

垂类AI应用专题： 领先的独立大模型公司智谱，产品矩阵快速扩张

行业研究 · 行业快评

计算机 · 人工智能

投资评级：优于大市（维持评级）

证券分析师：熊莉

021-61761067

xiongli1@guosen.com.cn

S0980519030002

联系人：侯睿

hourui3@guosen.com.cn

- **国内领先的AI公司，产品进入快速迭代阶段。**北京智谱华章科技股份有限公司成立于2019年，是中国领先的AI公司，致力于追求通用人工智能创新。公司为机构客户及个人用户提供通用大模型服务，已为逾八千家机构客户、约80百万台设备提供支持。以2024年收入计，公司是中国最大的独立大语言模型厂商及第二大大语言模型厂商，市占率达6.6%，2022、2023、2024、2025H1公司日均token消耗量分别为5亿、21亿、0.2万亿及4.6万亿。
- **公司收入高速增长，净利润短期承压。**2025H1公司实现收入1.91亿元，同比增长35.03%。分业务来看，本地化部署的收入由2024H1的0.27亿元增加503.5%至1.62亿元，占公司收入比重达84.81%，主要由公司持续为客户提供更复杂的大模型服务，并将其扩展到更多客户群体推动；云端部署的收入由2024H1的0.18亿元增加60.7%至0.29亿元，占收入比重达15.18%，主要得益于客户数量及服务产品应用场景增加。2025H1公司实现净利润-23.51亿元，主要由算力服务费增加、研发开支等成本增加导致。
- **客户快速涵盖广泛，下游集中度高。**公司拥有广泛且多元化的客户群体，并实现用户的快速扩张。公司的客户包括企业、公共部门实体及个人用户。2022、2023、2024、2025H1，来自五大客户的收入分别占公司总收入的55.4%、61.5%、45.5%及40.0%，来自最大客户的收入分别占总收入的15.4%、14.7%、19.0%及11.0%。
- **模型能力持续提升，达到全球领先水平。**2025年12月，公司正式发布GLM-4.7并开源，新版本面向Coding场景强化了编码能力、长程任务规划与工具协同，并在多项主流公开基准测试中取得开源模型中的领先表现。在主流基准测试表现中，GLM-4.7的代码能力对齐Claude Sonnet 4.5：在SWE-bench-Verified获得73.8%的开源SOTA分数；在LiveCodeBench V6达到84.9%的开源SOTA分数，超过Claude Sonnet 4.5；SWE-bench Multilingual达到66.7%（提升12.9%）；Terminal Bench 2.0达到41%（提升16.5%）。
- **下游应用场景广，兼具市场规模与成长空间。**据沙利文数据，预计至2030年人工智能将赋能全球至少20%的日常商业决策，为全球80%的消费者主流智能设备提供支持，创造逾20万亿美元的经济影响。中国人工智能市场规模由2022年的937亿元增至2024年的1607亿元，期间复合年增长率为31.0%；预计于2030年，中国人工智能市场规模将进一步增至9930亿元，2024年至2030年的复合年增长率为35.5%。
- **风险提示：**产品研发未达预期风险、关键技术人员流失风险、核心技术泄密、持续亏损与盈利拐点不确定、经营现金流为负与外部融资依赖、算力与技术服务成本上行、毛利率波动、国际贸易摩擦加剧风险、规模扩张后的管理风险、行业政策变更的风险。

- 【 01 】 领先的AI公司，产品进入快速迭代阶段
- 【 02 】 模型能力持续提升，全球AI市场快速扩张
- 【 03 】 募集资金投入产品，维持行业优势
- 【 04 】 风险提示

国内领先的AI公司，产品进入快速迭代阶段



- 北京智谱华章科技股份有限公司成立于2019年，是中国领先的AI公司，致力于追求通用人工智能创新。公司为机构客户及个人用户提供通用大模型服务，已为逾八千家机构客户、约80百万台设备提供支持。以2024年收入计，公司是中国最大的独立大语言模型厂商及第二大大语言模型厂商，市占率达6.6%，2022、2023、2024、2025H1公司日均token消耗量分别为5亿、21亿、0.2万亿及4.6万亿。成立至今公司大致经历以下发展阶段：
 - ◆ 创业初期，产品逐步丰富（2019-2023年）：2019年公司成立后，2020年启动预训练框架通用语言模型（GLM）的开发，2021年公司发布首个百亿参数预训练大模型GLM-10B，同步推出模型即服务（MaaS）产品开发及商业化平台，2022年公司发布开源大模型GLM-130B与高性能代码模型CodeGeeX，2023年发布对话模型ChatGLM，并于3月发布开源ChatGLM-6B，发布中国首批通过监管备案的大模型产品智谱清言；
 - ◆ 产品快速迭代，公司进入快速发展阶段（2024年至今）：2024年，公司产品扩张进入快车道，发布具备智能体编排协同能力的基座模型GLM-4以及GLM-4-Plus，并在智谱清言上线AI视频通话功能、发布视觉理解基座模型GLM-4V、发布视频生成模型CogVideoX和移动应用程序智谱清影、发布端到端情感语音生成模型GLM-4-Voice及用于自主设备操作的基座智能体模型AutoGLM、发布处理高级推理任务的反思模型GLM-Z1，2025年公司发布端到端模型GLM-Realtime以及AI Agent AutoGLM Ruminatiion，公司最新基础模型已迭代至GLM-4.7。

图：公司发展历程

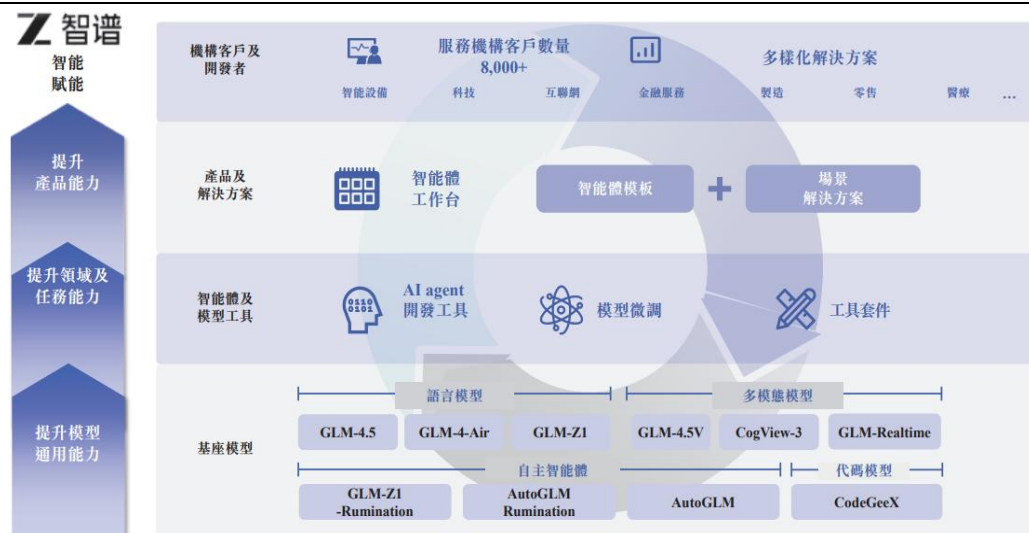
2019年	• 本公司成立。
2020年	• 我們啟動預訓練框架通用語言模型(GLM)的開發。
2021年	• 我們發佈首個百億參數預訓練大模型GLM-10B。
	• 我們推出模型即服務(MaaS)產品開發及商業化平台。
2022年	• 我們發佈開源大模型GLM-130B。
	• 我們於9月發佈高性能代碼模型CodeGeeX。
2023年	• 我們於3月發佈十億參數基座對話模型ChatGLM，並於3月發佈開源ChatGLM-6B。
	• 我們發佈中國首批通過監管備案的大模型產品之一智譜清言。
2024年	• 我們發佈具備智能體編排協同能力的基座模型GLM-4。
	• 我們發佈視覺理解基座模型GLM-4V。
	• 我們發佈下一代視頻生成模型CogVideoX和移動應用程序智譜清影。
	• 我們發佈GLM-4系列中最新、最先進的基座模型GLM-4-Plus，並在智譜清言上線AI視頻通話功能。
	• 我們發佈端到端情感語音生成模型GLM-4-Voice及用於自主設備操作的基座智能體模型AutoGLM。
	• 我們發佈處理高級推理任務的反思模型GLM-Z1。
2025年	• 我們發佈端到端模型GLM-Realtime，支持實時視聽互動、哼唱、2分鐘記憶保留及函數調用功能。
	• 我們發佈首個集深度研究和操作執行於一體的AI Agent AutoGLM Ruminatiion，可實現自主多步驟推理和行動。
	• 我們發佈GLM-4.5、GLM-4.5V、AutoGLM的更新版本(亦稱為「AutoGLM 2.0」)及GLM-4.6。

资料来源：公司招股意向书，国信证券经济研究所整理

国内领先的AI公司，产品进入快速迭代阶段

- 1) 全面的模型矩阵：公司已建立全面的AI模型矩阵，在语言、多模态、智能体及代码能力方面展现出行业领先的性能；
- 2) 丰富生态：公司利用机构客户大规模地触达其自身客户，从而以间接但高效的方式扩大对该等终端用户的影响；
- 3) 便捷的自定义部署：公司模型可托管于云端，通过应用程序编程接口（API）存取，进行本地化部署以计算私有数据集，或预安装于端侧设备上。公司亦提供开箱即用的模板及插件库，并提供模型微调、部署及智能体开发等标准化集成工具；
- 4) 安全性及可靠性：公司以领先的安全性能为支持，开发了安全可靠的模型产品，构建了安全可扩展的架构。
- 凭借公司的MaaS平台，公司形成了连接算力资源提供商、智能设备制造商、机构客户、开发者及个人客户的网络，从真实世界的模型部署中获得丰富的深度洞察，帮助公司更好地理解人们在不同应用场景中如何实际运用并受益于AI，并有针对性地优化训练策略，从而形成良性的洞察力飞轮效应。

图：微软资本开支与现金流复盘

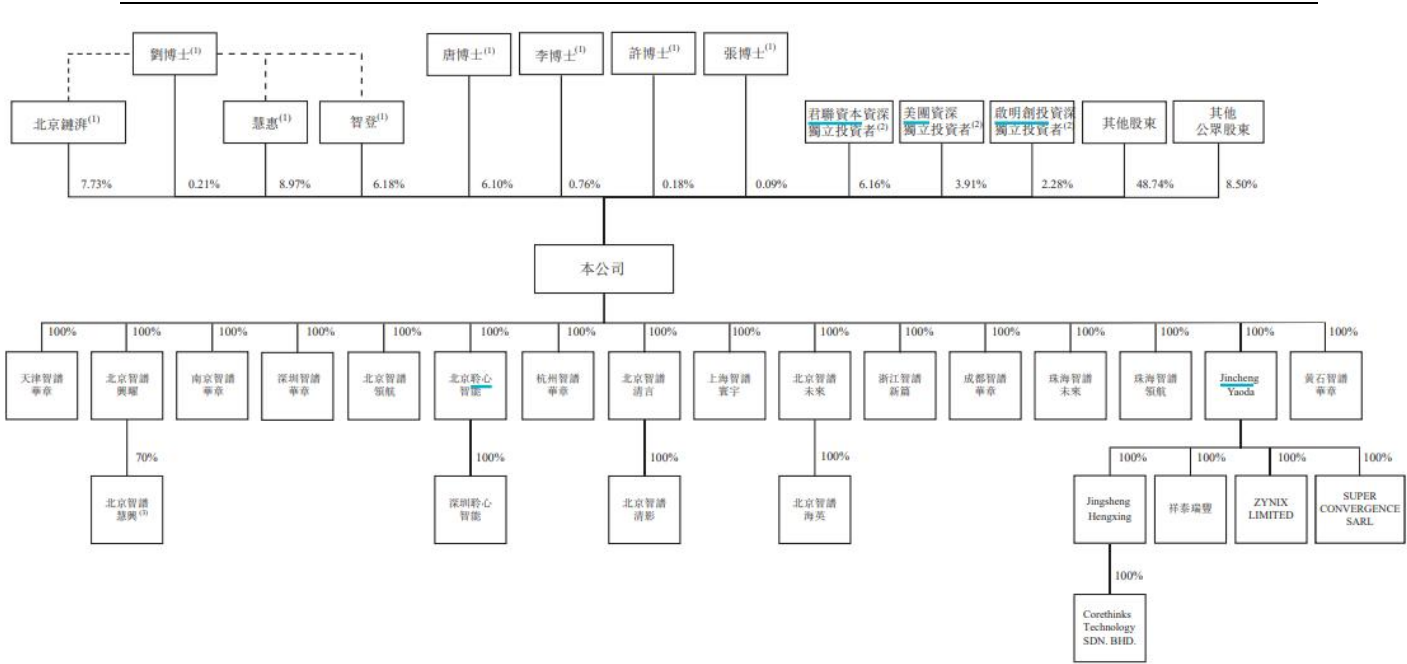


资料来源：公司招股意向书，国信证券经济研究所整理

国内领先的AI公司，产品进入快速迭代阶段

- 刘德兵博士为实际控制人，合计持股达23.09%。北京链湃科技、刘德兵博士、唐杰博士、李涓子博士、许斌博士、张鹏博士、珠海横琴慧惠及珠海横琴智登为一致行动人士，并为公司的控股股东集团，北京链湃、慧惠及智登因刘博士担任其各自普通合伙人而受刘博士控制，合计控股达23.09%，为公司实际控制人。公司领航资深独立投资者包括君联相道、君联锦帆及社保中关村创新基金、美团资深独立投资者天津三快，其他投资者包括启明融乾及启明融凯。

图：公司股东结构（截至2026年1月）

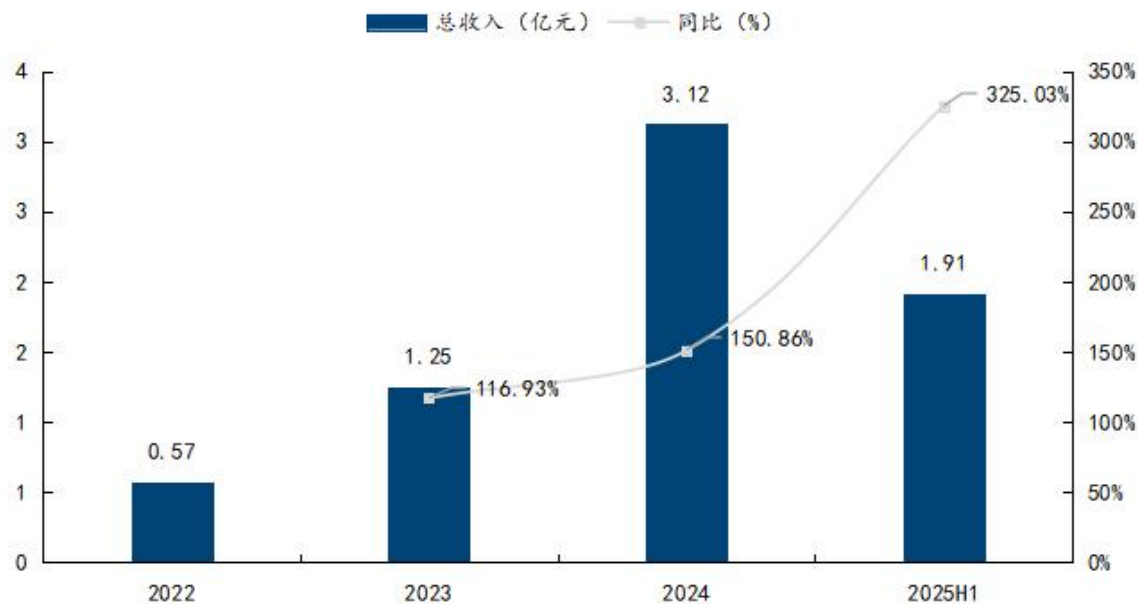


资料来源：公司招股意向书，国信证券经济研究所整理

公司收入高速增长，净利润短期承压

- 2025H1公司实现收入1.91亿元，同比增长35.03%。分业务来看，本地化部署的收入由2024H1的0.27亿元增加503.5%至1.62亿元，占公司收入比重达84.81%，主要由公司持续为客户提供更复杂的大模型服务，并将其扩展到更多客户群体推动；云端部署的收入由2024H1的0.18亿元增加60.7%至0.29亿元，占收入比重达15.18%，主要得益于客户数量及服务产品应用场景增加。2025H1公司实现净利润-23.51亿元，主要由算力服务费增加、研发开支等成本增加导致。

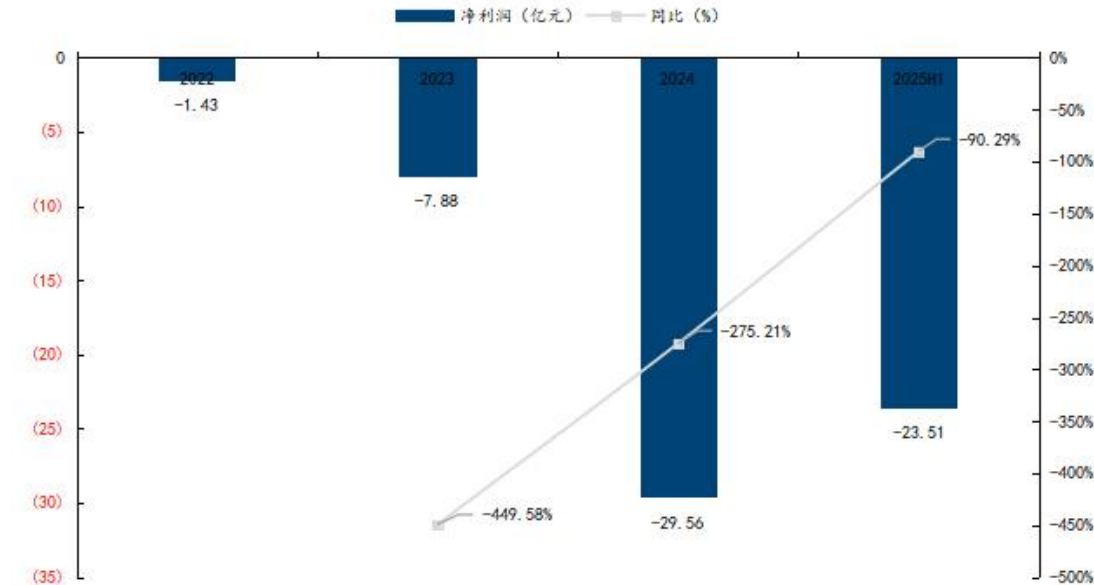
图：2025H1公司实现收入1.91亿元



资料来源：公司招股意向书，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

图：2025H1公司实现净利润-23.51亿元

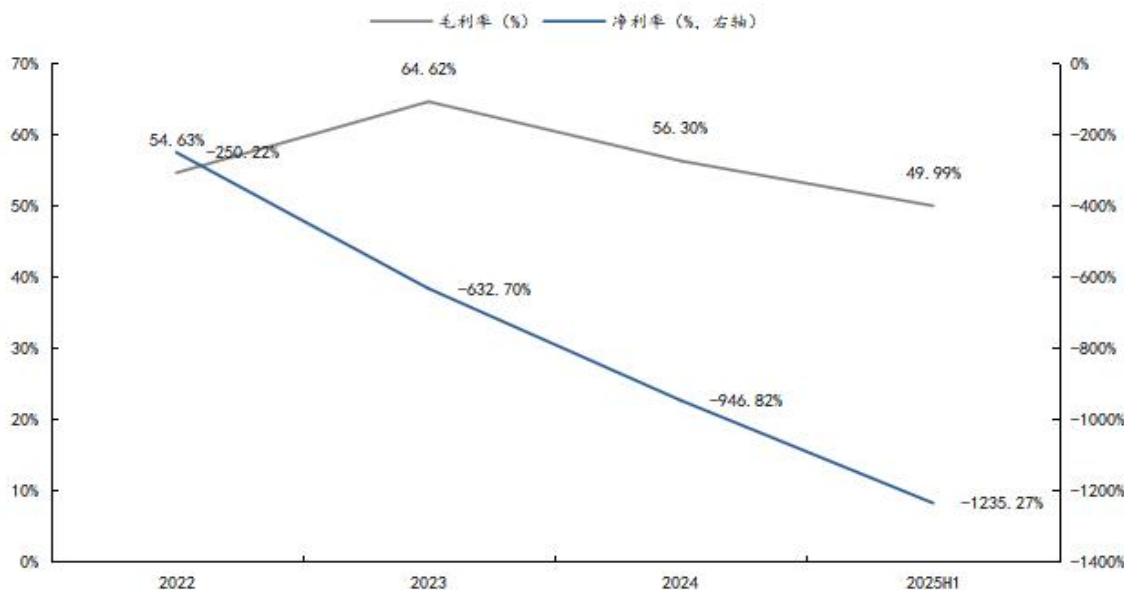


资料来源：公司招股意向书，国信证券经济研究所整理

公司收入高速增长，净利润短期承压

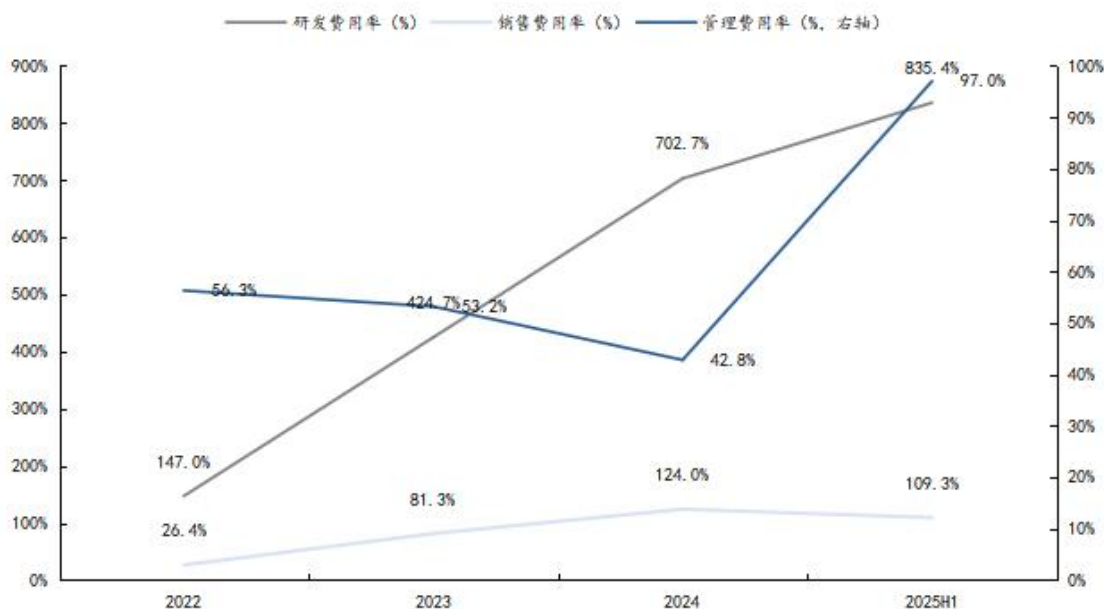
- 公司毛利率同比增长，业务拓展推动期间费用率增长。2025H1公司实现毛利率50%，同比提升1.1个pct，其中本地化部署的毛利率达59.1%，同比持平；基于云端部署的毛损率为0.4%，去年同期为28.6%，主要由于基于云端部署价格下降的市场趋势，以及公司以有竞争力的价格快速获得市场份额的策略导致。2025H1，公司实现研发/销售/管理费用率分别为835.4%/109.3%/97%，研发费用提升主要由于公司扩大研发团队，并对基座模型迭代及推进技术基础设施投入导致；销售费用提升主要由于战略性地作出更多广告投资以把握新兴市场机遇，广告及营销开支大幅增加导致；管理费用受成本控制措施以及提升运营效率推动逐步下降。

图：2025H1实现销售毛利率49.99%



资料来源：公司招股意向书，国信证券经济研究所整理

图：公司期间费用率情况



资料来源：公司招股意向书，国信证券经济研究所整理

客户快速涵盖广泛，下游集中度高



- 公司拥有广泛且多元化的客户群体，并实现用户的快速扩张。公司的客户包括企业、公共部门实体及个人用户。2022、2023、2024、2025H1，来自五大客户的收入分别占公司总收入的55.4%、61.5%、45.5%及40.0%，来自最大客户的收入分别占总收入的15.4%、14.7%、19.0%及11.0%。

表：2025H1智谱前五大客户

客户	主要业务	业务关系 年期	收入（人民币千 元）	占总收入的百分 比
A	艺术相关的学习服务、直播电商、 文化旅游研学、智慧教育服务及 人工智能教育	2	20,977	11.00%
B	电信服务	2	18,980	9.90%
C	电信基础设施、多媒体通信以及 信息及通信技术服务	1	17,927	9.40%
D	信息系统集成服务、技术服务、 计算机软件及辅助设备零售	1	9,623	5.00%
E	信息系统集成服务及软件开发	1	9,026	4.70%
合计			76,533	40.00%

资料来源：公司招股意向书，国信证券经济研究所整理

- 【 01 】 领先的AI公司，产品进入快速迭代阶段
- 【 02 】 模型能力持续提升，全球AI市场快速扩张
- 【 03 】 募集资金投入产品，维持行业优势
- 【 04 】 风险提示

- 2025年12月，公司正式发布GLM-4.7并开源，新版本面向Coding场景强化了编码能力、长程任务规划与工具协同，并在多项主流公开基准测试中取得开源模型中的领先表现，GLM-4.7在编程、推理与智能体三个维度实现突破：
 - 1) 更强的编程能力：显著提升了模型在多语言编码和在终端智能体中的效果；GLM-4.7可以在Claude Code、TRAE、Kilo Code、Cline和Roo Code等编程框架中实现先思考、再行动的机制，在复杂任务上有更稳定的表现；
 - 2) 前端审美提升：GLM-4.7在前端生成质量方面明显进步，能够生成观感更佳的网页、PPT、海报；
 - 3) 更强的工具调用能力：GLM-4.7提升了工具调用能力，在BrowseComp网页任务评测中获得67.5分；在 τ^2 -Bench交互式工具调用评测中实现87.4分的开源SOTA，超过Claude Sonnet 4.5；
 - 4) 推理能力提升：显著提升了数学和推理能力，在HLE（“人类最后的考试”）基准测试中获得42.8%的成绩，较GLM-4.6提升41%，超过GPT-5.1；
 - 5) 通用能力增强：GLM-4.7对话更简洁智能且富有人情味，写作与角色扮演更具文采与沉浸感。

客户快速涵盖广泛，下游集中度高

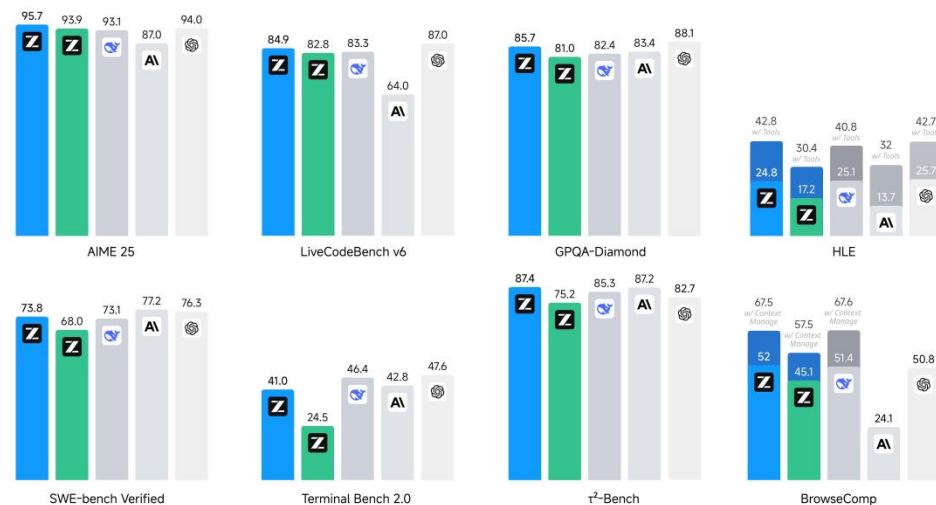
- 在主流基准测试表现中，GLM-4.7的代码能力对齐Claude Sonnet 4.5：在SWE-bench-Verified获得73.8%的开源SOTA分数；在LiveCodeBench V6达到84.9%的开源SOTA分数，超过Claude Sonnet 4.5；SWE-bench Multilingual达到66.7%（提升12.9%）；Terminal Bench 2.0达到41%（提升16.5%）。
- 据SimilarWeb数据，公司日活在2025年快速增长，当前日活稳定在20万人左右。2025年7月28日，公司发布GLM-4.5，在包含推理、代码、智能体的综合能力达到开源SOTA，推动日活大幅提升。随着公司产品加速迭代，公司2025年11月30日日活数量达到19.97万人。

图：公司GLM-4.7性能全面提升

LLM Performance Evaluation: Agentic, Reasoning and Coding

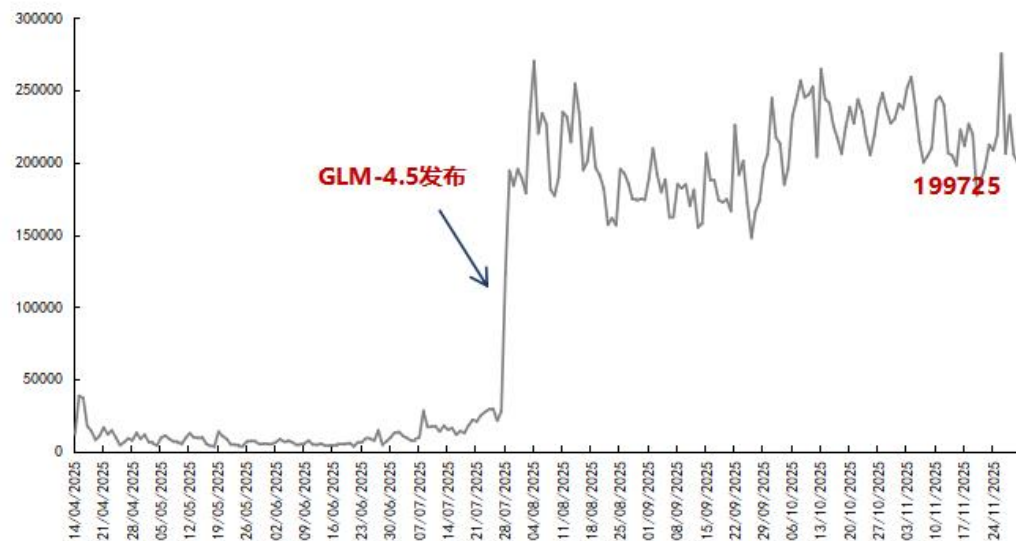
8 benchmarks: AIME 25, LiveCodeBench v6, GPQA, HLE, SWE-bench Verified, Terminal-Bench, τ^2 -Bench, BrowseComp
(Evaluation results under 128K context length)

GLM-4.7 GLM-4.6 DeepSeek-V3.2 Claude Sonnet 4.5 GPT-5.1 (High)



资料来源：公司官网，国信证券经济研究所整理

图：谷歌ROIC数据复盘

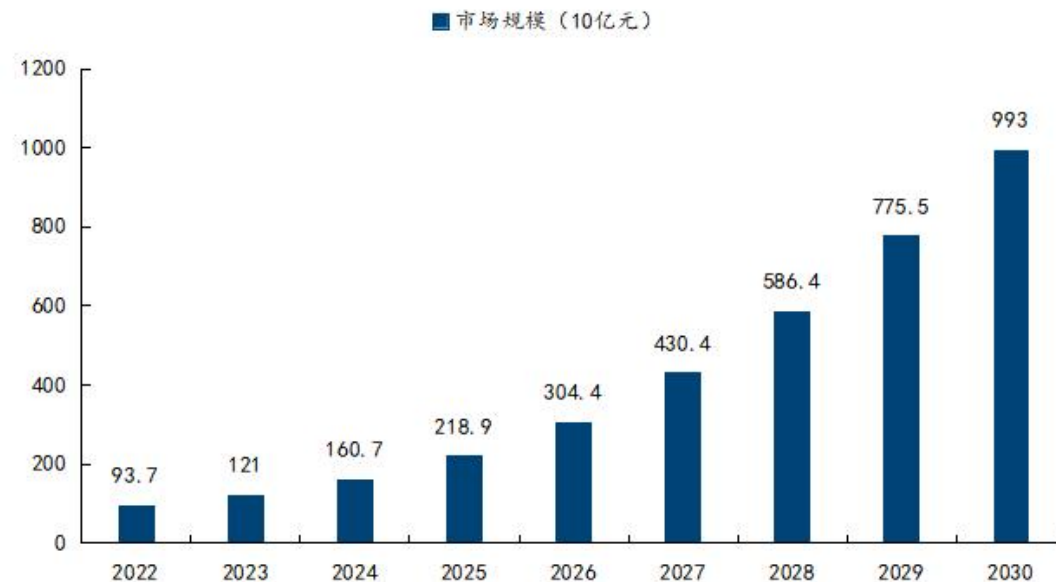


资料来源：公司模型日活快速提升，国信证券经济研究所整理

下游应用场景广，兼具市场规模与成长空间

- 人工智能是计算机科学的重要分支，指通过模拟和延伸人类的智能，使机器具备感知、理解、推理、学习和自主决策等能力的技术。人工智能的发展正在深刻地改变着经济增长、企业运行及人类生活。据沙利文数据，预计至2030年人工智能将赋能全球至少20%的日常商业决策，为全球80%的消费者主流智能设备提供支持，创造逾20万亿美元的经济影响。中国人工智能市场规模由2022年的937亿元增至2024年的1607亿元，期间复合年增长率为31.0%；预计于2030年，中国人工智能市场规模将进一步增至9930亿元，2024年至2030年的复合年增长率为35.5%。

图：中国人工智能市场规模

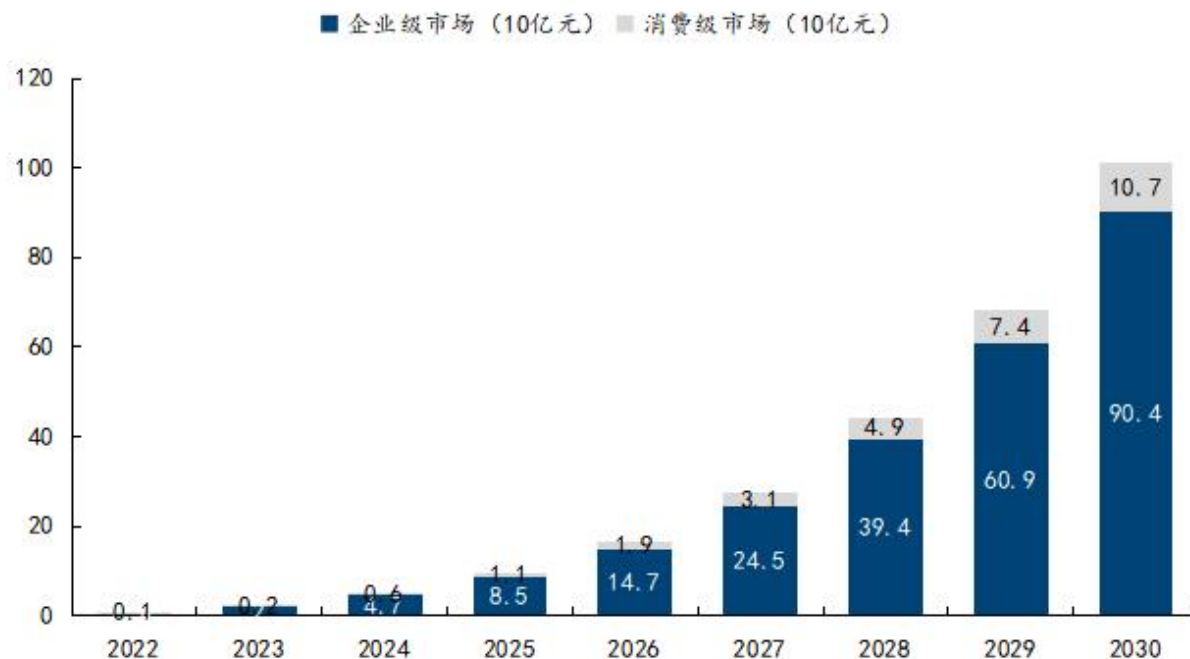


资料来源：沙利文，国信证券经济研究所整理

下游应用场景广，兼具市场规模与成长空间

- 在商业化进程方面，大语言模型在中国的客户市场仍处于早期阶段。尽管面向消费者的应用逐渐涌现，但用户对大语言模型的付费意愿仍处于较低水平。相比之下，企业级场景是中国大语言模型市场增长的主要驱动因素。企业用户对于大语言模型的部署有着明确的实际需求，在其业务运营中采用大语言模型也展现出更高的支付能力和落地效率。大语言模型正日益融入企业业务流程，以提升跨职能部门的效率及生产力。2024年中国大语言模型市场规模已达到53亿元，其中机构客户贡献47亿元，个人客户贡献6亿元，预计到2030年该市场规模将增至1011亿元，2024年至2030年的复合年增长率为63.5%。机构客户仍将是市场增长的核心驱动力，预计到2030年中国企业级大语言模型市场规模将达到904亿元，2024年至2030年的复合年增长率为63.7%。

图：中国大语言模型市场规模



资料来源：沙利文，国信证券经济研究所整理

- 【 01 】 领先的AI公司，产品进入快速迭代阶段
- 【 02 】 模型能力持续提升，全球AI市场快速扩张
- 【 03 】 募集资金投入新产品，维持行业优势
- 【 04 】 风险提示

募集资金投入产品，维持行业优势



- 假设超额配股权未获行使，经扣除全球发售应付的包销佣金及其他估计发售开支后，按发售价每股股份116.20港元，公司将从全球发售中获得约41.73亿港元的净募集资金。公司拟将全球发售所得款项按下述金额用于下述用途：
- 1. 约70.0%将用于持续增强在通用AI大模型方面的研发能力，其中约34.5%、26.0%及9.5%预计将分别于2026、2027及2028年使用：
 - (1) 约60.0%将分配用于训练基础设施：约20.0%将用于预训练大模型的持续研发；约20.0%将用于深度推理模型的持续研发；约10.0%将用于AI Agent的持续研发；约10.0%将用于探索前沿新兴技术。
 - (2) 约10.0%将分配用于改进GLM框架，优化数据处理平台并建设高质量语料库：约5.0%将用于增强GLM框架；约5.0%将用于优化数据处理平台及建立高质量语料库。
- 2. 约10.0%将用于通过提供最新的基座模型以及训练/推理工具及基础设施持续优化MaaS平台，其中约6.0%、3.0%及1.0%预计将分别于2026年、2027年及2028年使用：
 - (1) 约5.0%将分配用于持续部署不同参数规模的大模型，该等模型旨在兼容广泛的计算资源和终端设备：约2.5%将通过优化大模型并确保其与广泛计算资源及终端设备兼容，以加强平台组合的部署，从而促使模型更好地匹配终端设备计算能力并减少硬件限制；约2.5%将分配用于在消费级硬件上运行的大模型产品的研发，包括应用程序及消费电子等硬件产品。
 - (2) 约5.0%将分配用于加强销售及营销团队并开展更多营销活动，以进一步提升品牌知名度、提高市场渗透率，并与目标市场客户维系稳固的关系以为客户提供服务。
- 3. 约10.0%将用于发展业务合作伙伴网络以及进行战略投资，其中约4.0%、4.0%及2.0%预计将分别于2026年、2027年及2028年使用。
- 4. 约10.0%将用于营运资金及其他一般企业用途。

表：智谱募投项目（单位：万港元）

序号	项目名称	总投资金额	拟投入募集资金
1	持续增强通用AI大模型研发能力（70.0%）	292,140	292,140
2	通过提供最新基座模型及训练/推理工具与基础设施，持续优化MaaS平台（10.0%）	41,730	41,730
3	发展业务合作伙伴网络及进行战略投资（10.0%）	41,730	41,730
4	营运资金及其他一般企业用途（10.0%）	41,730	41,730
合计		417,340	417,340

资料来源：公司招股意向书，国信证券经济研究所整理

- 【 01 】 领先的AI公司，产品进入快速迭代阶段
- 【 02 】 模型能力持续提升，全球AI市场快速扩张
- 【 03 】 募集资金投入产品，维持行业优势
- 【 04 】 风险提示

- 风险提示1：技术上，产品研发未达预期风险、关键技术人员流失风险、核心技术泄密风险；
- 风险提示2：财务上，持续亏损与盈利拐点不确定、经营现金流为负与外部融资依赖、算力与技术服务成本上行、毛利率波动的风险；
- 风险提示3：经营上，国际贸易摩擦加剧风险、规模扩张后的管理风险、行业政策变更的风险。

免责声明

国信证券投资评级			
投资评级标准	类别	级别	说明
报告中投资建议所涉及的评级（如有）分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后6到12个月内的相对市场表现，也即报告发布日后的6到12个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。A股市场以沪深300指数（000300.SH）作为基准；新三板市场以三板成指（899001.CSI）为基准；香港市场以恒生指数（HSI.HI）作为基准；美国市场以标普500指数（SPX.GI）或纳斯达克指数（IXIC.GI）为基准。	股票投资评级	优于大市	股价表现优于市场代表性指数10%以上
		中性	股价表现介于市场代表性指数±10%之间
		弱于大市	股价表现弱于市场代表性指数10%以上
		无评级	股价与市场代表性指数相比无明确观点
	行业投资评级	优于大市	行业指数表现优于市场代表性指数10%以上
		中性	行业指数表现介于市场代表性指数±10%之间
		弱于大市	行业指数表现弱于市场代表性指数10%以上

分析师承诺

作者保证报告所采用的数据均来自合规渠道；分析逻辑基于作者的职业理解，通过合理判断并得出结论，力求独立、客观、公正，结论不受任何第三方的授意或影响；作者在过去、现在或未来未就其研究报告所提供的具体建议或所表述的意见直接或间接收取任何报酬，特此声明。

重要声明

本报告由国信证券股份有限公司（已具备中国证监会许可的证券投资咨询业务资格）制作；报告版权归国信证券股份有限公司（以下简称“我公司”）所有。本报告仅供我公司客户使用，本公司不会因接收人收到本报告而视其为客户。未经书面许可，任何机构和个人不得以任何形式使用、复制或传播。任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以我公司向客户发布的本报告完整版本为准。

本报告基于已公开的资料或信息撰写，但我公司不保证该资料及信息的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映我公司于本报告公开发布当日的判断，在不同时期，我公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。我公司不保证本报告所含信息及资料处于最新状态；我公司可能随时补充、更新和修订有关信息及资料，投资者应当自行关注相关更新和修订内容。我公司或关联机构可能会持有本报告中所提到的公司所发行的证券并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或金融产品等相关服务。本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告意见或建议不一致的投资决策。

本报告仅供参考之用，不构成出售或购买证券或其他投资标的的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，我公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

证券投资咨询业务的说明

本公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询，是指从事证券投资咨询业务的机构及其投资咨询人员以下列形式为证券投资人或者客户提供证券投资分析、预测或者建议等直接或者间接有偿咨询服务的活动：接受投资人或者客户委托，提供证券投资咨询服务；举办有关证券投资咨询的讲座、报告会、分析会等；在报刊上发表证券投资咨询的文章、评论、报告，以及通过电台、电视台等公众传播媒体提供证券投资咨询服务；通过电话、传真、电脑网络等电信设备系统，提供证券投资咨询服务；中国证监会认定的其他形式。

发布证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。



国信证券
GUOSEN SECURITIES

国信证券经济研究所

深圳

深圳市福田区福华一路125号国信金融大厦36层

邮编：518046 总机：0755-82130833

上海

上海浦东民生路1199弄证大五道口广场1号楼12楼

邮编：200135

北京

北京西城区金融大街兴盛街6号国信证券9层

邮编：100032