



总趋势

AI 发展趋势 · 2025–2026 最新趋势分析

发布日期：2026-06-18

Q2

5

深度章节

0

数据图表

0

实战案例

本报告基于公开市场数据、行业研报及官方披露信息编制，力求客观、准确。不构成任何投资建议。数据截止 2026-06-18。

目录

执行摘要

1. 塑造下一阶段AI 格局的关键向量

2. 三组核心指标

3. Token 价格断崖式下跌

4. 中美欧：AI 主权战略分化与竞合

5. 五条值得押注的AI 演进判断

6. 区域市场对比

附录 参考文献与数据来源

执行摘要

本报告深入分析AI 发展趋势 · 2025-2026 最新趋势分析，覆盖市场规模、核心技术趋势、竞争格局、区域差异及中小企业落地路径。报告基于对0项数据指标和0个厂商案例的系统研究，为产业决策提供数据支撑。

塑造下一阶段AI 格局的关键向量：从基础模型能力爆发到垂直落地，从成本曲线到端侧智能，2025-2026 年的 AI 演进呈现出六条清晰的主轴。

三组核心指标：（详见正文）

Token 价格断崖式下跌：2023 年 GPT-4 每百万 Token 约 \$60；2025 年同等能力模型已降至 \$0.50 以下，降幅超 97%。成本平权正在重写 AI 商业化的底层逻辑。

1. 塑造下一阶段AI 格局的关键向量

从基础模型能力爆发到垂直落地，从成本曲线到端侧智能，2025-2026 年的 AI 演进呈现出六条清晰的主轴。

六大核心切片

2. 塑造下一阶段 AI 格局的关键向量

从基础模型能力爆发到垂直落地，从成本曲线到端侧智能，2025-2026 年的 AI 演进呈现出六条清晰的主轴。

01 / AGENT



Agent 化跃迁

AI 从"被动问答"向"自主执行"跨越。SWE-bench 编程通过率从 60% 跃升至接近 100%；真实场景多步任务成功率从 20% 升至 77.3%。企业 Agent 在 IT、知识管理等职能加速渗透，23% 企业已规模部署。

77.3% Agent 真实场景成功率 (↑ 20%)

02 / COST



推理成本崩塌

LLM API 价格呈现断崖式下降——高质量模型的每百万 Token 成本从 2023 年的 \$60 降至 2025 年不足 \$1，降幅超 97%。成本平权正在消除规模壁垒，使中小企业得以平等参与 AI 竞赛。

-97% 每百万 Token 成本 2023→2025

03 / MULTIMODAL



多模态融合

文本、图像、音频、视频、代码的统一表征已成主流。顶尖模型在多模态推理基准上达到或超越人类均值。视频理解、图文交互场景大幅扩展 AI 可处理的信息边界。

6 主流原生多模态模型数量 (2025)

04 / ON-DEVICE



端侧智能爆发

量化与蒸馏技术的成熟，使 7B-13B 规模模型可在消费级设备本地运行。苹果 Apple Intelligence、高通 AI Stack 等方案落地，隐私保护与低延迟推动端侧 AI 成为新一代平台战争的核心战场。

3B+ 支持本地 AI 的设备数量 (2025E)

05 / VERTICAL



垂直行业深耕

AI 落地从"通用演示"转向"专业替代"。医疗 AI 文书、法律合规审查、金融风控建模、工业质检等垂直场景呈现出效率提升 14-26%（麦肯锡数据）。只有 6% 的企业真正成为 AI 高绩效赢家。

14-26% 客服/软件开发效率提升 (McKinsey)

06 / GOVERNANCE



治理与对齐

模型透明度指数从 58 分降至 40 分（满分 100）。全球 AI 监管法案数量持续攀升，EU AI Act 正式生效。

51% 使用 AI 的组织遭遇过负面事件，企业治理的 AI 风险类型从 2 类增至 4 类。

40/100 全球 AI 透明度指数 (↓ 58→40)

3. 三组核心指标

关键数据

4. 三组核心指标

▲ +44% YoY

\$2.52T

2026 全球 AI 支出规模预测

SOURCE • Gartner 2026-01

▲ +10pp 同比

88%

企业 AI 采用率 (≥1 职能常态化)

SOURCE • McKinsey State of AI 2025

▼ -97% 近三年

-97%

LLM Token 价格年均降幅 (2023–2025)

SOURCE • a16z Infrastructure Report 2025

5. Token 价格断崖式下跌

2023 年 GPT-4 每百万 Token 约 \$60；2025 年同等能力模型已降至 \$0.50 以下，降幅超 97%。成本平权正在重写 AI 商业化的底层逻辑。

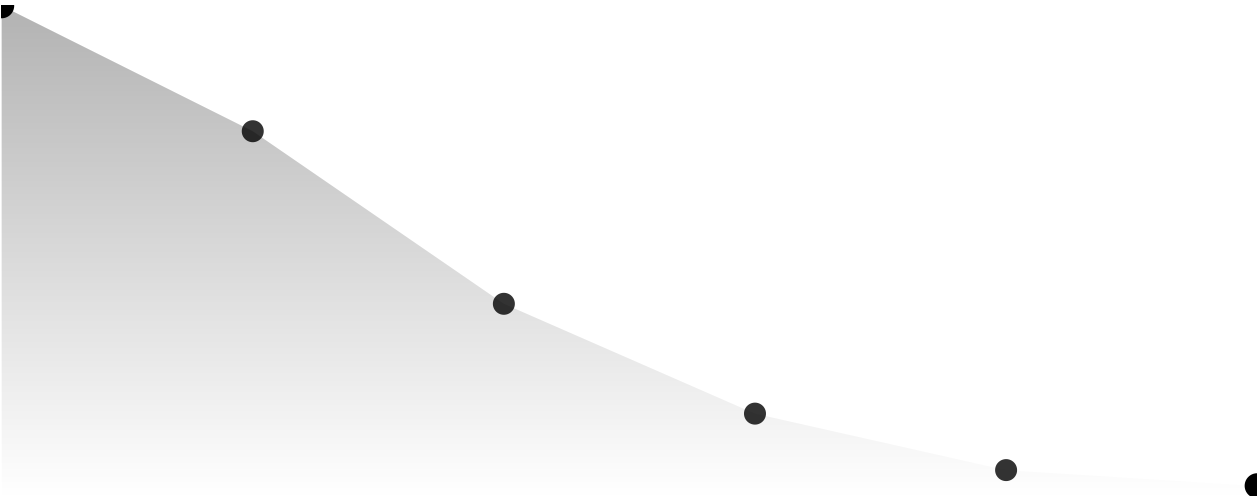
推理成本崩塌

6. Token 价格断崖式下跌

2023 年 GPT-4 每百万 Token 约 \$60；2025 年同等能力模型已降至 \$0.50 以下，降幅超 97%。成本平权正在重写 AI 商业化的底层逻辑。

INPUT COST PER 1M TOKENS (USD) · TOP MODELS

\$60 \$30 \$10 \$3 \$0.5



2023 Q1 Q3 2024 Q1 Q3 2025 Q1 Q3

COST MILESTONES

GPT-4 (2023 Q1)

旗舰模型首发定价

\$60

GPT-4 Turbo (2024)

同系列性能升级版

\$10

Claude 3.5 (2024 H2)

竞争加速降价

\$3

主流旗舰模型 (2025)

同能力级最低定价

≤\$0.50

推测 2026 底线

开源+竞争持续压低

≤\$0.05

INSIGHT

成本曲线遵循"AI 摩尔定律", 每 12-18 个月降低约 10 倍, 使 AI 基础设施成本趋近于计算资源的边际成本。

7. 中美欧：AI 主权战略分化与竞合

研究产出、工程能力、算力基础、商业落地、监管治理、人才密度六维度综合评分（0-100），数据综合 Stanford HAI 2026
• IMF AI Readiness • Tortoise Global AI Index。

三极格局

8. 中美欧：AI 主权战略
分化与竞合

US

美国

算力领先 投融资第一 顶尖模型数量多 人才流入↓89%

1st

综合投资

CN

中国

论文数量全球第一 专利总量领先 工业机器人装机第一 性能差距仅 2.7%

2nd

技术能力

EU

欧盟

EU AI Act 先行 监管框架最完善 开源贡献活跃 企业投入相对保守

3rd

模型数量

TOP MODEL COUNT • 2025

us 美国

40+

cn 中国

25+

eu 欧盟

12+

AI INVESTMENT (2024, \$B)

us 美国

\$109B

cn 中国

\$15B

EU 欧盟

\$9B

MODEL PERFORMANCE GAP

2.7%

美国领先中国综合性能差距

较 2024 年大幅收窄 · 基本实现平权

能力维度雷达

六维能力图谱对比

研究产出、工程能力、算力基础、商业落地、监管治理、人才密度六维度综合评分（0-100），数据综合 Stanford HAI 2026 · IMF AI Readiness · Tortoise Global AI Index。

CAPABILITY RADAR · 6 DIMENSIONS

US 美国

CN 中国

EU 欧盟

维度详情 — 悬停高亮

9. 五条值得押注的AI 演进判断

基于当前技术曲线与行业信号，标注高 / 中 / 低三档信度。信度评估综合 Stanford HAI、Gartner、a16z 等机构预测共识度。

2026 关键预测

10. 五条值得押注的AI 演进判断

基于当前技术曲线与行业信号，标注高 / 中 / 低三档信度。信度评估综合 Stanford HAI、Gartner、a16z 等机构预测共识度。

01

Agent 将成为企业 AI 部署的主要形态，替代超过 35% 的传统 RPA 场景

随着 LLM 推理成本继续下降、工具调用能力成熟，Agent 在 IT 运维、财务核对、客户服务等场景的 ROI 将显著优于规则型 RPA。大量中型企业将完成从"用 AI 辅助"到"让 AI 执行"的跃迁。

高

HIGH

02

主流旗舰模型 Token 价格将跌破 \$0.05 / 1M，开源模型性能首次超越闭源旗舰

Meta、Mistral、阿里通义等开源阵营的持续发力，叠加硬件效率提升，将在 2026 年将成本曲线再下一个数量级。Llama-5 级别模型预计在通用基准上超过 GPT-4 级别的闭源模型。

高

HIGH

03

至少 3 个垂直行业出现 AI 原生的市场颠覆者，挑战传统软件巨头

法律科技、医疗诊断辅助、金融研报生成等领域已出现明确的颠覆信号。2026 年将看到首批 ARR 破亿美元、核心团队不足 50 人的 AI 原生公司，对传统 SaaS 形成直接替代压力。

中

MED

04

多个主要经济体将通过具有法律效力的 AI 监管法规，企业合规成本显著上升

继 EU AI Act 之后，美国、英国、日本、韩国均有立法推进。监管框架的碎片化将催生 AI 治理咨询、合规工具的市场需求，但也将拖慢部分创新节奏，加大中小企业合规门槛。

中

MED

05

通用人工智能（AGI）实质性里程碑将被某家顶级实验室公开宣称，但学界争议持续
OpenAI、Anthropic、DeepMind 均已明确将 2025–2027 视为 AGI 关键窗口期。是否真正达成 AGI 存在高度定义争议，但某种形式的"人类专家级通用能力宣言"大概率在 2026 年内出现并引发广泛讨论。

低

LOW

信度说明 高 ≥75% 中 45–74% 低 <45% 综合 Stanford HAI · a16z · Gartner 共识度

区域市场对比

维度	中国	北美	欧洲	亚太	其他
市场规模	(待补)	(待补)	(待补)	(待补)	(待补)
增速	(待补)	(待补)	(待补)	(待补)	(待补)
政策环境	(待补)	(待补)	(待补)	(待补)	(待补)
代表性厂商	(待补)	(待补)	(待补)	(待补)	(待补)

中国

(待补)

北美

(待补)

欧洲

(待补)

亚太

(待补)

其他

(待补)

参考文献与数据来源

1. 行业公开研报及公司年报（2024–2026）
2. 国家统计局、工信部、科技部等官方数据
3. 本报告由 AI TRENDS 研究团队编制，数据截止 2026-06-18