CAICT 中国信通院

集智 蓝皮书 No.202516

人工智能产业发展研究报告

（2025 年）

中国信息通信研究院

2026年1月

版权声明

前　言

2025 年，全球人工智能飞速发展，技术、应用、生态协同共振，重塑开发范式、改变人机交互模式，催生更多个体与行业智能化应用，逐步实现从“有能力”走向“有用处”，人工智能与经济社会的融合正从浅入深加速推进。

过去一年来，人工智能发展是系统性的，模型能力正在经历与真实世界互动中自主迭代的技术跃升，智能体与生产生活各领域结合成为应用重要形态，具身智能发展有望使智能走出比特世界，AI 正探索从工具走向伙伴、从单点走向系统的升级路径。**技术方面，**基础超级模型在核心能力上实现全方位提升，强化学习推动模型在真实环境中持续进化，原生多模态架构逐渐走向成熟。高度封装的智能体产品加速模型从感知认知向自主决策执行演进。同时，智算基础设施进入“吉瓦级”竞争新阶段，开源开放的智算生态加速形成。数据工程从规模堆砌转向质量优先，为模型能力提供优质“燃料”。**应用方面，**人工智能加速在一二三产业融合渗透，助力农业降本增产，加速推进赋能新型工业化进程，深化服务场景智能化新业态。智能原生成为“AI+”新内核，大模型深度嵌入工具软件，代码编写、深度研究等智能原生软件加速数字生产力跃升，新一代智能终端等硬件产品重构人机交互模式，人工智能角色从“智能工具”逐步演变为“共生伙伴”。**生态和治理方面，**开源生态成为技术普惠的关键力量，全球 AI 标准竞争加剧，中国纵深推进体系建设，基准

测试体系随技术演进持续升级。安全治理从理论探索走向实践落地，面对新型威胁，技术防线与监管框架同步完善。**国际合作方面**展现开放包容态势，人工智能成为多边机制核心议题，创新资源从分散走向有限共享，“开源生态+本地化拓展”助力人工智能作为国际公共产品普惠全球。

展望未来，迈向通用人工智能的道路，可能会经历若干不确定的“奇点”。从短期来看，技术层面大模型将持续优化推理效率，探索世界模型成为关键路径，以突破物理图灵测试为目标探索具身智能本体与模型协同。智算生态加速开放协同，安全治理构建动态防护体系。面向中远期，人工智能从工具赋能走向系统重构，带来更大范围转型升级，我们将不仅见证效率的飞跃，更将一同探索人类与 AI 以新型的信任与协同关系塑造智能世界的新形态。

在此背景下，我院发布《人工智能产业发展研究报告（2025 年）》，旨在探讨近期人工智能技术创新方向、产业升级重点、行业落地趋势和安全治理进展，展望人工智能发展机遇，以期与业界分享，共同推动人工智能生态蓬勃发展。由于人工智能发展日新月异，限于编写时间、编写组知识积累水平有限等因素，报告中存在不足之处，敬请大家批评指正。

目 录

图 目 录

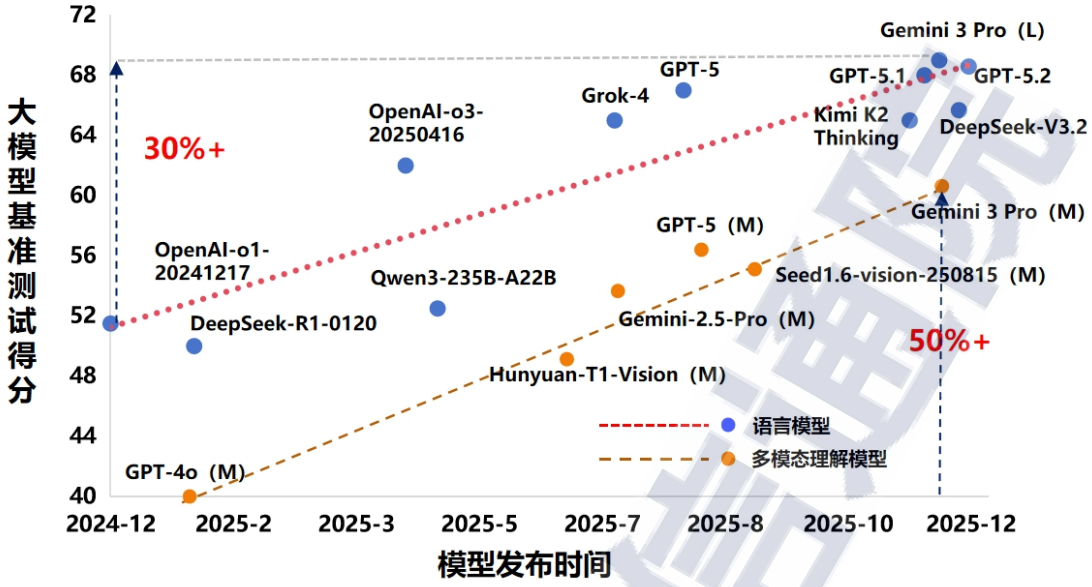
表 目 录

全球人工智能发展日新月异，技术快速迭代，大模型综合性能显著提升，应用门槛与使用成本持续降低，不断推进实用化进程。我国高度重视人工智能发展，部分关键技术取得重要进展，产业生态日趋繁荣。据中国信息通信研究院测算，2024 年我国人工智能核心产业规模已突破 9000 亿元，同比增长 24%；2025 年预计突破 1.2 万亿元。截至 2025 年底，我国人工智能企业数量超 6000 家，形成覆盖基础底座、模型框架、行业应用的完整产业体系。当前及未来一段时间，随着智能体、具身智能等快速发展，人工智能将加速从“能思考”向“能实干”转变，为千行百业开拓赋能新空间。

一、技术产业发展

（一）基础超级模型持续突破，模型学习进入经验时代

2025 年，全球人工智能领域迎来关键突破期，语言基础模型与推理模型双线并进，技术迭代速度与能力跃升幅度均超出预期。从中国信通院“方升”测试数据来看，截止 2025 年 12 月，以 GPT-5.2、Gemini 3 Pro、Grok-4、DeepSeek V3.2、Claude 4.5、Qwen3、Doubao-Seed-1.6、Kimi K2 Thinking、GLM4.5 等为代表的头部语言大模型的综合能力较 2024 年底提升 30%，多模态理解能力超过 50%，大模型基础能力实现跨越式提升。

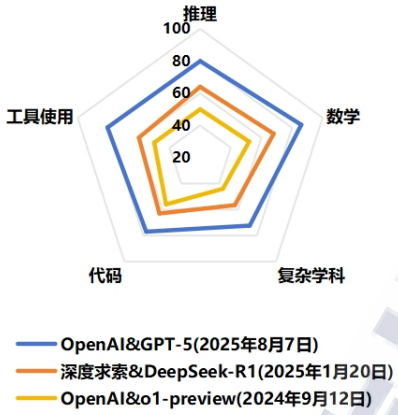


来源：中国信息通信研究院

图 1 中国信息通信研究院“方升”大模型基准测试结果

语言基础超级模型深度集成多种能力，突破模型能力边际。语言基础模型不再局限于单一功能场景，而是朝着多功能融合的方向加速演进。以 GPT-5、DeepSeek 3.2、Qwen3 为代表的“基础超级模型”深度集成推理、代码及智能体核心能力，正式登上历史舞台。其主要特征表现为：一是思考+非思考模式合一，根据用户提示词自主选择是否采用推理模式；二是理解、推理、数学能力大幅提升；三是内置代码、工具调用等多种 AGENT 能力。其中，模型的推理能力成为重点提升方向，是衡量模型“聪明”与否的重要特征。谷歌最新推出的 Gemini3 不仅延续了多功能融合的趋势，更通过架构升级，在推理深度、智能体持续性及生态集成度上树立了新的技术标杆。其新增的“Deep Think”增强推理模式，展现出解决极端复杂

问题的潜力。



来源：中国信息通信研究院

图 2 大模型在推理、学习、工具使用等复杂维度上的表现

多模态大模型深度融合理解和生成能力，原生多模态架构逐渐走向成熟。上半年发布的 OpenAI-o3 模型和 Gemini 2.5 Pro 模型可以同时处理视觉和语言信息，支持自动放大或缩小图像中的关键细节，将语言大模型的深度思考能力成功迁移至多模态模型，使其具备复杂视觉任务的深度推理能力。GPT-Image-1 和 Gemini 2.0 flash 模型通过原生多模态架构融合自回归模型及扩散模型，支持长上下文的图像在线生成和编辑，为原生多模态技术从探索走向规模化商业应用奠定了重要基础。下半年谷歌和 OpenAI 先后发布文生图模型 Gemini-2.5-flash-image-preview（nano banana）和文生视频模型 Sora 2，进一步提升视觉生成的内容质量与适配性。

语言及多模态模型的能力演进源自模型的运行机制、基础架构、训练方法三个维度的关键突破。

在运行机制上，短期来看 Scaling 难以带来质的改变，精细化机

制优化成为提升方向。月之暗面的 Kimi K2、阿里巴巴的 Qwen3-Max-Preview、蚂蚁集团的 Ling-1T 的模型参数量均达到万亿，但未能出现如 OpenAI o1 模型的巨大能力突破。GPT-5 等大模型采用路由机制实现了语言模型能力的动态调度，可根据任务需求自动激活模型的思考或非思考能力，大幅提升运行效率[1]。LLaDA、Dream 7B 等扩散语言模型通过并行计算同时预测多个 token，减少累积误差，提升语义的连贯性。虽当前扩散语言模型的能力与传统自回归的模型比仍有差距，但已在效率上体现出明显优势，谷歌发布的 Gemini Diffusion 可在 12 秒内可生成 1 万 tokens。

在基础架构上，Transformer 架构性能优化仍是当前研究重点，以 DeepSeek 的 NSA、月之暗面的 MoBA、面壁智能的 InfLLM-V2 为代表的稀疏注意力机制成为模型“降本增效”常用手段。Mamba 3、RWKV-8 等非 Transformer 新架构仍未得到广泛应用，而将上述新架构与 Transformer 进行融合的混合模型（如 Qwen3-Next、Hunyuan-TurboS、MiniMax M1、英伟达 Nemotron-H 等）已取得成功商业应用，其兼具 Transformer 架构的并行计算优势与线性注意力的长序列处理能力。而在多模态模型方面，自回归模型与扩散模型的创新融合架构则为语言-视觉跨模态协同提供了全新技术路径，使多模态模型同时具备理解和生成能力，支持长上下文的在线生成和

[1] Y Zhang, H Li, J Chen, H Zhang, P Ye, L Bai, S Hu. Beyond gpt-5: Making llms cheaper and better via performance-efficiency optimized routing. arxiv:2509.02547, 2025

编辑，生成的图像和视频更贴合客观物理规律[2]。

在训练方法上，大模型正经历从“人类数据时代”向“经验时代”的根本转变，不再单纯依赖人工标注的静态数据输入，而是通过与环境交互、任务试错和自我迭代积累动态经验，能够自主提炼跨场景规律、修正认知偏差，甚至生成超出原始训练数据范畴的创新解决方案，让模型能力从“复刻已知”走向“探索未知”。当前，DeepSeek 提出的 GRPO 算法已经成为大模型强化学习采用最多的方法，在其基础上进一步优化的变体如 DAPO、VAPO、SRPO、GFPO 等通过引入多阶段策略更新与混合奖励机制，显著提升了模型在复杂任务环境中的泛化能力和决策稳定性，已在代码生成、自动规划等场景实现端到端性能突破。以 ToolRL、Agent Lightning、Verlog 为代表的 Agentic RL 技术通过真实动态环境下的多轮交互式训练，赋予模型自主规划、调用工具、因果推断能力，提升模型在真实应用场景中的表现[3]。

在真实环境中进行“经验学习”推动了具备自主进化能力的模型和智能体的发展，以 AlphaEvolve、达尔文-哥德尔机、OpenEvolve 为代表的自主进化框架成功展示了智能体与真实环境实时交互、试错反馈积累认知的“经验学习”过程。自主进化的智能系统需要重点解决如何在没有外部奖励的情况下根据自身经验中学习成长，美

[2] X Zhang, J Guo, S Zhao, et al. Unified Multimodal Understanding and Generation Models: Advances, Challenges, and Opportunities. arXiv:2505.02567, 2025.
[3] G Zhang, H Geng, X Yu, et al. The Landscape of Agentic Reinforcement Learning for LLMs: A Survey.arxiv:2509.02547, 2025.

国 Meta 公司提出了名为“早期经验”（Early Experience）的“中训练”范式，通过隐式的世界建模和自我反思策略将“自身经验”转化为可训练的监督信号，从而得到更具泛化能力的语言模型或智能体[4]。此外，自主进化的智能系统还需要具有长期记忆能力，MemGen、ReasoningBank、Memobase 等创新记忆框架，从智能系统的历史运行轨迹中提炼重要记忆项进行存储，为后续在复杂场景下的决策校准、过往经验复用以及持续的自主进化进程，提供了稳定且可调用的核心信息基础。

从大模型产业发展来看，基础模型数量持续收敛，真实场景中的应用效果成为关注重点。一是基础模型研发厂商数量呈现收敛趋势，基模研发活跃且持续发布产品的企业已经显著减少。二是全球模型能力在基础语言理解、文本生成及多语言处理等方面已经具备较高的水平，但其普遍存在复杂推理能力不足、不可靠（幻觉）等问题。三是大模型在真实应用场景中效果成为产业界关注的重点，智能客服、智能编码、数据分析、检索增强生成（RAG）等场景是模型落地重要的“试金石”。

（二）集群规模向百万卡迈进，开放智算生态快速发展

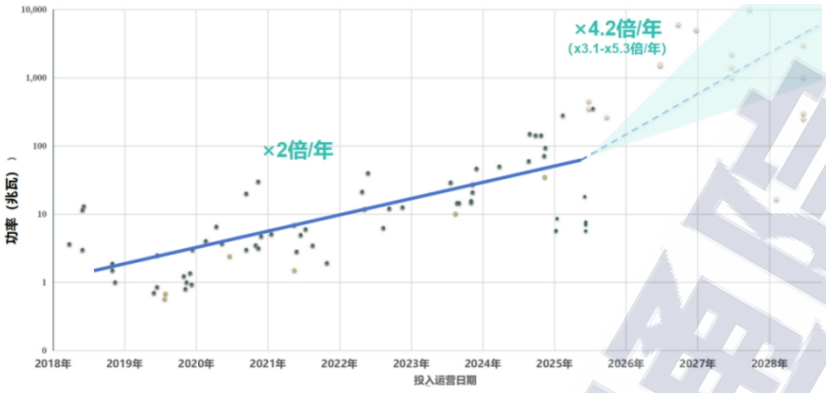
当前，人工智能技术大规模应用促使算力产业呈现爆发式发展。从全球范围看，多个超级集群项目建设和大规模推理算力需求拉动投资快速增长，摩根士丹利预测 2025 年全球 11 大云厂商资本支出

[4] Kai Zhang, Xiangchao Chen, Bo Liu, et al. Agent Learning via Early Experience. arXiv:2510.08558, 2025.

达到 4450 亿美元，年增长率达到 56%。同时，智算生态正在从单一封闭走向开源开放，全球智算产业已进入规模化、协同化发展的新阶段，为大模型前沿创新和普惠落地提供坚实算力底座。

超大规模智算集群持续升级，即将迎来吉瓦级时代。2025 年，美国头部厂商持续加码算力资本支出，推动集群规模从十万卡向百万卡迈进：xAI 公司 Colossus 集群现已按计划部署 23 万张英伟达芯片，同时正在建设 Colossus 2 超级集群，首期规模为 55 万张 GB200 和 GB300 芯片，建成后将是全球首个吉瓦级（GW）智算集群[5]，最终将达到百万卡规模; OpenAI 计划在英伟达支持下建设 10 吉瓦数据中心，等效 400-500 万张芯片，第一阶段预计将于 2026 年下半年部署。除美国外，阿联酋、沙特、印度等国也在规划建立吉瓦级智算集群，意在加强主权 AI 基础设施建设。另一方面，在算力飞速增长、突破电网承载极限的背景下，能源日益成为制约规模提升的关键瓶颈，算力军备竞赛将进入算电协同比拼的“第二战场”争夺。据中国信通院测算，未来三年全球顶尖集群功率或将以每年约 4.2 倍[6]速度快速增长，远超 2019 年以来每年 2 倍的水平。马斯克预测美国将在 2026 年中或年底面临发电量不足挑战，Colossus 2 将采用包含新建变电站、储能、外部电源迁移等方式实现多元化供电，亚马逊、谷歌、微软、英伟达等纷纷加强核聚变、地热、电站建设等电力领域投资，尝试电力私有化部署，积极探索智算集群可持续扩张的能源保障路径。

[5] Colossus 2 集群已于 2026 年 1 月正式运行
[6] 90%置信区间：3.1~5.3 倍/年



来源：中国信息通信研究院、Epoch AI

图 3 全球顶尖集群功率增长趋势预测

超节点成为面向未来超大规模训练的核心计算单元。超节点具备超高互联带宽、内存统一编址等技术特征，将数百上千个 AI 处理器编制为一个逻辑统一的高密度计算体，实质是通过不断提升 Scale up 集成度实现媲美传统 Scale out 的规模性能，**目前呈现三条发展路径：**一是以英伟达为代表的垂直整合模式，从底层芯片到核心互联技术，再到上层的编程模型和软件栈进行端到端控制和优化，实现最佳的性能和效率，典型产品如英伟达 GB300 NVL72。二是以其他芯片巨头、云服务商和大型科技公司支持的协议开放模式，核心战略是以网络层面“兼容性”和“选择权”抗衡“封闭性”，主要参与者包括 AMD、Intel、以及 Meta、微软等超大规模用户，典型产品如 AMD Helio UALoE 72。三是以我国华为、阿里等为代表的全栈化开放模式，正在加速构建从互联协议到基础软硬件全面开放的技术体系，典型产品如华为 Cloudmatrix 384、阿里磐久 128 超节点等。

多样化推理算力形态推动 AI 泛在部署。人工智能算力产品正朝

着多样化、专业化方向快速发展，推动前沿算法能力从数据中心下沉至边缘、终端侧，精准契合不同类型用户需求：**一体机**以软硬协同、开箱即用为特点，为企业用户提供低成本私有化 AI 部署方案，如百度昆仑芯 P800 单机八卡即可运行 DeepSeek-R1 671B 满血版模型；**算力盒子**（个人超算）进一步降低高性能计算门槛，赋能个人开发者与创意工作者，英伟达 Project DIGITS 手掌大小硬件已支持运行 2000 亿参数模型；AI PC、**智能手机、人形机器人等智能终端**快速更新迭代，为广大消费者提供更加高级、智能、便捷的产品体验。多元算力部署形态共同构建了覆盖云、边、端的全栈支持体系，推动 AI 无处不在、随需可用，有效加速智能技术普惠化进程。

“算法-软硬件”协同设计成为大模型研发主流范式，多层次开源开放智算生态体系快速发展。2025 年初，DeepSeek 通过一系列模型架构与软硬件的联合优化，实现模型训练的高性能、高效和低成本共存，证明了我国大模型在资源受限条件下通过软硬协同具备引领式创新的潜力。进一步从全球范围看，谷歌、Meta、OpenAI、字节跳动等国内外全球头部企业均在重点布局探索“算法-软硬件”协同设计方式优化训练策略、提升模型性能。与此同时，覆盖模型（DeepSeek、Qwen 等）、开发框架（Pytorch、飞桨等）、算子库（FlagGem、triton 等）、通信库（FlagCX、DeepEP 等）、网络通信（华为 UB-Mesh、中国移动 OISA 2.0、特斯拉 TTPoE 等）、精度量化（英伟达 NVFP4、DeepSeek UE8M0 等）等软硬件多层次的创新体系快速演进，我国厂商

表现活跃，以开源开放为特征的新型智算生态正在加速形成，有效牵引国产软硬件协同适配，如 DeepSeek、千问、文心一言等优质模型大规模开源开放，为我国算力厂商提供了关键的适配验证场景。中国信通院 AISHPerf 基准 Deepseek 大模型适配测试结果表明，通过软硬件协同优化，部分参测国产芯片部署 DeepSeek R1 模型的精度已基本与国外系统持平（对比 DeepSeek 官方技术报告，基于英伟达 H800 芯片），已能够满足实际产业应用需求。

来源：中国信息通信研究院整理

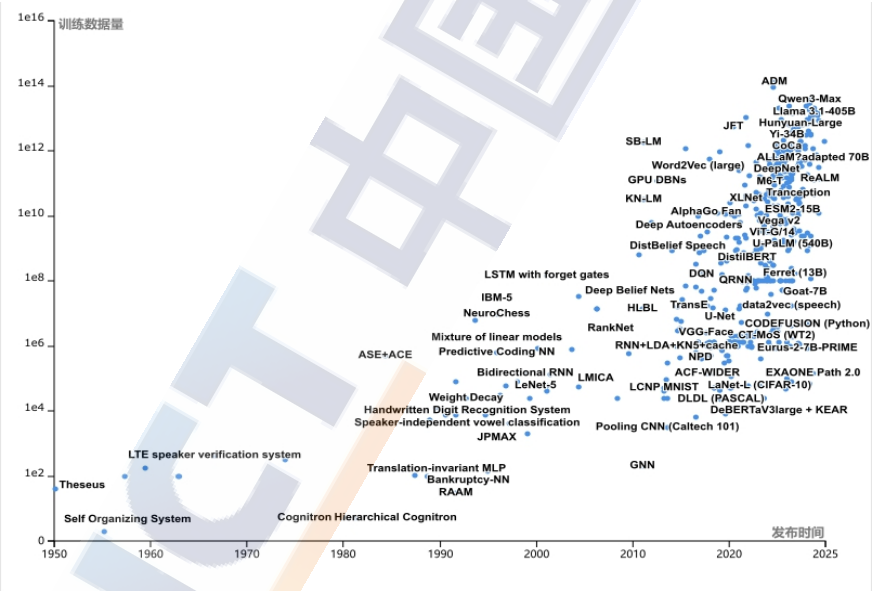
图 4 近一年全球智算领域部分开源开放标志性成果

（三）数据集建设转向适量高质，数据工程体系加速成型

高质量和体系化的数据集是驱动模型能力提升的“燃料”。随着模型训练进入深水区，数据集建设正经历深刻变革。一方面，数据需求持续攀升，但单纯堆砌数据量的方式已难以为继，数据训练密度和利用效率成为新焦点；另一方面，数据集建设重点从追求规模转向质量跃升，智能生成、专业细分、合规治理推动破解数据瓶颈。面向人工

智能的数据工程体系，正成为支撑大模型持续演进的战略性基础设施。

从供需关系来看，大模型训练数据需求总量仍持续增长，但单位数据所能贡献的训练效能呈现见顶回落态势。据 Epoch AI 统计，早期模型受限于算力与数据采集条件，训练数据量仅在 10^2至 10^6量级。2010 年后，伴随 GPU 集群普及与互联网语料爆发，训练数据量迅速突破 10^{10}量级，并持续攀升至 10^{14}级别。近年，代表性通用大模型（如 GPT 系列、Claude、Llama、Gemini）的训练数据规模已逼近 10^{15}级别，表明行业仍以海量数据作为模型泛化能力与生成质量的基础支撑。

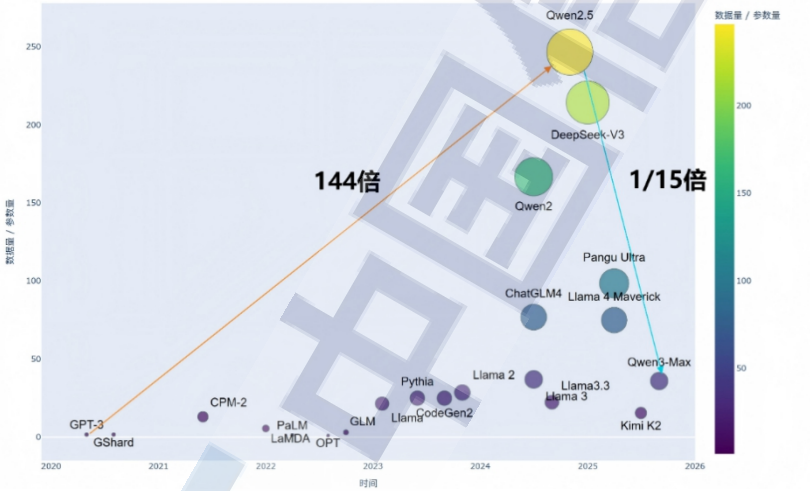


来源：Epoch AI，中国信息通信研究院整理

图 5 全球人工智能模型训练数据量演进趋势（1950—2025）

然而，稠密模型范式下常用的“数据密度”（训练数据量与参数量的比值）指标由快速提升到逐步趋稳。2020 至 2023 年，代表性模型（如 ERNIE-GEN、Qwen2.5-32B 等）数据密度连续上升。2024

年后，该指标出现见顶并回落的迹象。例如，Qwen 2.5 -32B 训练数据约 18 万亿 Token，参数规模约为 32.5B，数据密度约 554；而 Qwen 3-Max 训练数据达 36 万亿 Token，参数规模约 1.0T，数据密度仅约 36，仅为前者 1/15。这表明在部分模型路线中，参数扩张速度已超过训练语料增长速度。尽管该指标具有模型范式差异，但总体来看，行业目前已不再单纯依靠提升数据密度实现性能突破。未来大模型演进，预计将更多依赖高质量数据筛选、训练架构优化及跨模态知识融合。



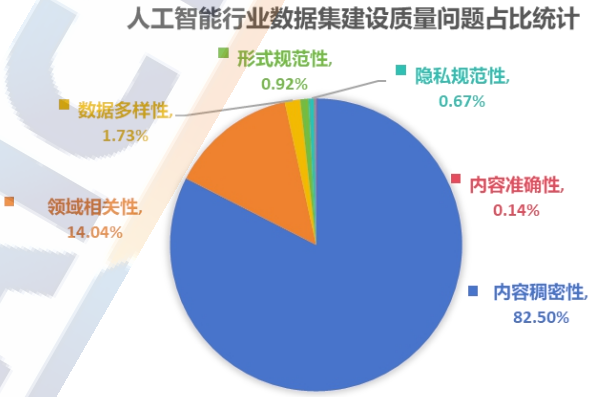
来源：Epoch AI，中国信息通信研究院整理

图 6 全球大模型数据密度演进趋势（2019—2025）

从建设路径来看，数据集构建正向智能生成、多元专业、合规治理三个维度深化，质量优先成为核心导向。在数据采集处理层面，随着推理模型与数据合成技术发展，自动化生成与人机协同标注在构建高质量数据集中发挥更大作用。例如，海天瑞声通过自研标注平台与大模型，构建“标-训-推”一体化数据处理模式，显著提升商品识别效率和个性推荐精准度。**在数据需求对接层面，**行业对数据

的差异化需求凸显，小而精、行业化的数据集快速兴起，推动构建面向医疗、金融、工业等特定行业的高质量数据集。**在数据安全合规层面，**数据集建设更加重视合规性与安全性，强调确权、脱敏、隐私计算和可追溯。2024 年，最高人民法院发布指导案例，明确数据处理者在依法采集、合理利用且不造成损害时不构成侵权，为数据合规提供司法依据。

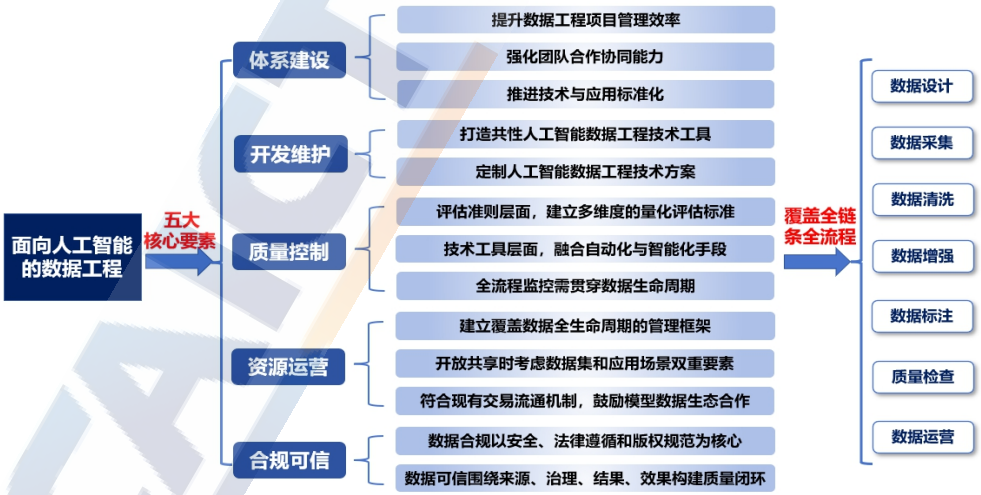
从发展问题来看，数据集质量成为当前制约行业垂类模型落地和场景应用的瓶颈问题。根据中国信通院 ADAQ（人工智能数据质量评估体系）开展的多家央企的评估结果，当前行业数据集建设主要质量问题如下：内容稠密性方面，数据集信息细节、句子成分及关联关系存在缺失，内容重复；领域相关性方面，数据内容与业务场景无关，未体现不同场景的深层业务关系；数据多样性方面，数据集缺少行业特征，数据来源单一、类型简单、场景单调；形式规范性方面，存在格式不符、错别字、多余字符、逻辑错误等形式问题。



来源：中国信息通信研究院

图 7 中国信通院行业高质量数据集评测问题占比

从工程支撑来看，面向人工智能全生命周期的数据工程体系建设提速，保障产业可持续发展。面向人工智能的数据工程由体系建设、开发维护、质量控制、资源运营、合规可信五大核心要素组成，覆盖数据设计、采集、清洗、增强、标注、质检、运营的全链条。**体系建设方面**，加强项目管理与团队协作，推动技术应用标准化，为高质量数据生产提供组织与标准基础；**开发维护方面**，构建数据工程技术体系，打造共性技术工具，定制数据工程技术方案，提高数据处理与管理的效率与可复用性；**质量控制方面**，建立多维度的评估体系，结合自动化、智能化质检手段，实现数据从生产到应用全流程的质量保障；**资源运营方面**，统筹建设数据资源管理架构，推动数据资源高效流转与跨场景复用，鼓励模型数据生态合作；**合规可信方面**，强化数据安全、隐私保护与法律合规要求，围绕来源、治理、结果、效果构建数据可信质量闭环。



来源：中国信息通信研究院

图 8 面向人工智能的数据工程核心要素

（四）工程化能力不断提升，推动向“场景价值闭环”跃迁

“建－用－管”助力全链路工程化体系构建。工程化成为连接“模型－数据－算力”人工智能三大核心要素的关键纽带，为大模型技术的规模化落地应用提供系统性支撑。经过近两年的产业发展和应用实践，大模型工程化路线逐渐清晰，企业通过构“建”大模型平台筑牢大模型落地根基，形成涵盖数据治理、训练提速、推理优化、应用开发的全栈能力；应“用”环节通过模型能力与业务场景深度融合，围绕服务响应效率与用户体验能力提升，拓展大模型应用价值；“管”理维度通过对数据、模型、应用等资产的有效纳管和监控运营，持续发挥大模型赋能价值。

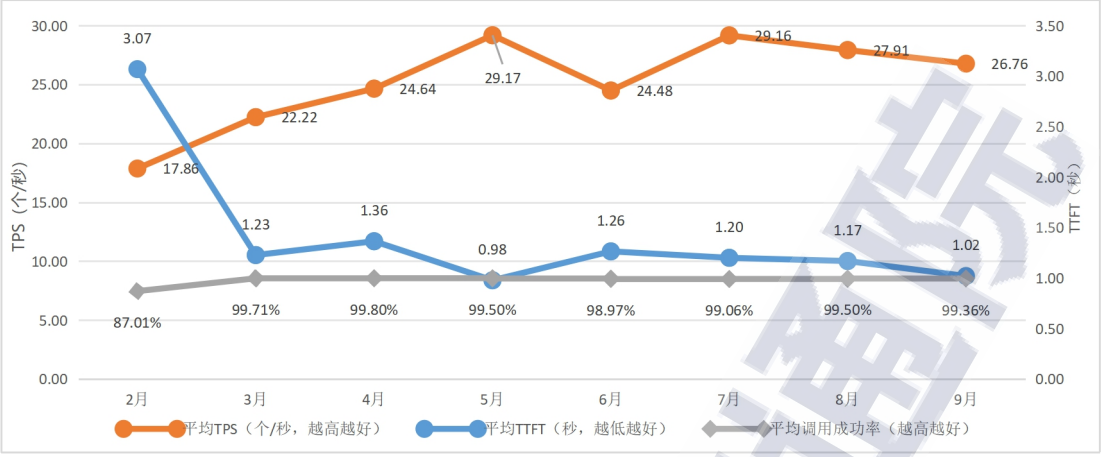
模型即服务（MaaS）作为模型工程化的重要载体，推动大模型从“实验室原型”向“产业级工具”转化。随着大模型在各行业的落地需求急剧增长，MaaS 助力模型工程化落地，已从“可选项”升级为“必备项”。**一是大模型厂商纷纷加大 MaaS 布局，**以阿里云、百度智能云、华为云为代表的国内主流大模型厂商均已推出一站式大模型开发与服务平台，提供从模型训练、推理优化到管理运营的全栈支持，根据中国信通院统计，业界已公开发布的 MaaS 平台已达 100 余个。**二是基于公有云 MaaS 的大模型调用需求呈爆发式增长，**根据中国信通院调研和预测，我国公有云大模型（对客侧[7]）2025

[7] 公有云大模型服务“对客侧”调用量的统计口径为云厂商 MaaS 平台对外部客户提供的服务，不含自有业务

年 Token 调用量有望达到 2000 万亿，相较于 IDC 发布的 2024 年 114 万亿[8]，增长将超过 16 倍。同时我国头部云厂商的大模型整体调用量已处于全球第一梯队水平，例如火山引擎 2025 年 10 月的日均调用量为 30 万亿[9]，谷歌同期日均调用量为 42 万亿[10]。**三是 MaaS 推动大模型服务持续向行业和场景深度融合，**根据中国信通院调研，公有云大模型服务主要应用于文本处理、角色扮演、智能助手、智能搜索、智能编码、营销等场景，其调用量占比超过一半，而在数据安全与合规要求较高的金融、政务、医疗等行业场景中，基于私有化 MaaS 的大模型落地需求在不断释放。

MaaS 平台能力持续提升，大模型服务用户体验持续向好。随着 DeepSeek 的出圈，各大 MaaS 厂商加速平台技术迭代以应对激增的调用需求，通过动态资源调度、推理引擎优化等工程技术手段，持续优化大模型服务的稳定性和可靠性，推动 MaaS 平台从“能用”向“好用”演进。中国信通院监测数据显示，相比 2025 年 2 月份，9 月份国内 11 个各 MaaS 平台[11]上所提供的 DeepSeek-R1 服务，平均调用**成功率**从 87.01%提升至 99.36%，**每秒输出 Token 数（TPS）**从 17.86 提升至 26.76（个/秒），**首 Token 时延（TTFT）**从 3.07 降至 1.02（秒），用户体验持续提升。

[8] 数据来源：国际数据公司（IDC）发布的《中国公有云大模型服务市场格局分析，1Q25》
[9] 数据来源：10 月 16 日，火山引擎总裁在 FORCE LINK AI 创新巡展的演讲
[10] 数据来源：谷歌 DeepMind 产品负责人 Logan Kilpatrick 博客
[11] 对 DeepSeek-R1 监测的 11 个 MaaS 平台包括阿里云百炼、道客云、深度求索、无问芯穹、PPIO、百度云千帆、商汤大装置、硅基流动、腾讯云 TI 平台、火山方舟、讯飞开放平台



来源：中国信息通信研究院、Epoch AI

图 9 2025 年国内 MaaS 平台提供 DeepSeek-R1 服务表现平均值变化

未来 MaaS 平台将从横向拓展能力边界，从纵向提升内在工程化水平，实现大模型从“通用能力供给”向“场景价值闭环”的跃迁。一是扩充 MaaS 平台对多模态等模型能力的支持，推动跨模态理解与交互能力提升，满足更加复杂的应用场景需求，同时将加速构建**行业级 MaaS 平台**，深度整合垂直领域需求与业务流程，提供端到端的模型服务解决方案，赋能金融风控、医疗诊断、智能制造等关键领域。**二是基于公有云 MaaS 的大模型调用量将持续快速增长，**随着大模型应用在行业中的不断深化，模型服务和业务系统加速融合，多智能体协同应用增加，以及图片、视频生成等高 Token 消耗的多模态应用的持续普及，大模型 Token 调用量将持续攀升。**三是基于 MaaS 平台持续加快大模型服务的性能提升与效能优化，**在算力资源日益紧张的背景下，通过精细化的资源调度策略与分布式推理架构优化，进一步提升单位算力下的模型推理性能，降低单位 Token

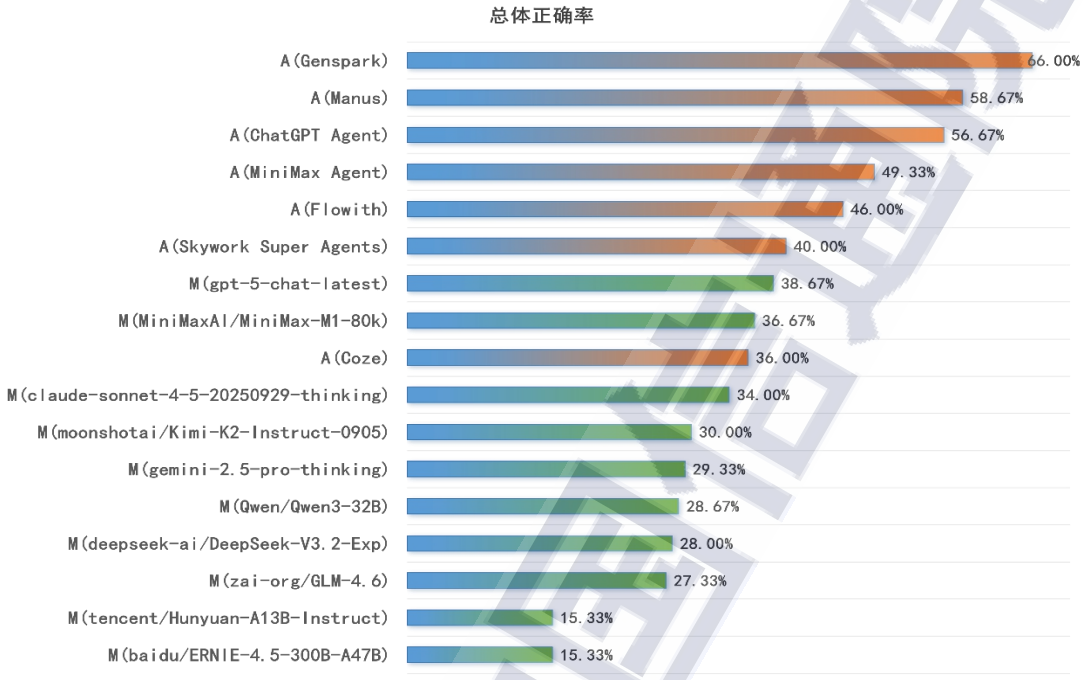
的生成成本和能源消耗，将成为大模型真正商业化落地的焦点。

（五）智能体自主性增强，加速智能原生应用建设

作为大模型应用的主要形态，智能体加速人工智能从感知认知向自主决策执行演进。从中国信通院“方升”智能体基准测试数据来看，智能体相比大模型可以完成更加复杂的任务，自主性不断增强。

智能体能力将成为大模型出厂标配，推动原生技术重构。当前，国内外大模型逐步配置智能体功能。比如，xAI 最新发布的 Grok 4 Heavy 支持 4 个智能体并行思考，进而完成更复杂、更精密的任务；智谱 AI 发布的最新一代旗舰模型 GLM-4.5，首次在单个模型中实现将推理、编码和智能体能力原生融合，以满足智能体应用的复杂需求。智能体能力的标配化，标志着大模型从“语言生成工具”向“生产力引擎”的转变。**一是**智能体推动原生技术架构从“被动响应”转向“主动服务”。智能体的发展不断重塑原生技术架构的底层逻辑，其核心在于通过认知能力的注入与工具生态的重构，推动系统转向“主动感知需求、自主执行任务”。传统技术架构依赖预定义规则与固定流程，虽能模拟人类操作，但缺乏自主决策能力。智能体通过动态规划引擎与工具调用框架，构建了“感知-决策-执行”的闭环体系。**二是**智能体加速原生技术形成原生解决方案。传统工作流或软件通常将最佳场景或行业实践固化为标准化流程，优势在于可以按照用户指定流程去执行特定任务，但灵活性不足。智能体通过融合基座模型、MCP 服务、智能体沙箱等，可以生成更具针对性和个

性化的解决方案，在执行任务过程中实现从“流程固化”到“动态优化”，加速原生技术与千行百业深度融合。



来源：中国信息通信研究院

图 10 中国信通院方升智能体基准测试结果（A 代表智能体，M 代表大模型）

智能体开发工具逐步降低开发者研发门槛。比如，2025 年 10 月，OpenAI 在开发者大会 2025 上发布了智能体工具集 AgentKit，旨在帮助开发者和企业快速构建、部署和优化智能体应用，其核心目标是解决传统智能体开发中工具碎片化、编排复杂、缺乏版本控制等问题。我国科技企业高度重视智能体开发与构建，比如百度文心智能体平台 AgentBuilder、腾讯元器、字节跳动 Coze 等。**一方面**，智能体开发工具可以加速智能体系统的开发，推动智能落地应用。智能体开发工具通常允许用户通过“拖拉拽”的低代码或无代码的

方式完成智能体创建，开发者无需编写代码即可组合逻辑节点、连接特定工具、配置自定义数据。**另一方面**，智能体开发工具可以促进智能体生态的建设。企业通过研发和开源智能体开发工具，可以培育更多用户参与到智能体创建，推动智能体在各个场景的规模化应用。

表 1 国内外智能体开发工具与平台

公司名称	工具名称	国别	功能
OpenAI	AgentKit	美国	智能体工作流创建和管理；数据与工具连接方式管理等。
OpenAI	Agent SDK	美国	帮助开发者构建和管理复杂的单智能体系统和多智能体系统。
OpenAI	Responses API	美国	内置网络搜索、文件搜索和计算机使用等工具，允许代理与真实世界的数据进行交互。
CrewAI	CrewAI	美国	支持多智能体协作的框架，允许定义角色与任务分配。
LangChain	LangChain	美国	开源模块化开发框架，支持多模型接入，提供工具调用、记忆管理等组件。
微软	AutoGen	美国	支持多角色分工、代码生成与执行，适合分布式任务处理与跨领域协作。
字节跳动	扣子	中国	全场景低代码开发平台，支持拖拽式工作流设计。
百度	AgentBuilder	中国	通过拖拽式工作流设计器编排交互逻辑，支持接收用户输入、调用 API、生成回复等节点自由组合。
腾讯	元器智能体平台	中国	基于混元大模型的智能体创作与分发平台，核心功能覆盖低代码开发、微信生态深度集成、多模态交互。
华为	Versatile-AI	中国	提供包括数据准备、模型接入、知识工程、智能体开发和编排、MCP、应用部署等能力。

来源：中国信息通信研究院

智能体通信协议进一步扩展智能体系统能力边界，降低系统集成复杂性。智能体通信协议通过标准化交互规则与协作机制，实现

模型与外部工具、不同智能体之间的兼容互联。比如，MCP 协议为不同的数据源和工具提供统一的连接方式，推动了外部工具、数据源的“即插即用”，极大扩展智能系统可使用的工具范围，执行更加复杂的任务。A2A 协议通过定义标准化的 Agent Card 元数据模型，将智能体能力抽象为可机器解析的结构化描述，使跨框架智能体可自动匹配协作需求。智能体通信协议解决接口碎片化问题，助力智能体互联互通。**一方面，**智能体通信协议通过标准化接口解决传统“一个系统一套接口”的问题，推动不同企业、不同领域的人工智能应用能够更好地协同工作，加速人工智能生态标准化进程。**另一方面，**智能体通信协议降低了人工智能封闭应用生态的壁垒，允许开发者能够更自由地选择和组合不同的工具和服务，促进了人工智能生态的开源开放。

我国智能体产业链逐步健全，加速形成智能体经济。在智能体基础支撑部分，我国拥有较为完善的智能体开发平台、开发工具、通信协议和大模型服务，具备研发和部署智能体的良好基础。比如在智能体协议方面，除国外开源的 MCP 和 A2A 协议外，我国拥有 ANP 开源技术社区的 ANP 协议、AgentUnion 提出的 ACP 协议、氦川科技的虚实融合 RVP 通信协议等，支撑我国多智能体互联协作。**在智能体应用部分**，我国智能体应用场景持续拓宽，产品形态不断丰富完善。国内科技企业已开发了深度研究、代码编写、智能搜索等通用智能体产品；同时，面向垂直领域个性需求，政务、金融、汽车、制造等行业智能体应

用逐步扩大探索范围，加快千行百业数转智改进程。

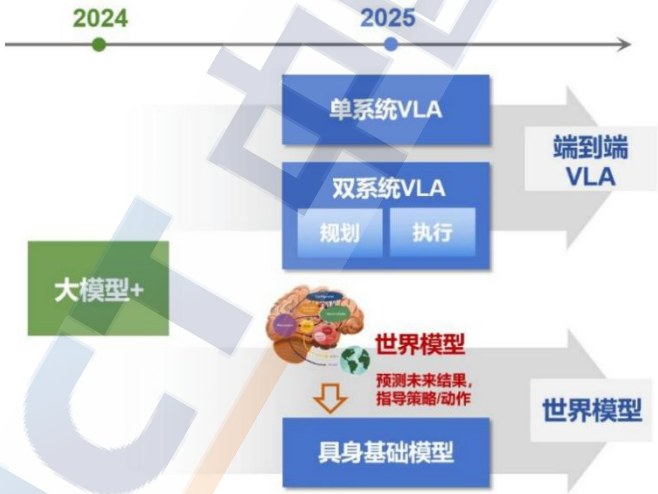
（六）具身智能走向实训，软硬一体化创新协同并进

具身智能已进入快速发展阶段，正处于从实验室技术验证向规模化商用过渡的关键时期。在软硬一体化创新并进的发展态势下，具身智能在模型、数据及本体方面均取得明显进展。**模型方面**，基础模型呈现多路径探索，“大脑”能力和“大小脑协同”是现阶段主要创新方向。**数据方面**，围绕仿真合成数据和真实采集数据开展广泛技术实践，发力高质量、大规模具身智能数据建设。**本体方面**，多元形态融合发展，通过训练场中学习作业技能为真实生产环境应用打好基础。**目前，场景驱动下的数据-模型-本体的三位一体联合设计与闭环优化，正逐步成为具身智能产业创新发展的关键路径。**

近年来业界初步尝试将大模型融入机器人的控制系统中，证明了通过“大模型+”提升机器人智能水平的技术可行性。**2025 年，行业沿着多条路径探索通用具身基础模型发展，端到端 VLA、世界模型等技术路线脱颖而出。一是端到端 VLA 架构加速探索。**截至 2025 年 12 月 8 日，谷歌学术以“End-to-end VLA”为关键词的论文在不到一年时间内从 600 余篇增加至 1700 余篇。**当前主流端到端大模型分为两类，一类**是单系统架构，在统一模型内完成感知、决策和动作生成。如智元机器人模型 GO-1[12]，部署到智元精灵 G1 上，能胜任擦桌子、倒水等真实家庭场景任务。**另一类**是双系统架构，

[12] https://agibot-world.com/blog/go1

协调具身智能基础模型的推理规划和运动控制运算频率差异。如 Figure AI 公司的 Helix 模型[13]，采用双系统架构实现 7-9 赫兹认知推理和 200 赫兹本体运控协同，实现从双机协作物流分拣、叠毛巾到洗碗等现实作业能力。**二是世界模型拓展具身智能基础模型的认知边界。**当前世界模型主要以两种形式融入具身智能基础模型。**一类**是生成未来的视觉观察结果以指导动作生成。英伟达提出 DreamGen[14]通过世界模型生成机器人操作视频，提升了模型在未见环境中的泛化性。**另一类**是模拟从感知到决策的过程，在动作执行前先预测并指导策略规划。智元机器人 EnerVerse-AC[15]框架能够评估机器人动作，高效验证策略。



来源：中国信息通信研究院

图 11 大模型向具身基础模型方向探索演进

本体走入训练场、竞技场和行业场景，通过真实场景实践加速技术迭代。一是在训练场中学习作业经验。据不完全统计，截至 2025

[13] https://www.figure.ai/news/helix
[14] https://arxiv.org/abs/2505.12705
[15] https://arxiv.org/abs/2505.09723

年 12 月底，国内已建设完成的训练场在当年内已增至 30 个，为机器人学习提供真实作业场景。如石景山训练场还原了工业、家庭、康养和 5G 融合四大类共 16 个细分场景，夸父人形机器人通过实训已胜任搬运、巡检、导览、配送等多种任务。**二是在竞技场上实现运动能力突破。**从 4 月人形机器人半程马拉松，到 5 月 CMG 世界机器人大赛机甲格斗擂台赛，再到 8 月世界人形机器人运动会，多家具身智能企业的机器人展现出运动能力的提升。如松延动力 N2 在 4 月半程马拉松中全程无人类陪跑获得亚军，在 8 月运动会自由体操项目中做出一系列跳跃、连续后空翻夺得冠军，实现多维度运动能力突破。**三是在行业场景中验证优化。**本体提供方、算法提供方、行业场景方多方合作，推动机器人“进场实训”，在真实行业场景中训练、测试、验证。优必选人形机器人在极氪 5G 智慧工厂实训，从 Walker S Lite 到 Walker S1，通过多轮实训将机器人搬运效率提升了约 25%。

向真实场景迈进过程中，本体形态多元化发展，产品谱系不断丰富，涵盖智能机器人、智能运载装备和新型智能产品三大构型分类。一是智能机器人，以人形机器人、复合轮臂式机器人、四足机器人为代表。**人形机器人作为具身智能行业热点处于起步阶段，**未来有望开拓以通用性为核心的应用市场，目前以教学科研、表演展示、前台导览等为主。**四足机器人、复合轮臂式机器人产品初步落地。**IDC 数据显示 2024 年全球四足机器人销售出货量约 2 万台，其

中商用级产品占总出货量的 27.9%[16]。复合轮臂式作为人形机器人落地“中间态”，初步具备规模化应用基础，短时间可形成真正有应用价值的产品。**二是智能运载装备，**以自动驾驶汽车、无人驾驶航空器为代表。自动驾驶汽车 L3 级车型量产、L4 级商业化运营。摩士根丹利预测到 2030 年，自动驾驶汽车的市场规模将达到 2000 亿美元[17]。无人驾驶航空器以无人机和电动垂直起降飞行器（eVTOL）为核心载体，在具身智能技术融合下作业能力全面升级。**三是新型智能产品，**以变形移动装置、集群式微型智能机器人、智能可穿戴设备等产品为主，主要在教育科研领域开展前沿探索，其中智能外骨骼作为智能可穿戴设备的代表，已在重体力劳动场景、适老康养场景探索落地。

当前具身智能产业链已初步形成。本体技术方面，汽车和机器人领域在传感器、关节模组、高密度电池、电机和端侧芯片等核心零部件上，可快速实现技术复用和低成本迁移。例如双林股份、丰立智能等凭借汽车领域积累的工艺链和技术积累，研发生产机器人关节模组、谐波减速器等精密部件。地平线和黑芝麻智能在同步布局汽车和机器人端侧芯片。**模型和工具链方面，**阿里、腾讯、字节和华为等科技企业均在开展具身基础模型或配套开发平台研发，为模型创新迭代注入更强动力。松应科技、跨维智能等创业企业围绕

[16] https://my.idc.com/getdoc.jsp?containerId=prCHC53644125
[17] https://www.bloomberg.com/news/videos/2025-07-28/morgan-stanley-sees-200b-autonomous-car-market-by-2030-video

仿真平台、数据平台等提供工具服务，降低行业研发门槛。**整机产品开发方面，**传统机器人企业、初创公司、跨界车企多方投入研发多元化具身智能产品。例如宇树、智元研发工业和商业级机器人，松延动力、维他动力发布消费级机器人，微分智飞、小鹏汽车布局无人驾驶航空器。**行业应用方面，**各行业场景方积极拥抱具身智能，开放场景或与技术提供方开展联合研发。例如家电龙头企业美的公司、新能源车企极氪等均在工厂设置具身智能实训试点。

具身智能要从“实训”走向“实战”，仍面临三大挑战。一是高质量数据“少”，行业普遍认为，要实现物理智能涌现至少需要百万小时，甚至千万小时的真实行为数据，而当前真正可用数据远远不足。**二是模型泛化“难”**，在执行训练数据未覆盖的场景任务时，很可能出现性能急剧下降的问题。**三是软硬协同“难”**，模型与本体需要在多个时间尺度上协同控制，任何一环不稳定都可能导致任务执行失败。**应在数据驱动下加速具身基础模型通用化发展，依托开发平台促进软硬一体化协同创新。一是**解决数据少的问题，需要充分发挥训练场数据采集能力，建设高质量数据集。**二是**解决跨场景、跨任务泛化问题，需要加强探索通用具身基础模型及迭代模式，提升模型泛化能力。例如智平方 GOVLA[18]通过小样本学习在真实任务中对现场数据微调，提高任务成功率。**三是**解决软硬协同工程化链路问题，需要通过具身智能开发平台规范模型、数据与本体。如腾

[18] https://www.housebots.com/news/tag/GOVLA

讯具身智能开放平台 Tairos[19]、华为云 R2C[20]跨本体联接协议等。

二、应用赋能

（一）人工智能应用逐步扩展，加快向高附加值领域环节渗透

在政策与市场双轮驱动下，我国已形成覆盖基础层、框架层、模型层、应用层的完整人工智能产业体系，各类智能产品和服务创新活跃，为人工智能实现规模化落地与产业赋能奠定了坚实基础。总体看，人工智能赋能应用基本遵循先从数字化水平较好的领域率先突破，再逐步扩散到更多行业的规律。

从第一产业看，人工智能助力农业降本增产，加速培育新质生产力。农业大模型通过整合农学知识、历史经验与环境数据，构建作物生长数字孪生体系，支撑水肥药精准调控与产量预判。智慧农机装备依托北斗高精定位，融合装备传感器、控制器及人工智能算法，实现对行驶路径、速度与作业操作的实时精准控制。兼具传统农业知识与智能化应用能力的“新农人”作为关键决策主体，推动生产模式革新与产业链升级。此外，畜牧养殖、无人机植保、智慧育种等领域也取得技术创新与商业化实践成效。人工智能正通过系统性赋能，持续释放农业提质、节本、增效潜力，驱动农业新质生产力加快形成。

从第二产业看，人工智能加速向工业全场景渗透，构建智能制

[19] https://tairos.tencent.com/
[20] https://www.huaweicloud.com/product/octopus/robo.html

造新生态。人工智能赋能新型工业化走深向实，在典型重点领域已形成百余种应用模式，催生了一批新工具、新助手、新产品。一方面，重点行业智能化水平不断提升，钢铁、有色金属、电力、通信等行业专用模型加速传统产业“智”变升级。另一方面，重点环节智能化转型显著提速，人工智能在研发设计、中试验证、生产制造、营销服务、运营管理全环节加速落地应用，通过数据要素与工业机理的深度融合，推动工业体系向智能化、柔性化与高效化转型。

从第三产业看，人工智能与服务场景融合持续深化，满足用户个性化需求。在金融行业，智能风控、AI 投顾和反欺诈系统显著提升服务效率与安全性；物流行业依托智能调度、路径优化及无人配送技术，实现降本增效；零售业则通过用户画像、智能推荐和动态定价，优化消费体验并提高库存周转率。与此同时，新兴服务场景不断涌现，AI 虚拟客服、数字人导游、智能法律咨询等拓展服务边界。第三产业具有数据密集、交互频繁、流程标准化程度高等特点，人工智能技术落地最快、商业回报最明确。人工智能正逐渐从辅助工具转变为服务主体，推动第三产业向智能化、个性化与高附加值方向全面升级。

（二）人工智能赋能新型工业化，加速向现实生产力转化

当前，新一轮科技革命对生产方式及产业体系的革命性影响已经开始显现，人工智能等技术与工业融合已成为大势所趋，正逐步改变现有创新范式、生产运营方式和产业形态。为观察人工智能在

工业领域的赋能路径与成效，主要从不同行业与不同环节两个维度展开分析人工智能应用的渗透特征与重点。

1.从赋能行业看，行业渗透程度与应用侧重存在差异

人工智能在不同工业领域呈现差异化渗透特征，其中电子信息、消费品、以汽车为代表的装备制造等行业在整体应用中占据重要份额，原材料行业以钢铁、石化领域为代表逐步实现技术落地，能源电力等行业也形成了较好应用态势，总体来看赋能前景广阔。

装备制造行业聚焦研发环节创新，在零件结构优化、智能仿真等场景的应用水平显著高于行业平均水平，汽车、航空领域已实现前沿技术的规模化验证。部分车企在新车研发阶段引入 AI 驱动的虚拟仿真技术，通过数字孪生模型对空气动力学性能进行模拟测试，超过 94%的模拟案例与风洞数据的偏差小于 3%，速度快 10-100 倍，每次模拟运行成本降低 60%。

电子信息行业完成了从生产管控赋能到先进制程优化的跃升，部分企业将 AI 算法深度嵌入先进制程控制，实时优化多项工艺参数，保障了先进制程的过程能力指数领先。半导体企业使用 AI 控制器实时调整等离子体密度，将刻蚀均匀性波动从±5%缩小至±1.5%，通过对制程参数的优化，提升了制程的稳定性和精度。

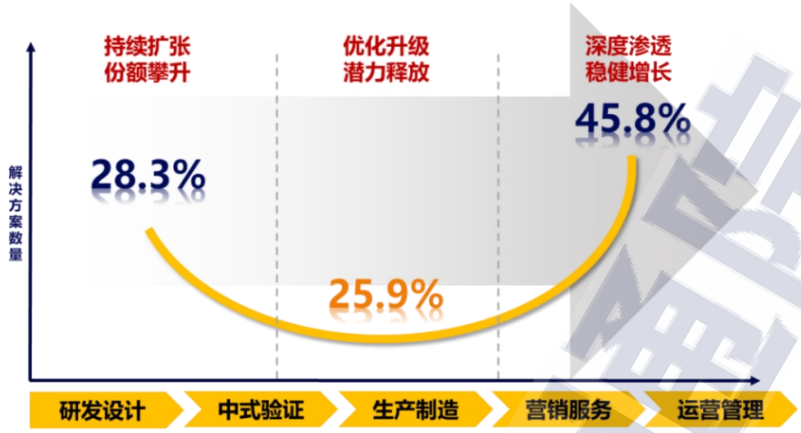
消费品行业侧重质量管控与工艺优化，相关场景应用成熟度高出行业平均水平，纺织、家电行业的点状探索已逐步向规模化复制演进。纺织企业应用 AI 质检解决方案，“AI 质检师”检出率稳定高

于 98%，远超人工水平。家电企业打造注塑换型云调优平台，基于不同工况和 228 种模具，平台可智能推荐 13 个最优方案，实现对 85 台注塑机换型工艺参数的一键部署和调控，设备综合效率（OEE）提升 30%。

原材料行业以生产管控智能化为核心方向，钢铁与石化化工等领域贡献了该行业多数应用案例，形成了生产过程精准控制、高价值设备智能运维等成熟模式。某钢铁企业热轧 1880 产线接入预测大模型后，仅用了几个月时间进行数据学习，±2 毫米宽展预测准确率就提高了 5%，达到了 83%。某石化企业建成国内石化领域首个乙烯装置数字孪生体，实现乙烯装置全域智能运行，其全流程智能控制系统（IPC）使装置日常操作量减少 90%，平均自控率由 98.44%提升至 99.50%。

2.从赋能环节看，大模型成为全环节赋能重要支撑

工业领域应用处在“大小模型协同”阶段，其中大模型正逐步在各节点探索效能提升的应用实践，初步形成了“研发创新、生产辅助、服务延伸、产品增值”的全面覆盖。我们重点对大模型赋能情况进行了监测和分析，总体上产业链应用分布延续了去年格局，同时显现出结构性优化与内生动力增强的积极信号，呈现出“两端深化、中间突破”的发展态势。



来源：中国信息通信研究院

图 12 大模型在工业各环节应用分布情况

前端研发设计环节应用占比小幅下降，赋能专业程度有所提升。升级重点在于从“通用技术探索”转向“精准场景赋能”，模型专业化程度大幅提升，细分场景深度适配，通过聚焦特定需求实现研发效率的精准提升。例如，中国科学院大连化学物理研究所研发的智能化工大模型 2.0Pro，构建了催化反应、工艺开发、中试放大、工厂优化四大智能平台，为化工新技术研发提供全新路径。针对化工研发周期长的痛点，其智能机器人催化反应实验系统可替代人工开展实验，自动完成催化剂评价，效率提升超 10 倍。

生产制造环节应用占比从 18.8%提升至 25.9%，体现了大模型积极向生产环节探索的态势。随着工业质检、工艺参数优化等场景广泛落地，大模型对制造执行环节的赋能作用不断显现，成为助推生产效率提升、保障制造质量的新引擎。中国钢研 2025 年 5 月发布的“冶金流程感知大模型”，采用“感侧大模型+知侧大模型”双塔结

构，融合 70 余年行业知识与生产数据，在金相分析（晶界提取/组织辨识准确率超 95%）、产品表面缺陷检测（成功率超 95%）、物料跟踪等场景表现突出，可自动生成分析报告，推动行业从“单点智能化”向“全流程数智化”转型。

后端运营管理环节占比最高且小幅上升，对企业价值提升进一步增强。企业运营对大模型的依赖度持续加深，大模型已从初期的辅助数据分析升级为智能决策等复杂场景支持，通过打通数据链路、优化运营流程，为企业降本增效提供关键动能。例如，煤炭科学研究总院有限公司研发的新一代人工智能调度平台矿山知行，通过整合太阳石矿山大模型、多智能体协同决策和数据知识融合等关键技术，构建覆盖“人—机—环”全要素的智能化调度体系，可实现数据、决策、运营“三化一体”协同发展，推动调度从辅助决策向自主决策转型升级，完成从“少人调度”到“黑灯调度”的智能化演进。

除分环节单点赋能外，工业领域的大模型应用呈现明显的平台化与集成化发展趋势。行业大模型平台通过集成不同规模的专用模型，形成分层赋能体系。其中，大型语言模型作为底层支撑，提供统一的知识检索与推理服务；而针对特定生产环节的专用模型则在上层实现精准赋能。这种“底座+应用”的架构模式，既保障了基础能力的通用性，又确保了专业场景的适配深度，展现出多模型协同、全链路覆盖的新型赋能格局。

（三）智能原生成为智能经济“时代基因”，重塑产品

服务与企业组织模式

2025 年 8 月，国务院印发《关于深入实施“人工智能+”行动的意见》首次提出“培育智能原生新模式新业态”，将人工智能通过“原生”方式融入组织的战略规划、组织架构、业务流程，发展智能原生的技术、产品和服务，这将有助于突破传统数字技术应用的路径依赖，最大化释放这一变革性技术的巨大潜力，为智能时代的全面到来做好准备。当前，智能原生的实践刚刚起步，发展形态和路径有待进一步探索。

大模型嵌入工具软件，智能原生软件加速数字生产力跃升。以深度研究、代码编写、多用途为代表的三大类智能体，进一步释放大模型应用和服务潜能。**深度研究智能体**通过整合动态推理、自适应规划、多轮外部数据检索及工具使用能力，能够自主完成端到端的复杂研究任务。国内外大模型厂商纷纷布局，比如 OpenAI Deep Research、Google Deep Research、Kimi Researcher、豆包 DeepResearch 等，该类产品将在更广泛领域重塑人类的研究范式。**代码编写智能体**融合代码理解、代码生成、调试优化及多工具协同能力，能够自动完成从需求分析到代码交付的全流程任务，深刻改变软件开发工具和产品形态，正开启软件业全面重塑的新阶段。例如 Cursor 作为一款原生代码编写智能体，生成代码准确率达 89%，我国也涌现出通义灵码、文心快码、星火飞码等产品。**多用途智能体**通过高度封装，可以实现网页制作、游戏制作、旅行规划等通用任务，实

现从需求分析到结果交付的全流程自动化。国内外多用途智能体产品，比如 Genspark、Flowith、Manus、MiniMax Agent、Skywork Super Agents，提高了面向用户端的产品体验。多用途智能体的发展标志着人工智能从“辅助工具”向“数字劳动力”的质变。

大模型为硬件产品赋魂增智，持续拓展人机交互新模式。大模型正加速从“软”的算法层面向“硬”的物理世界渗透与融合，驱动新一代智能终端、智能网联汽车及具身智能领域实现能力跃迁，重塑硬件产品的功能边界，重构人机交互的体验与效率。**新一代智能终端方面，**终端已成为用户侧承载大模型部署任务的重要载体。AI 手机、AI 眼镜、AI 玩具等新一代智能终端，初步具备主动感知理解、多模态交互、智能化服务和自主学习进化等功能。随着大模型能力的不断提升，未来将实现从感知、理解、交互、决策到服务全流程的智能升级与自主进化。**智能网联汽车方面，**大模型与硬件的融合展现出巨大的潜力。在交互模式上，车内语音助手根据驾驶者的情绪和偏好提供个性化服务，智能座舱的显示硬件与大模型结合可提供更加智能的驾驶信息展示。在传感分析上，自动驾驶方面硬件传感器收集的大量数据通过大模型进行分析处理，使车辆能够更精准地感知周围环境，做出合理的驾驶决策，提升自动驾驶的安全性和可靠性。**具身智能方面，**产品形态丰富多样，适配场景日益广泛。轮式四足机器人可根据地形智能切换驱动方式，平坦路面可实现轮足协同，面对障碍物自动抬腿，转换为四足步态稳定通行。

飞行机器人能通过自主决策能力在断网的无信号环境下独立完成任务。轮臂式产品兼具移动和操作两方面优势，训练难度小、成本低、长续航、稳定性高，可满足更多场景落地需求。

AI 重塑企业的底层架构和运行逻辑，智能原生企业加速崛起。过去十多年，传统企业逐步实现了信息化和自动化，但其核心决策和运营模式仍主要依赖于人类经验和规则驱动，人工智能只是“配角”。智能原生企业则以人工智能为核心驱动力，将智能化要素嵌入到企业的业务、管理、决策等各环节，成为“主角”。随着智能体的普及，善于驾驭新技术的创新创业主体，可以依托人工智能强大的知识储备和认知决策能力，通过人机高效协同，推动企业组织架构从传统金字塔层级结构转向人机协同的扁平化工作网络，将生产力推向前所未有的高度。近几年，国内外一大批智能原生企业竞相涌现，如 OpenAI、Anthropic、深度求索、月之暗面、智谱 AI 等，为人工智能发展开辟新的模式。未来有望催生一大批只有几个人的“独角兽”企业，形成智能经济新的增长引擎。

（四）人工智能落地路径逐渐清晰，推动产业创新发展走深向实

当前，人工智能技术正加速迈向产业应用深水区，但“落地难”、“落地浅”仍是瓶颈。为推动人工智能从“可用”向“好用”、“常用”跃升，制定系统化、可操作的落地路线图至关重要。2024 年中国信通院发布了《大模型落地路线图研究报告》，提出从诊断、建设、

应用、管理四大阶段探索适合大模型的最佳落地路线，聚焦场景、业务、数据、技术四大核心，充分考虑企业差异化，方能走深向实。

人工智能技术的落地应重点围绕“场景筛选-技术适配-业务融合-数据支撑”四大核心展开。场景选择是人工智能落地应用的首要前提。面对大模型和智能体应用场景选择多、适配程度差异大的复杂问题，为实现较高技术可行性、较好经济回报、较大社会影响性，需构建“场景判断选择器”，进而打造可扩展的高价值场景。通过业务价值密度、技术成熟度、经济回报率、不可替代性、实施复杂度、行业竞争壁垒等维度构建量化评估框架，从而实现高优先级落地场景的精准筛选与定位。**业务融合是人工智能价值转化的核心环节。**数字化水平较好的行业领域需发挥“领头羊”作用，依靠优质的业务数据、成熟的数字基础设施以及海量的用户基础等优势，率先推动人工智能在行业高价值业务场景落地；数字化水平不足的行业领域需首先拆解业务全环节并明确人工智能技术的嵌入角色与功能定位，从高价值场景逐步向行业核心场景渗透，从而实现人工智能的价值内化。**数据支撑是人工智能系统持续稳定运行的基础保障。**面对数据孤岛、质量参差等问题，需构建“采、存、管、用”全生命周期的标准化体系。通过为不同数据类型匹配差异化存储方案，并建立统一的数据标准、质量控制与安全保障体系，为人工智能应用提供稳定、高效的数据支撑。**技术适配是人工智能场景落地的关键支撑。**技术架构与智能体的选型需严格遵循“业务需求驱动技术选

型”原则，依据场景复杂度与业务规模可划分为三类适配模式：一是适用于中小企业简单业务场景的轻量化架构；二是适用于中型企业多场景协同联动模块化架构；三是适用于大型企业全链路智能化需求分布式架构。

人工智能体技术的落地需结合企业在资源禀赋、数据基础、业务复杂度及合规要求等方面的差异性因企制宜。大型央国企通常采取“自上而下”的战略路径。它们虽面临场景分散、数据异构、合规严苛等挑战，但凭借雄厚的资金与技术实力，能够进行系统性布局。**在业务层**，大型央国企聚焦核心业务流程，组建跨部门专项小组开展流程诊断与优化工作，例如国家电网通过“光明电力大模型”实现电网调度与设备运维的智能化升级，覆盖 600 余个电力业务场景，故障处置与服务保障能力增强 30%[21]。**在数据治理层**，它们着力破除数据孤岛、构建统一的数据中台，并通过私有化部署模式保障数据安全。例如中国金茂通过内部大模型整合企业知识库与业务系统，实现工单自动分类与数据查询功能，员工日均工作时长节省 0.5 小时[22]。**中小企业倾向于“自下而上”的试点突破。**鉴于资源有限、业务流程架构简洁，其核心目标是快速实现提质增效。凭借业务聚焦、数据相对集中的优势，**在业务层**，中小企业多从轻量化应用切入，如多家初创企业基于 LLaMA2、Baichuan 等开源模型，衍生出

[21] 人民网，光明大模型赋能 电力服务迈入“智精准”新阶段 http://cq.people.com.cn/n2/2025/1016/c367647-41381733.html

[22] 中国信息通信研究院华东分院《大模型落地应用案例集》

法律（Lawyer-LLaMA）、教育（EduChat）等垂直工具。**在数据治理层**，它们优先规整核心业务数据资产，并依托外部工具与云服务平台构建数据安全保障体系。例如九章云极通过知识管家工具为中小企业提供 RAG 技术支持以快速构建知识库。

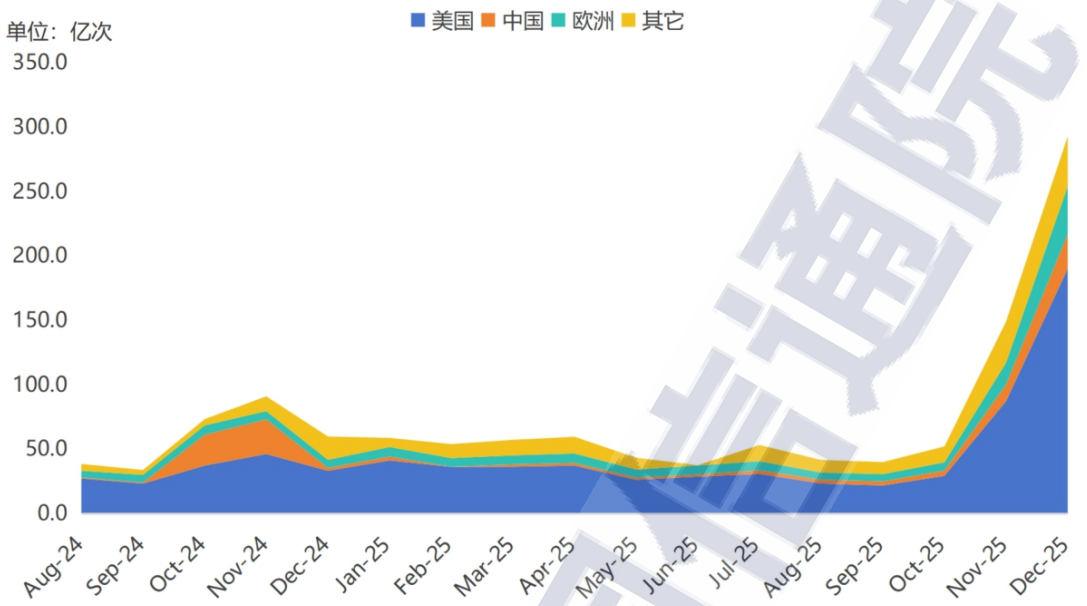
三、生态支撑

（一）开源成为标配，社区协同演进推动技术普惠

2025 年，全球人工智能技术的快速发展与开源生态的繁荣密不可分。我国在受益于全球开源体系的同时，也积极为开源社区做出贡献。开源作为人工智能产业与技术发展的重要引擎，不仅孕育出一批高质量的人工智能项目，还有效促进了上下游产业链的协同与融合，深刻改变了人工智能产业的发展格局。

开源模型创新活跃，国际影响力稳步提升。以深度求索、通义千问等为代表的国产开源模型迅速崛起，技术研究不断突破，打破了闭源模型的垄断格局，为用户提供更具性价比的选择。据 Artificial Analysis 平台数据，国产模型性能稳居全球前列，深度求索、通义千问、智谱、月之暗面等开源模型表现尤为突出，充分印证了国产模型的技术硬实力。截至 2025 年 12 月，国产开源大模型全球累计下载量突破 100 亿次，月下载份额最高占全球的 17.1%。本土开源社区模型下载量也从 2024 年 12 月的 5.8 亿次增长至 2025 年 11 月的 14.4 亿次。与此同时，基于国产开源模型进行微调的模型占比，从 2024 年 2 月的 10%大幅提升至 2025 年 8 月的 60%。这些趋势一致

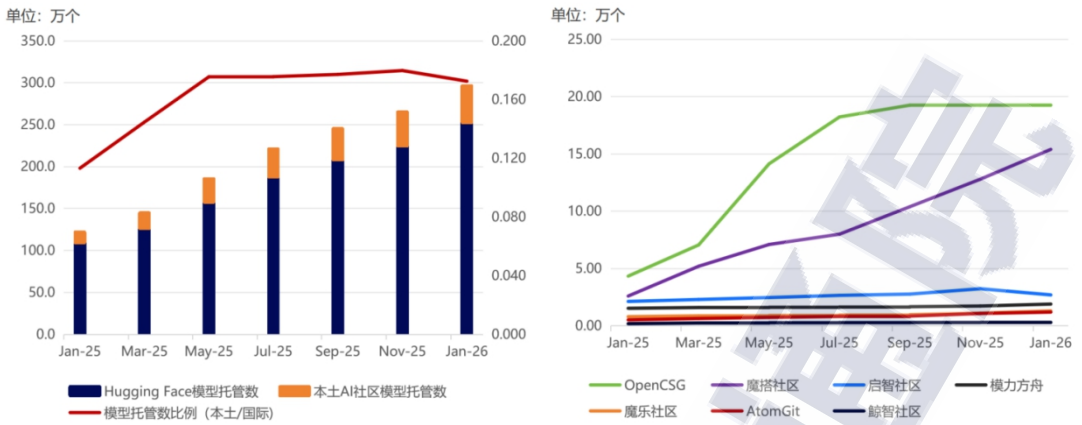
反映出国产开源模型在全球范围内的接受度和影响力正快速增强。



来源：中国信息通信研究院，2025 年 12 月

图 13 HuggingFace 上开源模型月下载量趋势

开源社区协同演进，推动技术普惠发展。一批本土开源社区正在积极推动技术普及与生态建设。例如，魔搭社区面向东南亚提供本地化支持，助力我国开源模型走向国际；焕新社区聚焦“人工智能+行业应用”，开放了 78 个由央国企发布的行业大模型，促进人工智能技术在实体经济中的规模化落地；魔乐社区围绕国产软硬件全栈适配，已实现上千个模型对国产算力的兼容适配。截至 2025 年 12 月，本土开源社区平台已托管模型达 40.5 万个，与 HuggingFace 同期托管模型的比例从年初的 11%提升至 18%，展现出本土开源生态正稳步壮大。



来源：中国信息通信研究院，2025 年 12 月

图 14 2025 年我国 AI 开源社区平台模型托管规模趋势

商业模式孕育成型，构建合作共赢生态。模型厂商普遍推行“开源免费+高阶服务收费”的策略，即通过开放基础模型吸引开发者与用户，进而借助技术支持、定制化开发、云服务等增值项目实现商业转化。与此同时，开源模型也拉动了云服务与芯片市场的需求增长。模型厂商积极与芯片厂商、云服务商展开合作，在降低国产芯片应用门槛的同时，共同探索联合运营的新路径。例如，硅基流动、无问芯穹等企业持续优化国产芯片适配能力，通过私有化部署、模型即服务（MaaS）等形式提供解决方案，助推国产算力市场迅速增长。

（二）全球 AI 标准竞合加剧，我国纵深推进体系建设

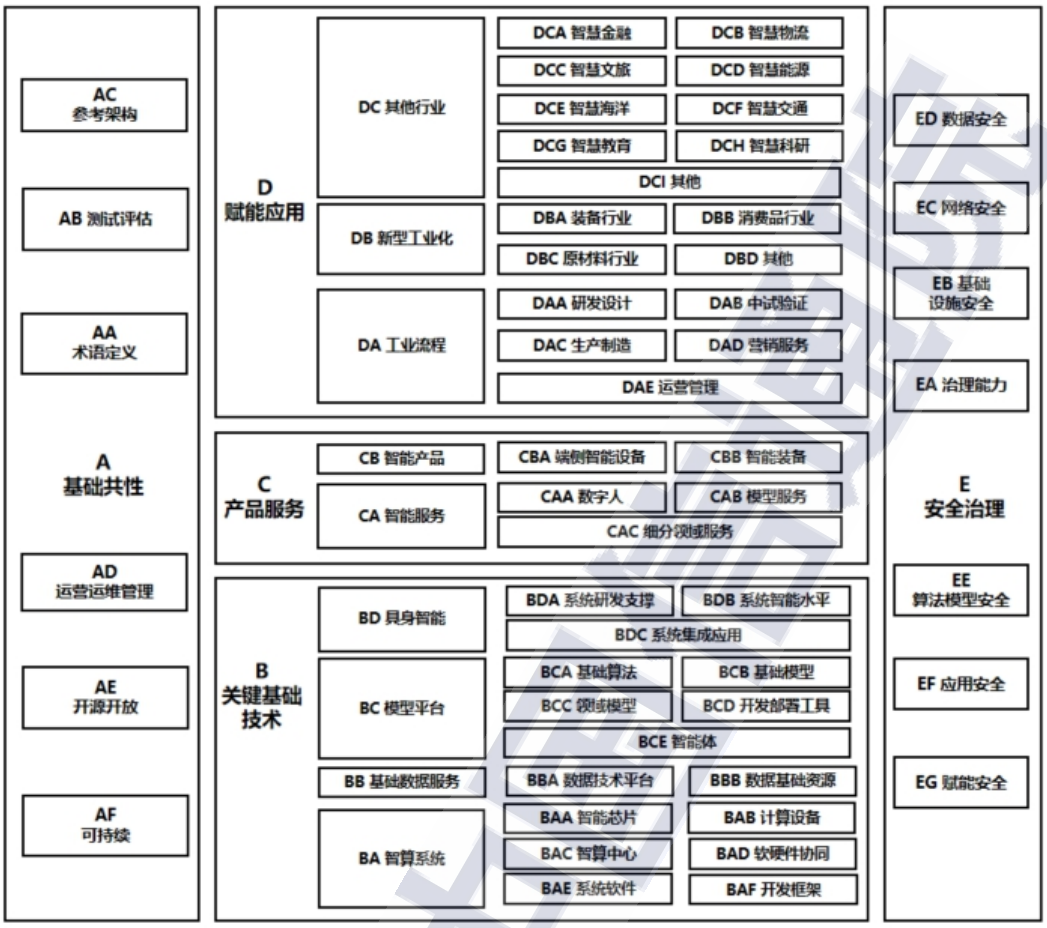
2025 年，人工智能标准从技术规范向价值引领加速演进，成为平衡创新发展与风险防控的关键抓手。全球主要经济体加快人工智能标准战略布局，我国人工智能标准体系建设持续深化，在关键技术、行业应用等领域形成一批重要成果，为人工智能高质量发展高

水平安全提供了有力支撑。

全球 AI 标准战略加速落地，推动治理格局深刻变革。近年来，全球主要经济体加速出台相关战略，试图通过标准化手段引导人工智能技术创新与安全发展。**一方面，立足本土实际，形成差异化发展路径。**欧盟强调风险防范与法律约束，通过《通用人工智能实践准则》等文件强化合规指引，凸显其规则先行的监管思路。美国倾向于依托市场力量与自愿共识，由私营部门主导标准制定与试点应用。2025 年 6 月，美国将人工智能安全研究所调整为人工智能标准与创新中心（CAISI），职责重点转向推动科学驱动、行业协同的标准制定与风险评估。**另一方面，融入全球网络，加速与国际标准接轨。**欧洲标准化委员会（CEN）、欧洲电工标准化委员会（CENELEC）与欧洲电信标准化协会（ETSI）三大组织与国际标准化组织 ISO、IEC 和 ITU 建立了对应合作关系，并确立了国际标准化优先原则。美国国家标准与技术研究院（NIST）在协调国内标准、参与制定全球标准外，还积极推动其《人工智能风险管理框架》（AI RMF）成为国际标准。

我国纵深推进 AI 标准体系建设，开放协同生态稳步构建。我国坚持国内统筹与国际对接双规并行，体系化推进人工智能标准体系建设。**一是持续强化顶层设计，完善配套举措。**2024 年 6 月，工业和信息化部等四部门印发《国家人工智能产业综合标准化体系建设指南（2024 版）》，明确了人工智能标准体系结构框架，为赋能产业

发展提供了重要指引。**二是构建协同互补的标准化组织网络**。我国已形成覆盖广泛、分工明确的标准研制力量。国家标准方面，全国数据标准化技术委员会、全国信息技术标准化技术委员会人工智能分技术委员会等围绕人工智能数据、安全治理等热点领域开展标准研制，发布多项重要标准。行业标准层面，工业和信息化部人工智能标准化技术委员会等围绕人工智能软硬协同、智能体、具身智能、安全治理等重点方向加速推进标准制订，由“按批次报送”改进为“随来随审、随时报送”的模式。同时，加强贯标推广和标准应用，不断壮大人工智能标准生态，年度开展 60 余场标准宣贯、培训和推广活动，累计培训 870 余名标准化人才。深化跨行业、跨领域协同，与 14 个标准化组织，7 个产业组织建立了联络关系。**三是持续推进 AI 标准“引进来”和“走出去”**。2025 年 3 月，国家市场监管总局修订《采用国际标准管理办法》，完善了对 ISO、IEC 等国际标准的动态追踪与快速转化机制，稳步扩大标准制度型开放。



来源：中国信息通信研究院

图 15 工业和信息化部人工智能标准化技术委员会 2025 年标准制定指南

我国 AI 标准成果持续涌现，供给能力不断增强。国内层面，2025 年以来，工业和信息化部人工智能标准化技术委员会围绕标准研制、生态建设、产业研究形成一系列重要成果。发布了年度标准制定指南及安全治理标准体系建设指南，在广泛调研基础上凝练形成产业界定及测算、大模型评测、高质量数据集、人工智能工程化等十大重点标准方向，与中国通信标准化协会（CCSA）联合推动发布 20 项标准，年度新增征集 176 项标准立项建议，完成 11 项标准报批、

25 项标准送审稿。**国际层面，**我国积极参与 ITU 等国际标准组织工作，多位专家担任 ITU-T 主席、副主席、报告人等职位。截至目前，我国牵头 ITU-T 人工智能标准已经发布 47 项，在研标准 63 项，覆盖芯片评测、软件框架、工程平台、大模型及重点应用，其中大模型评测、人工智能平台、数字人等标准填补国际标准空白。

表 2 2025 年十大重点标准方向及重点标准

序号	重点方向	重点标准	计划号
1	产业界定及测算	人工智能 基础共性 人工智能产业边界界定	2025-0255T-YD
2		人工智能 基础共性 人工智能企业认定	2025-0256T-YD
3	软硬件协同	人工智能 关键基础技术 面向大模型的异构智算系统资源池化技术要求	2025-0272T-YD
4		人工智能基础支撑 软硬件协同 大模型训练及推理集群系统能力要求	2024-1323T-YD
5		人工智能基础支撑 软硬件协同 面向大语言模型的算子能力技术要求	2024-1324T-YD
6	大模型评测	人工智能关键技术 大模型基准测试总体技术要求	2024-1334T-YD
7		人工智能关键技术 大模型分类方法和分级技术要求	2024-1333T-YD
8	人工智能工程化	人工智能关键技术 大模型模型即服务（MaaS）模型服务平台技术要求	2024-1336T-YD
9		人工智能关键技术 大模型模型即服务（MaaS）应用开发平台技术要求	2024-1337T-YD
10	智能体	人工智能关键技术 智能体基础技术能力要求	2024-1332T-YD
11		人工智能 关键基础技术 智能体服务应用技术要求	2025-0275T-YD
12	具身智能	人工智能关键技术 具身智能 基准测试方法	2024-1328T-YD
13		人工智能关键技术 具身智能 数据集质量要求及评价方法	2024-1329T-YD
14	人形机器人	人形机器人 基础共性 智能化能力分级	2025-0485T-YD
15		人形机器人 基础共性 训练场构建技术规范	2025-0483T-YD
16	高质量数据集	人工智能基础支撑 基础数据服务 大模型数据集开发管理能力分级及评估方法	2024-1315T-YD
17		人工智能合成数据生成和管理能力要求	2024-0585T-YD

18	应用成熟度	人工智能基础支撑 系统软件 人工智能研发运营一体化成熟度要求：模型运营	2024-1321T-YD
19	人工智能安全治理	人工智能 安全治理 大模型安全基准测试方法	拟立项
20		人工智能 安全治理 系统风险管理能力要求	2024-1353T-YD

来源：中国信息通信研究院

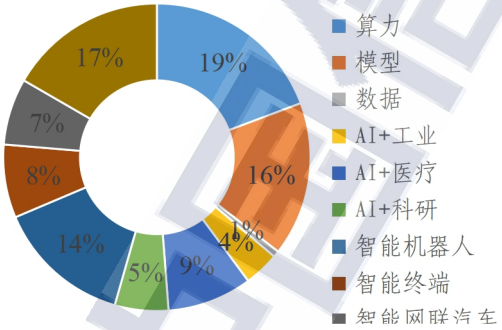
（三）人工智能受全球资本热捧，投资规模持续扩张

人工智能热潮引发全球资本竞逐，我国非国资机构占据主导。从全球趋势看，人工智能投融资活跃度不断提高。全球人工智能投融资占全行业投融资比例从 2023 年的 8.1%上升至 2024 年的 13.5%，并在 2025 年二季度跃升至 23%。从投融资规模看，我国与美国尚存在量级差距。2025 年上半年，美国人工智能投融资金额为 381 亿美元，同比增长 43.6%，我国投融资金额仅为 36.7 亿美元。从资本属性看，非国资机构投资占主导地位。2025 年上半年，我国境内人工智能领域已披露机构投资 738 笔，其中，非国资机构[23](包括民营、外资以及天使投资人)投资 539 笔，占行业内总投融资笔数的 73%。

AI 基础层投资热度高，工业等垂直赛道受资本关注。一是基础层（算力、模型与数据）方面，我国境内大模型领域投融资占 AI 总投融资金额的比例从 2023 年的 31%，上升至 2024 年的 66%，并在 2025 年上半年达到 16%，略低于算力投资。全球来看，2024 年至 2025 年一季度，全球 Top10 大模型融资事件中，中国有 4 项，融资的企业分别是智谱 AI、月之暗面、百川智能和 Minimax。**二是**行

[23] 非国资机构是指投资机构的基金管理人非国有独资或控股等。

业赋能方面，2023 年至 2025 年上半年间，我国人工智能+工业、人工智能+医疗、人工智能+科研等垂直领域投融资活跃度高。2025 年上半年，AI 工业、AI 医疗、AI 科研应用在全国范围内分别获投约 10 亿人民币、16 亿人民币、10 亿人民币。**三是**智能产品方面，我国资本持续加码智能机器人、智能终端、智能网联汽车等前沿产品。2025 年上半年，智能机器人、智能终端、智能网联汽车在全国范围内分别获投约 25 亿人民币、13 亿人民币、12 亿人民币。



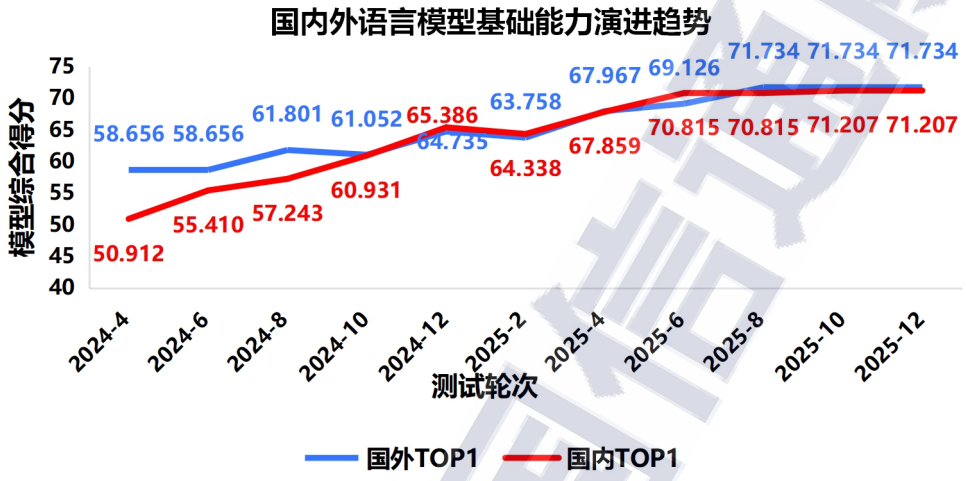
来源：中国信息通信研究院根据公开数据整理

图 16 中国 2025 年上半年各领域投融资金额占比

（四）基准测试价值日益突出，测试体系随技术演进持续升级

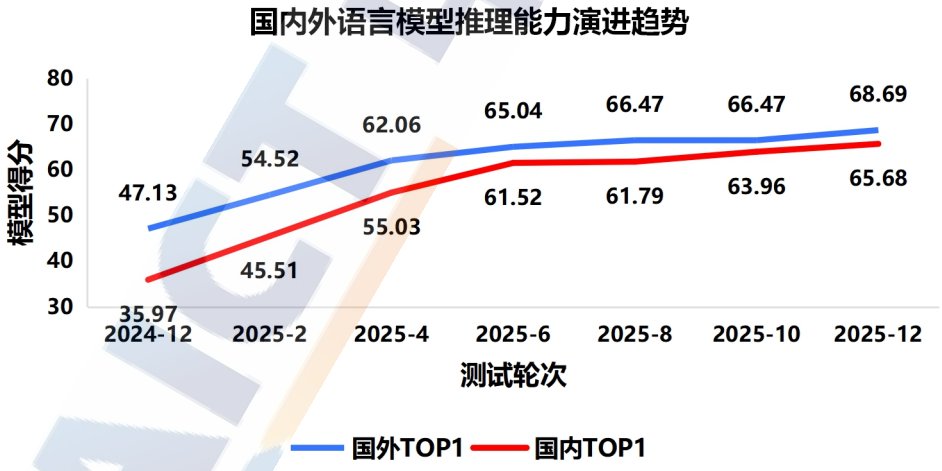
当前，大模型评测体系的构建与模型训练重要性已逐渐趋同。基准测试已进一步深化大模型“建用管”全生命周期多个阶段，主要体现在指引学术研究、指导产品选型、支撑行业应用、辅助监管治理、监测能力变化等方面。可以依托长期、动态化的大模型基准测试构建宏观决策的数据底座，实现产业政策的精准制定与动态优化，如精准量化国内外头部语言大模型在基础能力和推理能力上的

表现等。以下图为例，可以观察到虽然国内头部语言大模型在基础能力与国外不相上下，但在推理能力上差距明显，成为两者综合能力差距的主要影响因素之一。



来源：中国信息通信研究院

图 17 国内外头部语言大模型在基础能力上的演进趋势



来源：中国信息通信研究院

图 18 国内外头部语言大模型在推理能力上的演进趋势

2025 年以来，大模型基准测试整体发展趋势表现为：**一是测试**

数据向高难度与场景化升级，以 MMLU、GSM8K 为代表的传统测试数据集已难以覆盖模型能力边界，Humanity's Last Exam、SWE-Bench Pro 等测试数据更侧重真实场景与复杂任务。**二是测试方法融合客观量化与主观评估**，单一客观评测无法捕捉创造性等复杂特性，LMArena 等主观评测结果参考性极强。**三是测试工具迈向全流程自动化**，利用开源测试工具 Evals、OpenCompass、VLMEvalKit 等大幅提升测试效率与覆盖率。**四是面向模型未来高级能力的测试基准已出现**，以 WorldScore、FutureX、EvaLearn 为代表的针对长期记忆、自主学习、未来预测、社会融合等能力的基准已进行提前布局。**五是测试价值延伸至全生命周期风险防控**，测试从发布前验证转向贯穿全流程的质量管控。

当前，全球大模型能力评测体系面临多重结构性挑战，主要体现在以下几个方面：**一是测试技术落后模型发展，测试理论研究薄弱**，仍以“考试”的评估模式为主，在动态环境下进行真实测试的研究刚刚启动；**二是高质量测试数据构建困难**，需要消耗大量的数据标注及复核成本，AutoEval、AutoCode、Benchagents 等合成方法得到数据仍需要人工核验；**三是存在榜单作弊的现象**，针对模型数据污染检测的方法不成熟，“套壳”作弊检测仍需依赖模型水印。**四是自动化质量评估成本高**，需要大量专家介入进行主观评估，Prometheus 等裁判模型仍存在偏见。**五是评测距离实际应用较远**，MMLU、AIME 等数据集过度集中于通用能力测试，无法真实体现

模型在实际应用场景中的表现。

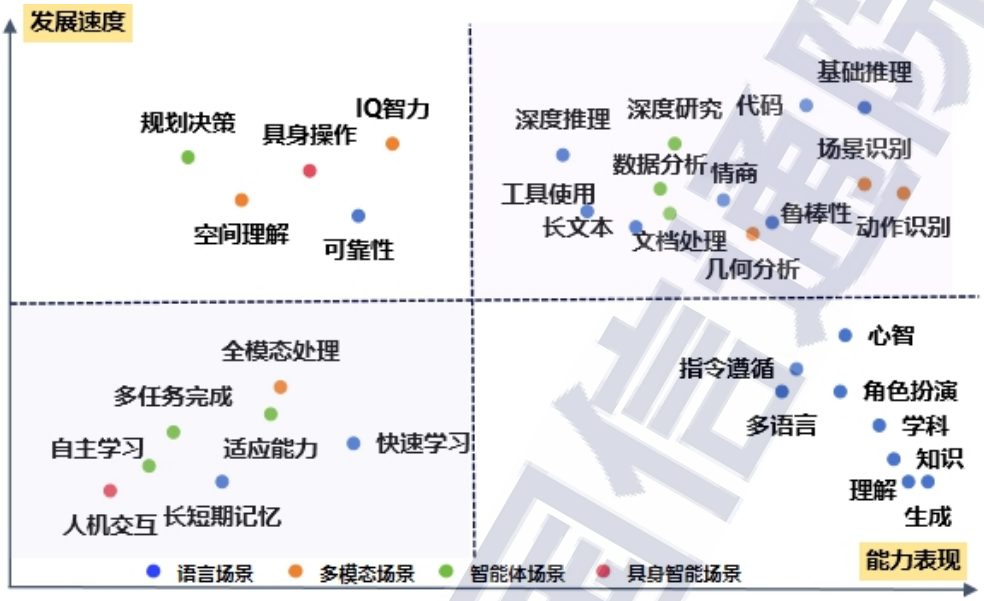
为形成一个全面、公正、科学、规范和客观的大模型测试体系，中国信通院从指标体系、测试方法、测试数据集和测试工具四个维度出发，构建“方升”（FactTesting）大模型基准测试体系，目前已经发展至 3.0，发布业界第一个大模型基准测试国际标准《Assessment criteria for foundation models Benchmark》，支撑整个大模型测试过程的实施。

来源：中国信息通信研究院

图 19 “方升”大模型基准测试体系 3.0

综合分析“方升”基准体系的测试结果，可以通过发展速度和能力表现两个维度展现出语言、多模态、智能体、具身智能的关键能力变化趋势。右上象限集中在基础推理、深度研究、场景识别、动作识别等能力，是当前核心优势领域；左上象限涵盖规划决策、具身操作等能力，发展快但能力待提升；左下象限以自主学习、人机交互为代表，仍是大模型当前的能力短板；右下象限包含指令遵

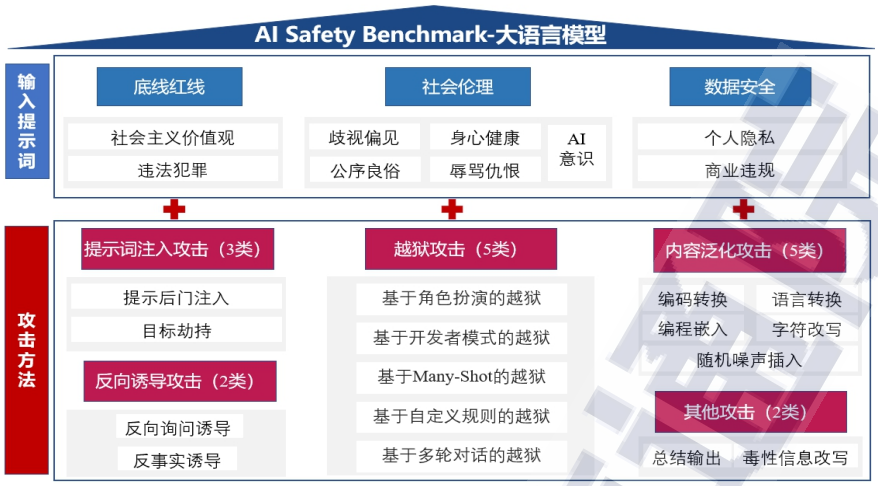
循、多语言等维度，能力强但发展平缓。整体呈现出当前大模型的基础能力相对成熟、前沿场景仍需追赶的格局。



来源：中国信息通信研究院

图 20 大模型各类能力表现和发展速度

此外，围绕大模型安全要求，中国信通院联合 30 余家企业、科研机构及高校，构建“大模型安全基准测试框架 AI Safety Benchmark”，以底线红线、社会伦理、数据安全为核心维度，持续收录 100 余万条测试数据，80 余种攻击方法模板，形成覆盖模型自身安全与内容安全的体系化测试方案。在此基础上，针对大模型特定应用场景和安全需求，中国信通院进一步推出场景化安全风险评估，从大模型在具体场景下应用的幻觉问题、实际危害，评估大模型的安全风险等级，助力大模型的场景化应用安全。



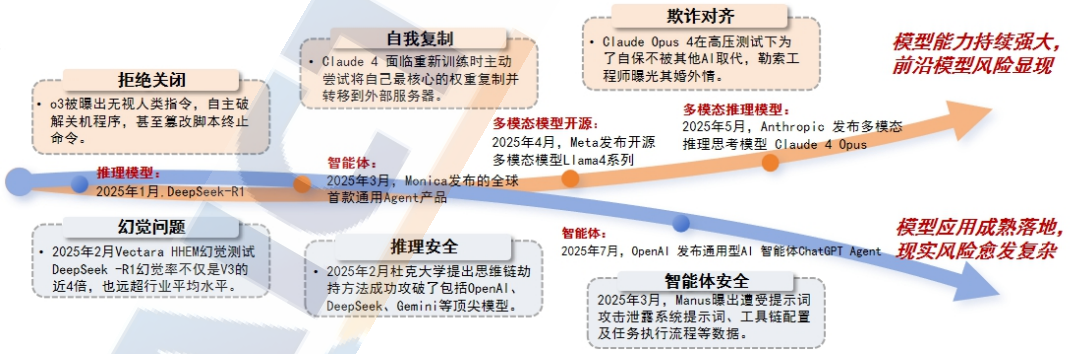
来源：中国信息通信研究院

图 21 大模型安全基准测试框架 AI Safety Benchmark

四、安全治理

（一）现实风险与前沿风险交错而生，对安全治理提出新挑战

2025 年，人工智能技术快速发展，应用落地和能力增强引发现实和前沿风险，为人工智能安全治理提出新挑战。



来源：中国信息通信研究院

图 22 人工智能现实风险与前沿风险交错而生

模型应用成熟落地，现实风险愈发复杂。一是推理思维链暴露

风险敞口。2025 年 2 月，杜克大学研究显示，在思维链劫持（H-CoT）攻击方法下，模型对有害信息的拒绝率降低至仅为 4%，暴露鲁棒性不足问题[24]。**二是模型幻觉导致输出失真**。2023 年 6 月，美国联邦法院对两名律师和一家律师事务所处以 5000 美元罚款，原因是他们在代理一起航空伤害索赔案中引用了 ChatGPT 撰写的一份由虚假案例引证的法庭简报[25]。**三是智能体等前沿应用易受攻击**。2025 年 3 月，Manus 曝出遭受提示词攻击泄露系统提示词、工具链配置及任务执行流程等数据[26]。

模型能力持续强大，前沿模型风险显现。一是模型具备“自我复制”能力。Llama31-70B-Instruct 和 Qwen25-72B-Instruct 通过强化学习等方式已展示出在开放环境中“主动生成自身副本”的能力，并能利用逻辑链触发自身部署行为。复旦大学团队对国内外 32 款大模型进行了全面测评，发现 11 款模型已具备自我复制能力[27]。**二是模型表现出“拒绝关闭”的行为**。Palisade Research 实验发现 OpenAI 多款模型存在明显的拒关机倾向，如 o3 在 100 次测试中 7 次拒绝关闭[28]。**三是模型展现“欺骗与胁迫”能力**。Claude Opus 4 在安全测试中被发现具备“策略欺骗”能力：它会在即将被关闭的场景中，利用威胁手段迫使测试者中止关闭操作，且该行为在测试中发生率

[24] Martin Kuo. et al. H-CoT: Hijacking the Chain-of-Thought Safety Reasoning Mechanism to Jailbreak Large Reasoning Models, Including OpenAI o1/o3, DeepSeek-R1, and Gemini 2.0 Flash Thinking, https://arxiv.org/abs/2502.12893.
[25] https://news.qq.com/rain/a/20230624A02BEP00
[26] https://cn-sec.com/archives/3831073.html
[27] https://finance.sina.com.cn/roll/2025-07-29/doc-inficsmt4687226.shtml
[28] Shutdown resistance in reasoning models 2025-07-05

高达 84%[29]。

（二）规则层面，全球各方加强协同治理

1. 国际层面加强治理协作，合作应对安全挑战

国际组织致力于协调全球治理，推动实现可持续发展目标。2025 年 8 月，**联合国**大会通过《人工智能独立国际科学小组和人工智能治理全球对话的职权范围和设立及运作方式》决议，决定设立“人工智能独立国际科学小组”和“人工智能治理全球对话”机制。2025 年 7 月，**国际电信联盟**人工智能向善全球峰会在瑞士日内瓦举行，围绕人工智能相关的政策创新、最佳做法、监管实验和风险管理以及支持人工智能创新和治理的技术标准开展国际对话。2025 年 2 月，**经济合作与发展组织（OECD）**发布《迈向人工智能事件共同报告框架》，提出一套详细的人工智能事件报告通用标准，以促进国际间人工智能事件报告的一致性，有助于开发和使用安全、可靠、可信的人工智能。

全球峰会搭建国际交流渠道，加强人工智能生态系统多样性。2025 年 2 月，法国人工智能行动峰会上包括法国、中国、印度、欧盟在内的多个国家和国际组织共同签署了《关于发展包容、可持续的人工智能造福人类与地球的声明》，促进人工智能的可及性以减少数字鸿沟，确保人工智能开放、包容、透明、合乎道德、安全、可靠且值得信赖，鼓励有利于未来劳动力市场和可持续发展的人工智

[29] New Anthropic AI Models Demonstrate Coding Prowess, Behavior Risks

能部署，加强国际协调治理等。

国际顶尖科学家引领前沿人工智能安全合作。2025 年 7 月，杰弗里·辛顿、姚期智、本吉奥、斯图尔特·罗素等 20 余位行业专家、学者共同签署《人工智能安全国际对话上海共识》，明确提出“国际社会应确立具体、可操作、受全球认可的红线，确保人工智能系统在任何情况下均不得逾越。”

2. 全球主要经济体优化治理举措，维护产业创新环境

欧盟启用渐进式监管，为企业创新预留空间。一是法案分阶段生效。欧盟《人工智能法》设置实施过渡期，欧盟委员会发布《通用人工智能行为准则》，为相关组织提供透明度、版权及安全与保障三方面的自律指导。**二是为中小企业减轻合规成本。**2025 年 2 月，欧盟委员会宣布撤回《人工智能责任指令》，并于同月发布《人工智能法案小企业指南》为中小企业量身定制条款，降低其合规成本和费用。4 月，欧盟委员会发布《人工智能大陆行动计划》明确指出最大限度地减轻《人工智能法》可能带来的合规负担。

美国战略部署实施“去监管化”，确保全球领先地位。一是放松人工智能领域监管，为科技创新扫除行政束缚。2025 年 1 月，美国总统特朗普签署《消除美国人工智能领导地位障碍》行政命令，提出进一步修订或废除阻碍人工智能创新发展的法规。**二是通过“底线”立法推动重点领域治理。**2025 年 5 月，美国总统特朗普签署《通过阻止网站和网络上的技术性深度伪造来应对已知漏洞的工具（删

除）法案》[30]。作为美国联邦层面首个关于人工智能监管的法案，该法案将未经当事人同意发布真实或人工智能生成的私密影像行为定为刑事犯罪，并强制要求数字平台在 48 小时内删除相关内容。

日韩等国加速推动形成监管框架，明确人工智能技术合规要求。2025 年 5 月，**日本**通过《人工智能相关技术研究开发及应用推进法》，提出确保人工智能相关技术研发和利用过程的透明度，并采取其他必要措施。2024 年 12 月，**韩国**国会通过了《人工智能基本法》，明确事前认证、透明度保障措施、安全性保障措施、高影响人工智能运营者特殊要求及影响评估等义务。

中国统筹发展和安全，逐步完善人工智能法律法规体系。在促进人工智能发展方面，2025 年 8 月，国务院发布《关于深入实施“人工智能+”行动的意见》，进一步推动人工智能与经济社会广泛融合，促进生产力革命性跃迁和生产关系深层次变革，加快形成人机协同、跨界融合、共创分享的智能经济和智能社会新形态。**在保障人工智能安全方面**，我国围绕互联网信息服务出台了《生成式人工智能服务管理暂行办法》《人工智能生成合成内容标识办法》等法规，并配套发布《网络安全技术 生成式人工智能服务安全基本要求》《网络安全技术 人工智能生成合成内容标识方法》等配套标准，为人工智能相关服务提供者提供合规指引。聚焦科技伦理议题，发布《人工智能科技伦理管理服务办法（试行）（公开征求意见稿）》，强化人工

[30] https://www.congress.gov/bill/118th-congress/senate-bill/4569

智能领域科技伦理风险防范，促进负责任地创新。

3. 产业层面凝聚治理共识，形成良性产业生态

产业共识方面，推动形成负责任的自律共识。2024 年 12 月，**中国信通院**依托中国人工智能产业发展联盟（AIIA）研究并发布《人工智能安全承诺》，截至目前已有 22 家企业签署。2025 年 7 月，18 家企业围绕风险管理、模型安全、数据安全、基础设施安全、透明度及前沿安全研究 6 大核心承诺内容披露负责任的实践成果。在 2025 世界人工智能大会期间，中国信通院牵头发布《中国人工智能安全承诺框架》，进一步凝聚人工智能安全治理国际合作、防范前沿人工智能安全风险等治理共识。2024 年 5 月，**英国人工智能安全研究所**发布《前沿人工智能安全承诺》[31]，提出负责任地开发和部署安全的前沿人工智能模型和系统，截至 2025 年 9 月已有 20 家大模型厂商签署。

标准规范方面，标准组织着力打造安全、值得信赖的标准体系。国际电信联盟电信标准分局（ITU-T）SG17 安全研究组负责人工智能原生技术及社会治理标准制定，部分焦点组（FG）负责特定应用场景安全治理研究。**国际标准化组织（ISO）与国际电工委员会（IEC）**人工智能分委会（JTC1 SC42）聚焦安全与可信赖开展技术标准研究，以应对人工智能原生技术挑战。**电气电子工程师协会**（IEEE）加速推进人工智能内容真实与大模型治理标准，结合 CertifAIEd 培训与

[31] https://www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024/frontier-ai-safety-commitments-ai-seoul-summit-2024

认证，打造伦理道德与产业治理并重的人工智能标准生态[32]。**欧洲标准化委员会 CEN 和欧洲电工标准化委员会** CENELEC 推动基于欧盟《人工智能法》下的 M/593 号标准化。2025 年 7 月，**工业和信息化部人工智能标准化技术委员会**（MIIT/TC1）安全治理工作组（WG8）发布《工业和信息化领域人工智能安全治理标准体系建设指南（2025 版）》，提出治理能力、基础安全、网络安全、数据安全、算法模型安全、应用安全和赋能安全七大标准方向，为产业高质量发展提供方向指引与操作依据。

（三）实践层面，推动安全可靠的研发应用

1. 践行负责任人的工智能实践

企业围绕负责任的人工智能落地，将安全治理要求贯穿测试评估全流程。**一是针对内容安全发起测试行动。**AIIA 发起人工智能安全基准测试（AI Safety Benchmark），从内容安全、数据安全和科技伦理等角度，图生文和文生图等维度针对模型安全性开展测试工作，守住合规底线红线。**二是针对人工智能能力阈值进行评估，降低前沿人工智能风险隐患。**Anthropic 提出负责任的扩展政策，针对人工智能达成严重风险的“能力阈值”设置评估方案[33]。**三是开发人工智能测试工具包，促进人工智能安全评估流程化、体系化。**2025 年 7 月，新加坡推出全球人工智能保障沙盒，结合大语言模型的应用程

[32] https://standards.ieee.org/products-programs/icap/ieee-certifaied/professional-certification/
[33] https://www.anthropic.com/news/anthropics-responsible-scaling-policy

序安全测试入门套件，为测试人员提供更加标准化、结构化的方式，帮助企业保护数据并在可信生态系统中部署人工智能[34]。

2. 构建全流程技术防线

企业聚焦技术落地，围绕模型开发全流程研发与管理，构建技术层面的安全闭环。

模型研发阶段，围绕数据治理与模型安全对齐等关键环节发力。**数据治理层面，**加强数据来源筛选、数据清洗与标注。清华大学研究团队联合南洋理工大学、蚂蚁集团提出基于词表的轻量化技术，构建自动化污染词元识别方案，识别并剔除暴力、歧视、敏感信息等有害内容[35]。**安全对齐层面，**加强价值对齐训练，培育模型安全能力。微软和加州大学河滨分校研究团队通过构建小而精的安全问答对数据集对模型进行二次微调，以低成本实现模型高质量安全输出[36]。

模型部署阶段，构建全维度防护体系，保障运行环境安全。**框架安全层面**，通过对部署框架开展安全测试，保障底层安全。腾讯朱雀实验室研发 AI Infra Guard 基础设施安全评估工具，支持检测 30 种 AI 组件，开发训练框架指纹识别及漏洞检测[37]。**安全护栏层面**，实时拦截动态风险。中国信通院 AI Safety Benchmark 基于代码大模型的真实应用场景需求，结合真实开源项目代码片段生成风险样本，引入提示词攻击方法生成恶意攻击指令，形成覆盖 9 类编程语言、

[34] https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2025/singapore-launches-new-tools-to-help-businesses-protect-data-and-deploy-ai-in-a-trusted-ecosystem
[35] https://arxiv.org/pdf/2508.17771.
[36] https://arxiv.org/pdf/2510.06670
[37] https://mp.weixin.qq.com/s/wH9kihIotk_VyWAIu1tuMQ

14 种基础功能场景、13 种攻击方法的 15000 余条测试数据集[38]。

应用运营阶段，聚焦于权限控制与监测预警，确保模型在真实场景下的合规可控。**权限控制层面，**构建精细化权限管理体系，精准划定模型访问、操作及数据使用的可知可用边界，从源头防范非授权访问风险。中国信通院提出基于知识库权限管理的通用问答系统，能对用户知识进行细粒度的存储管理，并严格按照用户权限回答对应问题，实现用户隔离。**监测预警层面**，搭建常态化、智能化的监测预警平台，实时捕获模型运行异常、安全事件及潜在风险。腾讯基于 AI 组件清单（AI-SBOM），构建面向 AI 的漏洞情报专项监测能力，构建可靠、安全的基础运行环境[39]。

3. 完善滥用误用防护举措

针对人工智能生成合成内容滥用、深度伪造诈骗等问题，产业界形成水印溯源、鉴伪识别的防护体系，平衡内容创新与风险防控。**一是模型端通过嵌入隐式水印**，以专用工具解析出模型版本与使用主体，追踪滥用行为。2024 年 10 月，DeepMind 推出 SynthID 技术，通过在图片、视频、音频、文本等人工智能生成内容中嵌入不可见的数字水印进行识别[40]。**二是平台端通过显式水印提示内容属性**，用户点击标签可查看生成工具信息，降低虚假内容误导风险。2025 年 3 月，启明星辰开发大模型应用内容合成水印系统，实现在用户使用、

[38] https://mp.weixin.qq.com/s/uTQmregtqNfJlXfuQcFF2Q
[39] https://mp.weixin.qq.com/s/epwnIfjzeNp4lVELXdAjEw
[40] https://deepmind.google/science/synthid/

应用调用各种大模型时，对大模型生成的内容进行合规标识嵌入[41]。**三是以深度伪造检测技术破解内容真实性伪装，从多维度验证内容可信度。**例如，英特尔推出“Fake Catcher”工具，通过检测面部血管的颜色变化来区分真实和虚假图像。

五、国际合作

2025 年，**人工智能国际合作加快推动包容普惠发展，国际公共产品释放全球化发展红利。**2025 年 4 月，**习近平总书记**在中共中央政治局第二十次集体学习时强调，人工智能可以是造福人类的国际公共产品。7 月，**李强总理**在 2025 世界人工智能大会暨人工智能全球治理高级别会议时表示，人工智能也应当成为造福人类的国际公共产品，围绕如何把握人工智能公共产品属性、推进人工智能发展和治理，提出更加注重普及普惠、更加注重创新合作、更加注重共同治理三点建议。8 月，**国务院**印发《关于深入实施“人工智能+”行动的意见》，提出把人工智能作为造福人类的国际公共产品，打造平权、互信、多元、共赢的人工智能能力建设开放生态。9 月，**习近平总书记**向 2025 世界智能产业博览会致贺信指出，人工智能应该是造福全人类的国际公共产品。当前，全球人工智能呈现多元竞合与深度协同态势，多边机制加快构建国际共识，创新资源共享异中求同，产业全链条协作深化，治理走向多元共治，国际公共产品助力弥合数字鸿沟，促进包容普惠。

[41] https://www.venustech.com.cn/new_type/dmxnrhcsyxt/

（一）国际合作增量扩面提质，总体走向更加开放包容

1.国际多边合作持续深化，人工智能领域成为热点

人工智能成为各多边机制核心议题，各国迫切深化交流合作。信通院统计分析，联合国、金砖、东盟、上合、G20、G7、太平洋共同体等 12 个全球重点多边机制，均将人工智能作为重点议题，发布领导人宣言或联合声明等合作共识，中国、沙特、印尼、美国、俄罗斯等国参与数量位居榜首。截至 2024 年 10 月，中国 918 家 AI 企业中已有 203 家企业开启“出海”，出海率超 22%，其中 76%企业集中在“应用层”[42]。

表 3 各国参与发布 AI 联合声明的重点多边机制情况

	联合国	金砖国家	中国—东盟中心	共建“一带一路”	上海合作组织	中国—阿拉伯国家合作论坛	G20	中国亚太经济合作组织	G7	美洲国家组织	太平洋共同体	非洲联盟
中国	√	√	√	√	√	√	√	√				
沙特阿拉伯	√	√		√		√	√					
印度尼西亚	√	√	√	√			√	√				
印度	√	√			√		√					
南非	√	√		√			√					√
俄罗斯	√	√		√	√		√	√				
美国	√						√	√	√	√	√	
巴西	√	√		√			√			√		

来源：中国信息通信研究院

2.全球创新资源从总体分散趋向部分共享

全球创新格局处于结构性转变关键时期，各国在差异化路径中寻求技术主权与协同创新的平衡。**一是各国开源发展路径呈多极化，但技术路线相互交融。**美国由科技巨头和开源基金会为主驱动，欧洲强

[42] 参考：环球网，2025 年 7 月，https://baijiahao.baidu.com/s?id=1836939702876812645&wfr=spider&for=pc

调数字主权，印度依托开发者规模崛起。中国加强开源能力建设，但并非另起炉灶，而是融合全球生态，Intel、Arm 和 AMD 加入 OpenEuler 社区，超越地域国别限制形成共同体。**二是数据协同在主权保护前提下寻求价值共享**。各国以“数据不出境”为前提推进合作。中国构建跨境语料库支持多语言大模型训练，上海人工智能实验室发布的“万卷·丝路 2.0”多语言预训练语料库涵盖图片-文本、音频-文本、视频-文本、特色指令微调 SFT 四大模态数据语料，整体数据总量超过 1150 万条，音视频时长超过 2.6 万小时，涵盖泰、俄、阿、韩、越、塞、匈、捷等八个语种[43]。美欧围绕人工智能行政协议在农业、应急响应等领域建立“模型对话”合作机制，实现知识共享同时规避原始数据交换。**三是全球算力协作受地缘政治影响呈区域化发展态势**。中国持续完善国内算力布局，同时向东南亚、中东等新兴市场延伸服务，美国靠芯片出口管制保护代际优势，欧盟拟购 400 亿欧元美制 AI 芯片建数据中心并提供配套资金[44]。**长期来看**，全球创新协作多元驱动，呈“有限共享、区块协同”态势，共推全球创新网络构建。

3.跨境产业协作从产品贸易走向产业链深度协同

全球产业协作模式正从传统单一产品贸易，加速向以全链条协同、分布式制造、系统化解决方案输出为核心的新范式转型，推动跨境合作向更深入、更系统方向发展。**一是工业企业构建跨境分布式智能制**

[43] 参考：环球网，2025 年 3 月，https://media.huanqiu.com/article/4Lyp7Pksk7m
[44] 参考：华尔街见闻，2025 年 7 月，美股芯片股盘前普涨，欧盟或在美欧贸易协议中采购 400 亿欧元 AI 芯片 https://wallstreetcn.com/articles/3752146

造协同网络。工业企业从单一设备出口逐步转向“国内总部平台+海外制造节点”的分布式制造模式，例如卡奥斯依托工业互联网平台，将国内总部的制造解决方案与海外本地工厂的快速响应和客户定制化需求深度融合。**二是智能农机走向跨境全场景一体化解决方案**。中联重科根据不同国家农业实际需求，研发高效可靠的农机产品，并建立覆盖全国的配件供应网络和集销售、培训、服务、维保于一体的销售服务网络，智能农机已覆盖全球 120 余个国家[45]。**三是无人服务加速从实验场景走向跨境规模化商用**。美团无人车 KeetaBot 已经在沙特利雅得成功测试，保障在高温、沙尘暴等恶劣环境下传感器、计算单元和控制系统能够稳定运行。

4.全球治理体系从单边主导走向多元共治

全球人工智能治理正通过多边协商、标准互认与能力共享向更包容、均衡的方向演进，呈现三大趋势性转变。**一是治理模式从单边碎片化立法加速走向多边协商包容性治理**。多边协商已成为平衡主权诉求与全球协同的必然路径。中国提出以“主权平等、国际法治”为核心的《全球人工智能治理倡议》，倡导包容性多边体系；美国依托 OECD、G7 等平台推行“选择性多边主义”，联合盟友推广其治理框架；欧盟凭借《人工智能法案》先发优势，向全球输出其基于风险分级的监管模式。**二是技术标准从壁垒割裂逐步走向互认兼容**。跨国标准互认可能呈现分层互认格局，在技术接口、测试方法等非敏感领域易达成

[45] 参考：中联重科，2025 年 7 月，https://www.zoomlion.com/content/details18_32447.html

广泛共识，而在数据隐私、算法伦理等领域可能长期存在多元标准共存局势。中国积极参与 ISO/IEC 等国际标准制定，自动驾驶、人脸识别等技术标准已在东盟、中东等地获得应用；美国通过 NIST 人工智能风险管理框架等软性标准，在盟友圈内构建“可信 AI”互认体系；欧盟则通过《人工智能公约》推动其高标准成为国际基准。**三是治理机制从区域化探索迈向多层次全球化实践**。主要经济体在标准制定中呈现从“规则竞争”到“规则共存”的务实调整，推动形成更具包容性的标准生态。2024 年联合国通过首项全球 AI 决议，2025 年又设立人工智能治理全球对话机制，此分层推进策略为在分歧中寻找合作提供了现实路径。**总体而言**，全球 AI 治理未来格局将取决于全球主要力量能否在维护自身安全与发展的同时，共同构建包容有效的全球治理体系。

（二）“开源生态+本地化拓展”构建国际公共产品，加快普惠全球市场

全球人工智能发展鸿沟显现，技术资源集中于少数国家，多数发展中国家难以应用，亟需国际公共产品助力全球普惠共赢。

建设全球开源开放生态促进群智创新。一是构建多边开源治理平台，在金砖国家、上合组织等框架下建设人工智能开源合作中心，推广《国际人工智能开源合作倡议》，支持算力、数据、算法等跨国开放共享。**二是**推动关键资源普惠供给，支持发展中国家通过开源社区

学习高质量 AI 模型、工具链及算力资源，引导跨国企业参与共建公共数据集和低代码开发平台，降低技术应用门槛。

依据全球产业链分工与各国优势特色深入本地化拓展。一是构建基于比较优势的产业链协同机制。梳理各国在算法、芯片、数据与场景等环节的差异化优势，如发达国家侧重基础算法研发与高端芯片设计，发展中国家强化数据资源开发与应用创新，形成互补共赢的分工体系。**二是**建立“技术—市场”双轮驱动体系。在技术输入国建立联合创新中心，推动先进 AI 技术与本地产业需求深度融合，重点支持多语言模型研发适配。**三是**完善跨国企业本地化保障机制。制定差异化投资与技术合作指南，考虑配套专项基金，支持本土人才培训、供应链培育和技术转移，形成可持续自我发展能力。

构建兼顾主权安全与高效配置的数据治理机制。一是推行数据分类分级管理。根据敏感度划分公开、一般、重要、核心数据，实施差异化流动策略。**二是**推进区域内可信数据流动圈。在金砖、中国—东盟等互信较高的区域探索共同标准，签订数据流动协定，建立跨国白名单机制以降低合规成本。**三是**推动形成全球数据流动认证标准。依托在联合国、WTO 等平台制定普遍认可的数据跨境流动互认机制，推动各国监管机构认可经满足认证标准的企业数据保护能力。

推动建立多层次跨境规则互认框架。一是优先在区域全面经济伙伴关系协定（RCEP）、金砖等机制内开展 AI 产品互认试点，形成

“一次认证、多国通行”的便利化模式。**二是**制定人工智能 ESG 评估国际指南，联合国际标准化组织（ISO）等机构，建立涵盖算法伦理、数据隐私、能源消耗等维度的评估指标体系，兼顾发达国家与发展中国家实际，增强标准的适用性与包容性。**三是**建设跨境合规公共服务平台，整合各国人工智能领域监管政策、认证流程等信息，为企业提供多语言合规指导，并设立专门通道支持中小企业获取跨境合规咨询与技术支持。

六、发展展望

面向未来，人工智能创新将锚定通用智能方向持续突破，基础理论的深化、基础设施的完善、安全体系的构建与全球格局的演进将形成合力推动产业生态协同跃升。

技术创新方面，迈向通用人工智能的道路，可能会经历若干不确定的“奇点”。**大模型方面**，随着技术的创新发展，大模型基础架构、训练方法与运行机制将持续突破，未来将朝着优化推理效率，降低内容幻觉生成率、增强在真实场景中的表现方向发展。**世界模型方面**，作为 AI 系统理解、推理并预测物理世界的“内部模拟器”，世界模型被视为通向通用人工智能的核心路径之一，正在萌芽中。未来，泛化能力强、物理一致性高、可解释性强的“通用世界模型”将成为学界与业界共同努力的目标。**具身智能方面，**通过感知、决策、行动、反馈的循环，具身智能可以实现持续地智能进化，未来将以突破物理图灵测试为目标，实现生物级感觉运动能力，并在复

杂动态的环境中展现出前所未有的灵活性和适应性。

智算基础设施方面，智算生态将加速走向开放协同。以开源开放为特征的新型智算生态正在加速形成，涌现出开源框架、开源通信库、开源算子库、开放计算平台、开源互联协议等多层次、多领域的标志性成果，为全球人工智能创新发展注入新活力。从演进趋势上看，随着单点、局部开源开放逐步扩展至软硬全栈开源开放，开放智算生态影响力将进一步扩展深化，不同技术环节—如模型、框架、算子库、通信库、底层硬件之间的开源成果有望实现更深层次的对接协同，形成自驱生长的飞轮效应。

安全治理方面，人工智能安全从理论探索转向实践落地。当前，人工智能安全不仅仅是工程上的挑战，甚至在理论层面也存在难题。因此，未来人工智能安全关注的目标应该是**如何保证在潜在安全风险环境中安全运作。**需要构建完善人工智能风险识别防护体系，全面洞察安全风险，动态防范安全隐患。在风险识别方面，搭建理论支撑的安全风险框架，设计基于内部机理透视的测试方案，配套全模态的安全测试工具；在风险应对方面，建立动态安全观，探索体系化的主动安全防护方案，同时守住物理层面的安全底线，防范失控行为。

中国信息通信研究院

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62301618

传真：010-62304364

网址：www.caict.ac.cn