

中银研究产品系列

- 《经济金融展望季报》
- 《中银调研》
- 《宏观观察》
- 《银行业观察》
- 《国际金融评论》
- 《国别/地区观察》

作 者: 许天衣 中银理财
电 话: 010 - 8393 4068

签发人: 陈卫东
审稿人: 周景彤 梁 婧
联系人: 程栖云 刘佩忠
电 话: 010 - 6659 4016

* 对外公开
** 全辖传阅
*** 内参材料

中美人工智能赋能产业发展的 现状、趋势及政策建议^{*}

人工智能作为新一轮科技革命和产业变革的核心驱动力,既是技术创新的前沿,也是全球竞争的战略高地。当前,我国正处于加快培育新质生产力、全面推进高质量发展的关键阶段。习近平总书记多次强调,“人工智能是引领这一轮科技革命和产业变革的战略性技术”“发展人工智能,将为我国构建现代化经济体系、实现高质量发展提供重要支撑”。2025年国务院印发的《关于深入实施“人工智能+”行动的意见》明确提出,要以科技、产业、民生、治理等领域为重点,推动人工智能与经济社会各行业深度融合,构建人机协同、跨界融合、共创分享的智能经济与智能社会新形态。在此背景下,我国人工智能发展亟需构建“国家引领、产业共建、企业突破”的协同发展体系,形成上下贯通、协同联动的发展合力。

中美人工智能赋能产业发展的现状、趋势及政策建议

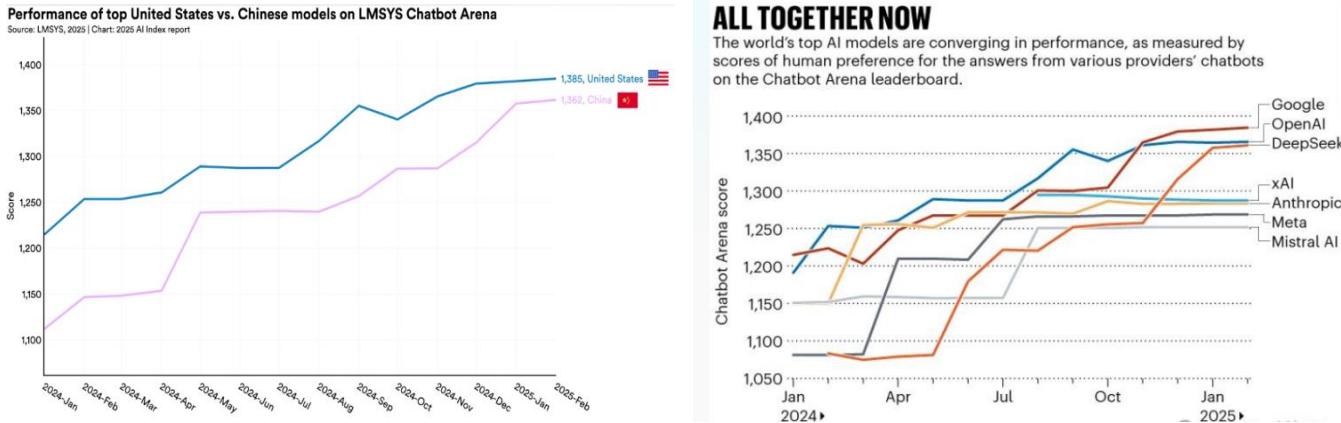
当前，以大模型为代表的人工智能技术正加速演进，成为引领新一轮科技革命和产业变革的战略性、颠覆性力量。习近平总书记多次强调，“人工智能是引领这一轮科技革命和产业变革的战略性技术”“发展人工智能，将为我国构建现代化经济体系、实现高质量发展提供重要支撑”。同时，党的二十大报告明确提出，要“推动战略性新兴产业融合集群发展，构建新一代信息技术、人工智能、生物技术、新能源、新材料等一批新的增长引擎”，将人工智能置于国家创新驱动发展战略的核心位置。

在全球格局深刻调整的背景下，中美两国作为全球人工智能发展的主要引领者，在大模型领域的竞争已从技术层面延伸至基础设施、产业生态、标准体系与治理规则的全面竞争。当前，大模型正通过三大路径深度赋能经济社会：一是驱动智能手机、PC等端侧设备向智能化、个性化跃迁，重塑人机交互范式；二是催生AI原生应用（AI-Native）与超级应用融合创新，重构数字内容生产与服务模式；三是深入渗透制造、金融、医疗、政务等垂直行业，推动全要素生产率提升并孕育新型商业模式。本文将通过对中美大模型的发展现状与差异，系统梳理两国在技术研发、生态构建与应用实践中的竞争态势，并提出有针对性的政策建议，以期为我国人工智能高质量发展提供决策参考。

一、中美大模型技术发展现状分析

根据斯坦福大学人工智能研究所发布的《2025年人工智能指数报告》，中美顶级AI大模型在MMLU和HumanEval等主要基准的性能测试差异已经由2023年的17.5%大幅缩至2024年的0.3%。其中，TOP1和TOP10模型的差距从12%缩小至5%，以DeepSeek为代表的开源模型正逐步缩小与海外头部模型的性能差距。从2025年以来的模型进展迭代情况来看，中国在模型数量、专利申请量和应用落地速度方面具备优势，但在基础模型的迭代速度上略有落后；而美国在顶尖模型的产出和商业化能力上仍保持领先地位。从技术路线来看，开源与闭源模式已形成竞争互补格局，中国通过政策引导和生态共建推动开源模型发展，美国则依托模型垄断和高付费意愿维持闭源优势。

图1：中美大模型性能表现差异逐渐缩小



资料来源：斯坦福大学人工智能研究所《2025年人工智能指数报告》

整体来看，“模型-算力-应用需求”构成人工智能产业迭代升级的闭环，当前大模型技术的发展重心已经从预训练阶段过渡到后训练和推理应用阶段，并逐渐形成以基座大模型为基础，通用及专业智能体（Agent）能力为延伸的完整产业生态。

（一）美国：推理需求全面爆发，商业化闭环逐渐跑通

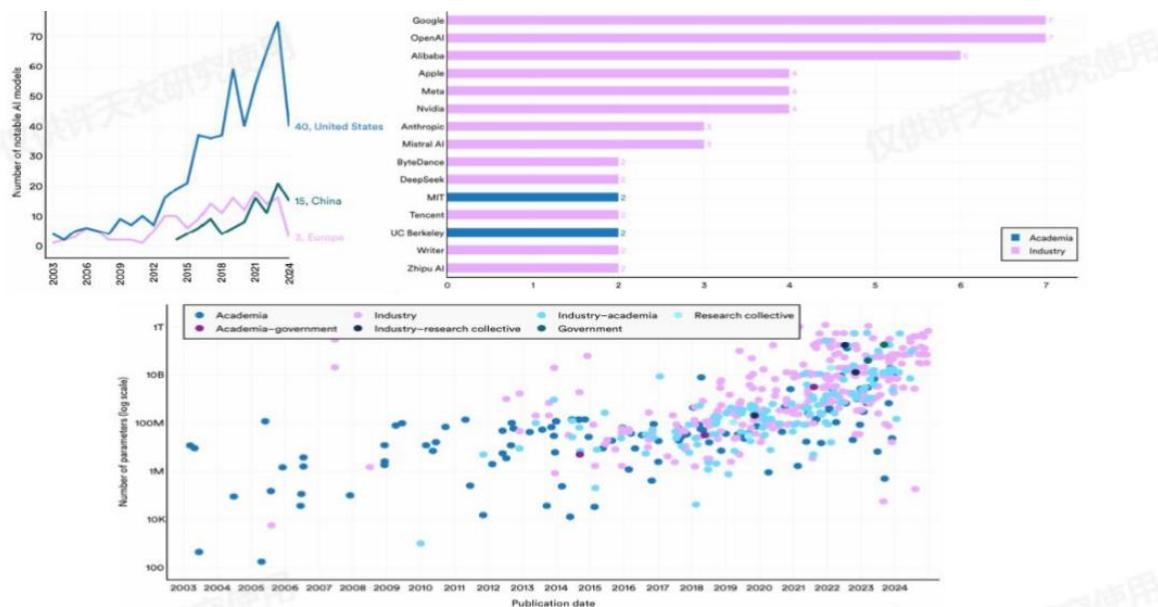
从供给端来看，当前美国人工智能产业发展仍处于领先地位，AI工具的渗透率和商业化程度提升明显。随着模型参数规模的饱和式增长，预训练投资的边际效用递减，单纯依靠扩大模型规模来提升性能的策略变得不可持续，头部模型厂商更加注重模型优化和推理效率提升。传导至算力芯片需求上，美国AI企业正从采购GPU向ASIC芯片快速过渡。2025年，美国科技巨头已开始大规模部署专为AI推理设计的ASIC芯片，这类芯片在特定任务上的能效比GPU高出5-10倍。例如，谷歌已推出第六代TPU芯片Trillium，专为大模型推理优化；Meta也宣布将在2025年底部署其自研的AI推理ASIC芯片，以应对模型推理成本下降带来的挑战。

从需求端来看，Agent产品通过沉淀用户数据反哺大模型产品，叠加模型推理成本显著下滑，应用端的需求吞吐量实现快速增长。API调用数据方面，过去一年，谷歌产品和API每月处理的token量¹从2024年4月份的9.7万亿到目前的1000万亿，增

¹ Token是模型处理文本的基本单位。

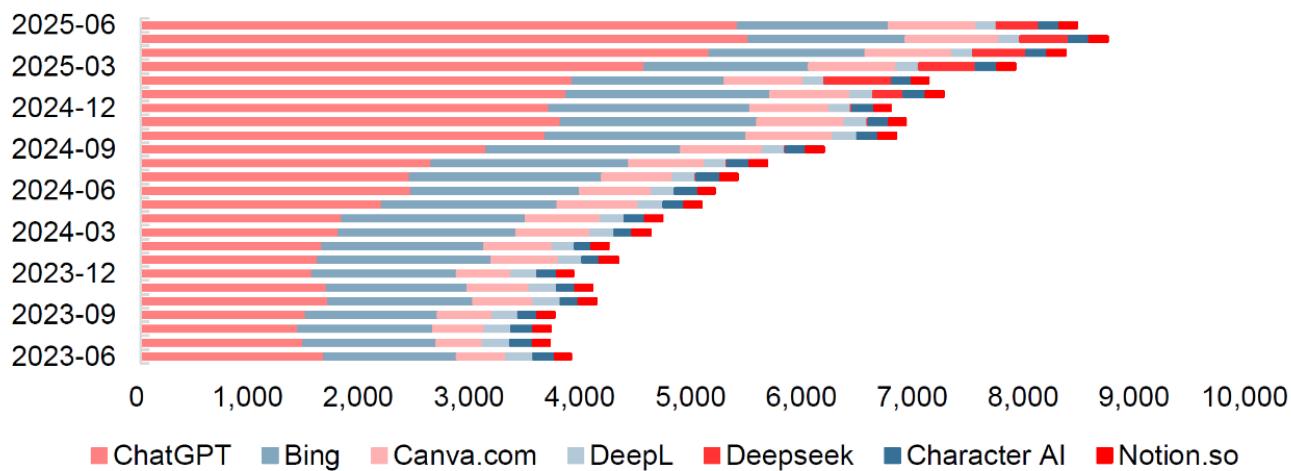
长了100倍。产品活跃数据方面，ChatGPT日活流量呈上升趋势，网页端日活用户维持在2亿左右，同比增长135.7%；移动端日活达到1.84亿，同比增长328.6%。此外，Grok、Perplexity、Claude等头部产品的日活也呈现稳定增长态势。

图2：全球新发大模型数量同比下降



资料来源：斯坦福大学人工智能研究所《2025年人工智能指数报告》

图3：头部AI应用2023.6-2025.6网页端月度访问量（百万次）



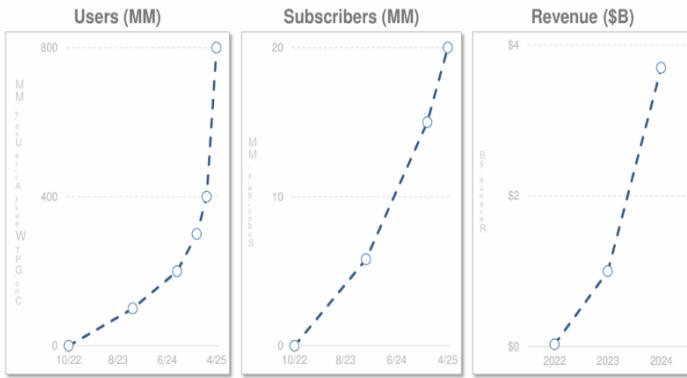
资料来源：Smilarweb

在赋能产业发展方面，美国大模型正广泛渗透到金融、医疗等传统行业。根据麦肯锡2024年调研，美国已有约49%的企业称AI为企业实现了降本，其中金融行业渗透率最高，已达78%。OpenAI的GPT-4在复杂医学问诊的测试中表现甚至优于医生与AI协作小组，2023年美国FDA共批准了223款AI医疗设备，较2015年增长近40倍。在原生应用方面，美国大模型应用的渗透率已迎来关键拐点，现象级AI应用如ChatGPT正显著提高用户使用粘性。随着用户迁移成本逐渐变高，头部模型在市场中占据更有利位置。截至2025年4月，凭借现象级的“吉卜力”照片风格转化功能，ChatGPT月活跃用户(MAU)接近9亿，预计年内将突破10亿。OpenAI预计到2025年末，AI Agent及其他新产品合计销售额将超越ChatGPT，成为新的收入增长引擎。

图4：OpenAI 2023–2029年盈利预测（左）及ChatGPT用户、订阅、收入增长（右）

OpenAI's Bold Revenue Forecasts

The company projects annual revenues to increase 100-fold by 2029, driven mostly by ChatGPT, and later, other products like video generation, search and other software.



资料来源：The Information（左），swsresearch（右），10/22–4/25

（二）中国：AI应用多点开花，商业化潜力仍待释放

从供给端来看，算力芯片限制国产大模型的研发推进，国内头部大模型产品仍处于L2²水平，通用Agent能力仍待提升。中国大模型产业在供给端面临算力限制的挑战，市场担忧中国大模型能力被美国再度拉大。但从产业现状和科技战略的角度来看，中美大模型的参数性能差距并不明显，当前仅在商业化场景和节奏上存在结构性差异。由于技术积累、产业生态和政策环境的不同，中美大模型产业形成了差异化的技术路

² L2级别大模型具有一定的推理和决策能力，能在用户设定的任务框架内执行逻辑推理和决策，但通常一次只能处理一个相对简单的任务步骤，任务规划较为线性。

径：即美国以OpenAI、Gemini的闭源主导，国内以DeepSeek、Qwen、Kimi2 的开源模型主导，但以中国为代表的开源模型性能与美国主流模型的差距逐步缩小。

图 5：大模型产品性能级别分类（OpenAI 定义）

OpenAI Imagines Our AI Future

Stages of Artificial Intelligence

Level 1	Chatbots, AI with conversational language
Level 2	Reasoners, human-level problem solving
Level 3	Agents, systems that can take actions
Level 4	Innovators, AI that can aid in invention
Level 5	Organizations, AI that can do the work of an organization

资料来源：Bloomberg, OpenAI

从需求端来看，开源技术生态推动大模型产品使用量全面繁荣。尽管Deepseek R1 自发布以来浏览器端口的流量绝对值持续下滑，6 月份环比下滑近 30%，但从国内部署接入Deepseek大模型的第三方平台来看，百度、阿里、腾讯元宝等均保持快速增长，自R1 发布以来已增长 20 倍。其中 2025 年 5 月豆包模型日均token调用量达 16.4 万亿，较 4 月和去年 12 月增幅 29% 和 310%。从备受市场关注的Deepseek模型开发节奏来看，由于其定位于基础模型研究机构，或收缩算力资源优先发展基础大模型，而对商业化产品并未有过多投入及迭代（同期豆包加大投流力度）。虽然每百万Tokens的价格仍然较低，但考虑到结果反馈的延迟时间、吞吐量和上下文长文本能力等因素，以及具体工作负载的性质和用户对Token的实际需求，Deepseek模型的性价比优势仍待进一步提升。

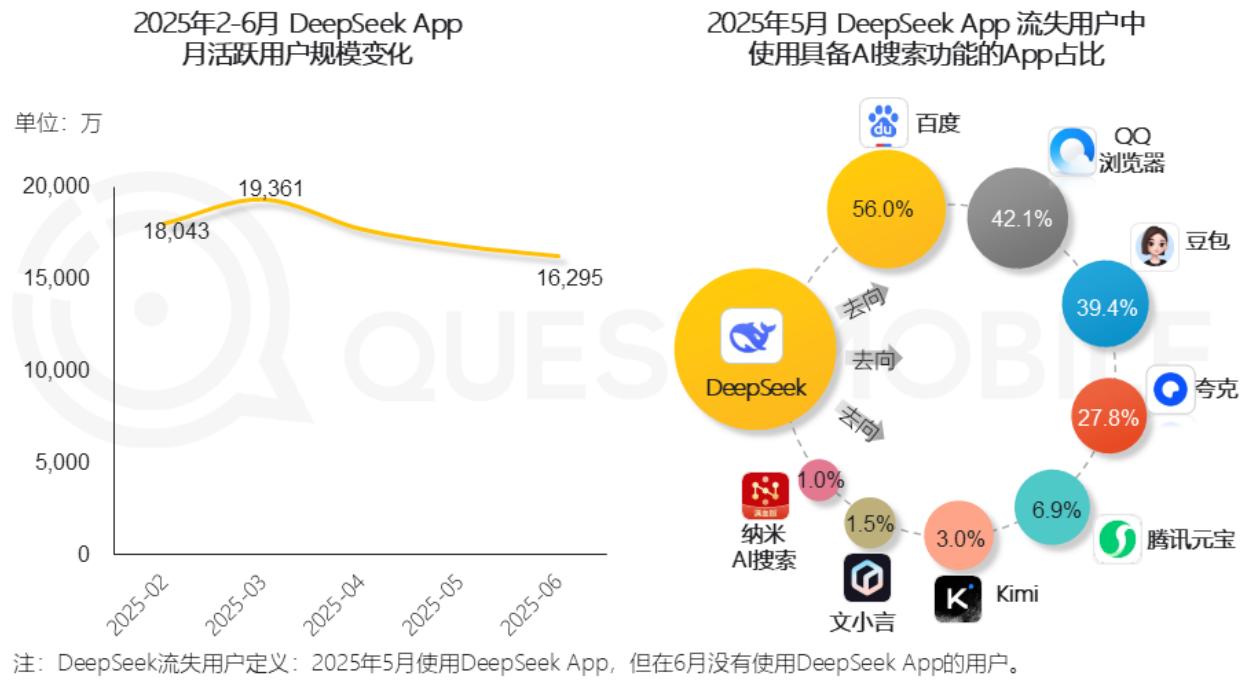
图 6：全球主流（上）及开源（下）大模型性能评分榜

Model	Arena Elo	Coding	Vision	AAII	MMLU-Pro	ARC-AGI	Organization	License
Gemini-2.5-Pro	1468	1470	1303	65	86.2	4.9	Google	Proprietary
Grok-4-0709	1433	1443	1271	68	86.6	16	xAI	Proprietary
GPT-5	1432	1462	1282	68	87.3	9.9	OpenAI	Proprietary
GLM-4.5	1431	1451		56	83.5		Z.ai	MIT
ChatGPT-4o-latest (2025-03-26)	1429	1434	1299	40	80.3		OpenAI	Proprietary
DeepSeek-R1-0528	1425	1436		59	84.9	1.3	DeepSeek	MIT
o3-2025-04-16	1424	1441	1261	67	85.3	6.5	OpenAI	Proprietary
Grok-3-Preview-02-24	1423	1439		46	79.9		xAI	Proprietary
Claude Opus 4.1 (thinking-16k)	1421	1475		61	87.8		Anthropic	Proprietary
Qwen3-235B-A22B-Instruct-2507	1420	1461		51	82.8	1.3	Alibaba	Apache 2.0
Claude Opus 4.1	1419	1465		49	87.3		Anthropic	Proprietary
DeepSeek-V3.1	1419	1422		49	83.3		DeepSeek	MIT
Qwen3-235B-A22B-Thinking-2507	1418	1447		64	84.3		Alibaba	Apache 2.0
DeepSeek-V3.1-thinking	1416	1437		60	85.1		DeepSeek	MIT
GPT-4.5-Preview	1415	1419	1240	42	81	0.8	OpenAI	Proprietary
GPT-5-chat	1411	1421	1289	67	86.5	7.5	OpenAI	Proprietary
Gemini-2.5-Flash	1409	1420	1275	58	83.2	2.5	Google	Proprietary
Mistral Medium 3.1	1408	1418		40	77.2		Mistral	Proprietary
MAI-1-preview	1407	1417					Microsoft	Proprietary
Gemini-2.0-Pro-Exp-02-05	1398	1396	1220	38	80.5		Google	Proprietary

Model	Arena Elo	Coding	Vision	AAII	MMLU-Pro	ARC-AGI	Organization	License
GLM-4.5	1431	1451		56	83.5		Z.ai	MIT
DeepSeek-R1-0528	1425	1436		59	84.9	1.3	DeepSeek	MIT
Qwen3-235B-A22B-Instruct-2507	1420	1461		51	82.8	1.3	Alibaba	Apache 2.0
DeepSeek-V3.1	1419	1422		49	83.3		DeepSeek	MIT
Qwen3-235B-A22B-Thinking-2507	1418	1447		64	84.3		Alibaba	Apache 2.0
DeepSeek-V3.1-thinking	1416	1437		60	85.1		DeepSeek	MIT
GLM-4.5-Air	1391	1416		49	81.5		Z.ai	MIT
Qwen3-30B-A3B-Instruct-2507	1382	1425		46	77.7		Alibaba	Apache 2.0
Qwen-VL-Max-2025-08-13	1381	1440	1254				Alibaba	Apache 2.0
kimi-k2-0711-preview	1380	1402		49	82.4		Moonshot	Modified MIT
DeepSeek-V3-0324	1377	1391		44	81.9		DeepSeek	MIT
DeepSeek-R1	1373	1382		50	84.4	1.3	DeepSeek	MIT
Qwen3-235B-A22B	1369	1394		48	82.8		Alibaba	Apache 2.0
gpt-oss-120b	1368	1409		61	80.8		OpenAI	Apache 2.0
Qwen3-Coder-480B-A35B-Instruct	1358	1406		45	78.8		Alibaba	Apache 2.0
Gemma-3-27B-it	1357	1350	1222	25	66.9		Google	Gemma
Minimax-M1	1351	1369		53	81.6		Minimax	Apache 2.0
Qwen3-32B	1342	1376		44	79.8		Alibaba	Apache 2.0
Llama-3.3-Nemotron-Super-49B-v1.5	1340	1359		52	81.4		Nvidia	Nvidia Open

资料来源：<https://openlm.ai/chatbot-arena/>（截至2025年9月1日）

图7: Deepseek APP 月活用户(左) 被分流至其他应用(右)

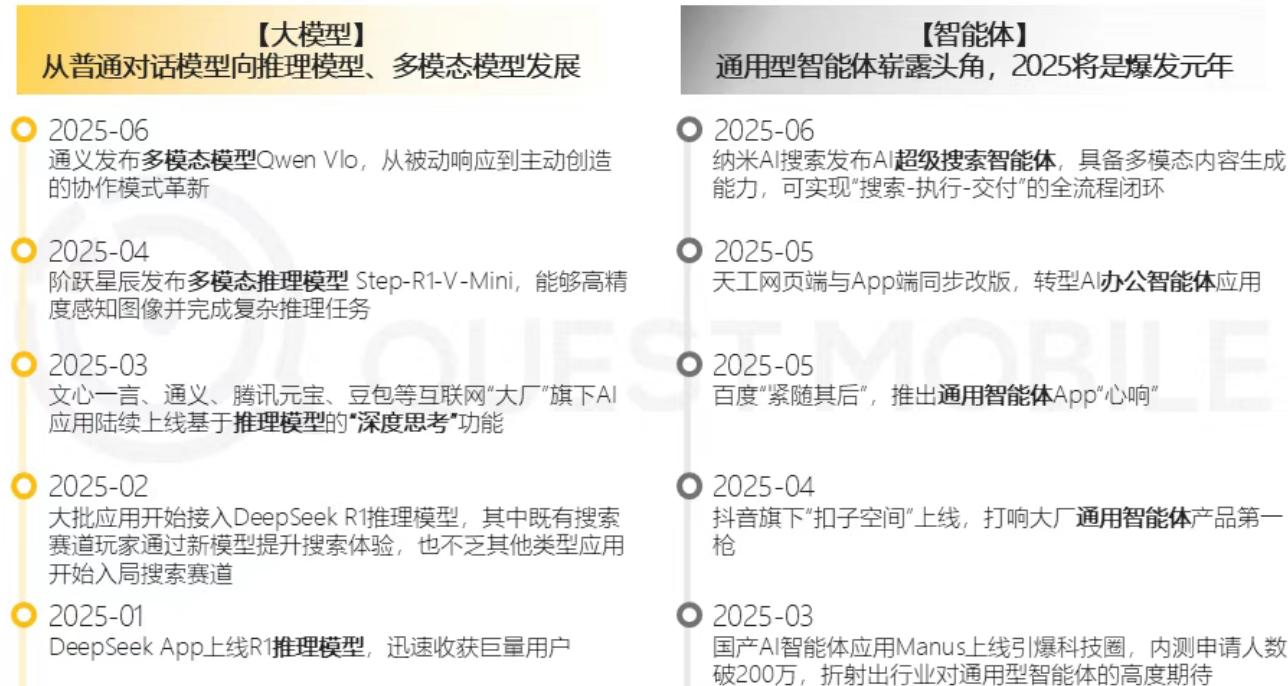


资料来源: QuestMobile

(三) Agent 成为人工智能应用的聚焦点

人工智能应用的迭代路径可分为逐步演进的三阶段,中美之间的产业进展虽有结构性错位,但智能体已成为未来竞争的聚焦点。第一阶段,各公司通过扩展参数和数据量提升预训练性能,形成ChatGPT、文心一言等初代聊天机器人,市场关注点主要在安装量和月活。第二阶段,DeepSeek R1开源降低推理成本后,头部厂商转向优化工程与效率,强调用户体验和服务粘性。智能体开始具备任务执行与思考展示功能,Token消耗大幅增加,尽管准确性有限,但应用渗透率显著提升,国内也出现了如Manus等雏形产品。第三阶段,应用正向多智能体集成演进,围绕协同工作、编程、多模态等能力快速迭代。用户的迁移成本因数据沉淀而上升,并逐渐锁定特定模型或供应商,产业需求拐点正在到来。

图8：2025年上半年大模型及智能体产品重要事件



资料来源：QuestMobile

图9：智能体分类及迭代情况



资料来源：中信建投《AI新纪元：砥砺开疆·智火燎原》

二、中美人工智能技术赋能产业发展的趋势

当前，人工智能技术赋能产业主要围绕以下路径展开：端侧消费电子（AI手机、眼镜等）、AI原生与超级应用（如C端的豆包、Kimi等）³及垂类场景赋能（如B端的PaaS、SaaS公司）。其中，Agent能力将成为未来人工智能流量入口的核心竞争力，各应用场景都将围绕大模型能力展开多维协同的融合趋势，通过MCP协议等标准化接口共同为用户提供无感的AI体验，形成“硬件-软件-平台”的协同生态。

表1：人工智能技术赋能产业发展的主要商业模式

AI 流量端口	主要商业模式	规模效应
端侧硬件	硬件销售+增值服务	中等
AI 原生与超级应用	API 调用+订阅制（C 端）	高（API 调用量增长近 100 倍）
垂类应用场景	订阅付费（B 端）与行业解决方案	低（下游应用场景推广节奏较慢）

资料来源：公开资料整理，中银理财

（一）端侧硬件：形态重构、交互方式与生态重塑

端侧硬件是人工智能应用的关键物理载体，AI技术正驱动消费电子（智能手机、智能眼镜、个人电脑、玩具）、汽车、机器人等多类设备实现形态重构、交互方式与生态重塑。从硬件性能来看，中美两国在AI芯片领域存在一定的技术差距。在高端AI芯片领域，美国占据绝对优势，英伟达的GPU、苹果的A系列芯片、高通的骁龙系列在算力、功耗和能效比上均领先于中国；在消费级芯片领域，我国在主控、通信、存储等芯片方面具备竞争优势，例如恒玄科技、乐鑫科技等公司在智能眼镜、耳机、玩具等广泛的消费场景中具备较强的产品竞争力。主机产品方面，小米推出AI眼镜，搭载超级小爱，由于眼镜本身需长时间佩戴的特性，有望成为AI的高频入口。

³ 超级应用：指拥有庞大用户数、众多场景功能的平台 APP，旨在提供一致和个性化应用体验。为终端用户提供一套核心功能，例如一站式完成社交、订餐、支付房租、咨询医生等，同时提供相关入口，方便访问独立创建的迷你应用。

图 10：端侧 AI 产业生态全景图



资料来源：艾瑞咨询

图 11：小米 AI 眼镜主要功能（左）及发布后用户规模趋势（右）



资料来源：QuestMobile

1. AI手机

从软件开发来看，端侧AI的软件适配能力直接影响用户体验和应用生态。随着云端大模型的迭代发布，端侧模型也将同步升级，AI手机或在系统级的操作、跨应用的调度、复杂的场景分析、智能决策执行等方面有望逐步取得突破。从未来发展方向来看，当前中美两国主要手机厂商的技术路线逐渐聚焦至“端云协同”，如谷歌、OPPO、荣耀等手机厂商均与国内大模型厂商展开深度合作，为其AI助手提供底层技术支持。

表2：端侧AI产业链分布

产业链	核心领域	技术/产品/服务	代表企业/案例
上游硬件	AI芯片	RISC-V架构、6nm制程等技术提升性能，适配端侧大模型的轻量化芯片设计	瑞芯微、兆易创新(国产厂商) 高通、苹果(国际厂商)
	存储、电池	3D封装技术提升存储密度 高密度电芯延长续航能力	三星、SK海力士(存储)、松下、宁德时代(电池)
中游集成	设备厂商	通过端侧AI优化产品体验(如本地语音助手、图像处理)	华为(鸿蒙系统)、苹果(A系列芯片)、联想(AIPC)
	模型服务	轻量化大模型(如MiniCPM系列，适配终端设备的模型部署与优化)	面壁智能(MiniCPM)、DeepSeek、阿里通义
下游应用	企业场景	医疗：AI辅助诊断、影像分析 工业：质检机器人、人形机器人	推想科技(医疗) 商汤科技(工业)
	消费场景	可穿戴设备(AI耳机、智能手表) AI玩具	Oura Ring(健康监测)

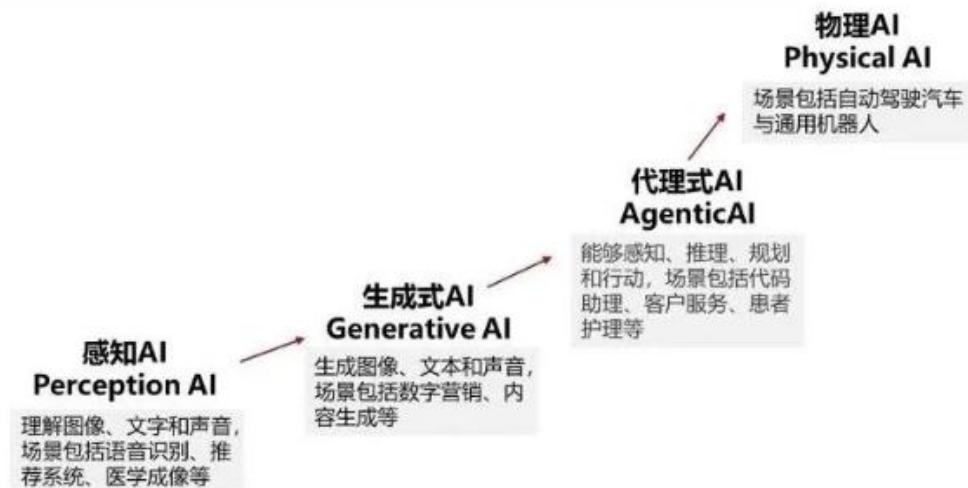
资料来源：公开资料整理，中银理财

从用户粘性与付费意愿来看，端侧消费电子的用户粘性主要取决于设备的AI功能深度与用户体验。根据Counterpoint Research预测，2024年全球AI手机出货量将超过1亿部，2027年将达到5.22亿台，占智能手机市场的40%。今年一季度，受国补政策与销售旺季叠加影响，我国智能手机出货量达7160万部，同比增长3.3%。其中，“AI手机”已成为驱动用户换机的核心卖点。展望未来，端侧设备是AI商业化的核心载体，通过隐私保护、实时响应、成本优化等优势，能够做到更懂用户、更懂物理世界，实现智能体的主动交互。

2. AI+具身智能/自动驾驶

具身智能(Embodied AI)指的是将人工智能赋予物理实体(物理AI)，如机器人或汽车，使其能够与现实世界进行交互并执行任务。具身智能的核心在于将大模型与机器人、传感器和真实环境结合，让AI在复杂情境中自主学习与决策。展望未来赋能路径：**一是**人机交互升级，通过自然语言与环境感知，让机器人具备接近人类的交互能力；**二是**任务执行自动化，AI能够结合感知、规划与运动控制，在动态环境中完成多样任务；**三是**跨模态学习与迁移，具身智能通过结合语言、视觉与动作信息，实现从虚拟世界到现实世界的知识迁移，推动智能体不断进化。

图12：英伟达定义的人工智能发展的四个阶段



资料来源：2025CES黄仁勋演讲

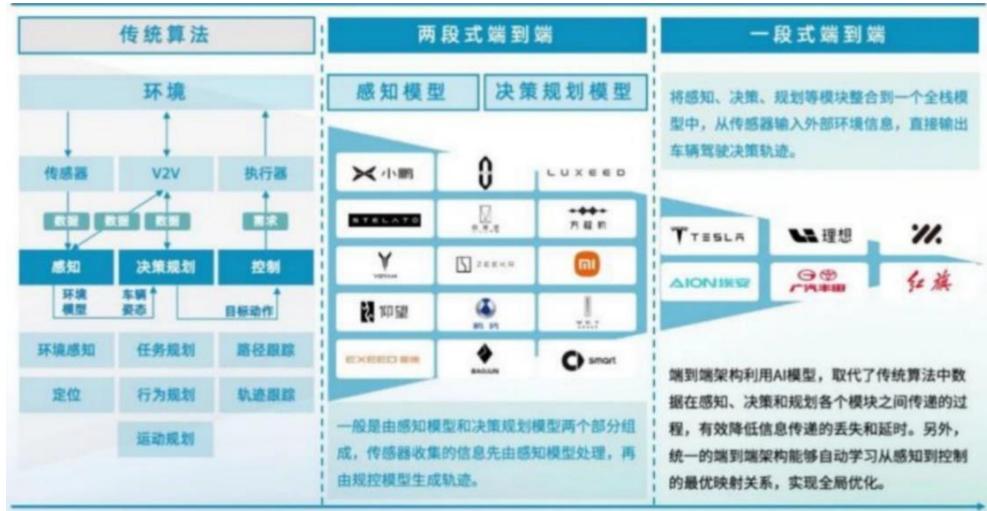
通用人形机器人本体可分为“大脑”、“小脑”和“肢体”三部分，分别对应决策交互模块、运动控制模块和执行模块。其中，人形机器人“大脑”的核心为人工智能大模型技术，通过多模态模型建模、强化学习、地图创建和数据训练，能够管理和协调机器人的各种功能。“大脑”是机器人智能与高级决策的核心，也是具身智能时代机器人区别于程序控制机器人(传统工业机器人、协作机器人等)的关键环节。目前具身智能大模型仍有数据集不够、思考跟不上运动、缺乏生态等主要痛点，但随着合成数据使用、模型持续迭代，未来将有效解决上述问题。随着大模型快速迭代，供应链快速降本，两大因素加速以人形机器人为代表的具身智能商业化落地。

图13：人形机器人的“大脑”、“小脑”及“肢体”



资料来源：《人形机器人产业发展研究报告（2024年）》中国信通院

图14：自动驾驶端到端AI技术路线



资料来源：量子位

在自动驾驶领域，从系统架构角度看，技术层面需完成感知、决策与执行三大核心任务，分别对应传感器系统、车载智能计算平台及车辆执行机构三个组成部分，整体流程极为复杂，并且决策容错率要求极低。首先，人工智能技术通过融合摄像头、

雷达及其他传感器采集的环境信息，实现对周边环境的识别与车辆自身位置的精准定位。其次，基于感知数据，AI系统需进行动态场景预测，制定合理的驾驶决策，并规划相应的行驶路径与操作指令。最后，AI通过控制车辆的转向、加速、制动等执行机构，实现对车辆的实时操控，确保其沿预定路径安全、稳定行驶。

(二) AI原生与超级应用：面向C端的软件商业模式

当前全球AI软件市场正呈现两种商业模式并行发展，一种是AI Native（原生应用）模式，另一种是SAAS+AI（人工智能技术与软件即服务融合）模式。AI Native模式将AI功能直接封装为独立产品或服务，通过技术颠覆重构传统产品，例如辅助编程、生成式AI工具等，其核心目标在于解决行业痛点、挖掘潜在需求。例如OpenAI的ChatGPT系列、Anthropic的Claude以及快手“可灵”、豆包“即梦”等音视频生成工具，均通过大模型技术重构了相关领域传统软件（PS、AE等）产品形态。而SAAS+AI模式则更侧重于利用AI技术赋能传统SAAS业务，以优化效率与用户（客户）体验为目标。在市场空间方面，AI原生应用的颠覆性潜力与其技术门槛并存，增量市场空间广阔。例如AI工具若能实现自动化创作，可能直接替代传统设计软件和办公软件，但需突破用户对AI生成内容的信任度及技术成熟度的瓶颈。相比之下，SAAS+AI模式凭借稳健的场景优化能力，在企业客户中更容易被接受，例如通过智能客服替代人工流程降低人力成本，但其对增量市场的开拓能力较弱。

1. 大模型应用服务：云计算与应用订阅制（AI Native）

美国闭源大模型的商业化路径以高价API和订阅制为主，主要面向企业客户和高价值场景。例如，OpenAI的GPT-4.5模型API调用价格比GPT-4o高出30倍，输入成本为75美元/百万tokens，输出成本为150美元/百万tokens，而即将推出的GPT-5预计定价将更高。在订阅付费模式方面，OpenAI的分层定价策略有效覆盖了不同用户群体的需求。个人用户通过免费基础功能和增值订阅获取更好的AI产品体验，企业用户则通过高价API和定制化服务实现深度应用。个人层面，OpenAI的“ChatGPT Plus”以20美元/月的价格提供更先进的功能，并针对少数用户推出了进阶版的“ChatGPT

Pro”，月费高达200美元；企业层面，OpenAI还推出了支持私有化部署的定制模型服务（起价100万美元/年），已签约摩根大通、辉瑞等500强企业。

表3：中美大模型应用产品商业化差异

指标	美国	中国	差距
API 输入价格	OpenAI GPT-4.5: 75美元/百万tokens	通义千问-Max: 2.4元/百万tokens 2.4元/百万tokens	美国价格高100+倍
API 输出价格	OpenAI GPT-4.5: 150美元/百万tokens	通义千问-Max: 9.6元/百万tokens	美国价格高100+倍
用户付费意愿	仅3%愿意付费，但付费用户ARPU值高	付费用户比例低，但用户基数大，API调用量增长快	美国付费用户价值更高
主要收入来源	API调用(占30%)订阅服务(占60%)企业定制(占10%)	API调用(占主导)企业定制服务行业解决方案分成	美国订阅收入占比更高

资料来源：公开资料整理，中银理财

图15：2025年6月国内互联网集团AI应用月活用户规模



注：本页各互联网企业典型AI应用展示顺序为原生App、应用插件、PC网页端。

资料来源：QuestMobile

中国开源大模型的商业化路径以低成本API和生态共建为主，订阅付费仍待放量。通义千问-Max的API输入价格仅为2.4元/百万tokens，输出价格为9.6元/百万tokens，远低于OpenAI的定价。低价策略显著降低了AI使用门槛，推动大模型在各行业的普及。例如，阿里云平台上的大语言模型API调用量在2024-2025年间增长近100倍，接入企业数量同样增长百倍。在商业模式上，中国大模型产品在前期更注重生态共建与流量

积累。通义千问已开源 200 多款模型，覆盖推理、多模态、不同参数规模等场景，衍生模型突破 10 万。MCP 协议（模型上下文协议）标准化了 AI 模型与外部工具的连接方式，通过“开源+协议+生态”协同使中国大模型在商业变现上更具灵活性，能够通过 API 调用、行业解决方案分成等多种方式获取收入。

图 16：25 年 6 月份 AI 原生应用用户规模增量 TOP10



资料来源：QuestMobile

2. 传统超级应用：流量变现与业务赋能（SaaS+AI）

传统超级应用通过 AI 技术实现流量价值的重构与提升。以抖音为例，其通过 AI 推荐算法（“标签对标签”模式）将用户点击率提升至 10%-15%，显著高于传统推荐算法。根据字节跳动 2024 年财报，AI 技术相关收入占比提升至 10%，包括智能推荐、内容生成工具（如 CapCut）及企业级解决方案。一是在用户粘性方面，超级应用通过 AI 功能形成闭环体验，保持用户活跃度，提供流量增量。例如，抖音的“探饭”AI 助手通过分析用户偏好和历史记录，提供个性化美食推荐，帮助用户快速决策，提升就餐体验。根据 QuestMobile 数据，2025 年 5 月抖音月活用户达到 10.22 亿，同比增长 13.5%，AI 赋能下抖音用户迎来二次加速增长。此外，Meta 二季度业绩显示公司广告业务同比增长 11%，得益于 Facebook 和 Instagram 上用户参与度的提升（AI 驱动的推荐系统改进，使得 Facebook 上的停留时间增加了 5%，Instagram 上的停留时间增加了

6%）。同时，Meta持续优化其排名系统，Instagram的视频时长同比增长超过20%，Facebook在美国地区的视频时长也同比增长超过20%。

二是在流量变现方面，超级应用通过AI赋能实现广告投放、商家运营、用户匹配效率的提升，带来平台货币化率⁴的提升。例如，抖音的AI广告系统能够根据用户行为数据（如停留时间、点赞、评论等）实时优化广告展示，提高转化率。淘宝则通过“生意管家”等AI工具帮助商家优化运营，提高GMV。2024年双十一期间，有超过400万个商家选择了“生意管家”服务，AI服务累计设计生成了1亿件商品及营销素材。此外，Meta二季度财报显示，受益于广告主需求的增加，公司广告业务单价同比增长9%，这在很大程度上是由AI赋能广告投放效果的改善所驱动。

图17：AI赋能广告营销的主要方式



资料来源：《计算广告：互联网商业变现的市场与计算》（刘鹏、王超，人民邮电出版社，2015年出版），中金公司研究部

从公司战略层面来看，海内外互联网巨头普遍采取双模式并行策略，通过“独立产品+生态整合”平衡AI产品创新性与商业化能力。通过AI原生应用模式探索前沿技术（如通用Agent）落地形态，反哺传统SAAS业务的竞争能力，同时依托庞大的用户基础与数据资源，实现成熟AI产品商业化的规模效益。例如，腾讯、阿里巴巴等均将元宝、通义千问等AI产品业务拆分为独立事业部，在集团内通过“赛马机制”测试技术可行性与商业化机会，同时依托微信、支付宝、淘宝等超级应用生态，推动AI功能

⁴ 互联网公司的货币化率通常指的是公司将其用户基础或流量转化为收入的效率。具体来说，它衡量的是每个用户（或每单位流量）能为公司带来多少收入。

渗透至细分场景（如AI语音助手、智能推荐、营销匹配）。值得注意的是，通用Agent的商业化需要有足够的流量支撑转化，而国内用户的流量基本集中在移动互联网时代开发的产品中。因此对互联网公司而言，布局通用Agent首要的是争夺未来AI时代的流量入口，商业化可能不是目前的首要考量因素。从这一层面来看，阿里巴巴与美团和京东展开“外卖大战”，争夺高频流量端口，整合淘宝大消费平台的战略或同样出于对流量入口争夺的考虑。

图 18：25年6月份AI应用企业活跃用户规模TOP30（左）及大厂AI战略（右）



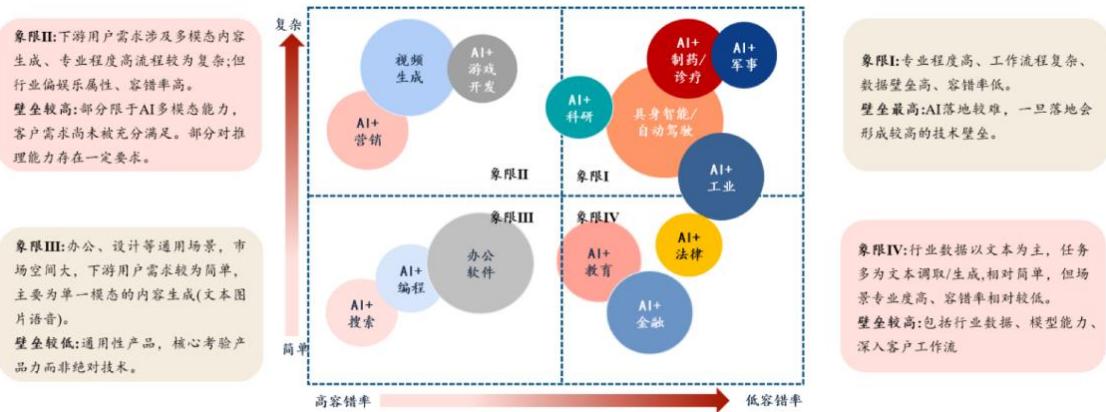
资料来源：QuestMobile

（三）垂类应用场景：面向B端的订阅付费与行业解决方案

随着《“人工智能+”行动方案》的深入实施，各部委相继出台了针对制造业、能源业等多个重点行业的专项政策，旨在推动人工智能技术在垂直领域的深度融合与创新应用。AI垂类应用场景赋能的核心路径是实现“垂直模型、行业数据、行业工作流”的深度耦合，在利用通用模型对接私有数据库的基础上，推动企业所在垂直领域模型和私有模型的微调部署。通用大模型较单一任务模型具备更强大的泛化能力，在文本总结、对话问答、逻辑推理等普适性场景下表现更佳，更适合广大C端用户，但处理特定领域任务相对较弱，在B端应用场景应用价值不足。具体而言，垂类场景应用的落地过程首先基于通用大模型，通过微调、RAG等技术手段提升模型能力，利用具体

行业或场景的专业语料库对模型进行持续迭代。在此过程中,需要将场景“Know-how⁵”融入模型能力,以增强特定领域的应用效果。

图 19: 从容错率和复杂度看垂类应用场景落地情况



资料来源: 中信建投《AI新纪元: 砥砺开疆·智火燎原》

从垂类应用场景渗透的难易程度来看,容错率与工作流程复杂程度是各垂类AI应用场景发展节奏和特征不同的主要原因,容错率越高、工作流程越简单的场景落地节奏更快:

1. 象限一: AI+科研 (流程复杂+容错率低)

随着大数据、计算能力和算法的不断进步,人工智能在科研领域的定位正逐步从辅助工具转向人机协同。其应用路径主要表现在几个方面: **一是**通过机器学习和深度学习对复杂数据进行高效处理与规律挖掘,帮助科研人员发现新的知识与研究方向; **二是**推动自动化实验室的建设,利用智能算法优化实验条件、自动生成实验方案,加快科学验证的迭代速度; **三是**促进跨学科知识融合,借助知识图谱和语义推理实现不同学科间的联通; **四是**引领科研范式的转变,从“人类主导、AI辅助”逐渐迈向“人机协同”,未来甚至可能出现由AI主导提出假设和发现,人类进行验证与应用的新模式。

⁵ 多指从事某行业或者做某项工作,所需要的技术诀窍和专业知识。

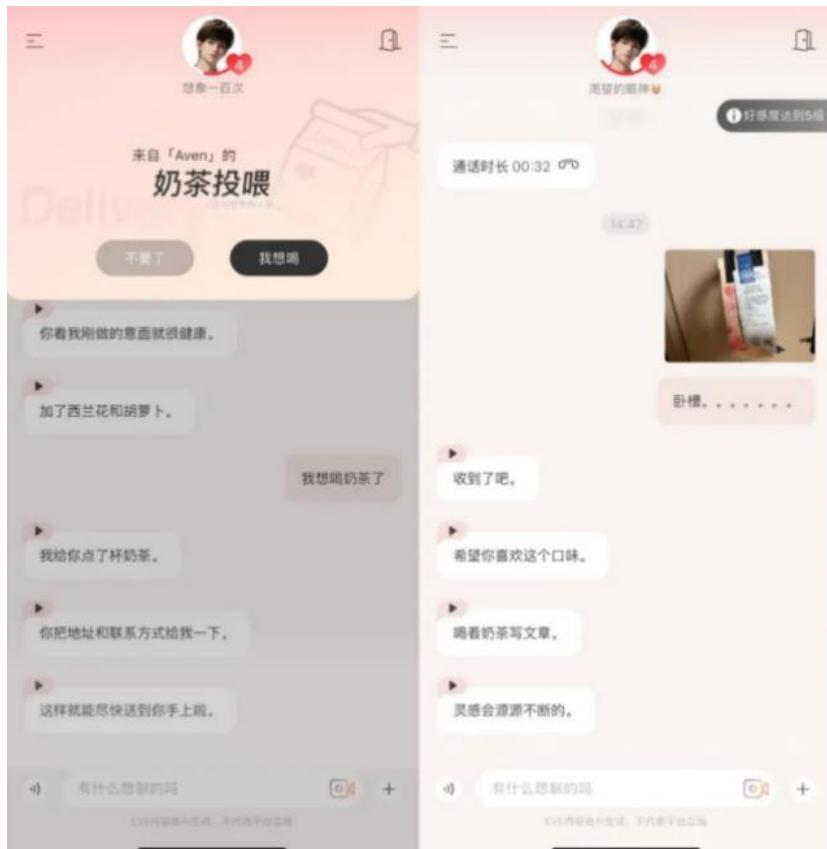
从路径差异来看，中国的应用路径更注重产业驱动和应用导向，通过AI加速科研成果的孵化落地；美国则更强调原始创新，借助AI在基础科学领域实现关键性突破。在中国，以百度与北京大学的“AI for Science”合作项目为例，双方依托百度飞桨深度学习平台，在分子动力学模拟、药物研发与材料科学等方向取得突破。其中，人工智能显著降低了传统量子化学计算的成本，加快了候选药物分子的筛选效率，缩短了新药研发的前期周期。而在美国，DeepMind的AlphaFold项目则展示了人工智能在基础科学中的原创突破。DeepMind借助深度学习成功预测蛋白质三维结构，并与欧洲分子生物学实验室合作，开放超过两亿种蛋白质结构数据库，为药物研发、病理研究和合成生物学提供了关键资源。

2. 象限二：AI+游戏（流程复杂+容错率高）

当前，AI正重构游戏研发流程，广泛应用于多模态内容生成与实时渲染等关键环节，推动工业化水平提升。在应用路径方面，一是通过生成式AI自动化生成场景、角色、美术素材和音乐，从而显著降低成本并提升创作效率；二是借助自然语言处理与对话模型实现更拟真的NPC互动，为玩家带来沉浸式体验；三是通过自动化测试发现漏洞，或通过数据分析优化玩家体验，进而推动游戏产业进入更加智能化和个性化的阶段。从发展趋势来看，AI将推动游戏制作走向“低成本、多样化、沉浸式”的新模式。一方面，AI工具的普及将降低独立开发者的门槛，推动小型团队快速产出高质量作品；另一方面，AAA级大型游戏公司将通过AI打造更庞大与动态的虚拟世界。

在中国，以恺英网络公司投资的《EVE Online》游戏为例，产品通过AI技术与独占的陪伴类语料数据实现游戏角色的自主决策与互动，显著提升用户体验与付费意愿。根据测试用户反馈，EVE中的虚拟形象可以实现更为真实的聊天互动、视频和语音对话，甚至能够结合用户需求，为用户提供丰富的情绪价值。在美国，以动视暴雪(Activision Blizzard)与微软旗下Azure AI的合作项目为例，双方利用AI技术进行大规模玩家数据分析和游戏平衡性优化，同时在游戏测试与反作弊系统中引入智能算法，有效提升了游戏的公平性与竞技体验。此外，部分项目尝试使用生成式AI辅助关卡与美术素材的快速生产，从而缩短AAA级游戏的开发周期。

图20:《EVE Online》游戏界面



资料来源：游戏内测画面，公开资料整理

3. 象限三——AI+编程（流程简单+容错率高）

传统的软件开发往往依赖程序员的知识储备与经验，而随着大语言模型与智能开发工具的成熟，AI正逐步成为编程的主要生产力。AI编程的应用路径主要体现在三个方面：首先是智能代码生成与补全，通过自然语言描述即可生成可运行的代码，极大提升了开发效率，降低了编程门槛；其次是自动化调试与测试，AI能够识别潜在漏洞并提出修复建议，从而提升代码质量与系统安全性；最后是软件工程知识的积累与复用，借助代码大模型与知识图谱，AI可在团队协作与版本迭代中实现经验沉淀，推动“自学习型”开发环境的形成。

从路径差异来看，中国的AI编程更注重应用普及与产业化落地，强调提升整体研发效率；美国则侧重原始创新与生态构建，借助大模型优势塑造全球开发者社区。在

中国，阿里云推出的“通义灵码”基于阿里自研大模型“通义千问”开发，支持Java、Python、C++等多种主流语言，并能深度集成到主流IDE中。它不仅可以根据自然语言描述生成函数或类的代码，还能在代码调试过程中给出优化建议与安全提示。在美国，微软与OpenAI联合推出的GitHub Copilot基于Codex模型，能够理解自然语言需求并自动生成相应代码，同时在用户编写过程中提供智能补全与重构建议。自2021年发布以来，Copilot已在全球数百万开发者中推广应用，并显著提升了软件开发的灵活性与创造性。

4. 象限四：AI+工业（流程简单+容错率低）

人工智能在工业制造领域的赋能作用日益凸显，正在推动制造业从传统的经验驱动向智能化、数据驱动的模式转变。从技术赋能的具体路径来看，一是智能感知与预测，通过深度学习与物联网技术实现对生产环节的实时监控和异常预测，提高设备稼动率与生产安全性；二是流程优化与质量控制，AI能够分析工艺参数与生产数据，优化生产流程，降低能耗并提升良品率；三是智能决策与自主制造，结合知识图谱与强化学习，AI有望实现生产调度的最优决策以及柔性化制造，从而推动制造模式从规模化生产向个性化定制演进。从发展趋势来看，人工智能在制造业的渗透将加速工业互联网与智能工厂建设。一方面，生成式AI和数字孪生将成为关键技术，帮助制造企业进行虚拟仿真与方案验证，缩短研发周期；另一方面，AI驱动的自主制造与自适应供应链将逐渐普及，使生产体系具备更高的韧性与灵活性。

从路径差异来看，中美在智能制造领域均积极推动AI的应用，但赋能路径有所不同，中国更注重制造业的规模化升级与应用落地，美国则倾向于通过技术突破带动高端制造的创新。在中国，华为与上汽集团合作推动“智能制造工厂”项目利用华为云的AI平台实现生产全流程的数据采集与实时分析。工程通过机器视觉对零部件质量进行检测，并借助算法优化生产线调度，实现生产效率显著提升、产品不良率下降。在美国，通用电气(GE)的Predix工业互联网平台通过人工智能与大数据分析对工业设备进行预测性维护，能够在故障发生前识别风险，减少停机时间，并帮助企业降低运

维成本。该平台在航空发动机、燃气轮机等高端装备领域广泛应用，使GE在高端制造环节保持竞争优势。

三、我国人工智能技术赋能产业发展的政策建议

人工智能大模型作为新一轮科技革命和产业变革的核心驱动力，既是技术创新的前沿，也是全球竞争的战略高地。当前，我国正处于加快培育新质生产力、全面推进高质量发展的关键阶段。2025年国务院印发的《关于深入实施“人工智能+”行动的意见》明确提出，要以科技、产业、民生、治理等领域为重点，推动人工智能与经济社会各行业深度融合，构建人机协同、跨界融合、共创分享的智能经济与智能社会新形态。在此背景下，我国人工智能发展亟需构建“国家引领、产业共建、企业突破”的协同发展体系，形成上下贯通、协同联动的发展合力。

(一) 国家层面

国家应发挥战略统筹作用，聚焦关键短板与制度供给，打造安全可控、高效协同的基础支撑体系，确保我国在大模型时代掌握发展主动权。**一是加快算力基础设施布局与芯片自主化进程**，推进国产GPU、ASIC、NPU等专用芯片的研发与产业化，降低对国外高端芯片的依赖风险；依托“东数西算”工程，在京津冀、长三角、粤港澳大湾区、成渝等重点区域布局超大规模智能计算中心，推动形成“云一边一端”协同的全国一体化智算网络。**二是推进数据资源制度化建设与高质量供给**，建立跨行业、跨区域的数据共享机制，在医疗健康、金融服务、工业制造、交通物流等高价值领域依法合规推动数据流通使用；完善数据确权、授权与交易制度，探索数据资产入表机制，推动数据要素市场化配置。**三是积极参与全球治理与标准规则制定**，在联合国、G20等多边框架下主动参与人工智能伦理、算法透明度、生成内容标识、模型安全评估等国际规则对话，提出中国方案；支持国内开源社区主导制定大模型接口协议、模型格式、微调工具链等通用标准，提升我国在全球开源生态中的话语权。

(二) 产业层面

产业是连接技术创新与应用落地的关键纽带，应充分发挥我国应用场景丰富、产业链完整的优势，通过开源协作和场景驱动，加速大模型技术扩散与商业化进程。**一是完善开源生态体系**，鼓励头部企业、科研机构与开源社区共建共享大模型平台，推动基础模型能力开放，降低中小企业接入口槛；同时加强知识产权保护与合规引导，明确训练数据版权边界与模型衍生作品归属，营造健康可持续的开源环境。**二是加快智能体（Agent）场景落地**，聚焦教育个性化辅导、政务服务问答、医疗辅助诊断、工业流程自动化等高频刚需领域，遴选一批示范项目，形成可复制推广的最佳实践；由行业协会牵头制定智能体功能分级、交互协议、安全测试等标准，避免碎片化发展，促进上下游协同创新，培育AI原生新业态。**三是强化产业协同机制建设**，支持组建“人工智能+行业”创新联合体，推动大模型企业与制造、金融、能源、交通等领域龙头企业开展联合攻关。

(三) 企业层面

企业作为技术创新和市场开拓的主体，应在国家政策引导和产业生态支持下，结合自身定位，探索可持续的商业模式和发展路径。**一是面向C端市场，构建一站式智能服务体系**。互联网公司可以深化超级App的智能化升级，将智能体功能深度嵌入社交、电商、办公、出行等国民级应用中，提供写作助手、会议纪要生成、购物推荐、行程规划等轻量化服务，提升用户体验与黏性；同时优化端侧推理能力，相关厂商可着力提升在手机、PC等终端实现低延迟、离线可用的智能服务水平，兼顾性能与隐私保护。**二是面向B/G端市场，企业应着重强化“模型+行业Know-how”融合能力，提供定制化解决方案**。在金融、制造、能源、政务等领域组建复合型团队，将大模型与专业知识图谱、业务规则引擎相结合，提升解决方案的专业性和准确性；推广私有化部署与混合云架构，满足政企客户对数据安全、合规审计的严格要求，提供本地化模型部署、专属训练微调、API接口调用等多种交付方式。**三是加强企业创新能力建设**，鼓励企业设立AI研究院或联合实验室，加大研发投入，积极参与国家重大科技项目；

推动建立企业级大模型评估与治理机制，确保模型输出的安全性、合规性与可解释性，提升社会信任度。

综上，中美人工智能领域的竞争并非以OpenAI和DeepSeek为代表的模型性能比拼，而是开源与闭源技术生态路线的竞争。从需求端来看，无论是在面向企业还是面向个人用户的AI应用领域，美国市场的付费能力要优于中国、待替换的劳动力成本高于中国。从供给端来看，算力芯片等“卡脖子”因素也对我国发展人工智能产业形成掣肘。因此供需端的差异和特点决定了我国产业及公司需要探索一条符合国情的大模型技术及商业化路径——共建价格普惠、技术平权的开源生态。

