



从能力到绩效，2026 年 AI 应用落地加速

—— 数字经济双周报（2026 年第 2 期）

2026 年 01 月 26 日

核心观点

- **本期焦点：在 2026 年达沃斯世界经济论坛上，关于人工智能的讨论明显转向“投入能否转化为绩效”。**

尽管 2025 年全球 AI 投资规模已接近 1.5 万亿美元，但行业层面来看，本轮 AI 热潮仍旧遵循着堆积算力与扩展基础设施的传统路径，应用端的兑现速度明显滞后。因此，近期市场讨论重心开始转向 AI 应用于真实业务的效果。各主要经济体高度重视，中国推动“人工智能+”行动、美国发布 America's AI Action Plan，都将讨论重心逐步转向应用部署与制度保障，强调将技术嵌入具体产业场景，力求把能力转化为生产力。

企业层面，头部厂商的商业化进程正在加速，OpenAI 营收快速增长，谷歌 Gemini 调用量显著攀升，阿里通过生态整合推动应用进入高频场景，AI 正从技术储备走向利润来源。然而，随着应用逐渐深入，AI 的规模化实施以及信任与合规压力日益凸显，在中小机构和基层场景中的渗透速度仍然有限。

- **中国动态：工业和信息化部发布《推动工业互联网平台高质量发展行动方案（2026—2028 年）》。**工信部明确以推动信息化与工业化深度融合为目标，通过完善多层次平台体系、扩大设备连接规模，并同步提升数据加工能力和人工智能应用水平，引导制造企业深化平台应用，加快制造业数字化、智能化转型，推动新型工业化和新质生产力加快形成。
- **美国动态：算力扩张已快于能源体系消化能力，人工智能竞争正从模型与芯片，进一步下沉到电力、输电和制造等关键约束环节。**1 月 13 日-16 日，Meta、OpenAI 与谷歌相继加码算力与能源布局，通过锁定核电、电网容量和本土制造能力保障长期供给；同时，美国电网接入与审批瓶颈日益突出，政府要求企业自行承担扩容成本。
- **其他国家：亚洲主要经济体正在加大对 AI 算力与芯片基础设施的集中投入力度。**韩国扩建高带宽存储封装产能，日本动员政企资金推进 2 纳米制程，印度布局大型数据中心并推广本地化 AI 服务。尽管投资重点各异，但均围绕算力底座与关键芯片展开布局，力图在全球供应链竞争中占据先机。
- **技术动态：英伟达更正数据中心用铜测算，将 1 吉瓦机架用量从 50 万吨修订为约 200 吨，从而显著扭转了市场对“AI 耗铜”的高预期。**按全球容量测算，数据中心对铜需求的直接拉动仍然有限。从更长周期看，铜需求的主要支撑仍来自电网升级、电动汽车普及等电气化趋势的叠加。
- **智库动态：世界经济论坛与麦肯锡联合发布白皮书，指出当前 AI 竞争力的关键已不再是完全自给自足，而是投资方向、能力布局方式以及合作网络的构建质量。**报告建议各国放弃追求全产业链覆盖的思路，转而聚焦自身比较优势，通过提升基础设施互操作性和推动区域战略联盟，在全球 AI 生态中占据更有价值的位置。

分析师

彭雅哲

☎：010-80927607

✉：pengyazhe_yj@chinastock.com.cn

分析师登记编码：S0130525020001

研究助理：杨晓

✉：yangxiao_yj@chinastock.com.cn

风险提示

1. 对政策理解不到位的风险
2. 政策落实不及预期的风险
3. 技术发展不确定性风险

目录

Catalog

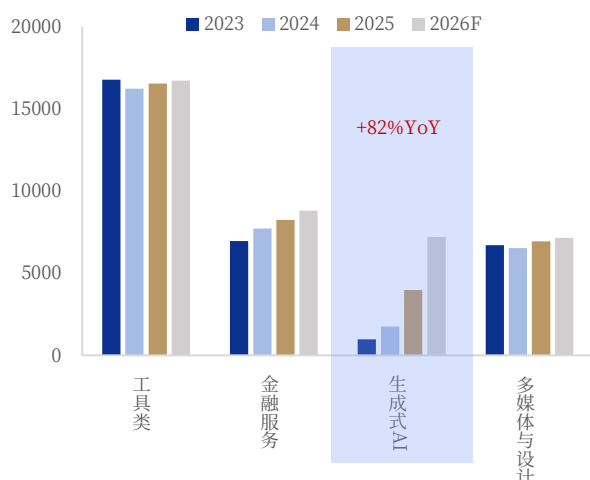
一、 本期焦点：2026 年 AI 应用落地加速	3
二、 中国动态：工业互联网高质量发展行动方案发布	5
三、 美国动态：人工智能发展已转向资源要素争夺	6
四、 欧洲动态：系统推进数字主权与基础设施自主	7
五、 其他国家：亚洲主要经济体加大 AI 硬件投资	8
六、 技术前沿：AI 拉动铜价预期大幅修正	9
七、 智库观点：“因国制宜”发展人工智能	10
八、 风险提示	13

一、本期焦点：2026 年 AI 应用落地加速

在 2026 年达沃斯世界经济论坛上，关于人工智能的讨论明显换了一个方向。相比过去围绕概念能力和技术前景的热议，今年更多聚焦在一个更现实的问题上：投入能否真正转化为绩效。尽管预计到 2030 年，人工智能应用的年度投资约为 1.5 万亿美元，但相当一部分企业在启动和扩展项目时依然进展缓慢。**行业层面来看，自 2022 年底启动的本轮 AI 热潮仍遵循着堆积算力与扩展基础设施的传统路径。**数据中心、专用芯片与算力网络吸纳了最集中的资本开支，然而应用端的兑现速度相比起来却有明显滞后。若大量前端投入长期无法转化为稳定的业务回报，这类资本支出便可能从战略性资产变成财务报表上的沉重项，市场也会开始以“泡沫风险”来重新评估其价值。因此，当下市场讨论的重心开始转向可量化的投资回报率以及真实业务场景中的效果。**国家层面来看，各主要经济体的关注点正在逐步转向应用部署与制度保障。**例如，中国推动的“人工智能+”行动强调将技术嵌入具体产业场景，力求把能力转化为生产力；美国通过 America's AI Action Plan，以优先级与定制化政策框架，激励 AI 工具的开发和加速采用。

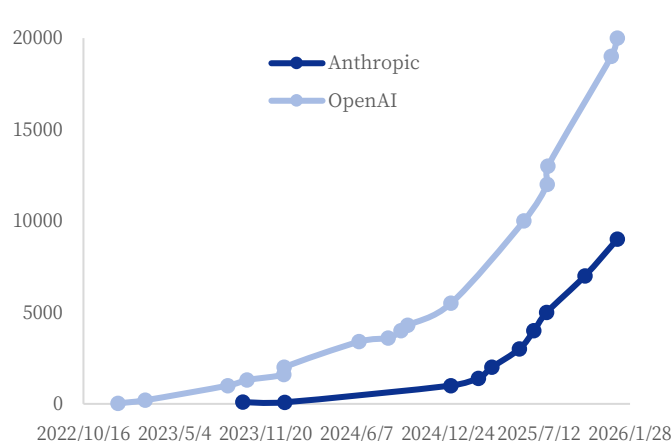
在 AI 应用加速推进过程中，首先表现出的是 AI 逐步从成本项变为明确的营收增长来源。1 月 18 日，OpenAI 首席财务官 Sarah Friar 在官方博客中披露，公司 2025 年年化营收已超过 200 亿美元，而 2024 年约为 60 亿美元，2023 年仅为 20 亿美元左右。在短短两年内完成这种级别的跃升，说明商业化进程已进入实质阶段。同样在 1 月 16 日，Ramp 发布的企业支出数据显示，OpenAI 在美国企业中的采用率于 2025 年 12 月升至 36.8%，付费企业占比达到 46.6%，均创下新高。这些变化与去年 12 月 GPT-5.2 推出的节奏基本同步，一定程度上反映出头部厂商对竞争压力的快速响应。谷歌方面的情况类似。The Information 在 1 月 20 日披露，自 2025 年 3 月 Gemini 2.5 版本推出以来，Gemini API 调用量在五个月内从约 350 亿次跃升至 850 亿次，增幅超过 140%；企业订阅用户数达 800 万，从而驱动 Google Cloud 第三季度营收至 152 亿美元，其中人工智能相关收入每季度贡献数十亿美元。AI 不再只是技术储备，而是开始直接进入利润表。这一趋势也体现在企业级服务的组织方式上。IBM 于 1 月 19 日推出的 Enterprise Advantage 服务，将人工智能工具与专业咨询相结合，助力企业构建和管理定制化内部平台，已在 Pearson 等企业实施，顾问效率提升达 50%。从执行层面看，这类“带方法论的落地服务”可能比单纯模型升级更关键。技术成熟后，市场开始倒逼 AI 从“潜力”走向“性能”，这一阶段的压力，显然比模型竞赛本身更现实。

图1：全球 iOS 与 Google Play 平台 APP 下载次数（单位：百万）



资料来源：Sensor Tower；中国银河证券研究院

图2：人工智能公司收入飙升（单位：百万美元）



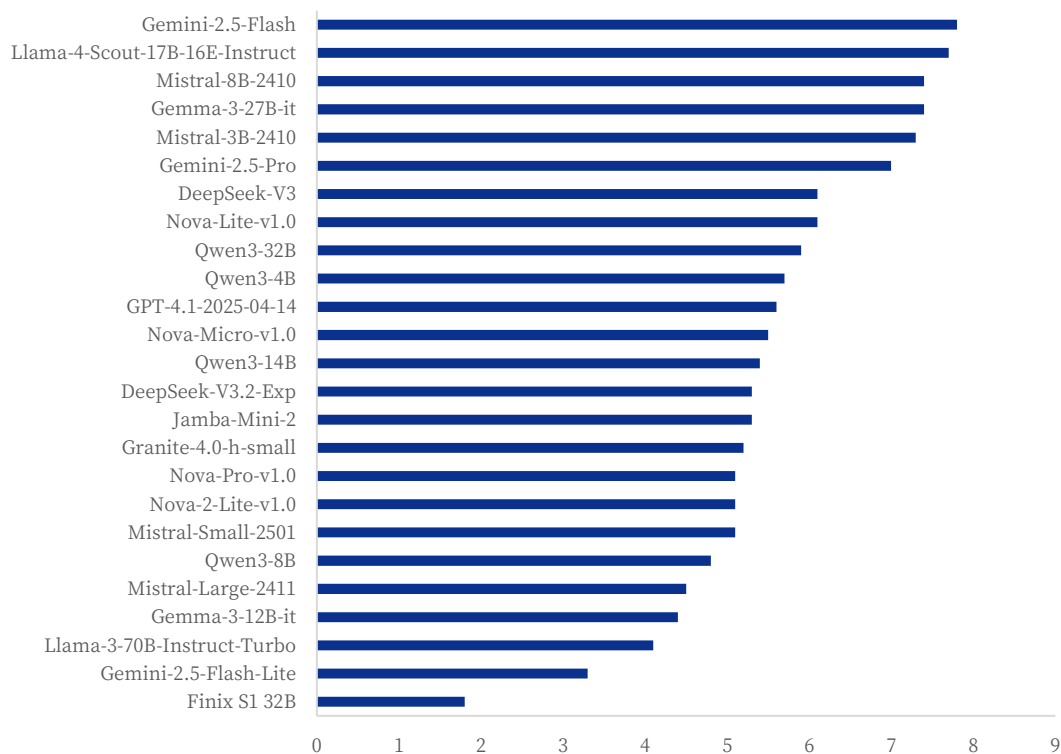
资料来源：EpochAI，中国银河证券研究院

其次，在消费领域和垂直行业的应用中，人工智能正通过生态系统整合实现更深入的落地。中国企业在此领域的表现尤为突出。阿里巴巴于1月15日在杭州举行的发布会上宣布千问 App 全面升级，接入淘宝、天猫、支付宝、飞猪和高德等平台，提供人工智能驱动的购物、外卖订购、机票预订等服务，并上线超过400项办事能力，向全体用户开放测试。该应用利用交易数据生成精准推荐，并计划每季度进行重大更新，优先关注用户满意度而非短期商业化。阿里健康于1月19日完成内测并开放下载的“氢离子”人工智能产品，针对临床和科研医生，提供基于循证医学的支持，其严重幻觉率与国际领先的 OpenEvidence 相当，首字生成时间缩短至3秒左右，且在三年内不考虑盈利模式，旨在构建“解决中国医生一切医学问题的医学人工智能”。智谱 AI 与滴滴于1月12日宣布的战略合作，聚焦通用人工智能关键技术及出行领域的智能体应用，利用滴滴的海量真实场景数据优化意图对齐与推理机制，推动代理在复杂物理环境中的验证。相较海外，中国厂商更多选择从高频场景入手，借助本土生态优势加速落地进程。

此外，企业级市场的整合也在加速。1月20日，OpenAI 与 ServiceNow 签署三年战略协议，将 GPT-5.2 和多模态能力原生集成进 ServiceNow 平台，构建语音智能体与“计算机使用”模型，并设置明确的收入承诺。这一合作看似是产品集成，实质上更像是渠道与信任体系的绑定。企业 AI 增长的前提是信任基础，从前线部署工程师到稳定运维流程，都是规模化不可回避的环节。当前多个厂商同时押注智能体形态，本质上是在寻找一种“可持续执行单元”，而不仅是更强的对话接口。

不过，进入落地阶段之后，约束条件也变得更加清晰。如何将 AI 落地规模化仍是首要难题，许多企业在试点阶段表现尚可，但一旦扩展到核心流程，复杂度迅速上升。同时，信任与合规压力日益突出，无论是 OpenAI 的商业化尝试，还是阿里医疗体系对幻觉率的控制，本质上都在回应同一问题：AI 是否足够可靠。更现实的一点是，普惠渗透仍然较为缓慢。目前，头部企业在 AI 应用落地方面跑得很快，但中小组织与基层场景的转化周期，可能远比预期要更长。

图3：全球 Top 25 大语言模型（LLM）幻觉率情况（单位：%）



资料来源：vectara，中国银河证券研究院

二、中国动态：工业互联网高质量发展行动方案发布

1. 工信部出台三年行动方案，推动工业互联网平台实现高质量发展

1月13日，工业和信息化部发布了《推动工业互联网平台高质量发展行动方案（2026—2028年）》。这份文件的核心目的是推动信息化和工业化更深层次融合，加速制造业向数字化、智能化方向转型。方案给出了明确的三年目标：到2028年，形成专业型、行业型、协作型并存的多层次平台格局，影响力较大的平台超过450家；重点平台的工业数据加工增值能力、模型积累能力和人工智能落地水平要有明显跃升，连接的工业设备总量突破1.2亿台套；规模以上工业企业的平台普及率达到55%以上，初步建成一个泛在连接、智能融合、协同高效、开放共享的工业互联网生态体系。这份方案的出台，实际上是为新型工业化和新质生产力培育提供了一个比较清晰的平台支撑框架。

围绕高质量发展，方案明确了四方面重点工作：一是建立从基础级到生态级的四级梯次培育机制，引导平台向专业化、特色化方向聚焦，避免低水平重复建设；二是系统推进工业数据资源整理和标签化工作，同步开发通用模型和行业模型，并加快工业场景智能体的研发与应用；三是推动更多制造企业把平台用深、用实，鼓励采用“先用后付”、订阅服务等灵活模式，打通大中小企业之间的融通链条；四是强化生态支撑，从开源社区建设、标准制定到网络安全防护全面发力，持续提升我国工业互联网平台的国际话语权和可控水平。

现阶段我国工业互联网平台建设已经积累了相当规模，影响力平台超过340家，连接设备也突破了1亿台套，但问题依然突出。平台之间同质化现象比较严重，很多应用场景的穿透深度不够，产业链上下游的协同效率也偏低。人工智能作为当前最活跃的战略技术，正带动制造业发生新一轮深刻变革。方案把“人工智能+制造”作为重要抓手，意在通过充分挖掘海量工业数据的价值，把前期积累的技术和连接优势尽快转化为实实在在的产业竞争力。

2. 地方层面正加快人工智能产业生态与应用场景的系统构建

各地在人工智能领域的布局策略正在转向构建完整、可持续的区域性生态体系建设，这不仅涉及技术研发的深化，其中包括政策引导、资金支持、人才集聚、基础设施建设和应用场景的深度融合。重庆人工智能湾区于近期正式启动建设，以“一湾”生态平台为核心，联动重庆人工智能学院和通用人工智能研究院，形成政产学研用金一体化的创新联合体，聚焦行业智能体、具身智能、社会智能等方向，建设概念验证中心、中试平台和众创空间，配套200亿元产业基金。江苏省于1月13日印发《江苏省“人工智能+”行动方案》，明确到2027年新一代智能终端和智能体普及率超过70%、到2030年产业规模超万亿元的三阶段目标，新增“人工智能+”新兴产业行动，覆盖人形机器人、低空经济、智能驾驶等领域。深圳市1月15日发布《深圳市打造人工智能OPC创业生态引领地行动计划（2026—2027年）》，计划到2027年底建成超过50万平方米OPC社区，培育千家高成长性人工智能创业企业，通过“智能券”“模型券”等工具降低研发门槛。浦东新区1月20日印发《浦东新区推进张江人工智能创新小镇建设工作方案》，目标到2027年集聚超800家垂类大模型应用企业、产业规模达650亿元，到2030年实现企业超1000家、规模超千亿元的双千目标，重点打造数字孪生小镇和标志性应用场景。

3. 具身智能与机器人领域资本投入和数据要素正同步加速聚集

具身智能赛道当前呈现资本与数据要素双向聚集的特征。投资机构的踊跃参与和高质量训练数据的快速积累正在推动行业从概念验证向实际应用加速转型。1月12日，自变量机器人完成10亿元A++轮融资，由字节跳动、红杉中国、深创投等联合投资，成为国内唯一同时获得字节、美团、阿里投资的具身智能企业，聚焦通用具身大模型和全栈自研硬件。1月15日，智元机器人与湖北人形机器人创新中心完成国内首例定制化人形机器人训练数据交易，交付数千小时针对搬箱、折衣等

细分场景的真机数据，依托 1.2 万平方米高仿真训练平台。

4. 国家政策与标准体系持续引导人工智能深度融入产业升级与民生服务

国家层面对人工智能的部署已形成从规划指引到标准规范的全链条支撑体系。这种从宽向深的推进为产业化和规模化应用提供了制度性支撑。1 月 12 日，工信部部长李乐成接受采访，阐述“十五五”开局思路，强调稳增长、扩需求、创动能、增活力，聚焦“人工智能+制造”专项行动、工业数据筑基、5G-A 演进，并布局量子、人形机器人、脑机接口等未来产业。1 月 13 日，上海市印发服务业提质增效和消费提振措施，覆盖金融、平台、文旅等 6 大领域，通过规范平台减佣降费、支持 AI 微短剧等激发消费新动能。同日，人社部表示将针对人工智能、新能源汽车、低空经济等领域实施技能提升专项培训。1 月 19 日，五部门联合印发零碳工厂建设指导意见，分阶段培育标杆工厂，覆盖汽车、锂电池、算力设施等重点行业。1 月 20 日，国家医保局发布辅助操作类立项指南，明确机器人辅助手术、远程手术定价逻辑。1 月 21 日，广东印发人工智能赋能交通运输措施，加强数据供给、算力保障和关键技术攻关，推动智慧公路、航道等应用。1 月 21 日，国务院新闻办发布会回顾 2025 年工业和信息化成效，人工智能核心产业规模超 1.2 万亿元，人形机器人进入量产元年。

5. 人工智能在商业、教育、养老等领域的标准化协议与融合应用稳步推进

人工智能的应用正从通用工具向特定领域的标准化协议和规范体系演进。在多个应用领域，AI 的落地呈现从零散试点向标准化协议与互操作框架转化的趋势，这为跨平台协作和规模化复制创造了条件。1 月 12 日，蚂蚁国际与谷歌合作推出通用商务协议 UCP，建立 AI 代理间通用交互语言，支持购物全流程无缝执行，已在 Gemini 和千问 App 中落地。1 月 13 日，团结香港基金发布 AI 在中小学教育研究，建议制定教学法为本的应用框架、建设全港 AI 资源平台。1 月 13 日，工信部等印发养老服务措施，提出研制人工智能终端智能化分级国家标准和移动应用适老化规范。1 月 16 日，支付宝联合伙伴发布中国首个智能体商业信任协议 ACT，打通电商、外卖等平台的 AI 任务执行。

三、美国动态：人工智能发展已转向资源要素争夺

1. 人工智能基础设施竞争已延伸至能源保障、电网接入与本土制造全链条

1 月 13 日，Meta 启动 Meta Compute 项目，由全球基础设施团队牵头，目标在 2030 年末前建成数十吉瓦级人工智能算力底座，并通过与 Vistra 签署为期 20 年的核电协议、参与小型模块化反应堆项目，提前锁定长期能源供应。1 月 14 日，OpenAI 与 Cerebras 签署超过 100 亿美元合作协议，部署 750 兆瓦低延迟专用算力，重点用于复杂查询、代码生成及代理任务的实时推理。1 月 15 日，OpenAI 发布采购方案请求，明确锁定美国本土在计算硬件、电力系统、冷却设备及机器人模组等环节的制造产能。1 月 16 日，谷歌公开表示，美国输电系统接入审批已成为数据中心并网的主要瓶颈，部分地区等待时间超过十年。同期，白宫与多州州长签署协议，要求数据中心运营商自行承担新增电网建设成本；敦促 PJM 电网运营商启动紧急电力拍卖，并为容量市场设定价格上限。

从这些动作可以看到，算力扩张的节奏已经明显快于能源体系与电网体系的消化能力。科技公司开始提前介入核电、专用芯片和本土制造环节，尽量把关键资源握在自己手中。政府的处理方式则更偏向于将增量成本回推给企业，避免电价压力外溢到居民端。人工智能竞争正在从模型和算力层面，进一步下沉到电力供应、输电能力和制造能力这些“硬约束”上。

2. 医疗健康领域已成为人工智能企业优先深耕的垂直方向

1 月 12 日，Anthropic 正式推出 Claude for Healthcare 工具集，基于 Claude Opus 4.5 模型接入美国医疗保险与医疗补助服务中心数据库、ICD-10 诊断代码库以及国家提供者标识符注册表，支持符合 HIPAA 标准的报告生成、事前授权审查等功能。同日，OpenAI 宣布完成对医疗初创

公司 Torch 的收购，将分散在不同医疗体系、影像机构及苹果健康应用中的患者数据整合为统一记忆库，直接服务于 ChatGPT 健康咨询用户。同样是在当天，英伟达与礼来制药在 J.P. Morgan 医疗健康大会上宣布建立人工智能联合创新实验室，计划五年内投入最高 10 亿美元，基于 BioNeMo 平台与 Vera Rubin 架构，将礼来的湿实验室与计算干实验室打通，用于药物发现、临床开发和制造环节的建模与优化。这一轮布局已经明显超出“健康助手”或对话工具的范畴，医疗正在成为少数几个最早承载“生产力级 AI”的行业之一。

3. 芯片出口政策在商业利益与国家安全之间寻求动态平衡

1 月 13 日，美国政府调整对华出口审批规则，批准英伟达 H200 芯片向中国客户销售，但附加第三方实验室性能验证、出口总量不超过美国采购量 50%、不得用于军事用途等多项限制条件。1 月 14 日，特朗普依据《贸易扩展法》第 232 条签署公告，对先进计算芯片及其衍生产品加征 25% 关税，首阶段自 1 月 15 日起实施，同时设置豁免通道，用于激励有利于美国本土供应链建设的进口行为。1 月 22 日，英伟达首席执行官黄仁勋确认，公司已超越苹果，成为台积电最大客户，人工智能需求在事实上主导了高端制程产能分配。从近期政策调整节奏看，美国对高端算力出口的态度更接近“有限放行、严格约束”。英伟达取代苹果成为台积电最大客户也反映出先进制程资源已开始围绕 AI 需求进行重新分配。

4. 人工智能硬件与具身智能正加速从概念验证转向量产落地

1 月 19 日，OpenAI 全球事务负责人确认，公司首款人工智能硬件设备预计于 2026 年下半年发布，形态可能以音频为核心。1 月 21 日，媒体披露苹果正在秘密研发代号为 AI Pin 的可穿戴设备，集成摄像头与麦克风，计划 2027 年首批量产 2000 万台。1 月 22 日，OpenAI 被披露已在旧金山设立机器人实验室，通过远程操作 Franka 机械臂采集家务任务数据，用于训练人形机器人。同日，马斯克在达沃斯论坛更新进度，特斯拉 Optimus 人形机器人计划于明年年底向公众销售，Cybercab 无人出租车预计今年年底在美国实现规模化推广。可以看出，企业正在同步搭建从模型到硬件、再到真实交互场景的闭环体系。相较此前更多停留在演示阶段，这一轮节奏已经开始直接对准消费级和商用市场。

5. 政府在人工智能战略资源配置与安全防护方面的角色持续强化

1 月 15 日，美国参议院通过拨款法案，大幅提高国家科学基金会、NASA 等机构科研经费，重点投向人工智能与量子计算方向。同日，多州州长与白宫签署协议，要求数据中心运营商承担电网新增建设成本。1 月 21 日，众议院外交委员会推进《人工智能监视法案》，将人工智能芯片出口的国家安全审查提升为国会层面的长期监督事项。整体来看，美国政府对人工智能的定位正在发生变化。早期更多强调放开创新和市场主导，而当前阶段开始同步强化资源配置、成本分担和安全审查。

四、欧洲动态：系统推进数字主权与基础设施自主

1. 欧盟正通过资金、法规与战略多管齐下构建数字主权体系

1 月 12 日，欧盟委员会启动《欧洲开放数字生态系统战略》证据征集，核心目标是通过开源政策框架降低对外部技术体系的依赖。1 月 15 日，“地平线欧洲”计划下公布两项新一轮征集，总额超过 3 亿欧元，重点投向可靠 AI 服务、机器人、量子、光子学以及开放互联网堆栈。1 月 19 日，EuroHPC 法规完成修订，扩大联合执行机构权限，为 AI 超级工厂和量子基础设施提供更明确的法律支撑，相关招标计划于 2026 年第一季度启动。1 月 21 日，《数字网络法案》提案正式提交，统一欧盟内部连接网络规则，推动光纤替换铜缆，并引入欧盟层级的频谱协调与防灾机制。

这一系列动作分布在研发、开源生态、算力设施与网络规则四个层面，欧盟正在尝试把“数字主权”拆解为可操作的工程路径。一端用开源补链，减少关键技术依赖；一端提前为超级算力与量子设施铺好法律与资金通道；同时，通过统一网络规则打通底层基础设施。从推进方式看，欧盟更强调制度先行和规范嵌入，这一取向使其自主路径与美国产业驱动型模式形成明显区分。

2. 欧美监管机构已在 AI 药物研发领域形成初步共识框架

1 月 14 日，欧洲药品管理局与美国食品药品监督管理局联合发布 10 项原则，用于规范人工智能在药物全生命周期中的应用，覆盖早期研究、临床试验、生产制造及安全性监测等环节，并明确该原则将作为未来监管框架的重要基础，支持跨国协作。欧盟健康专员随后指出，此举标志着跨大西洋在新型医疗技术领域的合作重新启动，与欧盟人工智能反思文件及相关战略保持一致，后续将持续细化。长期以来，AI 制药在不同司法辖区面临规则差异带来的落地障碍，此次欧美率先对从研发到临床的完整链条给出共用框架，也是在为技术进入真实医疗体系提前扫清部分制度摩擦。

3. 英国通过监管松绑积极释放机器人与国防技术潜力

1 月 16 日，英国监管创新办公室将机器人和国防技术列为优先支持领域，设立直接向监管层报告的绿色通道，用于集中清理制度障碍，并拨款 5200 万英镑在全国建设 5 个机器人应用中心，覆盖农业、医疗和制造等场景。1 月 20 日，政府发布《现代数字化政府路线图（2030）》，明确公共服务 AI 一体化路径，包括扩展 GOV.UK 应用体系，在 NHS 急诊预测、假释系统中推进 AI 规模化部署，并提出到 2030 年技术岗位占公务员比例达到十分之一的目标。与欧盟强调规则整合不同，英国的做法更偏向“先放开、再规范”，出发点都较为务实，重点放在尽快形成可复制的应用场景。

4. Swift 试验验证了传统金融与数字资产的无缝衔接路径

1 月 15 日，Swift 完成代币化债券互操作性试验，与多家银行合作实现跨平台结算，支持法定货币与数字货币在交付付款、利息支付及赎回环节的全流程对接。随后，Swift 宣布将在其现有架构中引入基于区块链的共享账本，支持 24 小时不间断的实时跨境支付，目前已有 30 多家机构参与测试。这项试验的关键，并不在于代币化债券本身，而在于接口方式的选择。Swift 刻意强调通过标准化集成接入区块链系统，而无需让传统机构直接处理底层链上复杂性。从长期看，这一路径更有可能被大规模金融机构接受。

五、其他国家：亚洲主要经济体加大 AI 硬件投资

1. 韩、日、印加大对 AI 算力与芯片基础设施的集中投入力度

1 月 13 日，SK 海力士宣布将在清州市投资约 129 亿美元建设一座先进芯片封装厂，计划于 2027 年底前投产，主要聚焦高带宽内存（HBM）生产，以满足 AI 计算对存储芯片的强劲需求。1 月 20 日，日本财务大臣在达沃斯世界经济论坛上表示，政府与民间部门将在 AI 及半导体领域共同投入超过 3300 亿美元，重点支持 2 纳米芯片量产以及“物理 AI”等相关项目。同时，信实工业（Reliance Industries）近期宣布将在未来五年内于古吉拉特邦投资约 776 亿美元（约合 7 万亿卢比），其中核心内容包括建设印度规模最大的 AI 就绪数据中心，并推广本地语言 AI 服务。

上述举措均高度聚焦于算力基础设施、存储芯片及先进制程等领域，体现了相关国家对 AI 硬件基础的持续强化。尽管各国投资侧重点有所差异，韩国着力解决存储瓶颈、日本借助政府资金撬动民间资本并瞄准下一代工艺、印度则注重数据中心建设与普惠型 AI 服务的结合，但其根本目标一致，即抢占 AI 时代基础设施的战略制高点。在全球 AI 供应链高度整合的格局下，这种资源分配趋同的现象表明，各国正围绕本土化竞争展开角逐。最终，谁能在执行效率、能源供给匹配以及产

业生态整合方面占据优势，谁就更有可能在实际算力供给上形成相对领先地位。

2. 中国与东盟在数字与人工智能领域的区域合作深化

1月16日，第六次中国—东盟数字部长会议在越南河内举行。会议通过了《中国—东盟建立可持续和包容性的数字生态合作行动计划（2026—2030）》以及《2026年中国—东盟数字合作计划》，明确将在年内成立中国—东盟人工智能产业创新中心和东盟数字学院。双方重点围绕数字基础设施建设、AI应用创新、数字治理等领域开展对接，旨在推动技术标准互认和跨国项目落地。

中国与东盟在数字及人工智能领域的制度化合作深化，反映出区域层面对技术治理与产业协同的现实需求日益上升。在全球技术体系趋于分化的背景下，单一国家难以独立构建完整的标准、人才和应用生态。通过区域机制建立稳定的合作框架，有助于有效降低制度性摩擦和技术壁垒。

3. 韩国建立全面人工智能法律框架

1月22日，韩国《人工智能基本法》正式生效。该法对生成式AI规定必须添加数字水印并确保内容可溯源，对涉及生命安全、金融、医疗、教育等的高影响力AI要求制定风险管理方案并保留人类监督环节。同时对全球主要AI企业实施本土代表制度，设置一年合规宽限期以减轻初期压力。

韩国在全球较早推出可执行的综合AI立法，其中，水印、可解释性和高风险场景监管等要求，直指当前公众信任度最低的几个痛点，本土代表制则强化了域内执法的有效性。这种框架在创新与治理之间试图取得平衡，但也暴露了监管强度与产业活力之间的内在张力，尤其对依赖海外技术的大型公司而言，合规成本的上升是显而易见的现实压力。

4. 沙特阿美利用AI实现年度最高50亿美元经济收益

1月20日，沙特阿美CEO阿明·纳赛尔在达沃斯世界经济论坛上表示，AI及先进技术2025年为公司带来超出预期的经济回报，预计“技术实现价值”达30亿至50亿美元。这一数字主要来自降低钻井和维护成本、提升油田生产效率等方面。沙特阿美依托75亿美元风险投资部门，并与主权财富基金PIF在国家AI项目Humain上合作。一方面利用前沿技术长期扩大油气开采能力，另一方面通过与超大规模云服务商合作，将AI应用商业化并推广至整个能源行业。纳赛尔指出，公司目标是推动能源行业更智能地运用AI，将算法优势转化为实际现金流和生产力。

六、技术前沿：AI 拉动铜价预期大幅修正

1. 英伟达大幅下调数据中心铜使用量估算

1月14日，市场注意到英伟达此前一篇技术论文中关于数据中心用铜量的测算可能存在严重偏差。论文称在传统架构下，1吉瓦机架需用铜50万吨，该数字迅速在市场中传播，并被用于推算未来数据中心可能带来的铜需求增量。然而，咨询机构Thunder Said Energy指出，原始数据单位应为50万磅，约合226吨。1月16日英伟达发布更正，将1吉瓦机架铜母线用量修订为约200吨，仅为原估算的0.04%。这一修正基本扭转了此前对“AI耗铜”的预期判断。如果按麦肯锡对2030年全球数据中心容量219吉瓦的测算，对应输配电系统用铜约为4.4万吨，仅占当年全球铜产量的0.15%左右。从数量级看，数据中心对铜需求的直接拉动仍然十分有限。

不过，从更长周期观察，铜需求的中枢仍将受到电气化趋势的持续支撑。电网向智能化和分布式方向演进，虽然可能提高单位容量的用铜效率，但同时材料纯度、导电稳定性和加工精度提出更高要求。综合来看，铜的长期压力更可能来自电网升级、新能源汽车普及、制冷需求扩张等多条路径在同一时间窗口内的叠加，而非某一单一应用场景的集中放量。

2. AI基础设施正在向“系统内生智能”方向演进

1月12日，DeepSeek团队公布条件记忆机制论文并开源Engram模块，核心思路是将静态知

识存储与动态计算过程分离，以缓解 Transformer 在知识检索环节的效率瓶颈。次日，摩尔线程联合智源研究院宣布，基于国产 MTT S5000 千卡集群完成 RoboBrain 2.5 具身智能模型的全流程训练。1 月 14 日，智谱 AI 与华为合作开源 GLM-Image，该模型完全在国产昇腾芯片上完成从数据处理到预训练的全流程，并在四天后登顶 Hugging Face Trending 榜首。1 月 20 日，阿里云在 PolarDB 开发者大会发布 AI 数据湖库 Lakebase，将大模型能力直接嵌入数据库内核，实现多模态数据统一管理和库内推理。AI 能力正从外挂式组件转向基础设施的内生组成部分。从结果看，GLM-Image 的榜单表现以及 RoboBrain 的训练效果，至少说明本土软硬件协同在关键指标上已经具备稳定可用性，开始进入可工程化阶段。

3. 推荐算法与感知模型开源开始重塑透明度边界

1 月 15 日，阶跃星辰发布原生语音推理模型 Step-Audio-R1.1，在全球权威语音推理榜单上以 96.4% 的准确率排名第一，首包延迟为 1.51 秒，并同步开放模型权重。当天稍晚，X 平台在 GitHub 开源其核心推荐算法，系统已全面转向由 Grok 驱动的 Transformer 模型，基本不再依赖人工特征与规则，内容排序完全由用户历史行为数据实时打分完成。这一轮开源的意义已不止于技术展示。X 的算法调整，相当于将内容分发逻辑直接交由模型主导；阶跃星辰则把实时“边听边想”的推理能力推进到可部署层面。两项进展共同缩短了算法从实验阶段到产品验证的周期，也使外部能够更直观地观察真实系统的演进路径。

4. 医疗 AI 正加速向可信、可落地的专业工具转型

1 月 13 日，谷歌更新医疗 AI 产品线，发布支持高维影像的 Med-Gemma 1.5 以及医疗语音转文本模型 MedASR，并同步开源。1 月 21 日，亚马逊 One Medical 在应用中上线 Health AI 助手，基于患者完整病历和用药记录提供 24 小时个性化指导与任务闭环。1 月 22 日，百川智能发布 Baichuan-M3 Plus 医疗大模型，通过六源循证机制将幻觉率降至 2.6%，并向国内医疗机构免费开放 API 接口。从应用链条看，医疗 AI 的覆盖范围正在由影像诊断和语音转写，延伸至日常健康管理以及证据可追溯的临床决策支持。

5. 硬件与物理基础研究持续拓展计算的物理边界

1 月 13 日，西安电子科技大学团队公布半导体散热方面的新进展，通过离子注入诱导成核技术将氮化镓器件功率密度提升 30%-40%。1 月 16 日，Neuralink 受试者披露其脑机接口系统已支持远程 OTA 升级，无需二次手术即可优化信号处理与系统稳定性。1 月 23 日，中科院物理所团队在《Science》报道，在萤石结构铁电薄膜中首次观测到一维带电畴壁。同日，星凡星启发布全国产 XH3118 卫星算力模组，在轨推理吞吐量提升四倍，显存占用明显下降。这些进展短期内对产业格局影响有限，但集中出现在散热、材料形态和边缘算力三个方向。它们更多是在为更高密度、更复杂场景下的计算需求提前铺路。

七、智库观点：“因国制宜”发展人工智能

1. WEF：《Rethinking AI Sovereignty: Pathways to Competitiveness through Strategic Investments》

世界经济论坛与贝恩公司 2026 年联合发布的这份白皮书，对 AI 主权的理解进行了重要调整。它不再把完全自给自足当作目标，而是指出真正的 AI 主权体现在一个经济体能否通过本地投资和可靠的国际合作，掌握对 AI 生态的战略控制权，同时保持足够的灵活性和抗风险能力。

当前全球 AI 价值链的投资格局目前仍由中美两国主导，双方合计占到 65% 的份额，资金主要

流向基础设施和应用服务两个方向。2010年至2024年间，AI基础设施累计投资已超过6000亿美元；预计到2030年，应用服务领域的年度投资将达到1.5万亿美元，成为扩张最迅速的部分。各国关注的重点已逐渐从技术突破本身，转向如何通过战略性布局和合作网络占据更有利的位置。

基础设施是AI竞争力的基石，涉及数据中心、算力资源、云平台等紧密关联的环节。过去以硬件直接采购为主的投资方式，正在被公私合作、区域联合融资等更多元模式取代，新一代云厂商和本土服务商的崛起尤为明显。白皮书根据价值链覆盖广度和关键要素成熟度两个标准，把全球经济体分为五种类型：以中美为代表的全球领导者，在全链条尤其是基础模型领域占据主导；新加坡这样的生态系统构建者，通过均衡投入和区域协作，形成了区域性AI中心；多数传统工业国家属于选择性参与者，主要集中在数据和垂直应用上，硬件与算力依赖外部；印度等新兴经济体被定位为应用加速器，以本地场景落地为切入点，逐步补强基础设施；基础较薄弱的经济体则归为新兴合作者，主要依靠外部支持，在民生领域尝试AI应用。

针对这些不同起点，白皮书给出了五条通向生态系统构建者的转型路径。已有较好研发和产业基础的选择性参与者，可以选择与全球领导者形成互补的细分方向，加大资本和产业协同力度；被归为应用加速器类型的国家则应利用混合融资扩大算力规模、建立国家级数据平台，把应用端的先发优势转化为全链条实力；资源受限的新兴合作者可先通过公私合作推进本地应用，站稳应用加速器位置后再向上发展。其中，资本、土地、能源条件较好的新兴经济体，应优先建设基础设施，成为区域算力节点并进入选择性参与者行列；同时具备能源、土地和数据联通优势的经济体，则可以直接同步推进数据、应用和基础设施建设。这些路径都从各经济体的实际情况出发，提供了清晰且务实的行动方向。

白皮书建议政策制定者把注意力集中在几个关键点上：一是根据自身比较优势制定务实战略，避免资源过于分散；二是把AI主权重新理解为一种战略性的相互依赖关系，通过区域联盟填补能力短板；三是集中资金培育本土AI生态；四是提高基础设施的互操作性以避免被单一供应商锁定；五是推动数据和应用在医疗、农业等领域的负责任落地。对绝大多数经济体而言，追求彻底的AI主权既不现实也不必要。真正可行的方式是通过聚焦自身优势、建立可信合作关系并推动负责任的规模化应用，在AI浪潮中获得持续价值，并逐步缩小与领先者之间的差距。

2. Anthropic: 《Anthropic Economic Index report: economic primitives》

基于2025年11月Opus 4.5发布前的Claude使用数据，报告引入“经济基本要素”框架，通过任务复杂度、所需教育水平、用途、AI自主程度和成功率五项指标，评估AI对经济的多维度影响。**从使用情况来看，Claude呈现高度集中化特征。**在Claude.ai平台，前10项任务占总请求的24%，在API流量中则升至32%。编码任务在Claude.ai占比34%（从去年3月的40%峰值有所回落），在API端高达46%。平台端更注重协作，52%的对话用于学习或迭代；API则偏向自动化，占比75%。从地域来看，美国、印度和日本的使用量居前，并且与人均GDP正相关。GDP每增1%，使用量约升0.7%。低收入国家多青睐课程学习，高收入国家则更侧重工作和个人事项。

在效率层面，任务复杂度与AI提速呈正相关，但成功率存在明显下降。大学级任务（约需16年教育）在Claude.ai可达到12倍提速的效果，高中级任务只有9倍。对于人类需超5小时的任务，API成功率仅45%。用户提示与输出教育水平高度相关，**用户技能能够直接决定AI效果。**

就业影响同样分化明显。数据录入、数据库架构等岗位AI成功率高，因此未来受到冲击较大。在考虑成功率的情况下，AI预计每年推动美国劳动生产率提升1.0至1.2个百分点。

3. WEF: 《Global Cybersecurity Outlook 2026》

报告提炼出2026年网络安全领域三大核心趋势，深刻影响全球组织的风险应对策略：

第一，AI正在加速网络攻防竞赛。94%的受访者认为，AI将是未来一年网络安全领域最具颠覆性的力量。在防御侧，77%的组织已经引入AI工具来加强钓鱼检测（52%）和入侵响应（46%）；

在攻击侧，87%的受访者把 AI 相关的漏洞列为 2025 年增速最快的风险，其中数据泄露（34%）和对抗性攻击能力的升级（29%）最受关注。虽然评估 AI 工具自身安全性的组织比例从 2025 年的 37% 上升到 2026 年的 64%，但仍有三分之一的组织在正式部署前没有完整的验证流程。

第二，地缘政治成为网络安全核心变量。64%组织将地缘政治驱动攻击（如基础设施破坏、间谍）纳入核心。超大型组织（员工超 10 万）中有 91%因为地缘政治波动而调整了安全策略，主要动作包括强化针对国家行为体的情报收集（36%）和加深与政府部门的合作（33%）。区域间的信心差距非常大：中东和北非有 84%的受访者对自己国家应对重大网络事件的能力有信心，而拉美和加勒比地区只有 13%。

第三，网络欺诈已成为全球普遍威胁。73%的受访者表示自己或身边人在 2025 年遭遇过网络欺诈，钓鱼攻击（62%）、支付欺诈（37%）、身份盗窃（32%）排在前列。撒哈拉以南非洲（82%）和北美（79%）的发生率最高。在关注焦点上，CEO 和 CISO 的分歧很明显。CEO 把网络欺诈排在首位，CISO 则仍更关注勒索软件和供应链安全。而在高韧性组织里，CEO 对 AI 漏洞的警惕度明显更高，低韧性组织则继续把欺诈视为头号风险。

4.WEF:《Proof over Promise: Insights on Real-World AI Adoption from 2025 MINDS Organizations》

报告基于 2025 年 MINDS 项目中数百家组织的真实实践，系统梳理了 AI 从概念承诺走向实际见效的落地路径。该项目覆盖 30 多个国家，横跨能源、医疗、金融、制造等多个行业。报告提炼出五项关键洞察。

第一，很多组织已经把 AI 上升到战略层面，并习惯把已有项目的收益再投入到新的场景，约 75%的项目遵循这种“再投资”逻辑。**第二**，人机协同能带来显著的效率跃升，比如 Phagos 公司通过 AI 与专家知识结合，把噬菌体疗法开发周期从两年压缩到两个月。**第三**，数据能力是落地的基石，宁德时代利用约 600TB 历史数据搭建电池设计平台，把准确率从 70%提升到 95%。**第四**，技术架构正向云+本地的混合模式倾斜，55%的申请组织选择了这种方式来平衡灵活性和合规需求。**第五**，负责任 AI 必须实施分层治理：低风险场景可以更多依赖自动化，高风险场景则需要保留较强的人回路，以维护信任和可控性。

这些要素协同产生的实际效果在案例中得到验证：富士康 AI 工厂在换线作业上减少 50%工作量，故障排查时间缩短 30%；蚂蚁集团医疗 AI 平台诊断准确率超过 90%，服务约 1.6 亿用户；国家电网城市级调度系统每年减排 51 万吨二氧化碳，节约成本约 1.12 亿美元；施耐德电气 AI 微电网在单一场景实现 28%的碳减排。

报告强调，AI 要实现规模化价值，靠的不是单一技术堆叠，而是战略、人才、数据、技术架构和治理体系的整体协同。往前看，提升模型可靠性、建立跨组织的数据协作机制、推动行业标准趋同，将成为释放更大潜力的关键方向。

5.RAND:《U.S.-China Competition for Artificial Intelligence Markets》

报告利用 2024 年 4 月到 2025 年 5 月覆盖 135 个国家的网站流量数据，追踪了主要美中 LLM 平台的访问情况，评估市场渗透、地理特征，以及 2025 年 1 月中国模型 DeepSeek R1 发布带来的冲击。**数据显示，2024 年 4 月到 2025 年 5 月全球 LLM 使用量增长迅猛。**主要平台的月访问总量从 2024 年 4 月的约 24 亿次，上升到 2025 年 8 月的近 82 亿次，增幅近三倍。美国模型在 2025 年 8 月仍占据全球访问量的 93%。但在 DeepSeek R1 的发布两个月内，中国本土 LLM 访问量增长 460%，全球市场份额从 3%增长到 13%。同时，中国模型已在 30 个国家渗透率超过 10%，在 11 个国家超过 20%，尤其在发展中国家以及与中国经济政治联系紧密的国家表现突出。

报告分析认为，价格、外语支持和外交推广目前都不足以完全解释采用格局。中国模型推理成本只有美国对手的六分之一到四分之一，但免费层让大多数用户没有直接感受到价差。美国模型在

多语言覆盖上的早期优势已被中国大幅追平。中国在 AI 外交上的高频接触可能对政府和大企业合作有帮助，但对普通用户的模型选择影响相对有限。报告指出，美国目前仍然凭借先发和技术优势保持主导，但 DeepSeek R1 的案例表明，只要切换成本不高，性能、价格或其他变量的变动就能够在短期内显著改变领先者的市场份额。

6.IMF: 《Bridging Skill Gaps for the Future: New Jobs Creation in the AI Age》

新技术正加速重塑劳动力市场，新技能（以 IT 和 AI 为主）从十多年前的罕见，迅速成长为发达经济体约十分之一、新兴市场约二十分之一的职位要求。这些技能最早在美国等发达国家显现，随后向全球扩散，主要流向专业、技术和管理岗位。IT 技能占比过半，AI 相关部分持续扩大，医疗等领域技能也稳步上升。

总体上，新技能需求正在逐步推动工资与就业的改善。相关职位薪资通常高出 3-3.4%；在美国地方市场，新技能职位占比每提高 1 个百分点，平均工资上升 2.3%、就业增加 1.3%；德国工资也有约 0.9% 的增幅。高技能和低技能群体获益最明显，中等技能群体基本未见改善，这也进一步加剧了职业的两极分化。目前，AI 相关新技能的发展尚未带动整体就业增长，反而对某些群体形成压力。在 AI 影响大但互补性较低的岗位中，高需求地区五年后就业水平下降约 3.6%，其中，对白领中层、年轻人和部分 IT 从业者冲击较大。

八、风险提示

对政策理解不到位的风险；政策落实不及预期的风险；技术发展不确定性风险。

图表目录

图 1： 全球 iOS 与 Google Play 平台 APP 下载次数（单位：百万）	3
图 2： 人工智能公司收入飙升（单位：百万美元）	3
图 3： 全球 Top 25 大语言模型（LLM）幻觉率情况（单位：%）	4

分析师承诺及简介

本人承诺以勤勉的执业态度，独立、客观地出具本报告，本报告清晰准确地反映本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与、未来也将不会与本报告的具体推荐或观点直接或间接相关。

彭雅哲：中国银河证券新发展研究院数字经济研究员。

免责声明

本报告由中国银河证券股份有限公司（以下简称银河证券）向其客户提供。银河证券无需因接收人收到本报告而视其为客户。若您并非银河证券客户中的专业投资者，为保证服务质量、控制投资风险、应首先联系银河证券机构销售部门或客户经理，完成投资者适当性匹配，并充分了解该项服务的性质、特点、使用的注意事项以及若不当使用可能带来的风险或损失。

本报告所载的全部内容只提供给客户做参考之用，并不构成对客户的投资咨询建议，并非作为买卖、认购证券或其它金融工具的邀请或保证。客户不应单纯依靠本报告而取代自我独立判断。银河证券认为本报告资料来源是可靠的，所载内容及观点客观公正，但不担保其准确性或完整性。本报告所载内容反映的是银河证券在最初发表本报告日期当日的判断，银河证券可发出其它与本报告所载内容不一致或有不同结论的报告，但银河证券没有义务和责任去及时更新本报告涉及的内容并通知客户。银河证券不对因客户使用本报告而导致的损失负任何责任。

本报告可能附带其它网站的地址或超级链接，对于可能涉及的银河证券网站以外的地址或超级链接，银河证券不对其内容负责。链接网站的内容不构成本报告的任何部分，客户需自行承担浏览这些网站的费用或风险。

银河证券在法律允许的情况下可参与、投资或持有本报告涉及的证券或进行证券交易，或向本报告涉及的公司提供或争取提供包括投资银行业务在内的服务或业务支持。银河证券可能与本报告涉及的公司之间存在业务关系，并无需事先或在获得业务关系后通知客户。

银河证券已具备中国证监会批复的证券投资咨询业务资格。除非另有说明，所有本报告的版权属于银河证券。未经银河证券书面授权许可，任何机构或个人不得以任何形式转发、转载、翻版或传播本报告。特提醒公众投资者慎重使用未经授权刊载或者转发的本公司证券研究报告。

本报告版权归银河证券所有并保留最终解释权。

评级标准

评级标准	评级	说明
评级标准为报告发布日后的 6 到 12 个月行业指数（或公司股价）相对市场表现，其中：A 股市场以沪深 300 指数为基准，新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准，北交所市场以北证 50 指数为基准，香港市场以恒生指数为基准。	行业评级	推荐：相对基准指数涨幅 10%以上
		中性：相对基准指数涨幅在-5%~10%之间
		回避：相对基准指数跌幅 5%以上
	公司评级	推荐：相对基准指数涨幅 20%以上
		谨慎推荐：相对基准指数涨幅在 5%~20%之间
		中性：相对基准指数涨幅在-5%~5%之间
		回避：相对基准指数跌幅 5%以上

联系

中国银河证券股份有限公司 研究院

机构请致电：

深圳市福田区金田路 3088 号中洲大厦 20 层

深广地区：

苏一耘 0755-83479312 suyiyun_yj@chinastock.com.cn

程曦 0755-83471683 chengxi_yj@chinastock.com.cn

上海浦东新区富城路 99 号震旦大厦 31 层

上海地区：

林程 021-60387901 lincheng_yj@chinastock.com.cn

李洋洋 021-20252671 liyangyang_yj@chinastock.com.cn

北京市丰台区西营街 8 号院 1 号楼青海金融大厦

北京地区：

田薇 010-80927721 tianwei@chinastock.com.cn

褚颖 010-80927755 chuying_yj@chinastock.com.cn

公司网址：www.chinastock.com.cn