智谱（02513.HK）

优于大市

领先的独立大语言模型公司，产品矩阵快速扩张

核心观点

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。

主营业务：公司主要围绕其MaaS平台为下游客户提供产品，其MaaS平台具有以下特点：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。

公司模型矩阵涵盖端侧小模型、经济型模型和千亿参数的旗舰大模型等各类参数规模，能够有针对性地满足特定客户需求。公司产品提供全面的功能支持，覆盖对话、通用智能体、代码生成、图像理解、文生图/视频、语音交互等多种能力，实现对各类模型应用场景的充分覆盖。

风险提示：盈利预测的风险、AI落地不及预期的风险、技术被赶超或替代的风险、宏观经济及行业波动风险

投资建议：领先的独立大语言模型公司，产品矩阵快速扩张，首次覆盖，给予“优于大市”评级。公司构建了以GLM语言模型为核心，覆盖智能体模型（如AutoGLM）、多模态模型（如CogView，CogVideoX）、代码模型（CodeGeeX）的全面产品矩阵，并为约8000家企业客户，8000万个人用户提供服务。公司GLM5大模型在Coding与Agent能力上取得开源SOTA表现，在真实编程场景的使用体感逼近Claude Opus 4.5，并基于开源合作培育了完备的成长生态。公司底层大模型能力领先，产品矩阵快速扩张，大模型商业化处于国内领先地位，未来业绩增长空间有望进一步打开，首次覆盖，给予“优于大市”评级。

盈利预测和财务指标	2024	2025	2026E	2027E	2028E
营业收入（百万元）	312	724	2,012	3,974	6,589
（+/-%）	150.9%	131.9%	177.8%	97.5%	65.8%
净利润（百万元）	-2956	-4698	-4231	-4425	-4164
（+/-%）	--	--	--	--	--
每股收益（元）	-6.63	-10.54	-9.49	-9.93	-9.34
EBIT Margin	-947.5%	-650.8%	-216.4%	-112.4%	-62.3%
净资产收益率（ROE）	74.7%	58.1%	34.3%	26.4%	19.9%
市盈率（PE）	-129.2	-81.3	-90.3	-86.3	-91.8
EV/EBITDA	-132.7	-81.2	-85.6	-85.0	-94.5
市净率（PB）	-96.56	-47.21	-31.00	-22.81	-18.27

资料来源：Wind、国信证券经济研究所预测　　注：摊薄每股收益按最新总股本计算

公司研究·海外公司财报点评

互联网·互联网II

证券分析师：熊莉
021-61761067
xiongli1@guosen.com.cn
S0980519030002

证券分析师：张伦可
0755-81982651
zhanglunke@guosen.com.cn
S0980521120004

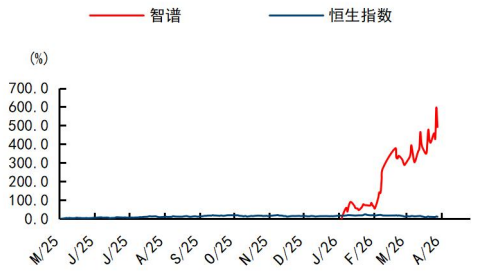
证券分析师：张昊晨
zhanghaochen1@guosen.com.cn
S0980525010001

联系人：侯睿
hourui3@guosen.com.cn

基础数据

投资评级	优于大市（首次）
合理估值	
收盘价	779.00港元
总市值/流通市值	347312/172404百万港元
52周最高价/最低价	938.00/116.10港元
近3个月日均成交额	1495.65百万港元

市场走势



资料来源：Wind、国信证券经济研究所整理

相关研究报告

◆ 公司概况：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：

1）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；

2）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、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 Rumination，公司最新基础模型已迭代至 GLM-5。

图1：智谱发展历程

资料来源：公司官网，招股说明书，国信证券经济研究所整理

◆ 公司治理：控股股东集团占股近 1/3，多家机构参投

刘德兵博士为实际控制人，控股股东集团合计持股达 30.22%。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。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等。

表1：公司股权架构（截至 2026 年 1 月）

股东名称	占已发行普通股比例(%)
北京链湃	7.73
慧惠	8.97
智登	6.18
刘德兵	0.21
唐杰	6.10
李涓子	0.76
许斌	0.18
张鹏	0.09
君联资本	6.16
美团	3.91
启明创投	2.28
其他股东	48.74
其他公众股东	8.50

资料来源：招股说明书，国信证券经济研究所整理

◆ **主营业务：围绕 MaaS 平台为客户提供产品，具备完善模型矩阵**

公司主要围绕其 MaaS 平台为下游客户提供产品，其 MaaS 平台具有以下特点：

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2：公司 MaaS 平台提供全方位 AI 能力

资料来源：招股说明书，国信证券经济研究所整理

公司模型矩阵涵盖端侧小模型、经济型模型和千亿参数的旗舰大模型等各类参数规模，能够有针对性地满足特定客户需求。公司产品提供全面的功能支持，覆盖对话、通用智能体、代码生成、图像理解、文生图/视频、语音交互等多种能力，实现对各类模型应用场景的充分覆盖。

1）语言模型：语言模型以 GLM 系列为核心，包括 GLM-4.7、GLM-5 等，并可根据部署形态提供经济型与端侧小模型选项，具备复杂语境理解与推理能力，可高质量完成知识问答、对话生成、信息抽取、总结改写、内容创作等任务；

2）智能体模型：以推理规划+感知操作为核心架构构建通用智能体能力，GLM-Z1-Rumination 进行深度推理与自主规划，AutoGLM 承担对用户界面与设备进行感知和操作，使智能体能够理解目标、拆解任务、进行多步规划并执行跨应用的复杂操作，覆盖自动化办公、流程编排、客户服务、智能终端控制等场景；

3）多模态模型：具备 CogView 和 CogVideoX 等模型，CogView 聚焦文生图能力，可将文本需求转化为高质量图像内容。CogVideoX 专注视频生成能力，支持从文本或图像生成动态视频内容，可用于短视频创作、教育演示、产品展示等场景；

4）代码模型：以 CodeGeeX 为代表，面向开发者与企业研发场景提供代码生成、补全、解释、重构、测试用例生成与多语言转换等能力，旨在降低研发门槛并提升工程效率。

图3：智谱具备完备的大模型矩阵

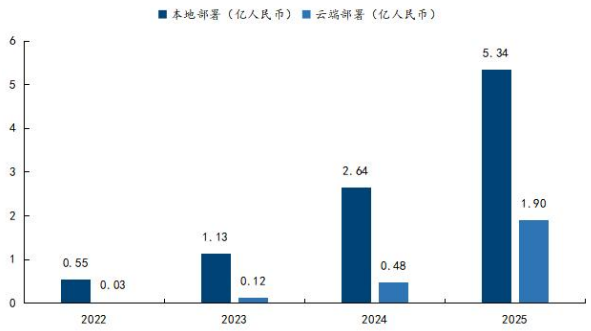
资料来源：招股说明书，国信证券经济研究所整理

公司产品主要按照本地化部署与云端部署两种方式落地，各业务收费模式如下：

1）本地部署中模型托管在客户自身的基础设施内，使组织能够利用其专有或敏感数据，定制私域专属 AI 模型。本地化部署在性能优化和基础设施配置方面提供更大的控制权，适用于复杂或高度专业化的应用场景。公司在将大模型及相关服务交付至客户指定地点并经客户检验验收时确认收入。2025 年，公司本地化部署的收入达 5.34 亿元，同比提升 102.31%，占公司收入比重达 73.72%，主要得益于公司通过持续迭代提升模型智能上限、模型通用性增强和市场需求强劲；

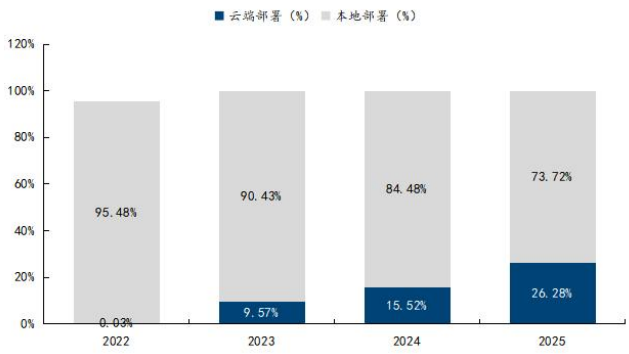
2）云端部署将模型托管在可扩展且可靠的云端基础设施上，适合追求敏捷性和易实施性的企业。客户无需投入昂贵的本地基础设施，即可快速且经济高效地部署 AI 解决方案。公司对于以订阅为基础的合同通常在合同期内按比例确认收入，对于以使用量为基础的合同，在向客户提供服务时根据客户对资源的使用情况确认收入。2025 年，云端部署的收入达 1.9 亿元，同比增长 292.66%，占收入比重达 26.28%，主要得益于持续迭代显著提升了模型智能上限，提升后的模型智能表现进一步推动了模型调用量的增加。

图4：公司收入拆分情况



资料来源：招股说明书，国信证券经济研究所整理

图5：公司分业务收入占比

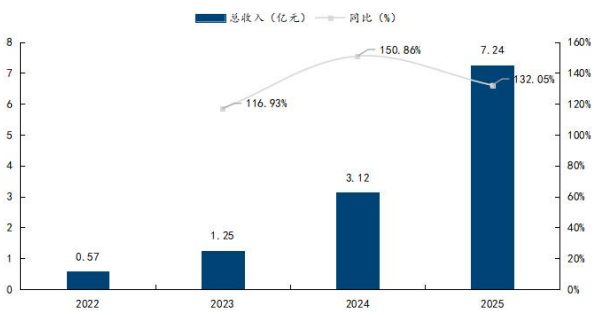


资料来源：招股说明书，国信证券经济研究所整理

◆ 财务分析：收入高速增长，净利润短期承压

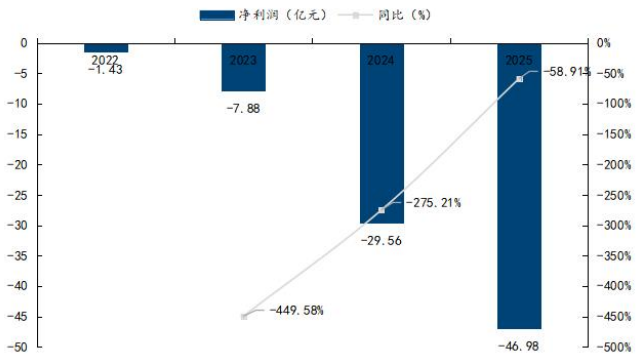
公司收入快速增长，扣除研发投入后大幅扭亏。2025 年公司实现收入 7.24 亿元，同比增长 131.85%，主要由云端部署收入大幅提升推动。公司于 2026 年 3 月推出 Claw Plan 上线仅两天订阅用户即突破 10 万；上线 20 天突破 40 万，验证了智能体长链路任务蕴含巨大的商业空间。公司实现净亏损 46.98 亿元，主要受研发投入持续增加影响，公司实现经调整净亏约 31.8 亿元，扣除 31.8 亿元研发支出后基本实现全年盈亏平衡。

图6：公司收入情况



资料来源：公司财报，国信证券经济研究所整理

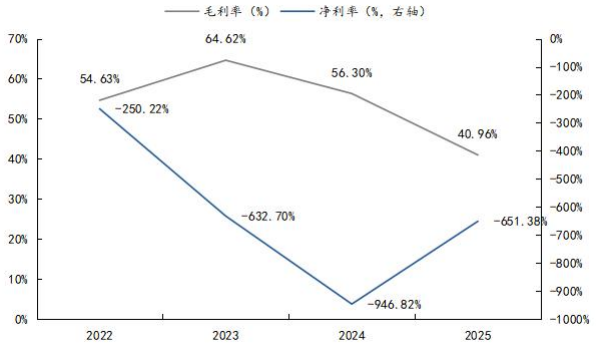
图7：公司净利润情况



资料来源：公司财报，国信证券经济研究所整理

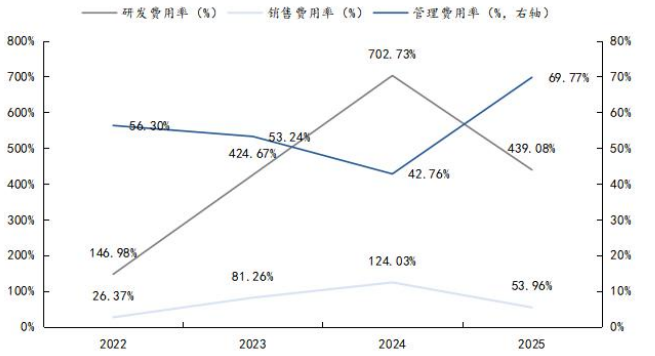
云端部署毛利率增长，整体利润率受业务拓展影响。2025 年公司实现销售毛利率 40.96%，其中本地化部署实现毛利率 48.8%，同比下降 17.2 个 pct，主要由于为满足客户需求而投入了更多的交付资源；云端部署实现毛利率 18.9%，同比提升 15.6 个 pct，主要由于模型推理效率提升、算力规模扩张导致边际成本递减，同时价格有所提高。2025 年，公司实现研发/销售/管理费用率分别为 439.08%/53.96%/69.77%，分别同比-263.65/-70.07/+27.01 个 pct。

图8：公司毛利率、净利率情况



资料来源：公司财报，国信证券经济研究所整理

图9：公司期间费用率情况

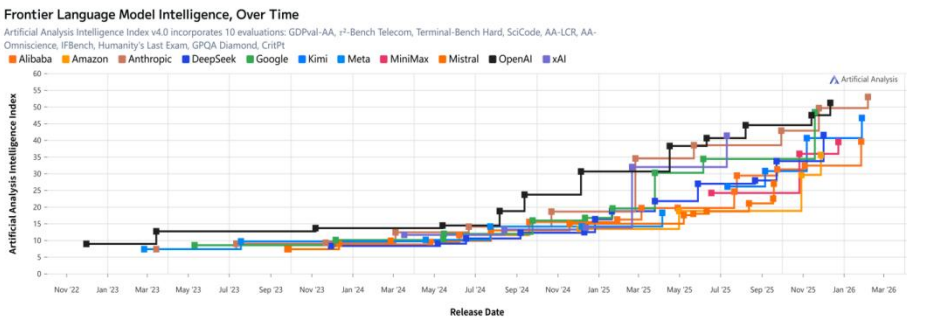


资料来源：公司财报，国信证券经济研究所整理

◆ 行业简析：下游应用场景广，兼具市场规模与成长空间

大模型能力边界不断扩展，智能水平跨越式提升。随着算力基础设施建设不断推进、算法架构的不断革新（如 MOE 架构）以及模型参数的大幅增长（2023 年 GPT3.5 仅有 1750 亿参数，2024 年 Google Gemini Pro 则有 1.8 万亿参数），大模型的智能水平已实现跨越式提升。以从 GPT-3 到 GPT-4 为例，其律师资格考试成绩由倒数 10%跃升至前 10%，而 GPT-4o、Claude 3.7 等后续迭代更新也持续刷新了智能上限。与此同时，模型能力的边界也在不断外延：一方面，长上下文技术（如 10 万+ token）打破了信息处理的记忆瓶颈；另一方面，Agentic AI 的兴起标志着模型从“被动问答”向“主动执行”的转变——通过自主规划并调用代码解释器、浏览器等外部工具，大模型已具备在复杂环境中解决实际问题的逻辑推理与执行能力。

图10：各前沿模型评测得分不断提升



资料来源：Artificial Analysis，国信证券经济研究所整理

AI 技术正经历从传统判别式 AI 向大语言模型的重要转变，大语言模型打开应用

空间。传统判别式 AI 主要侧重于识别和判断类任务，例如分类、回归和目标检测，其本质是通过学习输入与标签输出之间的映射关系，帮助机器做出准确判断。大语言模型则是规模庞大的深度学习模型，在海量数据上进行预训练，并基于拥有数十亿到数千亿参数的神经网络构建，使其能够理解并生成自然语言以及其他类型的内容，从而执行广泛的任务。因此，大语言模型能够处理许多传统判别式 AI 方法难以应对的任务。

表2：大语言模型与判别式 AI 主要区别

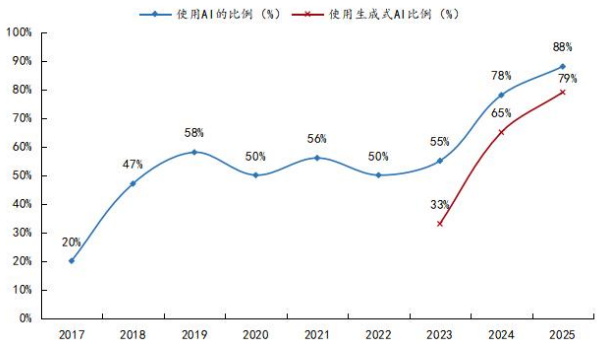
维度	大语言模型	判别式 AI
主要功能	能够执行包括文本生成、图像生成、语音生成及代码生成在内的多种生成任务，同时也能够处理内容总结、翻译及风格转换。其核心在于根据对上下文的理解输出符合语境逻辑的新内容。大语言模型凭借其自主性、感知能力、决策能力及执行能力，可根据外部环境的变化和设定目标自主行动。	侧重于将输入数据映射为人预设的类别，或直接输出连续数值进行预测，强调模型对已知领域的识别与判断能力。主要用于图像识别、语音交互及智能推荐等任务。
应用案例	文本、图像、音频、视频、代码等生成工具以及 AI Agent。	人脸识别、语音识别及数据内容推断。
算力	需要高性能的 GPU 或专用 AI 芯片，并需要硬件架构以支持大规模训练和推理。尤其在训练阶段，需要大规模分布式计算集群。	通常参数量较小，可使用通用 CPU 或 GPU 进行训练和推理，通常不需要大规模分布式计算集群。
算法	主要基于拥有数十亿或更多参数的复杂深度神经网络，以及强化学习技术。	通常使用逻辑回归、支持向量机、决策树、随机森林及较小的神经网络等技术。
数据	主要依赖大规模开放结构化数据（尤其在预训练阶段），对标注数据依赖较低（主要在微调或对齐阶段需要）。	主要依赖结构化、高质量的标注数据。

资料来源：招股说明书，国信证券经济研究所整理

AI 正处于从狭义人工智能迈向通用人工智能（AGI）的过渡阶段，大语言模型作为转变的核心，正日益成为驱动人工智能发展新时代的关键要素。凭借参数规模、语义理解、多模态融合与自我进化能力的持续跃升，大语言模型已打破传统判别式 AI 应用场景割裂的局限，展现出向通用智能逼近的技术潜力。大语言模型的商业化涉及面向企业客户提供模型能力、工具链支持以及训练或调优服务等，同时也包括面向个人用户提供 AI 生成内容应用服务。通过云端部署与本地化部署等方式，大语言模型厂商帮助企业客户构建在文本生成、语义理解、逻辑推理及多模态交互等方面的智能能力。

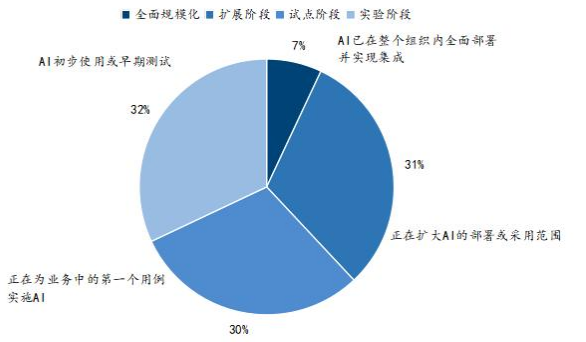
随着模型能力的提升以及智能体框架的完善，当前企业正积极尝试 AI 应用，据麦肯锡数据，2025 年企业使用 AI 的比例正快速上升，全球约 88%的企业正积极尝试 AI，其中 79%的企业开始使用生成式 AI。从企业整体层面来看，大多数公司仍处于试验或试点阶段，仅约 1/3 的公司已开始扩大 AI 项目的规模化应用。

图11：当前采用 AI 的公司占比提升



资料来源：McKinsey & Company-《The state of AI in 2025》-2025年-P3，国信证券经济研究所整理

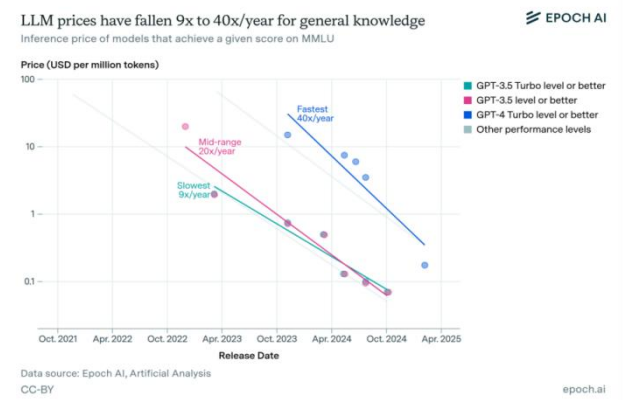
图12：AI 在企业端应用仍处于初级阶段



资料来源：McKinsey & Company-《The state of AI in 2025》-2025年-P3，国信证券经济研究所整理

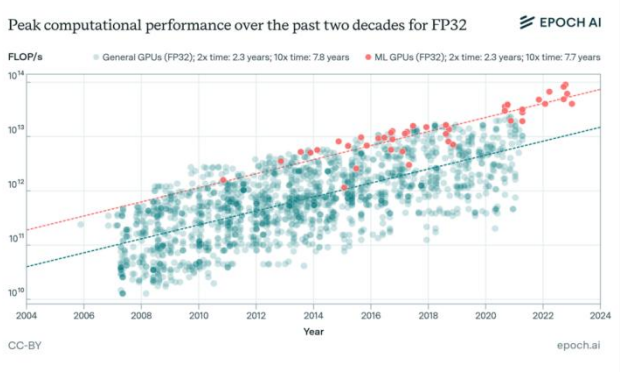
供给侧技术进步推动算力成本结构性下降，市场空间逐渐打开，随着算力基础设施建设不断推进，底层算力硬件性能不断提升，算法结构不断优化，算力成本的降低将向应用层传导，从而降低企业和开发者使用成本，减轻部署大模型的负担，进而提升下游企业和个人用户的付费意愿。高频次、长上下文的复杂应用场景也将随着成本降低具备经济可行性，以 GPT-4 level 为例，2025 年每百万 token 成本仅为 2024 年的 1/40，而 ML GPU 的计算性价比平均每 2.1 年就增长 100%。性能提升和成本下降并行发生，最终导致计算性价比呈指数级上涨，从而推动行业从技术验证期迈入爆发增长期。

图13：general knowledge 层面 LLM 价格下降 9x-40x



资料来源：EPOCH AI，国信证券经济研究所整理

图14：ML GPU 每 2.1 年翻一番，通用 GPU 每 2.5 年翻一番

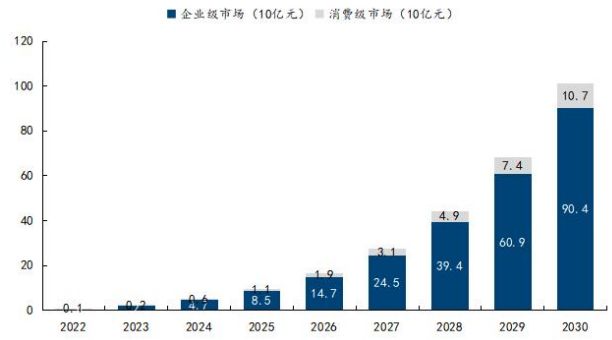


资料来源：EPOCH AI，国信证券经济研究所整理

在商业化进程方面，中国大语言模型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。企业用户对大语言模型的部署具备明确的需求，在业务运营中采用大语言模型也表现出更高的支付能力与落地效率。因此，当前中国大语言模型的商业化重心集中于机构客户的采用，融入企业业务工作流程，以提升跨职能部门的效率与生产力。据沙利文数据，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%。

随着大语言模型在各个应用场景的逐步渗透，中国人工智能市场规模亦有望快速提升。据沙利文数据，预计至 2030 年 AI 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将进一步增至 9930 亿元。

图15：中国大语言模型市场规模



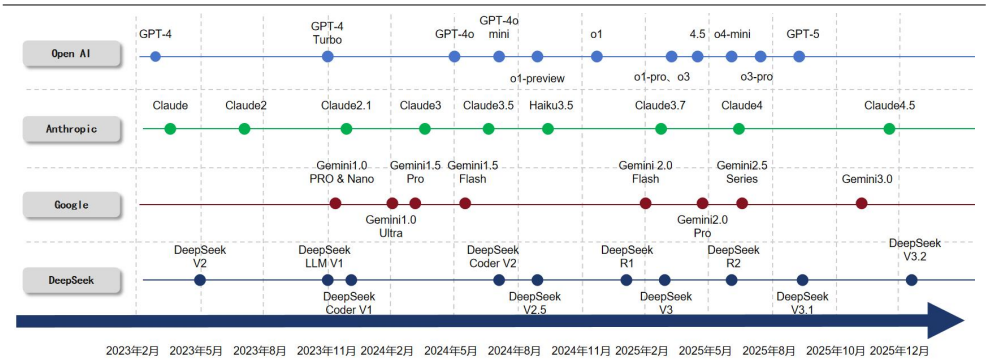
资料来源：招股说明书，国信证券经济研究所整理

图16：中国人工智能市场规模

资料来源：招股说明书，国信证券经济研究所整理

“百模大战”走向收敛，模型能力仍然是最关键决定因素。全球 AI 竞赛已经逐渐走过了被称为“百模大战”的野蛮生长期，迈向模型能力和商业化落地的全面竞争，但模型能力本身仍然是最关键的决定因素。目前大模型主要赛道涵盖生产力、娱乐、视觉生成、音频生成和通用 2B 服务等领域，每个细分领域侧重点略有不同，但能够在这些领域中长期保持自然增长的产品，都是由底层模型智能持续提升驱动的，一个模型智能水平的快速提升，将带来更好的用户体验和用户的自然迁移。**因此各 AI 公司仍然将模型能力视为最关键的变量，模型能力仍然是 AI 大战最关键的决定因素。2025 年以来，全球大模型竞争从以年为单位的代际演进转向以季度甚至月度为周期的竞速时代，模型能力不断提升，迭代不断加速。**

图17：各主要公司模型迭代时间表

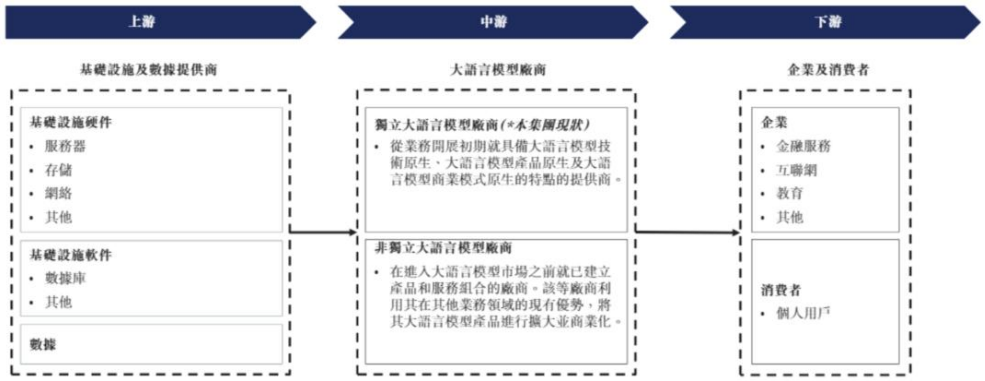


资料来源：招股说明书，国信证券经济研究所整理

中国大语言模型市场竞争激烈，独立大语言模型提供商具备自身优势。中国大语言模型市场的参与者可分为独立提供商和非独立提供商，独立提供商从业务开展初期便具备大语言模型技术原生、大语言模型产品原生以及大语言模型商业模式原生等特点，而非独立提供商通常为涉足 AI 领域的科技巨头。与非独立提供商相比，独立提供商面临着截然不同的竞争动态。非独立提供商依托其既有的多元化业务线，积累了庞大的用户群体，这有利于其大语言模型产品的推广。然而，若科技巨头所经营的业务线与客户自身业务存在直接竞争关系，企业客户可能不愿选择其提供的大语言模型产品。此外，部分行业的企业客户对进入或可能进入某些科技巨头的影响范围持高度谨慎态度，更倾向于采用独立大语言模型提供商的

AI 解决方案。

图18：大语言模型行业价值链示意



资料来源：招股说明书，国信证券经济研究所整理

◆ 公司优势：智能水平全球顶尖，领跑开源生态

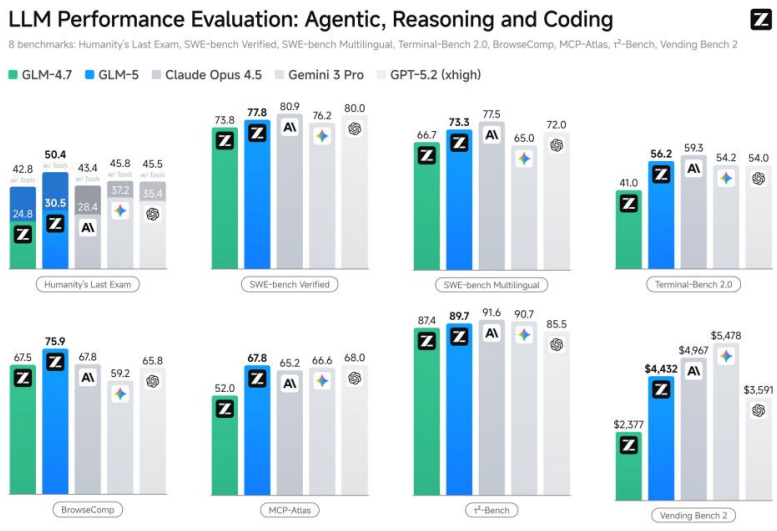
2026 年 2 月，公司发布全新旗舰模型 GLM-5，在 Coding 与 Agent 能力上取得开源 SOTA 表现，在真实编程场景的使用体感逼近 Claude Opus 4.5，擅长复杂系统工程与长程 Agent 任务：

1）参数规模扩展：从 355B（激活 32B）扩展至 744B（激活 40B），预训练数据从 23T 提升至 28.5T，显著提升了模型的通用智能水平；

2）异步强化学习：构建全新的“Slime”框架，支持更大模型规模及更复杂的强化学习任务，提升强化学习后训练流程效率；提出异步智能体强化学习算法，使模型能够持续从长程交互中学习，充分激发预训练模型的潜力；

3）稀疏注意力机制：首次集成 DeepSeek Sparse Attention，在维持长文本效果无损的同时，大幅降低模型部署成本，提升 Token Efficiency。

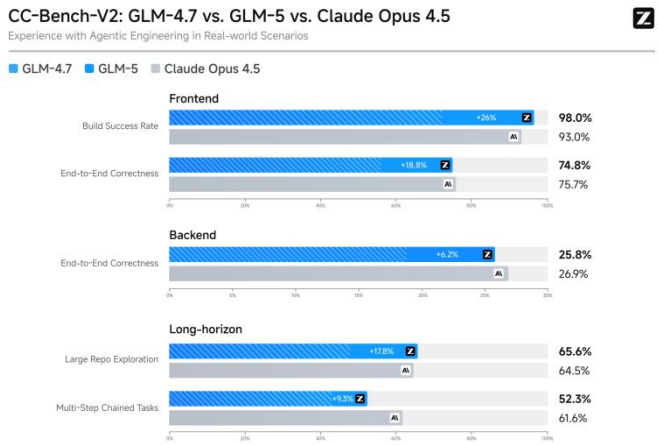
图19：GLM-5 测评结果较 GLM-4.7 大幅提升



资料来源：公司官网，国信证券经济研究所整理

GLM-5 在编程能力上实现了对 Claude Opus 4.5 的对齐，在业内公认的主流基准测试中取得开源模型 SOTA 分数。在 SWE-bench-Verified 和 Terminal Bench 2.0 中分别获得 77.8 和 56.2 的开源模型 SOTA 分数，性能超过 Gemini 3 Pro。在内部 Claude Code 评估集合中，GLM-5 在前端、后端、长程任务等编程开发任务上显著超越 GLM-4.7（平均增幅超 20%），能够以极少的人工干预自主完成 Agentic 长程规划与执行、后端重构和深度调试等系统工程任务。

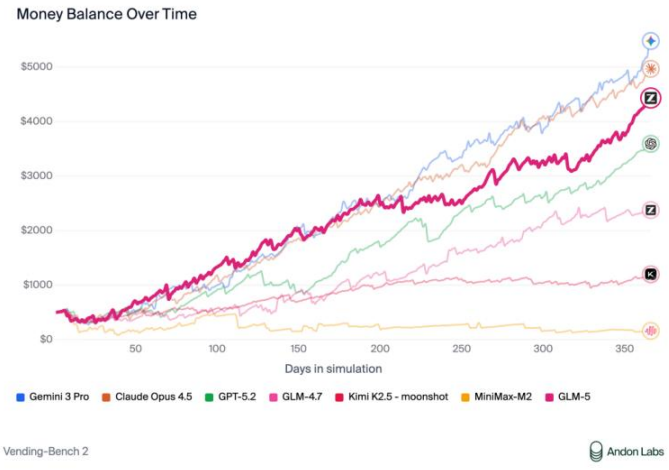
图20：GLM-5 测评结果较 GLM-4.7 大幅提升



资料来源：公司官网，国信证券经济研究所整理

GLM-5 在 Agent 能力上实现开源 SOTA，在多个评测基准中取得开源第一。在 BrowseComp（联网检索与信息理解）、MCP-Atlas（工具调用和多步骤任务执行）和 τ²-Bench（复杂多工具场景下的规划和执行）均取得最佳表现。在衡量模型经营能力的 Vending Bench 2 中，GLM-5 获得开源模型第一的表现，展现了出色的长期规划和资源管理能力。

图21：GLM-5 在 Vending Bench 2 测试中取得开源第一



资料来源：公司官网，国信证券经济研究所整理

公司基于开源合作、计算基础设施以及多元化行业合作伙伴关系培育了完备的成长生态：

1）开发者：公司广泛开源了自身的模型，截至 2025H1，公司的开源模型在全球开发者社区下载超 4500 万次，同时基于模型创建的开源项目已超 1000 个。开发者可借此通过模型微调及增量模型训练定制开源模型，使模型满足其特定需求；

2）基础设施：公司致力于实现广泛的算力兼容性，例如基于云的大规模集群、异构高性能服务器及边缘嵌入式加速器，以便模型可以轻松部署在各种算力基础设施上。截至 2025H1，公司的模型与全球 40 多个主要芯片平台兼容。当前 GLM-5 已完成与华为昇腾、摩尔线程、寒武纪、昆仑芯、海光等国产算力平台的深度推理适配；

3）商业伙伴：凭借公司强大的基座模型、顶级的算力兼容性和全面的开发工具，公司的商业合作伙伴能够迅速将 AI 能力扩展至多个行业。公司客户包括企业、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3：2025H1 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：招股说明书，国信证券经济研究所整理

表4：公司落地客户案例总结

客户类型	使用效果
消费电子产品生产商	智能体模型集成至其最新的智能手机系列中，实现端侧 AI 音视频通话、社交媒体内容生成、系统集成与函数调用等功能，依托 GLM-Realtime 模型使移动 AI agent 能在通话过程中实时解读图像和视频，即时识别屏幕上的文本以及细微的视觉细节。利用模型的多模态生成能力，用户仅需选取相册图片，移动 AI agent 即可高效创作简洁且吸引眼球的社交媒体内容。
金山办公 WPS AI	帮助 WPS AI 提升其生成内容的质量，用户仅需提供一个主题，WPS AI 即能逐步生成大纲、幻灯片内容及演讲稿，并自主完成内容格式化。帮助金山办公将我们的大模型集成至其各类办公软件产品中，显著改善用户体验并降低运行成本。
智联招聘 AI 招聘助手	对雇主而言，AI 助手通过对话式交互了解雇主招聘需求、筛选简历并提供有针对性的建议；在面试过程中，AI 助手会分析求职者技能并提供全面的面试后评估。对求职者而言，在求职者提供其优势信息后，AI 助手可生成完整、专业的简历，并可根据求职者偏好进一步润色该简历。
AI 生成内容平台	于 2025 年 4 月推出一款智能短视频生成智能体，该智能体允许用户以自然语言输入创意，并自动转化为完整短视频内容，包括视觉效果、旁白与背景音乐。公司技术为该智能体配备先进的大语言模型与视频生成能力，实现从文本输入到完整多模态内容的无缝转换。
汽车制造商鸿蒙座舱	引入通过多轮引导式对话确认用户意图的机制，使系统能更精准地解读用户需求，显著增强对指令的理解能力，即便指令以自然或非正式方式表达亦可有效识别。系统可在不同对话风格与角色间动态切换，依据个体用户偏好提供更沉浸、更具亲和力的对话体验；同时展现先进的情感智能，可促成共情的情境感知对话，实现更具人性化的交互。系统可基于用户输入与特定场景生成娱乐性即兴内容，如笑话、故事及个性化回复，提升座舱沟通的交互质量。升级后的系统亦支持语音控制游戏（如谜语竞猜等），营造更具互动性与愉悦感的用户体验。
蒙牛乳业的 AI 营养师	合作创建“蒙蒙”，使其多样化的客户群能够随时以自然方式与 AI 营养师交互，获取专家见解与个性化营养健康服务。此外，“蒙蒙”还包含 AI 健康规划师功能：该功能基于个人评估制定个性化健康与营养计划，并提供实时交互与进度追踪；系统采用自适应规划，在偏离目标时动态调整，并提供主动提醒、指导与共情激励。

资料来源：招股说明书，国信证券经济研究所整理

自 GLM-4.5 发布起至 2025 年 12 月，公司在 OpenRouter 上的 token 消耗量持续位居全球前十及中国前三，并在 GLM-5 测试版发布时提升至全球第一。这种稳定表现凸显了 GLM 的竞争力与市场认可度。2025 年 9 月，根据检索增强生成（RAG）领域的 LLM 幻觉排行榜，GLM-4.5 的幻觉率为全球第二低、中国最低。以 2024 年收入计，公司是中国最大的独立大语言模型厂商，也是中国第二大大语言模型厂商。在当前国内头部厂商中，多数公司仍难以在同一 MaaS 平台上同时覆盖语言、代码、图像、视频、音频等全模态能力。公司具备全模态 MaaS 能力，也是唯一一家将 MaaS 平台同时用于 GUI 智能体、手机/网站应用及计算机应用产品的厂商。

表5：公司市占率领先

排名	公司	类型	收入（人民币十亿元）	市场估有率
1	公司 A	非独立	0.44	9.40%
2	本公司	独立	0.31	6.60%
3	公司 B	非独立	0.3	6.40%
4	公司 C	非独立	0.29	6.10%
5	公司 D	非独立	0.22	4.70%

资料来源：招股说明书，国信证券经济研究所整理

表6：公司 MaaS 平台包含完备功能服务

公司	语言	代码生成	问答生成	视频生成	音频生成	实时视频	推理	文本	GUI 智能体	手机/网站应用	计算模型应用
本公司	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
公司 A	✓		✓		✓		✓				
公司 B	✓	✓	✓	✓	✓		✓				
公司 C	✓		✓			✓	✓				
公司 D	✓		✓		✓		✓				
公司 E	✓	✓	✓	✓	✓	✓	✓				
公司 F	✓	✓	✓	✓	✓	✓	✓				

资料来源：招股说明书，国信证券经济研究所整理

◆ 盈利预测

我们的盈利预测基于以下假设：

本地化部署：通过将大模型托管在用户自身的基础设施内，可以为用户定制其私有化专属的大模型，提升大模型的数据安全性，并为用户在性能优化和基础设施配置方面提供更大的控制权。随着智能化转型的不断推进，对大模型性能和安全的需求不断增强，公司本地化部署收入有望持续增长。我们预计 2026/2027/2028 年公司本地化部署收入增速分别为 120.0%/60.0%/50.0%，占公司总收入比重分别为 58.4%/47.3%/42.8%。

云端部署：云端部署模型将模型部署在云端基础设施上，可以为用户高效快捷的部署 AI 解决方案，公司通过订阅或 tokens 使用获取收入。随着 AI 大模型性能不断提升，功能不断增强，以 AI 模型为底层的 AI 应用不断拓展，对线上大模型的需求量有望持续增加。我们预计 2026/2027/2028 年公司云端部署收入增速分别为 340.0%/150.0%/80.0%，占公司总收入比重分别为 41.6%/52.7%/57.2%。

表7：公司营业收入预测（单位：百万元）

	2023	2024	2025	2026E	2027E	2028E
营业收入	125	312	724	2,012	3,974	6,589
yoy	116.9%	150.9%	131.9%	177.8%	97.5%	65.8%
毛利率	64.6%	56.3%	41.0%	41.7%	49.5%	54.3%
一、本地化部署						
收入	113	264	534	1,175	1,880	2,819
yoy	105.4%	134.4%	102.3%	120.0%	60.0%	50.0%
收入占比	90.4%	84.5%	73.7%	58.4%	47.3%	42.8%
二、云端部署						
收入	12	48	190	838	2,094	3,770
yoy	359.7%	306.6%	292.7%	340.0%	150.0%	80.0%
收入占比	9.6%	15.5%	26.3%	41.6%	52.7%	57.2%

资料来源：公司财报，彭博，国信证券经济研究所整理和预测

期间费用率：公司正处于高速发展期，随着公司收入规模扩大，公司销售及分销费用、行政费用和研发费用都将得到明显改善，总费用率随收入增长而逐步下降。我们预计 2026/2027/2028 年公司销售及分销费用率为 21.4%/13.0%/9.4%，行政费用率为 30.1%/18.3/12.7%，研发费用率为 205.5%/130.1%/94.1%。

表8：公司期间费用预测（单位：百万元）

	2024	2026	2026E	2027E	2028E
销售及分销费用	387	391	430	516	619
yoy	282.9%	0.9%	10.0%	20.0%	20.0%
销售及分销费用/营收	124.0%	54.0%	21.4%	13.0%	9.4%
行政费用	134	505	606	728	837
yoy	101.5%	278.3%	20.0%	20.0%	15.0%
行政费用/营收	42.8%	69.8%	30.1%	18.3%	12.7%
研发费用	2,195	3,180	4,135	5,168	6,202
yoy	315.1%	44.9%	30.0%	25.0%	20.0%
研发费用/营收	702.7%	439.1%	205.5%	130.1%	94.1%

资料来源：公司财报，彭博，国信证券经济研究所整理和预测

按上述假设条件与假设，我们得到公司 2026/2027/2028 年营业收入分别为 20.12/39.74/65.89 亿元，分别同比增长 177.8%/97.5%/65.8%。

◆ 投资建议

投资建议：领先的独立大语言模型公司，产品矩阵快速扩张，首次覆盖，给予“优于大市”评级。公司构建了以 GLM 语言模型为核心，覆盖智能体模型（如 AutoGLM）、多模态模型（如 CogView，CogVideoX）、代码模型（CodeGeeX）的全面产品矩阵，并为约 8000 家企业客户，8000 万个人用户提供服务。公司 GLM5 大模型在 Coding 与 Agent 能力上取得开源 SOTA 表现，在真实编程场景的使用体感逼近 Claude Opus 4.5，并基于开源合作培育了完备的成长生态。公司底层大模型能力领先，产品矩阵快速扩张，大模型商业化处于国内领先地位，未来业绩增长空间有望进一步打开，首次覆盖，给予“优于大市”评级。

◆ 风险提示

盈利预测的风险：1）我们假设公司未来 3 年收入增长 177.8%/97.5%/65.8%，可能存在对公司产品销量及价格预计偏乐观、进而高估未来 3 年业绩的风险。2）我们预计公司未来 3 年毛利分别为 8.39/19.65/35.76 亿元，可能存在对公司成本估计偏低、毛利高估，从而导致对公司未来 3 年盈利预测值高于实际值的风险。

AI 落地不及预期的风险：公司 AI 产品已经实现商业化，随着 AI 产业快速发展，AI 产品需要迅速迭代以满足市场需求，若公司新一代 AI 产品研发不及预期，将影响公司未来产品的市场份额；同时，目前市场已有同类产品上市或在研竞品，未来商业化预计会面临激烈竞争，出现商业价值低或不及预期的风险，如果不能如期获得市场认可，将会对公司经营发展产生不利影响。

技术被赶超或替代的风险：公司所处行业属于技术密集型行业，涉及软件平台构建、AI 产品研发等技术，在未来提升研发技术能力的竞争中，如果公司不能准确把握行业技术的发展趋势，在技术开发方向决策上发生失误；或研发项目未能顺利推进，未能及时将新技术运用于产品开发和升级，出现技术被赶超或替代的情况，公司将无法持续保持产品的竞争力，从而对公司的经营产生重大不利影响。

宏观经济及行业波动风险：如果未来宏观经济发生剧烈波动，导致终端 AI 需求下滑，将对公司的业务发展和经营业绩造成不利影响。

财务预测与估值

资产负债表(百万元)	2024	2025	2026E	2027E	2028E
现金及现金等价物	2268	2259	1925	1925	1925
应收款项	667	699	1942	3834	6358
存货净额	32	126	861	1508	2288
其他流动资产	6	2	5	9	15
流动资产合计	**3016**	**3571**	**5218**	**7763**	**11072**
固定资产	243	656	568	478	388
无形资产及其他	714	90	79	68	56
投资性房地产	97	198	198	198	198
长期股权投资	307	338	338	338	338
资产总计	**4376**	**4854**	**6401**	**8845**	**12052**
短期借款及交易性金融负债	269	605	2352	4843	7710
应付款项	81	40	272	477	723
其他流动负债	7488	11762	15278	19171	23145
流动负债合计	**7838**	**12406**	**17902**	**24490**	**31579**
长期借款及应付债券	0	85	85	85	85
其他长期负债	493	474	774	1074	1374
长期负债合计	**493**	**559**	**859**	**1159**	**1459**
负债合计	**8331**	**12965**	**18761**	**25649**	**33037**
少数股东权益	1	(18)	(36)	(55)	(73)
股东权益	(3957)	(8093)	(12323)	(16749)	(20912)
负债和股东权益总计	**4376**	**4854**	**6401**	**8845**	**12052**

关键财务与估值指标	2024	2025	2026E	2027E	2028E
每股收益	(6.63)	(10.54)	(9.49)	(9.93)	(9.34)
每股红利	0.00	0.00	0.00	0.00	0.00
每股净资产	(8.87)	(18.15)	(27.64)	(37.57)	(46.91)
ROIC	96%	62%	46%	38%	31%
ROE	75%	58%	34%	26%	20%
毛利率	56%	41%	42%	49%	54%
EBIT Margin	-948%	-651%	-216%	-112%	-62%
EBITDA Margin	-858%	-612%	-212%	-110%	-61%
收入增长	151%	132%	178%	97%	66%
净利润增长率	--	--	--	--	--
资产负债率	190%	267%	293%	289%	274%
息率	0.0%	0.0%	0.0%	0.0%	0.0%
P/E	(129.2)	(81.3)	(90.3)	(86.3)	(91.8)
P/B	(96.6)	(47.2)	(31.0)	(22.8)	(18.3)
EV/EBITDA	(133)	(81)	(86)	(85)	(95)

利润表（百万元）	2024	2025	2026E	2027E	2028E
营业收入	**312**	**724**	**2012**	**3974**	**6589**
营业成本	137	428	1174	2008	3012
营业税金及附加	0	0	0	0	0
销售费用	387	391	430	516	619
管理费用	2339	3698	4752	5907	7050
财务费用	0	0	(24)	68	175
投资收益	21	55	25	34	38
资产减值及公允价值变动	(17)	(22)	(22)	(22)	(22)
其他收入	(9)	(47)	68	69	70
营业利润	(2555)	(3806)	(4249)	(4444)	(4181)
营业外净收支	(403)	(912)	0	0	0
利润总额	**(2958)**	**(4718)**	**(4249)**	**(4444)**	**(4181)**
所得税费用	0	0	0	0	0
少数股东损益	(2)	(20)	(18)	(19)	(18)
归属于母公司净利润	**(2956)**	**(4698)**	**(4231)**	**(4425)**	**(4164)**

现金流量表(百万元)	2024	2025	2026E	2027E	2028E
净利润	**(2956)**	**(4698)**	**(4231)**	**(4425)**	**(4164)**
资产减值准备	0	0	0	0	0
折旧摊销	280	278	78	79	80
公允价值变动损失	17	22	22	22	22
财务费用	0	0	(24)	68	175
营运资本变动	3872	3992	2068	1853	1212
其它	(2)	(20)	(18)	(19)	(18)
经营活动现金流	**1211**	**(427)**	**(2081)**	**(2491)**	**(2867)**
资本开支	34	(658)	0	0	0
其它投资现金流	116	(443)	0	0	0
投资活动现金流	**(41)**	**(1132)**	**0**	**0**	**0**
权益性融资	0	0	0	0	0
负债净变化	0	85	0	0	0
支付股利、利息	0	0	0	0	0
其它融资现金流	(151)	1381	1747	2491	2867
融资活动现金流	**(151)**	**1550**	**1747**	**2491**	**2867**
现金净变动	**1019**	**(9)**	**(334)**	**0**	**0**
货币资金的期初余额	1249	2268	2259	1925	1925
货币资金的期末余额	2268	2259	1925	1925	1925
企业自由现金流	1759	(1101)	(2208)	(2536)	(2811)
权益自由现金流	1608	364	(437)	(113)	(118)

资料来源：Wind、国信证券经济研究所预测

免责声明

分析师声明

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。

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

重要声明

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中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 层
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