框架；组建专家组制定“人工智能就绪度评估方法”（RAM），推动伦理问题建议书在各国落地实施。**国际电联（ITU）** 长期开展人工智能标准化研究，迄今已与各方合作发布 220 余项标准建议书，覆盖网络、医疗、交通、应急等各领域，目前正与国际标准化组织（ISO）和国际电工委员会（IEC）共同推动数字水印标准研究；国际电联还定期举办“人工智能向善”峰会，发布《人工



智能向善：创新扩大影响案例集》等报告，为人工智能应用交流合作搭建舞台。^①世界卫生组织、国际劳工组织、粮食及农业组织等多个专门机构也加快了本领域的人工智能应用和风险研究，发布指南、案例集、建议书等多项务实成果，积极应对人工智能变革趋势。^②

^① “AI for Good --- Innovate for Impact Final Report 2024”, International Telecommunication Union, <https://www.itu.int/net/epub/TSB/2024-AI-for-Good-Innovate-for-Impact-final-report/>.

^② “United Nations System White Paper on AI Governance”, May 2 2024, <https://unsce.org/united-nations-system-white-paper-ai-governance>.

03

中国的人工智能发展与治理

- 中国人工智能技术和产业应用蓬勃发展
 - 中国加快构筑人工智能治理体系
- 中国积极参与全球人工智能治理体系建设

03

中国的人工智能发展与治理

近年来，中国始终积极拥抱变革创新，人工智能技术和产业发展取得长足进步。中国秉持以人为本、智能向善的理念，初步构筑了具有前瞻性、系统性的治理体系，以开放包容的姿态深度参与全球人工智能治理，积极推动人工智能造福所有人。

（一）中国人工智能技术和产业应用蓬勃发展

中国的人工智能发展起步于 20 世纪 70 年代，当时人工智能相关项目被纳入国家科研计划，这些基础性研究作为中国人工智能的后续发展奠定坚实基础。进入 21 世纪以来，在汇集国内外关键要素资源、政产学研有效联动的基础上，中国人工智能技术创新和产业应用得到蓬勃发展。

据中国信息通信研究院数据显示，2023 年中国人工智能核心产业规模达到 5787 亿元，同比增长 14%，人工智能企业数量达 4482 家。截至 2024 年 3 月，中国人工智能领域共有独角兽企业 52 家，



其中 13 家为一年内的新晋企业。^① 据《2023 全球人工智能创新指数报告》，在主要国家人工智能顶级论文数量上，中国占比 36.7%，在主要国家人工智能专利授权数量上，中国占比 34.7%，均居于全球首位，人工智能创新动能强劲。^②

1. 加强政策引导，营造良好发展环境

从 2017 年开始，中国政府根据不断发展的形势需要，持续推出人工智能发展政策（图 8）。2017 年 7 月，国务院发布《新一代人工智能发展规划》，从国家战略层面全面部署人工智能发展方向，涵盖了基础理论研究、关键技术攻关、产业应用等多个方面，激发了各类相关主体对人工智能技术的投入热情。同年 12 月，工业和信息化部发布了《促进新一代人工智能产业发展三年行动计划（2018-2020 年）》，按照“系统布局、重点突破、协同创新、开放有序”的原则，在深入调研基础上研究提出 4 方面重点任务，共计 17 个产品和领域。2022 年 7 月，科技部等六部门联合发布了《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》，强调通过场景创新推动人工智能技术的广泛应用，鼓励企业和社会各界开放应用场景，加速了人工智能技术在各个行业的落地应用。2024 年 3 月，中国政府工作报告首次提出开展“人工智能+”行动，标志着中国加快形成以人工智能为引擎的新质生产力，加快人工智能推动传统产业转型、提升居民生活品质的进程。

^① 《中国独角兽企业发展报告（2024 年）》，2024 年 4 月，https://www.gov.cn/yaowen/liebiao/202404/content_6948188.htm。

^② 《2023 全球人工智能创新指数报告》，2024 年 7 月，https://www.thepaper.cn/newsDetail_forward_27971972。

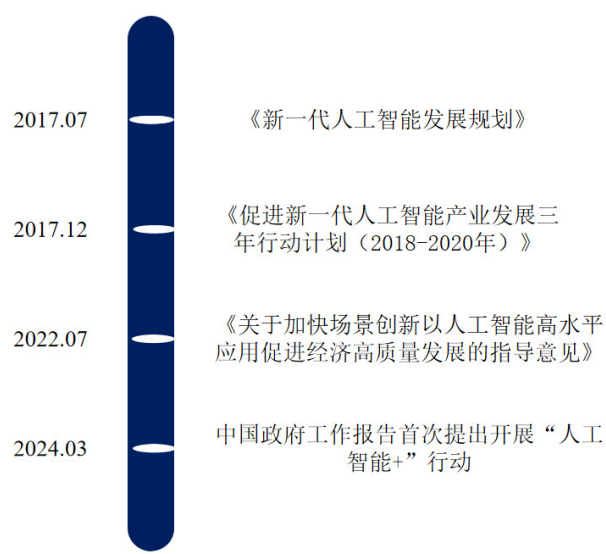


图 8 中国政府持续出台政策文件
来源：中国信息通信研究院

2. 加快算力设施部署，夯实发展基础

截至 2024 年 9 月，中国在用算力中心机架总规模超过 830 万标准机架，算力总规模达 246EFLOPS，位居世界前列。算力中心平均电能利用效率（PUE）降至 1.47，创建国家绿色数据中心 246 个，超过 140 个算力中心绿色低碳等级达到 4A 级以上。在人工智能技术应用前景的驱动下，中国兴起智算中心建设热潮。据 IDC 数据，2023 年智算服务市场同比增长 81.6 亿元人民币，其中生成式人工智能 IaaS 市场贡献 59%，智算集成服务市场贡献 38%。未来算力部署也将更加重视支持大模型赋能实体经济，推进产业融合运营。

3. 通用和垂类大模型协同发展，产业创新生态活跃

2023 年，中国人工智能厂商加速大模型商业应用，其中既包括具备一定通用性的基础大模型，也包括应用于具体领域的垂类大模型。



前者如百度发布的文心一言大模型、阿里巴巴发布的通义千问大模型、智谱华章发布的智谱清言大模型、字节跳动发布的豆包大模型等，是中国探索技术前沿的重要体现；后者如华为发布的盘古矿山大模型、盘古气象大模型、盘古海浪大模型、盘古金融 OCR 大模型等，有利于充分发挥中国的超大市场规模、丰富应用场景等优势。截至 2024 年 10 月，中国上线为公众提供服务的生成式人工智能服务大模型数量达到近 200 个。发布模型的企业不仅包括人工智能领域的头部企业，还包括在本轮大模型热潮中崛起的新兴独角兽，如智谱人工智能、月之暗面、Minimax 等，充分体现了中国人工智能行业创新生态的活跃度。

4. 促进技术应用，赋能经济社会发展

在中国，人工智能技术已在农业、制造业、服务业等得到广泛应用（图 9）。**在农业领域**，大模型以问答形式为用户提供针对农业领域的知识与信息。例如，一亩田集团发布的大模型“小田”学习了全国 2800 多个县农产品流通的大数据及多个农业细分领域的专业知识，覆盖新品种、新技术、供需行情、产销智能匹配等多个模块，针对不同用户群体提供咨询。**在制造业领域**，中国企业在研发设计、生产制造、经营管理、产品服务等各环节引入人工智能技术，提升自身智能化水平。例如，海尔的智能工厂利用机器视觉技术进行产品质量检测，减少了人为误差。**在服务业领域**，中国的服务业在移动互联网时代积累了丰富的交互数据，有助于以大模型技术提升服务品质和效率，充分发挥其为民众提供美好生活的能力，为企业和个人提供了更加便捷和安全的服务。在生活性服务业方面，以医疗健康行业为例，中国利用“腾讯觅影”等人工智能技术缓解了医疗资源紧张问题，实现远程诊断、远程会诊、辅助诊断等多种智慧医疗应用，提高了医疗服务的质量和可及性。在生产性服务业方面，以金融行业为例，蚂蚁集团基于自研基础大模型进一步定制了金融大模型，打造应用于理财顾问、保险代理、

投资研究、金融营销、保险理赔等金融服务各环节的人工智能业务助手。

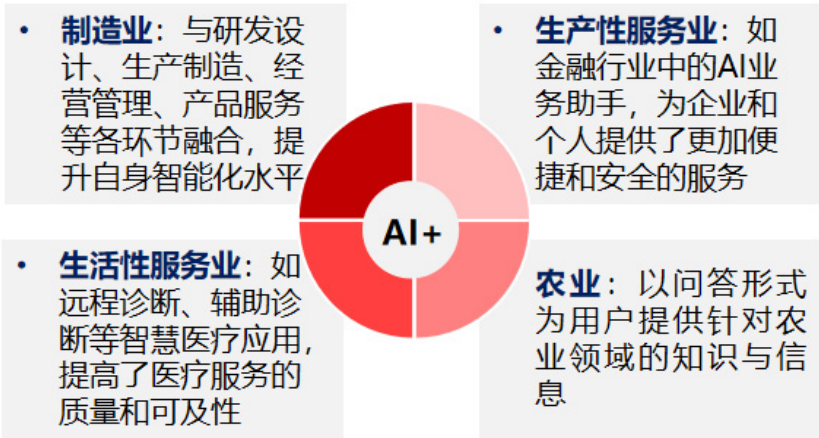


图 9 人工智能赋能各领域应用
来源：中国信息通信研究院

（二）中国加快构筑人工智能治理体系

中国的人工智能治理不断完善，已初步形成包括科技伦理、法律规制、标准指南等在内的整套治理体系（图 10）。

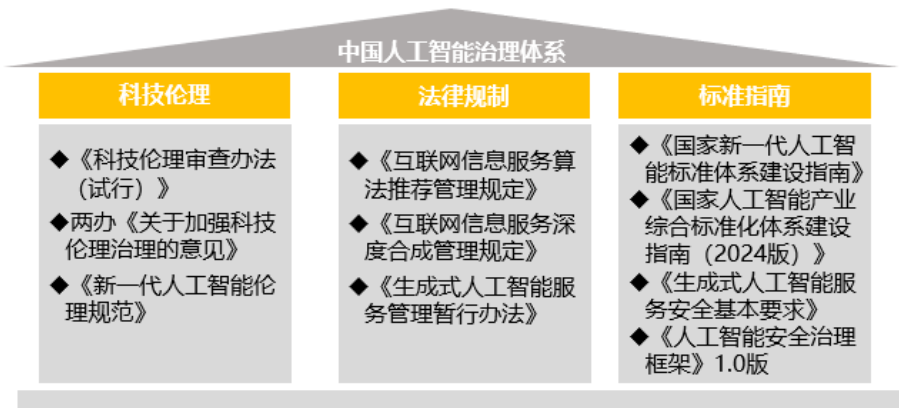


图 10 中国人工智能治理体系
来源：中国信息通信研究院



1. 注重人工智能安全治理，不断提升人工智能安全性、可靠性和可控性

中国在人工智能治理中坚持伦理先行。2021年9月，中国国家新一代人工智能治理专业委员会发布《新一代人工智能伦理规范》^①，旨在将伦理道德融入人工智能全生命周期，为从事人工智能相关活动的自然人、法人和其他相关机构等提供伦理指引。2022年3月，国务院办公厅印发了《关于加强科技伦理治理的意见》，将人工智能与生命科学、医学并列为需要制定科技伦理规范指南、设立科技伦理（审查）委员会的重点领域。2023年9月，《科技伦理审查办法（试行）》出台，为各方开展人工智能相关活动伦理风险审查评估提供明确指引。目前部分企业已将伦理审查作为产品研发立项的前置程序。

中国高度重视治理工具开发，促进人工智能产品安全可靠。例如，将数字水印等技术治理手段嵌入法律、伦理等规则设计中。工业和信息化部推动建立人工智能大模型公共服务平台，开发大模型安全测试数据集、“信元令”数字内容隐写标识系统等，助力人工智能产业安全健康发展。产业和学术界积极探索治理工具研发，如阿里巴巴与清华大学、浙江大学等合作完成开源项目 EasyRobust，可有效增强模型的鲁棒性，荣获2023年电气电子工程师学会（Institute of Electrical and Electronics Engineers, IEEE）开源科学奖。

2. 持续完善法律法规，探索敏捷治理路径

2021年12月，为应对推荐算法带来的系列问题，中国出台《互联网信息服务算法推荐管理规定》，对算法推荐服务做出全面规范，

^① 《新一代人工智能伦理规范》，2021年9月，https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html?ref=salesforce-research。

提出以算法备案方式增强推荐算法透明度。^① 算法备案有助于监管部门评估算法风险，提升算法对社会的透明度，推动广泛监督。在损害发生时，备案信息将为监管部门责任判定提供重要参考。2022 年 11 月，为防范深度合成技术的不合理应用，中国出台《互联网信息服务深度合成管理规定》，要求深度合成服务提供者建立健全管理制度和技术保障措施，如对使用者进行真实身份信息认证，建立健全辟谣机制和申诉、投诉、举报机制，对生成或编辑的信息内容添加标识等。^② 在生成式大模型推动人工智能进入新一轮发展浪潮后，中国于 2023 年 7 月出台了《生成式人工智能服务管理暂行办法》（以下简称“《暂行办法》”），成为全球首部针对生成式人工智能的立法。^③ 《暂行办法》立足于促进生成式人工智能健康发展和规范应用，明确了分级分类监管原则，延续了通过备案提高模型透明度的治理思路，同时明确支持人工智能技术在各行业和各领域的创新应用，为促进生成式人工智能创新规范发展释放了积极信号。此外，在《电子商务法》《个人信息保护法》《反垄断法》等法律的制定和修订中，也在各自的规制范围内增加了与人工智能算法相关的内容，和专门针对人工智能的部门规章一起形成了中国人工智能领域的法律法规体系。

3. 加快制定标准指南，推动规则规范落地实施

在人工智能这一技术性较强的领域，无论是法律法规等“硬法”

^① 《互联网信息服务算法推荐管理规定》，2021 年 12 月，https://www.gov.cn/gongbao/content/2022/content_5682428.htm。

^② 《互联网信息服务深度合成管理规定》，2022 年 11 月，https://www.gov.cn/zhengce/zhengceku/2022-12/12/content_5731431.htm。

^③ 《生成式人工智能服务管理暂行办法》，2023 年 7 月，https://www.gov.cn/gongbao/2023/issue_10666/202308/content_6900864.html。



还是伦理规则等“软法”，往往都需依托更为具体的标准、指南才能真正实现落地。2020年7月，国家标准化管理委员会等五部门印发《国家新一代人工智能标准体系建设指南》。2024年6月，工信部等四部门印发《国家人工智能产业综合标准化体系建设指南（2024版）》，提出基础共性标准、基础支撑标准、关键技术标准、智能产品与服务标准、赋能新型工业化标准、行业应用标准、安全/治理标准等7个重点方向，涵盖了人工智能产业链的各个环节，为产业高质量发展提供了全方位的标准化支撑。目前，中国网络安全标准化技术委员会已发布《人工智能安全治理框架》1.0版和《生成式人工智能服务安全基本要求》，针对人工智能内生安全风险和应用安全风险提出相应技术应对和综合防治措施，为人工智能安全开发和应用指引方向。中国产业界积极参与国际标准制定，如科大讯飞联合中国电子技术标准化研究院等牵头制定了国际标准ISO/IEC 5259-4:2024《人工智能 - 分析和机器学习的数据质量 - 第4部分：数据质量过程框架》，为人工智能领域数据质量提供标准依据。

（三）中国积极参与全球人工智能治理体系建设

1. 提出人工智能全球治理的中国方案

2023年10月，中国国家主席习近平提出《全球人工智能治理倡议》，阐述中国关于全球人工智能治理的原则和主张。**一是坚持以人为本、智能向善**，发展人工智能应以增进人类共同福祉为目标，以保障社会安全、尊重人类权益为前提，确保人工智能始终朝着有利于人类文明进步的方向发展。应遵守适用的国际法，符合全人类共同价值，共同防范和打击对人工智能技术的恶用滥用。**二是坚持发展和安全并重**，各国享有平等发展和利用人工智能的权利，打造开放、包容、普惠、非歧视的人工智能发展环境，让人工智能红利惠

及所有国家；同时要发展安全可控的人工智能，打造可监督、可追溯、可信赖的人工智能技术，不断健全法律规章制度，确保人工智能始终处于人类控制之下。**三是坚持多边主义**，增强发展中国家在人工智能问题上的代表性和发言权，支持联合国在人工智能全球治理中发挥主渠道作用，支持在联合国框架下成立国际人工智能治理机构，协调国际人工智能发展、安全与治理的重大问题，让各国平等参与、平等受益。

2. 回应发展中国家关切，做人工智能能力建设的倡导者和落实者

为帮助发展中国家从人工智能技术发展中平等受益，中方提出以能力建设为抓手，弥合数字鸿沟和智能鸿沟。中方主提的《加强人工智能能力建设的国际合作》决议在第 78 届联合国大会获协商一致通过，143 个国家参与联署。^① 决议聚焦人工智能能力建设主题，鼓励开展政策交流、知识共享、技术转让、人员培训、研究等相关合作，支持联合国在国际合作中发挥主渠道作用，实现人工智能包容普惠可持续发展。该决议获得协商一致通过，充分体现了各国特别是发展中国家参与人工智能发展治理、加强能力建设国际合作的迫切愿望。

中国积极推进决议后续落实工作。2024 年 9 月，中国举办了人工智能能力建设研讨班，包括泰国、新加坡、马来西亚、越南、印尼在内的 30 余国派员参加。同月，中国与赞比亚在第 79 届联大期间共同举办“人工智能能力建设国际合作高级别会议”，发布《人工智能能

^① “A/RES/78/311, Enhancing international cooperation on capacity-building of artificial intelligence”, July 2024, <https://digitallibrary.un.org/record/4053588>.



力建设普惠计划》^①，得到各方积极支持和响应。普惠计划围绕人工智能基础设施、产业赋能、人才培养、数据建设、安全治理等领域合作提出了“五大愿景”，提出中方将采取“十项行动”，包括促进人工智能和数字基础设施联通，开展模型研发和应用、语料建设等合作，加强人工智能战略对接和政策交流，积极分享技术实践，帮助发展中国家在人工智能发展进程中平等受益等。为了推进普惠计划的落实，中国还倡议成立人工智能能力建设国际合作之友小组。

3. 积极开展人工智能国际对话与合作，促进相互理解与信任

中美人工智能政府间对话机制首次会议于 2024 年 5 月在瑞士日内瓦举行，围绕人工智能科技风险、全球治理、各自关切的其他问题深入、专业、建设性地交换意见。

中新数字政策对话机制首次会议于 2024 年 6 月在北京召开，就可信商业数据流动和人工智能能力合作开展交流，促进在人工智能治理框架领域的相互了解。

《中法关于人工智能和全球治理的联合声明》决定深化两国关于国际治理模式的讨论、坚持智能向善宗旨，促进安全、可靠和可信的人工智能系统，主张人工智能必须服务公共利益、尊重文化和语言的多样性。

《2024 年中非互联网发展与合作论坛关于中非人工智能合作的主席声明》提出加强政策对话与沟通、推动技术研发与应用、促进产业合作与发展、开展人才交流与能力建设，以及筑牢网络和数据安全屏

^① “AI Capacity-Building Action Plan for Good and for All”, Chinese Ministry of Foreign Affairs, September 27, 2024, https://www.mfa.gov.cn/eng/wjzbzd/202409/t20240927_11498465.html.

障等合作建议，充分彰显双方深入协同的愿景、决心与信心。^①

5月30日，中国—阿拉伯国家合作论坛举行，会议通过《中国—阿拉伯国家合作论坛第十届部长级会议北京宣言》，双方强调坚持发展和安全并重原则，弥合智能鸿沟，共同做好风险防范，探讨形成以联合国为主渠道的人工智能国际治理框架。双方还将加强政策协调，建立对话机制，深化务实合作，形成优势互补。

中国与东盟于2024年10月共同发布《中国—东盟关于推动建立可持续和包容性的数字生态合作联合声明》，称将推动人工智能准则、发展、安全和治理最佳实践的交流沟通。^②中国总理李强宣布中国将面向东盟实施人工智能赋能发展科技能力提升行动。

4. 搭建开放包容、平等参与的人工智能多边合作平台

世界人工智能大会自2018年起已成功举办7届，逐步发展为全球人工智能领域最具影响力的行业盛会之一。2024年7月，世界人工智能大会暨人工智能全球治理高级别会议在上海召开，吸引了各国政府、国际组织、产学研等各界代表约1000人与会，会议发表《人工智能全球治理上海宣言》，强调共同促进人工智能技术的发展和應用，确保其发展过程中的安全性、可靠性、可控性和公平性，促进人工智能技术赋能人类社会发展的共同理念。

“中国—金砖国家人工智能发展与合作中心”于2024年7月正式

^① “Chair’s Statement on China-Africa Cooperation on AI”, 2024 China-Africa Internet Development and Cooperation Forum, April 3, 2024, https://www.cac.gov.cn/2024-04/03/c_1713731793842754.htm.

^② “ASEAN-China Joint Statement on Facilitating Cooperation in Building a Sustainable and Inclusive Digital Ecosystem”, October 10, 2024, <https://asean.org/asean-china-joint-statement-on-facilitating-cooperation-in-building-a-sustainable-and-inclusive-digital-ecosystem/>.



启动，秉持开放包容、合作共赢的金砖精神，围绕人工智能技术研发与创新、产业合作生态、治理实践等方面开展务实合作，为金砖及金砖+国家的人工智能发展增添活力、注入动力。

联合国工发组织全球工业人工智能联盟卓越中心也于 2024 年 7 月正式落户上海，围绕“全球人工智能产业加速器”和“国际人工智能会客厅”两个定位，促进在工业部门应用人工智能，帮助发展中国家共享人工智能红利。

04

东盟的人工智能发展与治理

- 东盟国家人工智能发展潜力巨大
- 东盟国家人工智能治理迅速起步

04

东盟的人工智能发展与治理

近年，东盟国家重视推进人工智能技术的应用与发展，不仅在国家层面出台人工智能发展战略，也在地区层面加大协调。同时，东盟还致力于为人工智能技术带来的各类风险挑战和伦理问题提供应对思路，并以开放的姿态开展有关人工智能治理的国际对话。

（一）东盟国家人工智能发展潜力巨大

1. 东盟国家出台政策，推动人工智能技术发展与应用

6个东盟国家制定了本国的人工智能国家战略（表2）。自2019年11月新加坡率先发布《国家人工智能战略1.0》后，其他东南亚国家根据自身国内需求和经济发展程度，也相继出台了国家总体战略或路线图，目的是通过发展人工智能技术加速国家数字化转型，提高生产力，推动经济增长，在全球科技竞争中占据更有利的地位。



表 2 东盟国家与人工智能相关的国家战略

国家	与人工智能相关国家战略	发布时间
新加坡	《国家人工智能战略 1.0》	2019 年 11 月
	《国家人工智能战略 2.0》	2023 年 12 月
印度尼西亚	《印尼 2020-2045 年期间人工智能战略》	2020 年 8 月
越南	《越南人工智能研发和应用国家战略 (2021-2030)》	2021 年 1 月
马来西亚	马来西亚《数字经济蓝图》，下设“国家 人工智能路线图”	2021 年 4 月
菲律宾	《国家人工智能战略路线图》	2021 年 5 月
	《国家人工智能战略路线图 2.0》	2024 年 7 月
泰国	《泰国国家人工智能战略和行动计划 (2022-2027)》	2022 年 1 月

** 其他东盟国家如柬埔寨、老挝、缅甸和文莱，虽暂未制定与人工智能相关的国家战略，也在逐步制定和推进数字经济总体规划和人工智能相关政策。

东盟已开始协调区域人工智能发展政策。东盟虽尚未形成统一的人工智能战略，但已开始借助现有的区域经济框架，如《区域全面经济伙伴关系协定》（RCEP），加强在数据流动、隐私保护和跨境电商等方面的交流与合作，夯实人工智能发展基础。在《东盟数字总体规划 2025》（ASEAN Digital Masterplan 2025）中，东盟提出通过区域合作推动人工智能发展的愿景，并计划在 2025 年完成《东盟数字经济框架协议》（DEFA）谈判，旨在减少地区数字政策的碎片化，推动区域内人工智能政策的整合。

东盟着力推进同域外国家的人工智能合作。随着域外国家日益重视东盟人工智能发展和应用的潜力，东盟国家也加大了与这些国家的合作力度，推动在人工智能关键领域的技术交流、政策对接和规则制定。2022 年 11 月，东盟与中国发布《中国—东盟全面战略伙伴关系行动计划（2022-2025）》，强调为迎接第四次工业革命，双方将强加在人工智能、网络和数据安全等领域的合作。2023 年 9 月，东盟与中国共同

提出《共同推进实施中国—东盟科技创新提升计划的联合倡议》，在人工智能、数字经济和智慧城市等领域探索合作模式，缩小数字鸿沟。2024年10月，东盟与中国共同发布《中国—东盟关于推动建立可持续和包容性的数字生态合作联合声明》，致力于合作构建可持续和包容性数字生态系统。

2. 推进数字基础设施建设，为人工智能发展打下基础

东盟国家加速改善数字基础设施并加快互联互通，为人工智能、云计算和其它数字技术的广泛应用创造条件。近年，东盟国家数据中心逐年增加（图11），为人工智能发展提供了必要的计算和存储支持，降低了数据传输至境外的隐私和安全风险。为满足对数字化服务和应用的庞大需求，东盟数据中心容量也取得显著增长。从2024年到2028年，东盟数据中心容量预计将从1677兆瓦（MW）增加到约7589兆瓦，增长约四倍。^①5G网络基础设施建设也在稳步推进。据联合国贸发会议预计，2020-2025年期间东盟每年在5G基础网络的投资规模约为140亿美元，主要用于升级通讯设施、网络和设备以满足快速增长的5G需求。^②为了支持数据中心对能源的巨大需求，同时满足绿色发展的目标，东盟国家通过《东盟能源合作行动计划》（APAEC）等框架，推动成员国加强区域内能源合作，加大清洁能源使用，并提升能源效率。东盟国家之间也有了更多数字基础设施方面的合作项目，如新加坡与马来西亚通过海底电缆和光纤网络连接，印度尼西亚与菲律宾、马来西亚加强海底电缆项目合作，有效部署区域内资源，加强东盟数字基础设施的互联互通。

^① “ASEAN Investment Report 2020–2021 Investing in Industry 4.0”, The ASEAN Secretariat and United Nations Conference on Trade and Development, September 8, 2021, <https://investasean.asean.org/files/upload/ASEAN%20Investment%20Report%202020-2021.pdf>.

^② 同上。

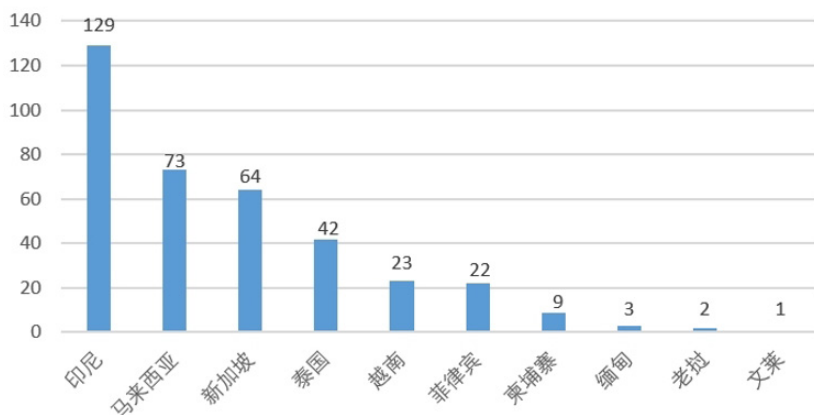


图 11 2024 年东盟国家数据中心数量
(截至 2024 年 10 月 30 日)

资料来源: <https://www.datacentermap.com/datacenters/>

3. 吸引国际投资，助力人工智能企业快速成长

谷歌、淡马锡和贝恩于 2023 年 11 月联合发布《2023 年东南亚数字经济报告》，显示自 2021 年以来，东南亚国家数字经济收入复合年均增长率达 27%。预计到 2025 年，东南亚数字经济收入将以 16% 的复合年均增长率增至 2950 亿美元，地区中产阶级人数将达到 3.63 亿人，占总人口的 52%，年轻劳动力的增加以及实际工资的增长将推动数字红利持续释放。^① 东盟预测，2030 年东南亚数字经济规模将达到 1 万

^① “e-Conomy SEA 2023”, Google, TEMASEK, BAIN & COMPANY, November 01, 2023, https://services.google.com/fh/files/misc/e_conomy_sea_2023_report.pdf & “Future of ASEAN, Forecast to 2025: GDP is forecast to increase from \$2.4 Trillion in 2015 to \$5.2 Trillion in 2025, largely due to the ongoing implementation of the AEC - Research and Markets”, April 28, 2017, <https://www.businesswire.com/news/home/20170428005791/en/Future-of-ASEAN-Forecast-to-2025-GDP-is-forecast-to-increase-from-2.4-Trillion-in-2015-to-5.2-Trillion-in-2025-largely-due-to-the-ongoing-implementation-of-the-AEC---Research-and-Markets>.

亿美元。^①一旦《东盟数字经济框架协议》（DEFA）生效，其渐进式规则将使东盟数字产业规模最高释放至 2 万亿美元。^②

东南亚地区数字经济发展的良好前景，吸引了外资不断涌入。根据乔治敦大学研究机构发布的报告，新加坡、菲律宾、泰国、印度尼西亚和马来西亚 60% 的投资交易均来自外资。^③ 中国企业，包括阿里巴巴的 Lazada、腾讯的 Shopee、字节跳动的 TikTok Shop 等在东南亚运营的电商平台，运用人工智能技术提升用户体验。阿里云、腾讯云和华为云在新加坡、泰国、印度尼西亚等地投资兴建数据中心，提供云计算和人工智能基础设施支持。阿里达摩院、科大讯飞和智谱华章都推出了支持东南亚多语言的人工智能大模型，服务本地人工智能技术的推广。

外资的持续流入为东盟地区的初创公司带来了显著的增长动力。根据联合国贸发会议（UNCTAD）与东盟在 2024 年 10 月发布的联合报告，从 2015 年到 2021 年，东盟筹集超过 100 万美元资金的初创企业数量几乎翻倍，达到 1920 家，到 2024 年上半年增长到 4710 家，增长 150%。^④ 在这些初创企业中，数字领域的初创企业数量快速增长，业务范围也在不断扩大，涵盖自动化、医疗保健、金融、内容创建、娱乐以及零售等各方面。

^① “Digital Economy Framework Agreement (DEFA): ASEAN to leap forward its digital economy and unlock US\$2 Tn by 2030”, August 19, 2023, <https://asean.org/asean-defa-study-projects-digital-economy-leap-to-us2tn-by-2030/>.

^② 同上。

^③ Ngor Luong, Channing Lee and Margarita Konaev, “Chinese AI Investment and Commercial Activity in Southeast Asia”, February 2023, <https://cset.georgetown.edu/publication/chinese-ai-investment-and-commercial-activity-in-southeast-asia/>.

^④ “ASEAN Investment Report 2024”, The ASEAN Secretariat and United Nations Conference on Trade and Development, October 2024, <https://asean.org/wp-content/uploads/2024/10/AIR2024-3.pdf>.



在资金的加持下，东盟的人工智能初创企业中涌现出许多“头雁”企业，为东盟整体经济注入了活力。例如，泰国的 Social+ 公司、新加坡的 Wiz、Kroolo、Ai Palette 和 Mindverse 公司、菲律宾的 Sprout Solutions 公司和马来西亚的 Shieldbase 公司，通过开发机器人或流程自动化工具提升客户互动与体验，显著提高公司生产效率。医疗健康领域也聚集了一批重要的人工智能初创企业，如新加坡的 KeyReply 公司和泰国的 Meticuly 公司，通过智能虚拟助手和智能调度系统等人工智能解决方案改善诊疗效果。此外，一些人工智能初创企业专注于在人力资源、金融、信息技术、媒体娱乐和电子商务领域的应用创新。例如，菲律宾的 Sprout Solutions 提供人力资源和薪资管理平台，新加坡的 Bluesheets 运用人工智能技术提升金融服务效率，新加坡的 Locofy.ai 致力于软件开发和信息服务，帮助加速应用程序开发，越南的 AI Hay 公司则专注于合成媒体和数字资产创建，推动 AI 在社交生活中的应用，新加坡的 Pebblely 公司和越南的 Modeli 公司通过人工智能技术实现个性化产品推荐、自动化库存管理和客户服务的智能化，为电子商务的创新和效率提升提供了有力支持。^①

4. 采取措施促进数字人才的引进和培养

人工智能的到来改变了传统的就业格局，熟悉新技术、掌握新技能的人才成为保持经济活力的重要因素。东盟国家采取了许多措施提升地区人工智能技术人才保有量。首先，通过签证政策吸引全球人才。新加坡通过多样化的签证计划，特别是创业和技术签证，吸引了全球顶尖的人工智能等高新技术人才，为技术创新和创业生态系统注入活力。马来西亚的“DE Rantau”签证计划为科技专业人才和数字游牧民

^① Laura Nguyen and Denning Tan, “ASEAN Genai Startup Report 2024”, September 2024, <https://genaifund.ai/wp-content/uploads/2024/09/ASEAN-GenAI-Startup-Report-2024-by-GenAI-Fund.pdf>.

进入马来西亚提供了便利，将尽可能多的全球高技能人才纳入马来西亚快速发展的数字经济。^①

其次，东盟国家加快培养本地技术人才，为劳动力提供技能转型机会。新加坡通过“SkillsFuture”项目为新加坡人提供贯穿一生的学习机会，帮助人们不断适应社会和经济变化，在职业生涯中不断提升自己的技能水平。该项目提供包含数据分析、技术支持服务、数字媒体、网络安全等超过 7000 门课程，帮助劳动者提升数字技能。^② 马来西亚通过《数字经济蓝图》强调了在自动化冲击较大的行业中进行再教育和培训的重要性。泰国增加教育支出，探索创新的教育模式，在基础教育中引入技术创新课程，以发掘人才潜力。东盟多个国家还加强了针对在职人员的技术技能培训，如柬埔寨政府与职业培训机构合作，为在职人员提供数字技能培训。^③ 越南培训大量的工程师和信息技术专业人员，推动建设人工智能创新的稳定人才渠道。^④

最后，东盟国家与国际企业和组织合作推进人才培养，推动技术

^① “Attracting Global Talents: Bringing Digital Nomads and the Highly Skilled into ASEAN”, Economic Research Institute for ASEAN and East Asia, November 2023, <https://www.eria.org/uploads/media/policy-brief/FY2023/Bringing-Digital-Nomads-Highly-Skilled-into-ASEAN-web.pdf>.

^② “New SkillsFuture Series To Train 50,000 Annually In Eight Priority And Emerging Skills Areas”, October 2017, <https://www.skillsfuture.gov.sg/newsroom/new-skillsfuture-series-to-train-50-000-annually-in-eight-priority-and-emerging-skills-areas>.

^③ “Driving ASEAN’s talent transformation through technology”, Accenture, November 2017, <https://cariasean.org/publications/advancing-asean-in-the-digital-age/driving-aseans-talent-transformation-through-technology/>.

^④ “Harnessing AI to rapidly accelerate Vietnam’s economic expansion”, September 2024, <https://www.thestar.com.my/business/business-news/2024/09/03/harnessing-ai-to-rapidly-accelerate-vietnams-economic-expansion>.



与国际标准对接。例如，东盟多家科技公司与世界经济论坛合作，通过“ASEAN Digital Skills Vision 2020”培训了数百万中小企业员工，重点提高其数字技能。2022年9月，东盟基金会和华为共同举办亚太数字人才峰会，讨论亚太地区构建面向未来的ICT人才库、释放数字化潜力的路径。^① 印尼的“数字人才奖学金计划”(Digital Talent Scholarship)中，华为、思科、谷歌等国际企业提供奖学金和实习培训机会，帮助年轻人掌握前沿技术，以支持印尼数字经济的转型和工业4.0的发展。^② 东盟企业也注重国际合作与人才交流，如印尼的科技巨头GoTo推出了多项国际人才培养计划，给公司全球范围内的员工提供包括人工智能技术、数据使用方面的培训。菲律宾Aboitiz集团通过与国际大学和企业合作，推动人工智能技术在公共政策和医疗领域的应用，同时为员工提供国际化的培训和工作机会。东盟国家的这些举措提升了在技术领域的人才竞争力，为区域内的数字化转型和人工智能应用打下初步人才基础。

总体来看，东南亚地区的人工智能发展呈现出蓬勃的生机和活力。然而，也需注意到，东南亚各国间人工智能水平的差异可能会加剧区域内发展不平衡，对地区包容性发展构成潜在威胁。首先，数字技能差距广泛存在。仍有许多劳动者和公民缺乏接受数字技术培训的机会，对人工智能技术的认知较浅，获取和利用数字服务的能力也不足。例如，许多东南亚农村地区的居民尚不具备使用智能手机和数字平台的条件和技能。其次，高昂的成本成为东盟不少国家发展人工智能的障

^① “Huawei and ASEAN Foundation Call for Action to Groom Digital Talent in APAC”, September 2022, <https://www.telecomreviewasia.com/news/industry-news/2957-huawei-and-asean-foundation-calls-for-action-to-groom-digital-talent-in-apac/>.

^② “Indonesia Government to Boost Digital Talent for Industry 4.0”, November 2021, <https://opengovasia.com/2021/11/18/indonesia-government-to-boost-digital-talents-for-industry-4-0/>.

碍。东盟部分国家发展较为落后，互联网、移动数据、计算中心等成本在许多低收入国家难以承受，尤其是在乡村和偏远地区。第三，缺乏多样化的本地服务与内容导致了“南橘北枳”的困境。在欠发达地区，许多数字平台尚未能提供与本地语言、文化和习惯相适应的内容，极大地限制了技术的普及率、拉低了用户的体验感，难以发挥出数字技术的潜在价值。最后，网络基础设施仍不够先进。尽管东盟国家的互联网覆盖率在提升，但在农村和偏远地区，网络连接不稳定和速度缓慢的情况依然普遍存在。上述不足若不能有效弥补，会大大掣肘东盟国家的人工智能技术应用和发展进度，加深地区国家内部和国家间的智能鸿沟，给东盟的包容、韧性和可持续发展战略蒙上阴影。

（二）东盟国家人工智能治理迅速起步

1. 东盟国家积极完善人工智能治理框架

东盟国家多将人工智能治理整合至人工智能国家战略，通过多层次的规制措施加快治理进程。比如，新加坡《国家人工智能战略 1.0》《国家人工智能战略 2.0》、印度尼西亚《2020-2045 年人工智能国家战略》、越南《2021-2030 年人工智能国家战略》、泰国《国家人工智能战略和行动计划（2022-2027）》都提及人工智能透明性、公平性、数据安全与隐私保护等治理要素，为人工智能开发与应用提供了政策、法律框架。部分东盟国家制定特定的法律法规，将治理进一步细化。例如，马来西亚制定《个人数据保护法》，并发布了“人工智能治理和道德准则（AIGE）”，确立了七项治理核心原则；印尼发布《人工智能伦理指南（草案）》强调算法透明、数据隐私和公平；菲律宾发布《数据隐私法》，以应对技术带来的隐私泄露风险；越南更新其《数据保护法（草案）》并拟于 2026 年发布，希更好地保护个人数据。部分国家，如柬埔寨、老挝等国，虽尚未有明确的人工智能治理战略和专门法规，但在促进技术发展的同时，正逐步建立人工智能监管框架。



东盟构建和完善区域层面的治理框架。2021年1月，东盟通过《东盟数字总体规划2025》，为东盟成员国在数字经济领域的发展提供统一指导。在数字治理方面，该《规划》提出，要通过制定相关的数据安全法规、加强打击网络犯罪等措施保障数字服务的可信度，要建立统一的跨境数字贸易规则 and 标准、推动跨境数字贸易，要增强数字经济参与能力、建设包容性数字社会等一系列方向。2024年2月，在第四届东盟数字部长会议上，《东盟人工智能治理与伦理指南》发布，这是东盟地区首个针对人工智能治理的指导性文件，旨在通过区域合作推动人工智能的负责任开发和使用。会议还决定成立人工智能治理工作组，以讨论和处理东盟人工智能治理中的关键问题，促进安全、负责任且符合伦理的人工智能应用。^①2024年6月，东盟科学、技术和创新委员会启动了2024-2025年的人工智能行动计划“ACT on AI”，补充和加强已有的治理框架。

2. 东盟重视弥合区域内“智能鸿沟”

根据牛津大学发布的《2023年政府人工智能准备指数》，新加坡在人工智能准备方面遥遥领先，获得了81.97分（满分100分），远高于东盟地区55.49分的平均水平（具体分数可参考图12）。由指数排名可以看出，东盟国家人工智能发展水平差距较大，新加坡、马来西亚、泰国等发展较好。

^① “ASEAN Ministerial Meeting on Science, Technology and Innovation (AMMSTI) Statement on Artificial Intelligence (AI)”, June 2024, https://asean.org/wp-content/uploads/2024/06/ADOPTED-AMMSTI-Statement-on-AI_7June2024.pdf.

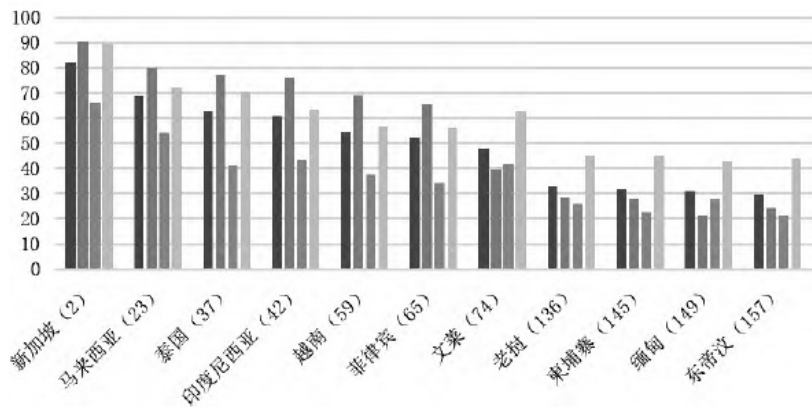


图 1 2023 年东南亚各国人工智能准备指数

注:括号内数字为国家人工智能准备指数排名。

数据来源: Emma Hankin, Pabul Fuentes Nettel, Livia Martinescu, et al. 2023 Government AI Readiness Index [R/OL]. (2023-12-20) [2023-12-29]. <https://oxfordinsights.com/wp-content/uploads/2023/12/2023-Government-AI-Readiness-Index-2.pdf>.

图 12

来源: 刘晓麒,《人工智能对东南亚地区秩序建构的影响》,《世界经济与政治论坛》2024 年第 4 期,第 90 页。

东盟积极整合资源,优化治理措施,致力于缩小区域内“智能鸿沟”。2021 年 1 月,《东盟信息通信技术总体规划 2025》(ASEAN Digital Master plan 2025, ADM2025) 发布,旨在引领东盟在 2021 至 2025 年的数字合作,帮助数字欠发达国家加强数字基础设施,推动区域互联互通及网络安全。东盟还通过区域合作机制和优先资助项目,如东盟基础设施基金、东盟催化绿色金融基金(由亚洲开发银行和东盟联合发起)等向数字基础设施不足的国家提供资金和技术支持。东盟积极深化国际合作,借助外部力量促进区域内的发展均衡。比如,联合国教科文组织高等教育创新中心(ICHEI)与东盟多国合作,开展高等教育数字化转型项目,提供在线课程和培训,提升教师和学生的数字技能,促进人工智能教育的发展,弥合各国在人工智能人才方面的差距。



东盟国家鼓励多主体参与弥合“智能鸿沟”。政府、企业和社会各界共同参与，提供政策引导、资金支持、技术培训等，帮助欠发达地区和群体更快接入智能技术，缩小数字差距。

3. 东盟国家着手开展数据治理

东盟国家高度重视数据风险管理，积极提升数据质量，并强化数据标准的对接与统一。《东盟人工智能治理与伦理指南》明确指出，在人工智能系统生命周期中，数据的收集、存储、处理和删除方式必须符合相关数据保护法、数据治理法规和伦理原则。不少东盟国家已着手制定数据保护相关法律，以降低人工智能系统生命周期中数据使用的风险，如马来西亚的《个人数据保护法》（2010年）、菲律宾的《数据隐私法》（2012年）、新加坡的《个人数据保护法》（2012年）、泰国的《个人数据保护法》（2019年）、印度尼西亚的《个人数据保护法》（2022年）和越南的《个人数据保护法（草案）》（2024年）等。

东盟国家是全球数据标注产业的重要基地。印尼、菲律宾和越南的公司，如 Kata.ai、Straive 和 Viettel 公司，分别在图片、文本和视频数据标注服务中发挥了关键作用，提高了数据的准确性和可靠性。

新加坡通过先进的数据收集机制，系统、实时地收集数据，应用于交通、医疗和物流等领域，优化各行业的运营效率。马来西亚通过《数字经济蓝图》推动公共和私营部门的数据共享，并开放政府数据，提升数据透明度，同时在社交媒体内容监管中，要求平台使用人工智能技术时必须确保不会过度收集或泄露用户信息。

东盟国家重视数据标准统一和对接以促进数据流动。RCEP 和正在推进的 DEFA 都将跨境数据流动作为重要议题推进，显示东盟希望通过数据互联互通和标准化增强区域经济竞争力的目标。此外，在 2024 年 6 月召开的东盟科学、技术与创新部长级会议上，各国达成共识，承诺加强数据治理和保护，推动地区标准对接，并保持数据的定期更新和审查。

4. 东盟国家开始对算法实施监管

人工智能算法利用数学和统计学原理帮助计算机从数据中学习、推断和执行任务，是人工智能技术应用的核心环节。但其内部逻辑规则的复杂性容易导致算法的“黑箱”效应，从而引发偏见、偏差和信息安全风险。《东盟人工智能治理与伦理指南》将算法透明度和可解释性列为重要治理原则，认为部署者有责任解释算法决策逻辑、执行统一的透明度标准，减少因“黑箱”效应带来的不信任感。新加坡发布的《人工智能治理框架模型》也强调在算法设计中引入可解释性，以提高算法的透明度并减少“黑箱”效应。^①

人工智能自动化分析系统对技术参数的变化极为敏感。一些社交媒体平台的算法推荐系统会因参数调整导致输出结果变化，可能会助长社会分裂或传播虚假信息风险。东盟一些国家已针对算法推荐的潜在风险采取措施，如印尼对国外社交媒体平台的算法推荐系统实施严格监管，防止其滥用算法影响公共舆论。

在金融和医疗等高敏感领域，东盟国家也在努力追求透明合规与信息安全。例如，新加坡金融管理局（MAS）通过风险框架和监管沙盒，支持企业在保护商业机密的同时，测试和开发创新的人工智能解决方案，确保合规与安全。泰国对金融和医疗等高风险领域的人工智能应用进行严格监管，最大程度地减少算法引发的风险与偏差。

5. 东盟国家关注人工智能系统是否稳健可靠

人工智能系统包括模型、产品、算力和供应链等关键部分，保障其稳健可靠对于降低整体风险和提升社会信任至关重要。《东盟人工

^① Model Artificial Intelligence Governance Framework (Second Edition), Information Media Development Authority (IMDA) and Personal Data Protection Commission (PDPC), January 21 2020, <https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf>.



智能治理与伦理指南》明确将“安全与保障”列为核心指导原则，强调对人工智能应用已知风险的评估和管理，确保系统能够抵御潜在的恶意攻击。一些东盟国家和企业已经采取了具体措施来验证和提高人工智能系统的稳健性。例如，印尼多服务及支付平台 Gojek 在部署机器学习模型之前，会进行离线基准测试，以验证模型的性能和可靠性；新加坡公司 UCARE.AI 与客户合作创建了一个验证框架，通过收集患者反馈，确保人工智能模型的准确性和可靠性。此外，新加坡在 2022 年 5 月推出人工智能验证（AI Verify）平台，通过系统安全审查清单对人工智能系统的算力、安全性和数据保护进行全面评估，帮助企业识别和应对潜在风险。

不仅如此，东盟各国对人工智能的具体应用场景（如金融和医疗领域）也加强了安全审查。例如，新加坡、马来西亚、泰国和越南都建立了严格的控制机制，对关键的数字基础设施进行定期检查，以确保其安全性和稳健性。

在人工智能系统的供应链方面，由于其通常涉及多个部门甚至多个国家，东盟国家采取了跨部门安全监控措施，并积极推动跨境合作。例如，东盟与日本合作建立了一个数字化的供应链生态系统，目标是提升供应链的透明性、追踪能力、连接性和可持续性。^①

6. 东盟着力提升在人工智能全球治理领域的影响力

近年，东盟成员国加大对人工智能全球治理问题的参与度。新加坡经常活跃在联合国和全球各大论坛关于人工智能治理问题的讨论中，并参与相关人工智能标准的制定。新加坡还与其他东盟国家合作，共同推动区域内人工智能政策与国际进程接轨。在联合国教

^① “Vision for the Digitalisation of Supply Chains in ASEAN and Japan”, Economic Research Institute for ASEAN and East Asia (ERIA), February 2024, <https://www.eria.org/uploads/Vision-for-the-Digitalisation-of-Supply-Chains-in-ASEAN-and-Japan.pdf>.

科文组织制定《人工智能伦理建议书》的过程中，东盟成员国（尤其是新加坡、马来西亚和泰国）积极参与，贡献了区域性的智慧。这些国家不仅协助推动该伦理框架在东盟区域内的实施，还通过推广多元、包容和可持续发展的地区价值观，推动人工智能全球治理体系的完善。

通过积极参与国际合作，东盟国家在人工智能全球治理中注入了独特的地区视角，倡导符合东盟价值观的治理原则，强调资源公平和社会伦理的重要性。这种以地区价值观为基础的参与方式，不仅促进了国际合作和跨文化理解，也为全球人工智能技术的负责任和可持续应用提供了新的参考，增强了东盟在人工智能全球治理中的影响力。



05

中国与东盟的 人工智能合作

- 中国与东盟人工智能发展合作基础深厚
- 中国与东盟人工智能治理合作前景可期

中国与东盟的人工智能合作以中国—东盟数字经济合作为基础，是数字丝绸之路的重要组成部分。中国秉持亲、诚、惠、容理念，与东盟国家开展人工智能产业和技术合作、推动治理和规则协调，着力构建区域人工智能生态圈，为缩小数字鸿沟、增强经济发展的包容性、可持续性和韧性持续贡献力量。

（一）中国与东盟人工智能发展合作基础深厚

1. 中国和东盟数字政策对接顺畅

中国与东盟关于人工智能发展合作的政策对接始于 2019 年。当年 11 月，中国—东盟领导人会议通过《中国—东盟智慧城市合作倡议领导人声明》，提出加强构建智慧城市生态系统的交流合作。2020 年被确立为中国—东盟数字经济合作年。2021 年，中国—东盟领导人会议通过《中国—东盟关于建立数字经济合作伙伴关系的倡议》，将中国



数字经济发展规划与《东盟数字总体规划 2025》相对接，同意积极推动双方在 5G、大数据、人工智能、智慧城市等领域合作，共同打造互信互利、包容、创新、共赢的数字经济合作伙伴关系。2024 年 10 月，第 27 次中国—东盟领导人会议发表《中国—东盟关于推动建立可持续和包容性的数字生态合作联合声明》，宣布从加强政策交流与战略对接、推动数字基础设施建设、加快新兴数字技术创新应用等方面，进一步促进中国—东盟数字生态系统合作。中国总理李强宣布中国将面向东盟实施人工智能赋能发展科技能力提升行动。

中国与东盟的数字经济管理部门合作非常密切。2021 年，首届中国—东盟数字部长会议召开。2022 年，第 2 次中国—东盟数字部长会议通过《落实中国—东盟数字经济合作伙伴关系行动计划 (2021-2025)》和《2022 年中国—东盟数字合作计划》，为双方在数字基础设施、数字贸易、数字技术创新、数字产业、数字人才培养等方面的合作制定了具体指导和规划。此后每年的数字部长会议，双方都会制定下一年的《中国—东盟数字工作计划》。在 2024 年召开的第 4 次中国—东盟数字部长会议上，双方签署了《中国—东盟通信和数字技术合作谅解备忘录》。

中国与东盟搭建了一系列人工智能和数字治理领域的对话和论坛机制。2023 年 7 月，首届中国—东盟人工智能合作论坛在广西南宁召开，中国科技部在论坛上发布《面向东盟的人工智能发展合作倡议》，呼吁各国建立稳定的合作机制，推进人工智能联合研发和应用合作，加强人工智能发展政策和治理原则的交流，健全具有广泛共识的治理框架和标准规范。2024 年 6 月，中国国家互联网信息办公室在广西主办中国—东盟数字治理对话，聚焦讨论人工智能治理、数据治理、网络安全应急响应合作等议题。

2. 中国和东盟人工智能产业合作活跃

数字基础设施合作推动东盟互联互通建设。数字基础设施主要包

括通信网络基础设施、数据中心和云服务基础设施、计算中心、卫星导航系统等，是人工智能产业发展的基石。中国企业积极参与泰国、菲律宾、柬埔寨等东盟国家国内的 4G 和 5G 通信基站、光纤宽带接入等基础设施建设。在数据中心和云服务方面，阿里云、万国数据、腾讯云等中资企业与泰国、印尼、柬埔寨、缅甸、老挝等国合作，在当地投资建设数据中心并提供质量上乘、价格合理的云计算服务。2023 年 9 月，中国—东盟人工智能计算中心正式上线。该中心拥有全球领先的人工智能多元算力架构，主要用于人工智能深度学习模型开发、训练和推理等场景，打造面向千行百业的公共算力服务平台。中国的北斗系统也已开始服务于东盟国家的国土测绘、精准农业、地质灾害监测等工程，为缩小“数字鸿沟”做出贡献。

智能产业合作促进企业提质、降本、增效。人工智能技术应用于生产制造、供应链管理、工程建设等领域，可显著提升企业生产效率和质量检测水平，并降低能源消耗和运维成本。中国—东盟技术转移中心等机构为中国与东盟国家的智能制造技术转移和合作提供了平台支持。中国—东盟（华为）人工智能创新中心推进“人工智能+工业”应用落地，在制糖、有色金属等领域启动了一批标杆项目，推动了相关产业的智能化发展。2018 年，中国海云数据与印尼美加达公司签署谅解协议，就人工智能产业聚合及孵化等领域展开合作。2019 年，商汤科技、马来西亚 G3 Global 公司同中国港湾工程公司签署战略合作协议，共建马来西亚首个人工智能产业园，这也是东盟第一个人工智能驱动的科技城。该产业园将作为人工智能解决方案开发的平台，推动马来西亚人工智能技术领域的生态建设与产业发展。

智慧城市项目优化城市治理。智慧城市是一种全新的城市建设理念，指在城市管理与运行中融入物联网、云计算等技术，构建集



产业发展、市政管理、人际交互于一体的综合创新应用体，实现城市运行效率和居民生活质量的提高，促进城市可持续发展。2018 年阿里云与马来西亚政府签订城市大脑项目，将人工智能技术应用到交通治理、城市规划和环境保护等领域。2019 年，中企参与建设的菲律宾马卡蒂市智慧城市项目入围 SCEWC 全球“创新创意奖”。中新天津生态城的智慧城市建设取得显著进展，已建立智慧城市场景应用研发基地、中新合作智慧城市研究中心等，并于 2023 年签署二期战略合作协议。深圳市和新加坡签署多份合作谅解备忘录并成立深圳—新加坡智慧城市合作联合执委会，加强电子贸易、数字身份互认、跨境仲裁等领域合作。此外，华为、浪潮、腾讯等企业也正从医疗、教育等不同角度，积极参与泰国、老挝、印尼等国的智慧城市建设。

人工智能大模型助力数字应用普及。2023 年是人工智能大模型集中投入应用的元年。6 月，科大讯飞携讯飞星火大模型与 C 端智能硬件登录新加坡，目前其生成式智慧驾驶舱、讯飞听见同传、讯飞智作、全屋智能语音面板等产品已进入东南亚市场。阿里巴巴国际站于同期发布人工智能外贸产品，覆盖智能商品发布、市场分析、视频聊天、实时翻译等环节。11 月，字节跳动在海外上线基于云雀大语言模型（现名为“豆包大模型”）的人工智能工具平台 ChitChop，为用户提供 200 余种工作、生活场景智能机器人服务。尤为值得注意的是，2023 年底，阿里巴巴达摩院推出首个基于东南亚语言训练的人工智能大模型版本 SeaLLM，一改东南亚地区缺乏本土语言大模型的面貌，必将加速人工智能技术在当地的普及，为在数字领域代表性不足的人群提供支持。

消费级人工智能工具方便和丰富大众生活。中国的智能家电、智能机器人、智能穿戴设备产品等，以其丰富的功能和合理的价格广受东南亚消费者的青睐。作业帮海外人工智能教育应用 Question.AI 在东

南亚市场表现亮眼，在菲律宾的周活跃用户规模已达近 120 万，在印尼教育类应用使用排行榜中位居第二。以 TikTok 为代表的短视频平台在东南亚地区的用户规模快速增长。腾讯推出的长视频流媒体服务平台 WeTV、爱奇艺国际站 iQiyi 通过优化翻译字幕、语境转换和本地化配音等方式，降低了观看门槛，受到当地用户欢迎。中国游戏厂商也在东南亚移动游戏市场取得良好成绩。另一方面，新加坡南洋国际俱乐部有限公司投资 20 亿元，在中国广西贺州建设智能医养健康中心项目。该项目由新加坡东盟人工智能中心研发设计，建设内容包括智能医养中心、智能乐龄公寓、智能乐龄医疗康复中心及智能医养住宅等，是东盟人工智能企业在中国医养健康领域的重要投资尝试。

（二）中国与东盟人工智能治理合作前景可期

1. 中国与东盟文化相通

中华优秀传统文化以“和合”理念为精神内核，崇尚以和为贵、和而不同、和衷共济、和合共生。此外还有仁、义、礼、智、信等多维度的价值追求。“仁”泛指友善、爱人，形成协调、和睦的社会关系。

“义”强调公平正义和社会整体利益至上，推崇先公后私、大公无私，在服务整体利益中实现个人最高价值。“礼”指以维护社会秩序为目标应遵循基本礼仪和社会行为规范。“智”突出强调重视理性、追求真理、明辨是非和创新进步。“信”指真实无妄、诚实守信以及忠诚于信念和原则。

基于这些传统价值理念，中国在国际上提出构建人类命运共同体，主张不同社会制度、意识形态、历史文化和发展水平的国家应平等相待、互商互谅、合作共赢，努力谋求在国际事务中利益共生、权利共享和责任共担。中国还陆续推出全球发展倡议、全球安全倡议和全球文明倡议，倡导平等有序的世界多极化和普惠包容的经济全球化，呼吁各



国为建设持久和平、普遍安全、共同繁荣、开放包容、清洁美丽的世界共同努力。

东盟国家与中国价值观相通。透过《东盟宪章》《东盟共同体愿景 2025》以及 4 份《东盟协调一致宣言》，东盟的三大核心价值观清晰展露。

一是珍视和平与稳定。东盟诞生于冷战中的 1967 年，却并未成为一个对抗性的军事组织，而是以促进本地区和平与稳定为宗旨。东盟的成立集中体现了东南亚国家对地区纷争的厌倦和对和平的向往。东盟于 1971 年发表《和平、自由与中立地区宣言》，强调不结盟和维护地区和平、稳定与安全的决心。50 多年来，东盟始终倡导宽容、温和的处世方式，强调协商解决分歧，国家间互信不断增强，地区和平持续巩固。

二是重视发展与共同繁荣。2015 年底，东盟宣布建成东盟共同体，同时宣告共同体的成立并不意味着建设进程的结束，而恰恰意味着开始。东盟强调要建设充满活力、可持续和高度一体化的经济，加强内部互联互通，通过务实互助合作减轻贫困、缩小内部发展差距，不让任何人掉队。东盟还展现出担当地区经济增长中心和世界经济增长引擎的雄心。

三是支持多样包容和平等对话。东盟拥有 10 个成员国，不仅语言、文化多样，而且社会制度、发展水平也存在较大差异。强调主权一律平等、协商一致、照顾各方舒适度是极有特色的“东盟方式”。包容性是东盟制定各项政策时严格遵守的原则。在对外关系中，东盟强调维护以《联合国宪章》和国际法为基础的多边主义，在东盟中心基础上发展同各国的友好包容、非歧视和互利关系。

2. 中国与东盟人工智能治理理念相近

中国和东盟的人工智能治理理念有三个共性特征：倡导包容普惠、强调以人为本和主张复合、开放的治理结构。

2.1 倡导包容普惠

中国认为人工智能治理攸关全人类命运，是世界各国面临的共同课题，联合国是最具权威性和代表性的政府间国际组织，理应作为全球人工智能治理的主渠道。全球人工智能治理应坚持广泛参与、协商一致的原则，构建全球范围内开放和包容的治理机制。中国尤其强调人工智能国际合作和全球治理要坚持开放、包容、普惠和非歧视的发展环境，不搞“小圈子”。要增强发展中国家在人工智能全球治理中的代表性和发言权，确保各国人工智能发展与治理的权利平等、机会平等和规则平等。为弥合智能鸿沟和治理能力差距，帮助发展中国家更有效地参与人工智能全球治理，中国于2024年7月推动联合国大会通过加强人工智能能力建设国际合作的决议，又于同年9月提出《人工智能能力建设普惠计划》，并身体力行在上海与联合国共同主办了首届人工智能能力建设研讨班。

包容、普惠地发展人工智能同样是东盟追求的重要目标。《东盟数字总体规划2025》提出，要将东盟建设成为领先的数字共同体和经济集团，在东盟建立数字化的包容性社会即是具体目标之一。《东盟人工智能治理与伦理指南》充分体现了东盟对兼容不同国情与发展水平、促进共同发展的重视，还特别提出须尊重各国主权和法律管辖权。2024年6月，东盟科学、技术与创新部长级会议发布《关于人工智能的声明》，称人工智能技术具备促进包容和公平的经济增长、缩小东盟内部社会经济差距的潜力，承诺将力争使人工智能利益在社会中公平分配。《声明》鼓励东盟内部合作及与对话伙伴、区域和国际组织等的合作，以促进包容性经济增长，造福所有东盟国家。

2.2 强调以人为本、智能向善

以人为本，是指发展人工智能，须以提升人类社会利益、增进人类共同福祉为目标。《全球人工智能治理倡议》指出，发展人工智能



应坚持“以人为本、智能向善”理念，以增进人类共同福祉为目标，以保障社会安全、尊重人类权益为前提，确保人工智能始终朝着有利于人类文明进步的方向发展。^①《东盟人工智能治理与伦理指南》将“以人为本”作为重要指导原则之一。称“人工智能系统应尊重以人为本的价值观，追求人类社会的利益”“确保人们从人工智能设计、开发和部署中受益，同时免受潜在危害”。^②无论是中国还是东盟，对人本价值的追求并不仅限于以上论述，而是贯穿于人工智能治理理念的始终。

以人为本、智能向善，体现在要求引导人工智能技术赋能人类工作，以提升质量和效率、改善人类生活品质、促进可持续发展上。从医疗、教育、养老等与大众生活息息相关的领域到环境保护、灾害预防、资源利用、能源管理等涉及更长远挑战的领域，人工智能技术都能做出正向贡献，推动人类社会不断进步。

以人为本、智能向善，体现在要求保障人类利益不受人工智能算法的伤害上。中国和东盟都强调应保护数据隐私，算法决策应避免产生针对不同人群的偏见和歧视、避免进一步侵害弱势群体的利益，算法不应被用于操控用户、诱导其做出原本不会做出的决定等。

以人为本、智能向善，还体现在关注人工智能对就业市场的干扰并努力降低人工智能对人类社会的远景威胁上。人工智能的广泛应用可能引发大规模失业。对人工智能技术的过度依赖可能导致人类思维能力下降、对技术的控制力断崖式下跌。为防止此类情形出现，必须尽快、尽可能多地展开人工智能教育和技能培训。要培养更多能够驾

^① 《全球人工智能治理倡议》，中国外交部，2023年10月20日，https://www.mfa.gov.cn/web/ziliao_674904/1179_674909/202310/t20231020_11164831.shtml。

^② “ASEAN Guide on AI Governance and Ethics”，https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf。

驭人工智能的专业人才；要重新设计工作任务、帮助受影响的人群转向更高价值的工作；还要普及人工智能知识、提高大众的技术素养，助其游刃有余地拥抱人工智能时代。

2.3 主张多元主体、全链条的治理结构

人工智能技术和产品具有广泛渗透性，且在设计、研发、部署、使用等环节蕴含数据安全、算法安全、网络安全、认知安全等多重风险。中国与东盟均提倡多元主体的人工智能治理结构，并对人工智能系统进行全链条、多种形式的监管。

中国在《人工智能安全治理框架》中提出，人工智能安全治理应构建各方共同参与、技管结合、分工协作的治理机制，建立完善技术研发机构、服务提供者、用户、政府部门、行业协会、社会组织等多方参与的综合治理制度规范，建立健全人工智能安全宣传教育、行业自律、社会监督等机制。针对人工智能技术开发、部署、使用等不同环节，实施灵活多样的治理方式，如制定数据收集和使用规则、安全开发规范、伦理审查准则、风险等级测试评估体系等各类规范，构建安全风险威胁信息共享和应急处置机制，推进人工智能可解释性研究、加大人工智能安全人才培养力度等。

《东盟人工智能治理与伦理指南》提出设立东盟人工智能治理工作组，由来自东盟各国、多学科背景的专家组成，推动和监督东盟地区的人工智能治理工作。另由专门机构制定各类标准、指南、工具和模板等，帮助企业和机构负责任地设计、开发和部署人工智能，并以法律、法规等形式明确参与人工智能系统设计、开发、部署的人员的责任。该《指南》建议发展公私合作伙伴关系，对人工智能系统的全生命周期进行管理，包括立项和风险界定、数据收集和处理、建模、成品分析、部署和监管等。

中国与东盟人工智能治理理念相近，双方就人工智能治理的诸多问题取得共识的前景可期。中国与东盟都主张不同社会制度、历史文



化和发展水平的国家平等相待、和平合作、对话解决分歧，因而更易在相互尊重和理解的基础上展开沟通，共同探讨人工智能发展和治理路径。中国和东盟均重视人工智能国际治理机制的包容性和不同发展水平国家的代表性问题。中国和东盟国家的文化传统和思维习惯中针对特定种族或宗教的偏见较少，积极推动人工智能系统更好服务于多元文化，避免引发歧视、煽动仇恨的观念深入人心。中国与东盟都秉持共同繁荣和可持续发展的价值理念，在人工智能治理合作中，均注重引导人工智能技术朝着有利于经济、社会和环境可持续发展的方向发展，且重视通过技术培训和能力建设等项目，促进人工智能技术资源和治理经验的共享，推动区域内人工智能能力和产业的均衡发展。

06

展望

- 推进基础设施发展，构筑智能转型基础
 - 加快技术应用，共建创新生态
- 加强治理对话与合作，筑牢安全和伦理屏障

06

展望

人工智能是人类发展新领域。当前，全球人工智能技术快速发展，给世界带来巨大发展机遇的同时也蕴含许多风险挑战。中国和东盟在人工智能领域有共同的发展意愿和相近的治理理念，应该加强合作，增加在各现有合作机制下关于人工智能问题的讨论与协调，并通过建立人工智能对话机制、举办研讨会等形式多样的交流活动，共同把握人工智能发展新机遇，携手应对新技术迅速应用带来的新挑战，合力推动建立包容和可持续的人工智能生态系统、分享人工智能技术进步带来的发展红利。中国和东盟可在相互尊重的基础上，基于各方意愿、能力和需求，采取以下举措推进人工智能发展与治理合作：

（一）推进基础设施发展，构筑智能转型基础

加强发展战略对接和政策交流。持续强化人工智能战略对接，深化政策制定者对数字基础设施支撑人工智能发展重要作用的理解。推动政策制定者加强投入和政策引导，制定中长期数字基础设施发展规

划，适度超前投资，为人工智能发展提供坚实的基础和健康的土壤。鼓励营造良好的市场环境，吸引更多企业参与数字基础设施的建设和应用发展，开发多种融资组合方式，为各方人工智能技术和产业的发展提供支持。因国施策、因地制宜。对于联接基础较弱的国家，应优先发展网络联接，弥合基础设施鸿沟。对于联接基础较好的国家，则可加快部署下一代网络基础设施。

持续推动网络覆盖升级。投资发展宽带网络，以确保广泛的互联网接入。应提升网络基础设施的覆盖率、可接入性和可负担性。优化网络容量，实现 4G 覆盖，并在有条件的区域推进 5G 部署。加强电力保障，确保数字基础设施具有稳定可靠的电力供应。

加快算力基础设施合作。鼓励双方产业界合作共建数据中心，为人工智能开发和应用提供充足、便利的算力服务，满足人工智能产业发展需求。制定兼容、互惠的数据中心和云计算监管政策，持续扩大相互投资。鉴于人工智能时代的算力需求远超以往，电力消耗与碳排放与日俱增，应鼓励围绕算力优化配置、产业技术创新、清洁能源等方向开展合作，推动算力基础设施可持续发展。

（二）加快技术应用，共建创新生态

共建开放、包容和可持续的人工智能生态系统。鼓励各方人工智能创新创业群体相互开放，生态加速融合。支持企业、高校和科研机构加强对接，开展人工智能技术研发合作，共建小语种语料库，开发基于本土语言的大模型。共同搭建人工智能应用服务平台，鼓励产业链不同环节企业通过提供基础模型和能力支持、对大模型进行调优等方式，合作开发符合产业需求、直击用户痛点的智能应用，挖掘垂直领域潜能，打造丰富的人工智能创新生态。

联合开展人工智能应用场景示范，更好赋能经济社会发展。建议联合向中国和东盟企业征集人工智能应用案例集，积极分享人工智能

在农业、教育、医疗、交通、智能制造等领域的优秀应用案例，激发、联动中国和东盟地区对人工智能产品和服务的市场需求。为帮助各方更好地了解人工智能技术及应用，建议中国和东盟选定一组城市，轮流举办人工智能产品应用展，也可考虑开设云展览，便利各国企业开展交流合作。

联合培育人工智能人才，加强能力建设。人才匮乏是制约人工智能产业发展的重要因素。鼓励中国和东盟的头部企业帮助区域内相对欠发达的地区培养人才，开展职业技能等级认证，开发大数据、物联网、人工智能等新兴 ICT 技术产业相关课程。企业是使用人工智能的关键行动者，建议邀请相对欠发达地区的企业高管到头部企业、科研机构等参观或进行短期培训，了解技术发展和应用情况，加深对人工智能如何提升企业效率、改变生产形态、变革组织架构的理解，进而推动企业智能化转型。

（三）加强治理对话与合作，筑牢安全和伦理屏障

加大数字基础设施的安全保障治理。人工智能模型的运行高度依赖数据和互联网，因此确保数据和网络的安全性与可持续性至关重要。中国与东盟可在数据中心的安全防护、数据加密和访问控制等关键领域展开合作，进一步加强保障措施。数据中心对稳定电力和网络有很高要求，可推动在东盟的供能网络中增加对数据中心的保障，提高能源利用效率，并建立完善的灾难恢复计划以确保业务连续性。网络安全方面，面对日益增多的跨国网络攻击、数据泄露和勒索软件威胁，双方还可共建安全标准框架，通过设立联合网络安全工作组，定期对网络安全标准、法规和合规要求进行审查和更新。双方还可在各自的互联网应急中心基础上建立合作机制，以快速响应跨境的网络紧急事件。

加强双方数据流动安全与跨境合作。中国和东盟可通过采取便利

数据跨境流动的措施，实现更高的数据利用效率，促进各行业的数字化转型，推动数字经济快速发展。双方可以《区域全面经济伙伴关系》（RCEP）、《促进和规范数据跨境流动规定》以及即将推出的《东盟数字经济框架协议》（DEFA）为讨论基础，进一步加强数据流动规则和安全标准协调。基于对数字主权的尊重，双方可商讨建立跨境数据流动的安全保障机制，推动数字治理规则的互信与互认，提升数据流动的安全性和便捷性。在此基础上，中国和东盟就可以数据保护法规进行深度交流，联合打击非法收集个人数据、网络诈骗及其它数据滥用和垄断导致的其他违法现象。同时，双方可进一步探索在多方框架下的数据流动治理路径，力求在保障数据安全的同时平衡数据的自由流动需求，构建一个更开放、安全的跨境数据生态系统。^①

推动人工智能安全标准与风险管理框架对接。为确保人工智能的安全应用，中国与东盟可讨论建立人工智能治理方面共识性的安全标准。双方可加强在技术标准、政策协调和技术替代方案等方面的讨论，防止标准割裂对地区整体发展的负面影响，保障产业链和供应链的稳定和可持续。中国与东盟可在全球人工智能治理规则的制定中协同合作，联合提出涵盖安全、隐私和伦理的人工智能国际标准，增强在国际规则制定中的话语权。这些协作将不仅推动双方在标准对接与风险管理上的一致性，还将为建立一个安全、可持续的区域性技术生态系统奠定基础。

深化人工智能伦理规范合作。面对快速发展的技术及其带来的社会伦理风险，中国和东盟应从前瞻性视角出发，推动人工智能伦理规范的制定，为未来智能社会建立坚实的安全基础。双方可成立由跨学

^① 张丽、孙菲阳：《数据跨境流动：全球治理趋势与我国规制策略》，电子工业出版社 2022 年版，第 282 页。

科专家组成的专家小组，提前识别并应对人工智能技术可能带来的社会和伦理挑战。为缓解大规模失业等潜在社会风险，双方可共同推动人工智能技能教育，帮助劳动者适应技术带来的就业转型。此外，双方可就数智化风险社会、医疗人工智能、深度合成技术、计算法学和人机共治等前瞻性问题的展开探讨，并结合东亚和东南亚的文化视角，在人工智能治理中引入人文主义立场，保障人类在技术发展中的主体地位。^①通过建立前置风险评估机制和公众反馈机制，实现从“灭火”到“防火”的治理转型，确保人工智能技术始终以造福人类为目标健康发展。

^① 刘正毅、梁正、郑烨婕：《黑镜与秩序 --- 数智化风险社会下的人工智能伦理与治理》，清华大学出版社，2022 年版，第 60-61 页。



07

附件

- 附件 1 全球人工智能治理倡议
- 附件 2 人工智能全球治理上海宣言
- 附件 3 人工智能能力建设普惠计划

附件 1

全球人工智能治理倡议^①

人工智能是人类发展新领域。当前，全球人工智能技术快速发展，对经济社会发展和人类文明进步产生深远影响，给世界带来巨大机遇。与此同时，人工智能技术也带来难以预知的各种风险和复杂挑战。人工智能治理攸关全人类命运，是世界各国面临的共同课题。

在世界和平与发展面临多元挑战的背景下，各国应秉持共同、综合、合作、可持续的安全观，坚持发展和安全并重的原则，通过对话与合作凝聚共识，构建开放、公正、有效的治理机制，促进人工智能技术造福于人类，推动构建人类命运共同体。

我们重申，各国应在人工智能治理中加强信息交流和技术合作，共同做好风险防范，形成具有广泛共识的人工智能治理框架和标准规范，不断提升人工智能技术的安全性、可靠性、可控性、公平性。我们欢迎各国政府、国际组织、企业、科研院校、民间机构和公民个人等各主体秉持共商共建共享的理念，协力共同促进人工智能治理。

为此，我们倡议：

——发展人工智能应坚持“以人为本”理念，以增进人类共同福祉为目标，以保障社会安全、尊重人类权益为前提，确保人工智能始终朝着有利于人类文明进步的方向发展。积极支持以人工智能助力可持续发展，应对气候变化、生物多样性保护等全球性挑战。

^① 习近平主席在 2023 年 10 月召开的第三届“一带一路”国际合作高峰论坛上发布。

——面向他国提供人工智能产品和服务时，应尊重他国主权，严格遵守他国法律，接受他国法律管辖。反对利用人工智能技术优势操纵舆论、传播虚假信息，干涉他国内政、社会制度及社会秩序，危害他国主权。

——发展人工智能应坚持“智能向善”的宗旨，遵守适用的国际法，符合和平、发展、公平、正义、民主、自由的全人类共同价值，共同防范和打击恐怖主义、极端势力和跨国有组织犯罪集团对人工智能技术的恶用滥用。各国尤其是大国对在军事领域研发和使用人工智能技术应该采取慎重负责的态度。

——发展人工智能应坚持相互尊重、平等互利的原则，各国无论大小、强弱，无论社会制度如何，都有平等发展和利用人工智能的权利。鼓励全球共同推动人工智能健康发展，共享人工智能知识成果，开源人工智能技术。反对以意识形态划线或构建排他性集团，恶意阻挠他国人工智能发展。反对利用技术垄断和单边强制措施制造发展壁垒，恶意阻断全球人工智能供应链。

——推动建立风险等级测试评估体系，实施敏捷治理，分类分级管理，快速有效响应。研发主体不断提高人工智能可解释性和可预测性，提升数据真实性和准确性，确保人工智能始终处于人类控制之下，打造可审核、可监督、可追溯、可信赖的人工智能技术。

——逐步建立健全法律和规章制度，保障人工智能研发和应用中的个人隐私与数据安全，反对窃取、篡改、泄露和其他非法收集利用个人信息的行为。

——坚持公平性和非歧视性原则，避免在数据获取、算法设计、技术开发、产品研发与应用过程中，产生针对不同或特定民族、信仰、国别、性别等偏见和歧视。

——坚持伦理先行，建立并完善人工智能伦理准则、规范及问责机制，形成人工智能伦理指南，建立科技伦理审查和监管制度，明确

人工智能相关主体的责任和权力边界，充分尊重并保障各群体合法权益，及时回应国内和国际相关伦理关切。

——坚持广泛参与、协商一致、循序渐进的原则，密切跟踪技术发展形势，开展风险评估和政策沟通，分享最佳操作实践。在此基础上，通过对话与合作，在充分尊重各国政策和实践差异性基础上，推动多利益攸关方积极参与，在国际人工智能治理领域形成广泛共识。

——积极发展用于人工智能治理的相关技术开发与应用，支持以人工智能技术防范人工智能风险，提高人工智能治理的技术能力。

——增强发展中国家在人工智能全球治理中的代表性和发言权，确保各国人工智能发展与治理的权利平等、机会平等、规则平等，开展面向发展中国家的国际合作与援助，不断弥合智能鸿沟和治理能力差距。积极支持在联合国框架下讨论成立国际人工智能治理机构，协调国际人工智能发展、安全与治理重大问题。

附件 2

人工智能全球治理上海宣言^①

我们深感人工智能对世界的深远影响和巨大潜力，认识到人工智能正在引领一场科技革命，深刻影响人类生产生活。随着人工智能技术快速发展，我们也面临前所未有的挑战，特别是在安全和伦理方面。

我们强调共同促进人工智能技术发展和应用的必要性，同时确保其发展过程中的安全性、可靠性、可控性和公平性，促进人工智能技术赋能人类社会。我们相信，只有在全球范围内的合作与努力下，我们才能充分发挥人工智能的潜力，为人类带来更大的福祉。

一、促进人工智能发展

我们愿积极推进研发，释放人工智能在医疗、教育、交通、农业、工业、文化、生态等各领域的应用潜力。鼓励创新思维，支持跨学科研究合作，共同推动人工智能技术突破与向善发展。共同关注和缓解人工智能对就业的影响，引导和促进人工智能赋能人类工作的质量与效率的提升。

倡导开放与共享的精神，推动全球人工智能研究资源的交流与合作。建立合作平台，促进技术转移与成果转化，推动人工智能基础设施公平分配，避免技术壁垒，共同提升全球人工智能的

^① 2024 世界人工智能大会暨人工智能全球治理高级别会议发表。

发展水平。

以高水平数据安全保障高质量数据发展，推动数据的依法有序自由流动，反对歧视性、排他性的数据训练，合作打造高质量数据集，为人工智能发展注入更多养料。

建立合作机制，大力推进人工智能赋能各行各业，率先在制造、物流、采矿等领域加速智能化，同步推进相关技术、标准的共用共享。

致力于培养更多的人工智能专业人才，加强教育培训与人才交流合作，提高全球范围内人工智能素养与技能水平。

呼吁各国秉持以人为本、智能向善原则，确保各国在开发和利用人工智能技术方面权利平等、机会平等、规则平等，不受任何形式的歧视。

尊重各国自主发展的权利，鼓励各国根据自身国情制定人工智能战略政策和法律法规，呼吁在人工智能技术、产品和应用的国际合作中，遵守产品与服务对象国的法律法规，遵守适用的国际法，尊重其经济社会制度、宗教文化传统和价值理念。

二、维护人工智能安全

我们高度重视人工智能的安全问题，特别是数据安全与隐私保护，愿推动制定数据保护规则，加强各国数据与信息保护政策的互操作性，确保个人信息的保护与合法使用。

我们认识到加强监管，打造可审核、可监督、可追溯和可信赖的人工智能技术的必要性。我们将以发展的眼光看问题，在人类决策与监管下，以人工智能技术防范人工智能风险，提高人工智能治理的技术能力。我们鼓励各国结合国情，制定相应的法律和规范，建立风险等级测试评估体系和科技伦理审查制度，在此基础上，鼓励行业制定更为及时和敏捷的自律规范。

我们愿加强人工智能相关的网络安全，增强系统与应用的真实性与可靠性，防止黑客攻击与恶意软件应用。在尊重运用国际国内法律框架前提下，共同打击操纵舆论、编造与传播虚假信息的行为。

合作防范恐怖主义、极端势力和跨国有组织犯罪集团利用人工智能技术从事非法活动，共同打击窃取、篡改、泄露和非法收集利用个人信息的行为。

推动制定和采纳具有广泛国际共识的人工智能的伦理指南与规范，引导人工智能技术的健康发展，防止其被误用、滥用或恶用。

三、构建人工智能治理体系

我们倡导建立全球范围内的人工智能治理机制，支持联合国发挥主渠道作用，欢迎加强南北合作和南南合作，提升发展中国家的代表性和发言权。我们鼓励国际组织、企业、研究机构、社会组织和个人等多元主体积极发挥与自身角色相匹配的作用，参与人工智能治理体系的构建和实施。

我们愿加强与国际组织、专业机构等合作，分享人工智能的测试、评估、认证与监管政策实践，确保人工智能技术的安全可控可靠。

加强人工智能的监管与问责机制，确保人工智能技术的合规使用与责任追究。

四、加强社会参与和提升公众素养

我们愿建立多元化的参与机制，包括公开征求意见、社会调查等方式，让公众参与到人工智能决策过程。

加强公众对人工智能的认知与理解，提高公众对人工智能安全性

的认识。开展科普活动，普及人工智能知识，提升公众的数字素养与安全意识。

五、提升生活品质与社会福祉

我们愿积极推动人工智能在可持续发展领域的应用，包括工业创新、环境保护、资源利用、能源管理、促进生物多样性等。鼓励创新思维，探索人工智能技术在解决全球性问题中的潜力与贡献。

致力于利用人工智能提升社会福祉，特别是在医疗、教育、养老等领域。

我们深知，这份宣言的落实需要我们共同努力。我们期待全球各国政府、科技界、产业界等利益攸关方能够积极响应，共同推动人工智能的健康发展，共同维护人工智能的安全，赋能人类共同的未来。

附件 3

人工智能能力建设普惠计划^①

为弥合数字和智能鸿沟，特别是帮助全球南方在人工智能发展进程中平等受益，中方认为要坚持联合国在国际发展合作中的统筹协调作用，坚持真正多边主义，基于主权平等、发展导向、以人为本、普惠包容、协同合作原则，通过南北合作、南南合作和三方合作等形式，切实落实联大加强人工智能能力建设国际合作决议（A/RES/78/311），推动落实联合国 2030 年可持续发展议程。为此，中国提出“人工智能能力建设普惠计划”，并呼吁各方对人工智能能力建设加大投入。

一、愿景目标

（一）促进人工智能和数字基础设施联通

完善全球可互操作的人工智能和数字基础设施布局，积极协助各国特别是全球南方发展人工智能技术和服务，助力全球南方真正接触到人工智能，跟上人工智能发展的步伐。

（二）推进“人工智能+”赋能千行百业

探索推进人工智能全方位全链条多场景赋能实体经济，推进人工

^① 中方在人工智能能力建设国际合作高级别会议上发布。

智能赋能工业制造、传统农业、绿色转型发展、气候变化应对、生物多样性保护等应用，因地制宜推动构建丰富多样、健康向善的人工智能发展生态。

（三）加强人工智能素养和人才培养

积极推动人工智能在教育中的广泛应用，开展人工智能人才培养和交流，加大分享通用专业知识和最佳实践，培育公众人工智能素养，保障和强化妇女和儿童数字和智能权益，共享人工智能知识成果和经验。

（四）提升人工智能数据安全和多样性

合作推动数据依法有序自由跨境流动，探索构建数据共享的全球性机制平台，维护个人隐私和数据安全。推动人工智能数据语料库平等多样，消除种族主义、歧视和其他形式的算法偏见，促进、保护和保全文明多样性。

（五）确保人工智能安全可靠可控

坚持公平和非歧视原则，支持在联合国框架下建立兼顾发展中国家利益的全球可互操作的人工智能安全风险评估框架、标准和治理体系。共同研判人工智能研发与应用风险，积极推进和完善应对人工智能安全风险的技术和政策，确保人工智能设计、研发、使用和应用促进人类福祉。

二、中国行动

（一）中方愿同所有国家开展人工智能领域南北合作、南南合作和三方合作，共同落实联合国未来峰会成果，积极同各国特别是发展

中国开展人工智能基础设施建设合作，共建联合实验室。

（二）中方愿开展人工智能模型研发和赋能合作，特别是推进人工智能赋能减贫、医疗、农业、教育和工业制造等，深化人工智能供应链国际合作，释放人工智能作为新质生产力的红利。

（三）中方愿同各国特别是发展中国家共同挖掘人工智能赋能绿色发展、气候变化应对、生物多样性保护等潜力，助力全球气候治理和可持续发展。

（四）中方愿搭建人工智能能力建设国际合作平台，中方人工智能产业界和产业联盟愿同各国特别是发展中国家开展多种形式的交流活动，共享最佳实践，以负责任态度共建人工智能开源开放社区，推动构建多层次多业态合作生态。

（五）中国政府将面向发展中国家举办人工智能能力建设中短期教育培训，共享人工智能教育资源，开展人工智能联合办学和访问交流等活动，助力发展中国家培养高水平人工智能科技与应用人才。

（六）中国政府愿加强同发展中国家的人力资源援助合作，在今年举办首届人工智能能力建设研讨班的基础上，将于 2025 年底前重点面向发展中国家再举办十期人工智能领域研修研讨项目。

（七）中方愿同各国特别是发展中国家共同培育公众人工智能素养，以线上线下相结合方式，多维度、多层次、多载体推广人工智能科普和专业知识，努力提高各自人民的人工智能素养和技能水平，特别是保障和提高妇女和儿童数字权益。

（八）中方愿同各国特别是发展中国家共同开展人工智能语料建设，采取积极举措消除种族、算法、文化歧视等，致力于维护并促进语言和文明多样性。

（九）中方愿同各国特别是发展中国家促进和完善数据基础设施，共同促进全球数据公平普惠利用。

（十）中方愿同各国特别是发展中国家共同加强人工智能战略对接和政策交流，积极分享在人工智能测试、评估、认证与监管方面的政策与技术实践，携手应对人工智能伦理与安全风险。



使用再循环纸张印刷



中国国际问题研究院

网址：<https://www.ciis.org.cn>

地址：北京市东城区台基厂头条 3 号院