

# 【广发计算机&海外】智谱(02513.HK)

## 深耕 AI 大模型领域，各场景落地拓展业务边界

### 核心观点：

- **智谱的模型产品矩阵完整，为其商业化奠定了良好基础。**作为国内领先的大模型服务提供商，智谱以自研 GLM 基座大模型为核心，形成了包括多模态、Agent 和 Coding 完整的模型矩阵，为企业客户提供 API 服务、本地化部署与行业解决方案。2022-2024 年收入体量快速放大，从 0.6 亿元扩张至 3.1 亿元，复合年增长率超 130%，主要系公司业务扩张和本地化及云端部署销量增加所致。2025H1 公司收入 1.9 亿元，在高基数背景下仍实现较快增长，显示公司商业化能力持续兑现。
- **公司收入以本地化部署为主，其中互联网行业客户占比较高。**公司模型的商业化落地主要分为云端和本地化部署两种方式，其中本地化部署占营收比重超过 80%。在本地化部署业务中，互联网与科技行业客户占比较高，从 2022 年到 2024 年稳定在一半以上的份额。
- **公司研发高投入，持续提升模型竞争力。**公司的研发费用自 2022 年的 0.8 亿元大幅提升至 2025H1 的 16.0 亿元，其中算力服务费占比较大。算力服务费由 2022 年的 0.2 亿元大幅提升至 2025H1 的 11.5 亿元。后续随着算力成本下行、模型架构优化带来的研发效率提升，叠加现有模型产品矩阵成熟度提高，研发费用率有望快速下降。
- **盈利预测与投资建议。**公司 25-27 年收入预计为 7.4 亿元、15.3 亿元和 28.8 亿元，营收增速预计为 138.1%、106.2%和 87.9%。预计 25-27 年 EPS 分别为-7.79 元/股、-8.14 元/股、-7.26 元/股。参考商汤和 MiniMax 两家公司估值水平，综合考虑公司 AI 基础能力、技术优势和行业地位，给予智谱 2026 年 70 倍 PS 估值，合理总价值为 1073.63 亿元人民币，对应每股合理价值 273.22 港元/股(参考汇率 1 港元=0.89 人民币)，给予“增持”评级。
- **风险提示。**下游企业客户对 AI 大模型付费意愿和能力不高的风险；研发投入巨大和成果转化的风险；科技巨头在 AI 大模型领域的投入及行业竞争加剧的风险。

### 盈利预测：

单位:人民币百万元	2023A	2024A	2025E	2026E	2027E
主营收入	125	312	744	1,534	2,882
增长率 (%)	116.9%	150.9%	138.1%	106.2%	87.9%
EBITDA	-626	-2,517	-3,421	-3,573	-3,156
归母净利润	-788	-2,956	-3,430	-3,583	-3,196
增长率 (%)	-	-	-	-	-
EPS (元/股)	-	-	-7.79	-8.14	-7.26
市盈率 (P/E)	-	-	-	-	-
ROE (%)	-	-	-	-	-
EV/EBITDA	-	-	-	-	-

数据来源：公司财务报表，广发证券发展研究中心

### 公司评级

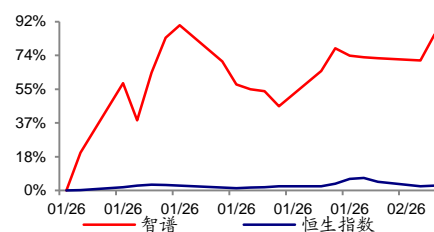
增持

当前价格	243.20 港元
合理价值	273.22 港元
报告日期	2026-02-03

### 基本数据

总股本/流通股本 (百万股)	440.23/215.70
总市值/流通市值 (百万港元)	107063.98/52458.65
一年内最高/最低 (港元)	250.00/131.50
30 日日均成交量/成交额 (百万)	4.05/774.65
近 3 个月/6 个月涨跌幅 (%)	84.94/84.94

### 相对市场表现



分析师：刘雪峰



SAC 执证号: S0260514030002

SFC CE No. BNX004



021-38003675



gfliuxuefeng@gf.com.cn

分析师：周源



SAC 执证号: S0260523040001



0755-23948351



shzhouyuan@gf.com.cn

请注意，周源并非香港证券及期货事务监察委员会的注册持牌人，不可在香港从事受监管活动。

### 相关研究：

## 目录索引

投资要点 .....	5
一、智谱：深耕 AI 大模型领域，技术积累深厚，模型能力全面 .....	6
（一）公司技术积累深厚，模型商业化持续推进 .....	6
（二）公司营收快速增长，研发投入较大 .....	9
二、AI 大模型技术升级驱动应用场景拓展 .....	15
（一）AI 技术持续升级，应用场景不断拓展 .....	15
（二）AI 大模型应用场景广阔，市场规模快速增长 .....	16
（三）AI 大模型行业竞争激烈，公司模型功能全面且具有中立性特点 .....	18
三、公司模型布局全面，基座模型能力突出 .....	20
（一）GLM 基座模型各维度能力处于行业领先水平 .....	21
（二）公司的多模态大模型覆盖语音、图像、视频等多个领域 .....	24
（三）面向 B 端和 C 端，公司 AI AGENT 产品赋能各场景 .....	28
（四）公司重点布局 AI 辅助编程赛道，模型和工具齐备 .....	32
四、盈利预测和投资建议 .....	36
五、风险提示 .....	39

## 图表索引

图 1: 智谱发展历程梳理.....	6
图 2: 北京智谱华章科技股份有限公司股权结构 (截止到 2026 年 1 月 8 日) ....	7
图 3: 智谱的一站式 MaaS 平台 .....	7
图 4: 智谱业务工作流程.....	9
图 5: 2022-2025H1 智谱收入 .....	9
图 6: 2022-2025H1 智谱收入按业务拆分 .....	9
图 7: 2022-2025H1 智谱本地化部署业务按行业拆分 .....	10
图 8: 2022-2025H1 智谱本地化部署业务按地区拆分 .....	10
图 9: 2022-2025H1 智谱毛利及毛利率.....	11
图 10: 2022-2025H1 智谱各项业务毛利及毛利率 .....	11
图 11: 2022-2025H1 智谱各项费用 .....	11
图 12: 2022-2025H1 智谱各项费用率 .....	11
图 13: 2022-2025H1 智谱研发费用结构.....	12
图 14: 智谱员工结构 (截至 2025 年 6 月 30 日) .....	12
图 15: 2022-2025H1 智谱流动资产 .....	12
图 16: 2022-2025H1 智谱银行存款及手头现金.....	12
图 17: 2022-2025H1 智谱流动比率 .....	13
图 18: 2022-2025H1 智谱资产负债率 .....	13
图 19: 2022-2025H1 智谱应收账款和应收账款周转率.....	13
图 20: 2022-2025H1 智谱账龄在一年以内的应收账款占总应收账款的比例 .....	13
图 21: 2022-2025H1 智谱净亏损和经营活动所用现金.....	14
图 22: 大语言模型行业价值链示意图 .....	16
图 23: 全球大模型市场规模预测.....	16
图 24: 中国大语言市场模型规模.....	18
图 25: 中国企业级大语言市场模型规模.....	18
图 26: 全球语言模型排行榜 (截至 2026 年 1 月 13 号) .....	21
图 27: GLM Coding Plan 全球用户使用数据展示.....	23
图 28: GLM Coding Plan 付费用户增长趋势 .....	23
图 29: GLM-4.6V 与同类模型的性能对比.....	25
图 30: GLM-4.6V 前端复刻能力案例: 小红书界面像素级复刻 .....	25
图 31: GLM-4.6V 视觉网页搜索案例: 电商以图搜款及导购清单生成 .....	25
图 32: GLM-Image 与同类模型的综合性能对比 .....	26
图 33: GLM-Image 绘制导演电影作品年表效果图 .....	27
图 34: Nano Banana Pro 绘制导演电影作品年表效果图.....	27
图 35: AutoGLM 外卖选购样例展示 .....	29
图 36: CoCo 政务解读示例.....	29
图 37: 千问 APP 宣布全面接入阿里生态场景 .....	30
图 38: AutoGLM 模拟人类操作外卖 APP 界面 .....	32
图 39: 千问 APP 独立界面点外卖交互.....	32
图 40: CodeGeeX4-ALL-9B 模型核心能力示意图.....	33

图 41: CodeGeeX 在 VS Code 中辅助生成单元测试代码示例 .....	33
图 42: 2022-2024 年 CodeGeeX 插件下载量与 GitHub Star 数增长趋势 (单位: 个) .....	33
图 43: Z Code 自动生成贪吃蛇游戏演示 .....	34
表 1: 智谱业务收费模式 .....	8
表 2: 大语言模型与判别式 AI 主要区别 .....	15
表 3: 云端部署与本地化部署主要区别 .....	17
表 4: 中国头部大语言模型厂商排名 (2024 年) .....	18
表 5: 中国头部大语言模型厂商的模型能力对比 .....	19
表 6: 智谱的 AI 模型及智能体 .....	20
表 7: 主流大语言模型 API 定价对比表 .....	21
表 8: GLM Coding Plan 订阅套餐详情 .....	22
表 9: GLM 基座模型商业化案例 .....	23
表 10: 智谱多模态模型功能与应用场景 .....	24
表 11: Nano Banana Pro 与 GLM-Image 大模型价格对比表 .....	27
表 12: 智谱多模态模型商业化案例 .....	28
表 13: 智谱 Agent 产品一览 .....	28
表 14: 智谱 AI Agent 商业化案例 .....	30
表 15: 智谱 AI Agent 和千问 APP 的功能和生态对比 .....	31
表 16: AI 辅助编程工具对比 .....	35
表 17: 2023-2027 年智谱营收拆分 (单位: 百万元) .....	36
表 18: 2023-2027 年智谱盈利预测 (单位: 百万元) .....	37
表 19: 智谱可比公司估值情况 (市值统计截止 2026.2.3 收盘) .....	38

## 投资要点

作为国内领先的大模型服务提供商，智谱以自研GLM基座大模型为核心，形成了包括多模态、Agent和Coding完整的模型矩阵，为企业客户提供API服务、本地化部署与行业解决方案。

**智谱的模型产品矩阵完整，为其商业化奠定了良好基础。**智谱围绕深度思考、认知世界和工具使用三方面能力构建了完整的模型产品矩阵，涵盖基座模型、反思及沉思模型、多模态模型、AI agent 与代码模型五大类。智谱完整的模型矩阵，不仅可满足B端金融、政务、教育等多元行业的定制化需求，还覆盖了C端用户的日常交互、内容创作、编程辅助等场景，大幅拓宽用户覆盖边界与商业落地空间。2022-2024年收入体量快速放大，从0.6亿元扩张至3.1亿元，复合年增长率超130%，主要系公司业务扩张和本地化及云端部署销量增加所致。2025H1公司收入1.9亿元，在高基数背景下仍实现较快增长，显示公司商业化能力持续兑现。

**从业务结构看，公司收入以本地化部署为主，云端业务增长提速明显。**公司模型的商业化落地主要分为云端和本地化部署两种方式，其中本地化部署占营收比重超过80%。在本地化部署业务中，互联网与科技行业客户占比较高，从2022年到2024年稳定在一半以上的份额。模型本地化部署主要源于政企及大型客户对数据安全、法规遵从性和定制化交付的需求，故本地部署是当前收入增长的核心来源，云端部署占比相对较低。但随着大模型迭代速度增快，云端部署的优势逐渐显现，即用户不需要跟随模型升级而重新购买服务进行部署，叠加自身的敏捷性和易实施性，云端部署业务的占比呈现逐步提升趋势，从2022年的不到5%爬升至2025H1的15%。

**公司研发高投入，持续提升模型竞争力。**尽管公司收入持续增长，但期间费用增速更快，其中主要是研发投入较大。研发费用自2022年的0.8亿元大幅提升至2025H1的16.0亿元，占收入比例达835%。目前，AI大模型市场竞争格局分散，公司在研发上大量投入有利于其增强其模型和产品竞争力，从而提升市场份额。**从研发费用结构看，算力服务费占比较大。**算力服务费由2022年的0.2亿元大幅提升至2025H1的11.5亿元，反映出公司在在大模型训练与迭代过程中需持续向第三方基础设施供应商采购更多算力资源。后续随着算力成本下行、模型架构优化带来的研发效率提升，叠加现有模型产品矩阵成熟度提高，研发费用率有望快速下降。

**盈利预测和投资建议：**公司2025-2027年收入预计为7.4亿元、15.3亿元和28.8亿元，营收增速预计为138.1%、106.2%和87.9%。预计2025-2027年EPS分别为-7.79元/股、-8.14元/股、-7.26元/股。参考商汤和MiniMax两家公司估值水平，综合考虑公司AI基础能力、技术优势和行业地位，给予智谱2026年70倍PS估值，合理总价值为1073.63亿元人民币，对应每股合理价值273.22港元/股（参考汇率1港元=0.89人民币），给予“增持”评级。

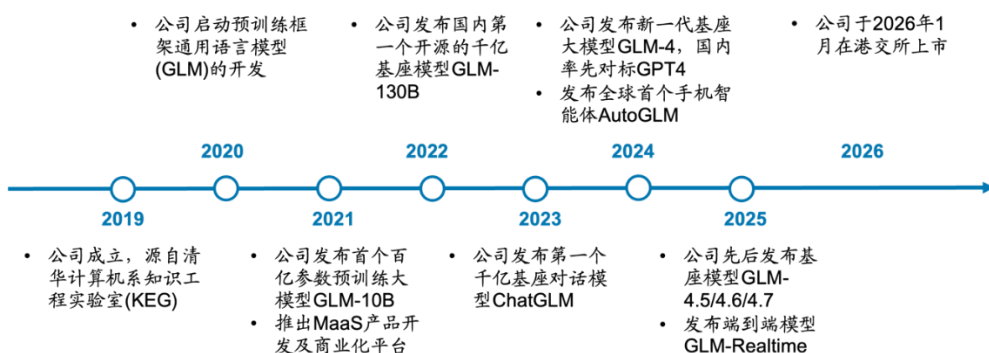
**风险提示：**下游企业客户对AI大模型付费意愿和能力不高的风险；研发投入巨大和成果转化的风险；科技巨头在AI大模型领域的投入及行业竞争加剧的风险。

# 一、智谱：深耕 AI 大模型领域，技术积累深厚，模型能力全面

## （一）公司技术积累深厚，模型商业化持续推进

智谱成立于2019年，定位为中国领先的人工智能公司，聚焦通用大模型（AGI）研发，为机构客户和个人客户提供通用大模型服务，稳步推进商业化落地，实现收入快速增长。2021年公司发布中国首个专有预训练大模型框架GLM，并推出模型即服务（MaaS）产品开发与商业化平台；2022年进一步开源千亿参数模型GLM-130B，强化技术影响力与生态建设。截至2025年6月30日，公司模型已支持逾8000家机构客户，并覆盖约8000万台设备；按公司招股说明书引用的弗若斯特沙利文数据，智谱以2024年收入计在中国独立通用大模型开发商中排名第一、在所有通用大模型开发商中排名第二，市场份额约6.6%。

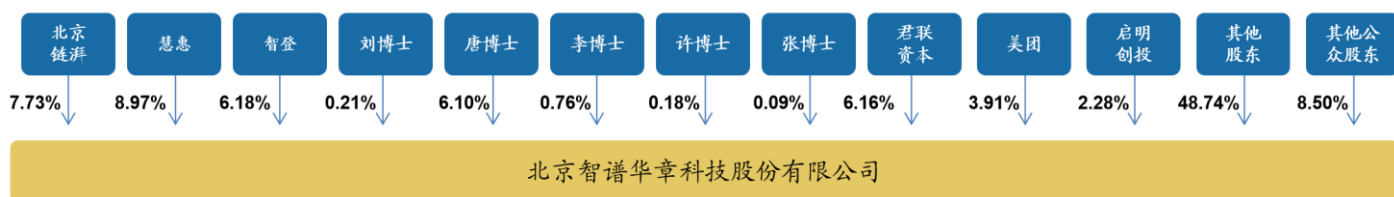
图 1：智谱发展历程梳理



数据来源：公司招股说明书、公司官网，广发证券发展研究中心

**公司股权结构较为多元。**北京链湃是创始人刘博士控制的持股平台，持有公司7.73%股份。慧惠、智登为刘博士控制的员工持股平台，分别持有 8.97%、6.18% 股份。刘博士、唐博士、李博士、许博士、张博士分别直接持股 0.21%、6.10%、0.76%、0.18%、0.09%。上述持股平台与核心创始人构成一致行动人士，形成控股股东集团，合计持股 30.22%。君联资本资深独立投资者（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）、启明创投资深独立投资者（启明融干、启明融凯）分别持股 6.16%、3.91%、2.28%。参与全球发售的公众股东持股 8.50%。其他股东如星连肇基和通智投资等合计持股 48.74%。

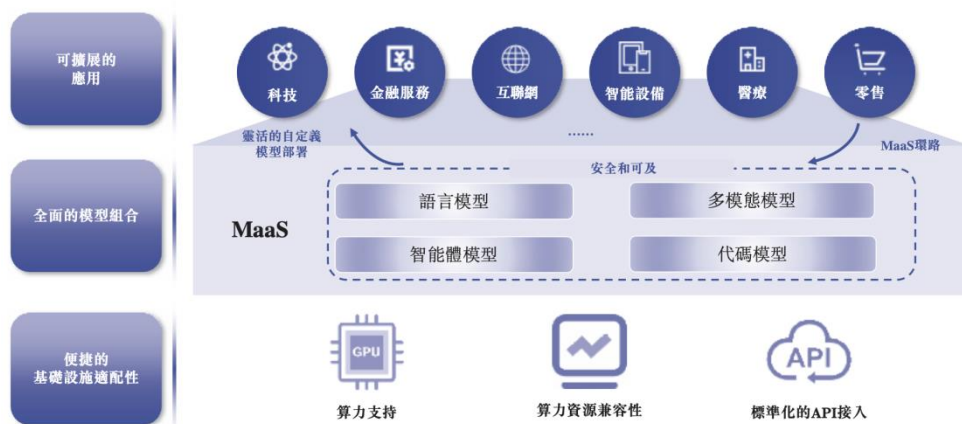
图 2: 北京智谱华章科技股份有限公司股权结构 (截止到2026年1月8日)



数据来源: 公司招股说明书, 广发证券发展研究中心

智谱依托一站式MaaS平台, 通过三层级平台架构整合打造产品与服务体系。首先以合作共建的先进算力设施为底层根基, 实现15亿至2300亿参数模型的跨云、跨芯片组部署, 让技术能力延伸至手机、电脑、智能汽车等终端设备。在此基础上搭建语言、多模态、智能体及代码模型矩阵作为核心支撑, 凭借多元模型能力适配不同客户的个性化需求。最终聚焦科技、金融服务、互联网、智能设备、医疗、零售等领域提供解决方案, 同时通过智能体工作区赋能客户快速定制专属工具, 助力机构优化运营决策, 实现技术商业化的高效落地。

图 3: 智谱的一站式MaaS平台



数据来源: 公司招股说明书, 广发证券发展研究中心

公司各业务收费模式如下:

1. 本地化部署业务主要聚焦机构客户需求, 提供定制化大模型解决方案, 采用套餐化交付模式。套餐价格结合模型规格、配套计算资源及项目实施成本综合核定, 支持一次性付费或按年付费, 灵活适配客户短期试用与长期合作的差异化需求。根据公司招股说明书, 典型案例方面, 中国综合性企业借助智谱大模型开发财务资本管理应用, 在金融新规问答场景实现超50%的人工智能自动化率。某综合性企业依托智谱CoCo企业级智能体, 高效构建集知识库与工作流程为一体的智能化办公系统, 大幅提升办公效率。截至2025年6月30日, 该业务已服务95个机构客户。

2. 云端部署业务覆盖机构与个人用户, 采用多元化收费模式。机构客户可选择按token消耗量计费, 或按订阅时长、模型规格及计算资源付费。个人用户可免费使用Z.ai、AutoGLM等平台基础功能, 高级功能需订阅。典型案例方面, 根据公司官方

微信公众号，GLM Coding Plan上线四个月内全球付费用户即突破15万。根据公司招股说明书，智谱清言平台活跃自定义AI agent数量超80万，用户粘性持续提升。智联招聘依托其模型开发AI招聘助手，为全招聘流程提供赋能支持。截至2025年6月30日，该业务已服务3061个机构客户。

**表 1：智谱业务收费模式**

业务	收费模式	客户情况	典型商业化案例
本地化部署	套餐形式一次性/按年计费	截至2025年6月30日有95个机构客户	1. 中国综合性企业利用智谱大模型开发财务资本管理应用，在金融新规问答场景实现超50%的人工智能自动化率； 2. 中国综合性企业办公系统依托智谱 CoCo 企业级智能体高效构建集知识库与工作流程为一体的智能化系统
云端部署	使用量计费、订阅制收费	截至2025年6月30日有3061个机构客户	1. 智谱 GLM Coding Plan 订阅套餐以每月最低 20 元上线，四个月内全球付费用户突破 15 万； 2. 智谱清言依托 GLM 系列模型打造生成式助手，平台活跃自定义 AI agent 数量突破 80 万； 3. 智联招聘依托智谱模型开发 AI 招聘助手为雇主及求职者提供招聘需求理解、简历筛选及专业简历生成功能

数据来源：公司招股说明书，广发证券发展研究中心

智谱的业务工作流程可以概括为一条“上游算力支持-中台服务交付-下游生态扩张”的链条。在上游，公司与基础设施供应商合作，一方面获取高弹性的云计算资源以支持客户按需使用，另一方面引入高性能本地化算力以支撑模型的本地增量训练与运行，并共同设计具备广泛适配性的基础设施以满足大规模、实时化部署需求；

在服务交付环节，公司通常先与客户深入沟通并围绕数据与应用场景两条主线进行需求分析：前者重点评估微调和增量训练所需的数据体量、类型以及数据安全合规要求，后者明确客户是单一场景还是多个场景统筹部署，从而决定模型规模与复杂度；在此基础上，公司设计定制化服务方案，进一步评估模型对目标场景的适配性并持续迭代方案，随后根据客户偏好选择云端部署或本地化部署，并通过模型微调、增量训练与提示工程完成定制化落地，部署后再基于客户反馈持续优化模型性能与服务质量。

在商业拓展的过程中，公司通过Z计划机制赋能行业业务合作伙伴，向其开放公司的大模型与基础设施使用权限并提供定制化技术支持，帮助合作伙伴扩大运营规模的同时，也推动智谱的大模型向更多的新行业新场景渗透，实现业务增长、生态扩张和合作共赢。

图 4: 智谱业务工作流程



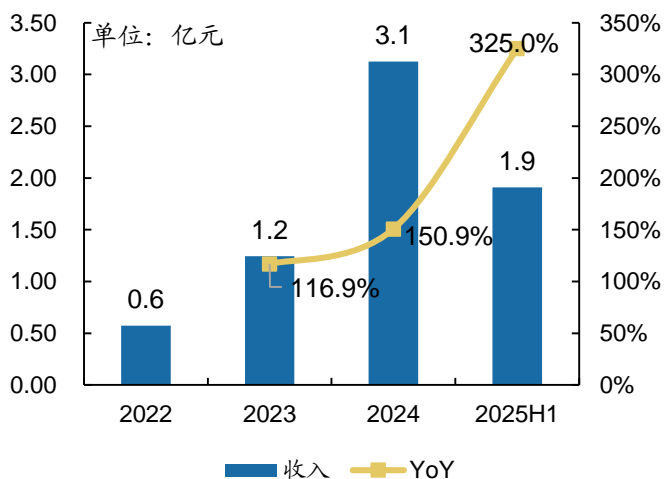
数据来源：公司招股说明书，广发证券发展研究中心

## (二) 公司营收快速增长，研发投入较大

2022-2025年上半年，公司收入规模保持高速增长。2022-2024年收入体量快速放大，从0.6亿元扩张至3.1亿元，复合年增长率超130%，主要系公司业务扩张和本地化及云端部署销量增加所致。2025H1公司收入1.9亿元，在高基数背景下仍实现较快增长，显示公司商业化能力持续兑现。

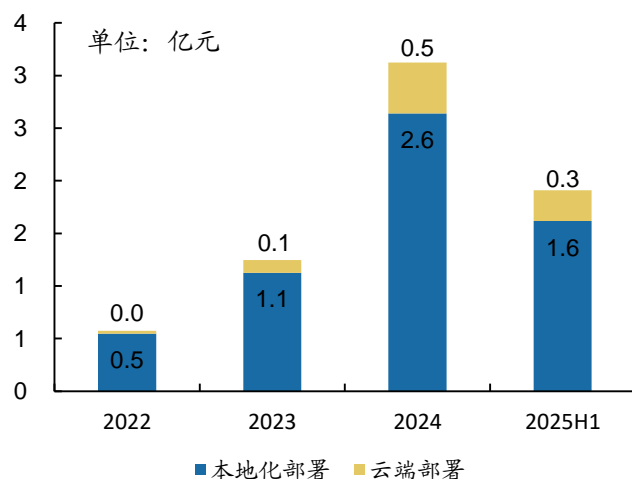
从业务结构看，公司收入以本地化部署为主，占比长期处于高位（80%以上），主要源于政企及大型客户对数据安全、法规遵从性和定制化交付的需求，故本地部署是当前收入增长的核心来源，云端部署占比相对较低。但随着大模型迭代速度增长，云端部署的优势逐渐显现，即用户不需要跟随模型升级而重新购买服务进行部署，叠加自身的敏捷性和易实施性，云端部署业务的占比呈现逐步提升趋势，从2022年的不到5%爬升至2025H1的15%。

图 5: 2022-2025H1 智谱收入



数据来源：公司招股说明书，广发证券发展研究中心

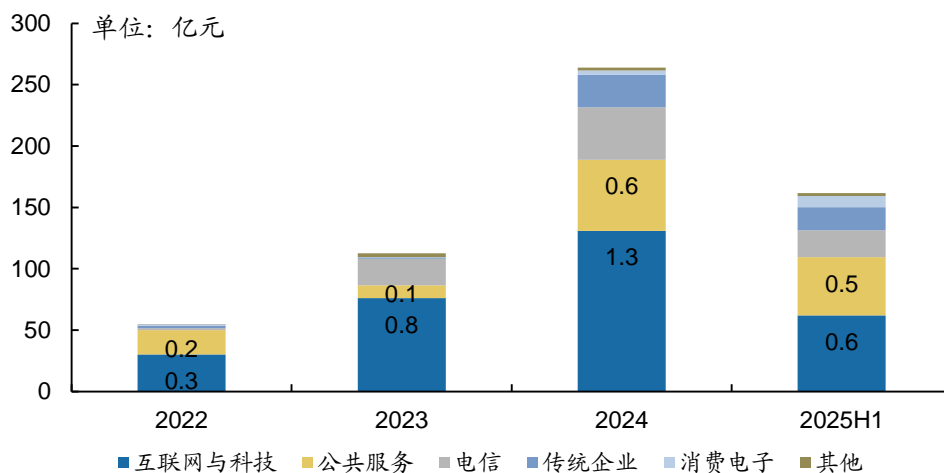
图 6: 2022-2025H1 智谱收入按业务拆分



数据来源：公司招股说明书，广发证券发展研究中心

在本地化部署业务中，互联网与科技行业客户占比较高，从2022年到2024年稳定在一半以上的份额，在2025H1小幅下降至约40%；其主要原因系互联网与科技公司对AI认知较深，具有较强的付费意愿。同时，公共服务类比例在2024年和2025H1有所提升，至2025H1占比近30%；其主要业务为与电网、低空经济相关的科技专项和出海主权大模型业务，例如马来西亚国家级MaaS平台。与传统To G端业务不同，该类业务所需人力资源少，回款账期短。未来随着公司出海业务的拓展，将会构建更多国家人工智能基础设施，该类业务占比有望稳定提升。由于云端部署的客户在注册和使用时不需要提供行业信息，故公司无法确定云端客户的所属行业。

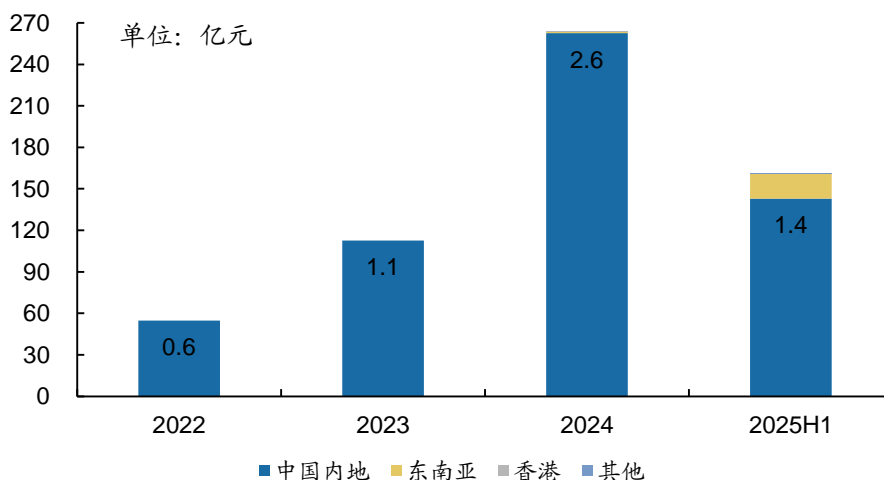
图 7: 2022-2025H1智谱本地化部署业务按行业拆分



数据来源：公司招股说明书，广发证券发展研究中心

公司本地化部署业务收入主要来自中国内地。自2024年起，公司的大模型本地化部署服务将范围拓展到了海外，其中海外客户主要为东南亚客户，此外也有少部分美国客户。公司积极推动大模型出海，重视AI发展的地域间合作，积极参与构建中国及东南亚等国家及地区的国家及城市级基座模型平台。

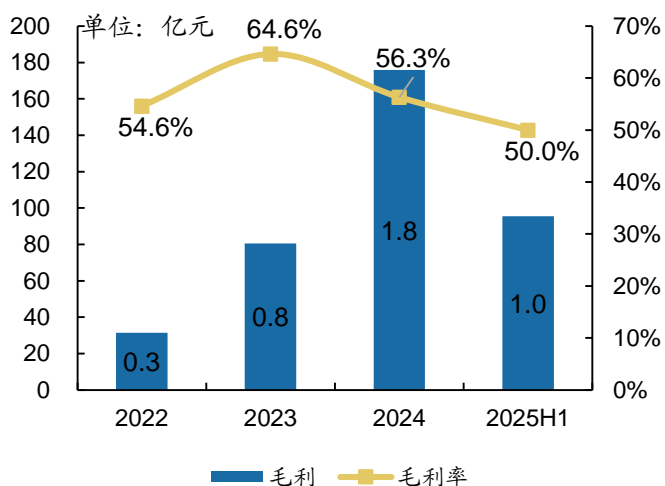
图 8: 2022-2025H1智谱本地化部署业务按地区拆分



数据来源：公司招股说明书，广发证券发展研究中心

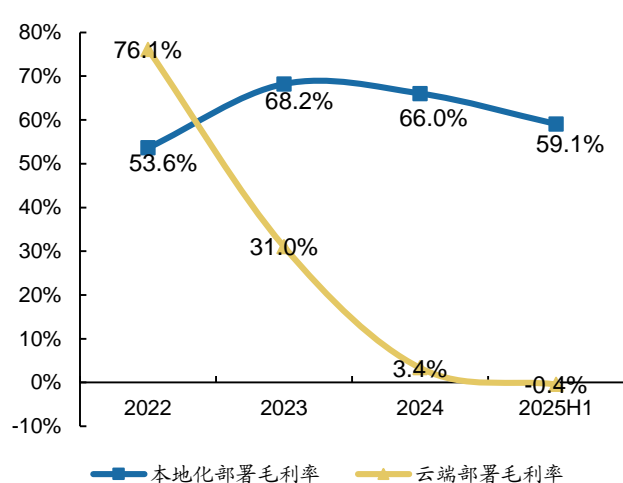
公司毛利率维持较高水平，但近年小幅下滑。从2022年到2025H1，公司的毛利率分别为54.6%、64.6%、56.3%、50.0%，整体保持较高水平，毛利润随业务销量和收入扩张而增长。公司毛利率有小幅下降的原因为，云端部署业务毛利率有较大下滑的同时该业务占收入比例逐渐提升。分业务看，本地化部署毛利率表现亮眼，体现出本地化部署业务具备的较强交付能力；云端部署毛利率呈逐年下降趋势，其原因主要系公司战略性降低服务价格、顺应市场趋势所致。预计后期伴随公司云端业务的市场份额逐渐提高和模型迭代进步，公司定价能力增强，毛利率有望提升。

图 9: 2022-2025H1智谱毛利及毛利率



数据来源：公司招股说明书，广发证券发展研究中心

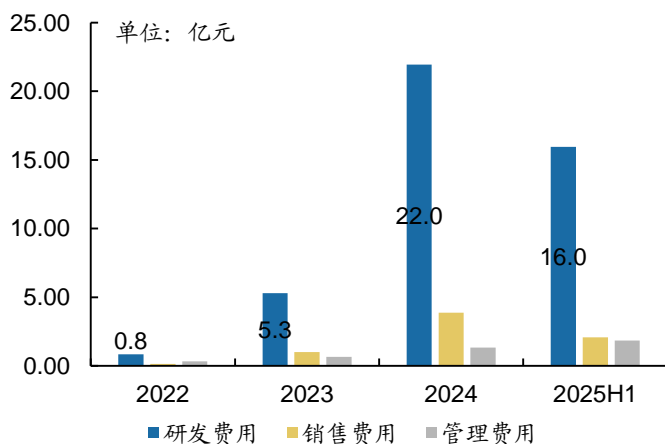
图 10: 2022-2025H1智谱各项业务毛利及毛利率



数据来源：公司招股说明书，广发证券发展研究中心

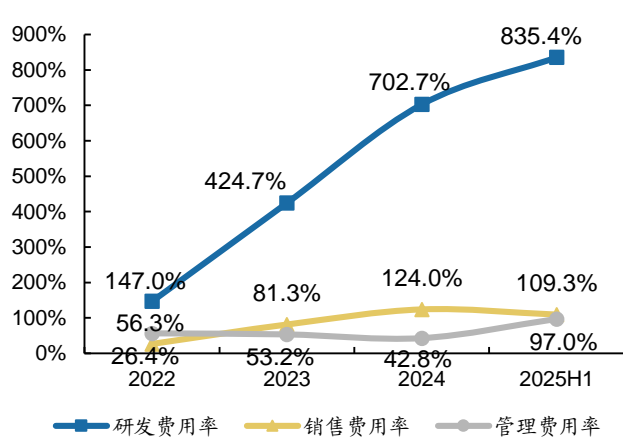
公司研发费用占比较大。尽管公司收入持续增长，但期间费用增速更快，其中主要是研发投入较大。研发费用自2022年的0.8亿元大幅提升至2025H1的16.0亿，占收入比例达835%。目前，AI大模型市场竞争格局分散，公司在新品研发上大量投入有利于其增强其产品竞争力，从而提升市场份额。管理费用比例较为稳定，销售费用比例则由2022年的26.4%提升至2025H1的109.3%。尽管公司在未来会继续产生较多的研发投入，但公司已经在控制其他开支、优化销售及营销策略和提升运营效率；随着公司未来商业化的落地和措施见效，公司的经营开支占收入比例预计会下降。

图 11: 2022-2025H1智谱各项费用



数据来源：公司招股说明书，广发证券发展研究中心

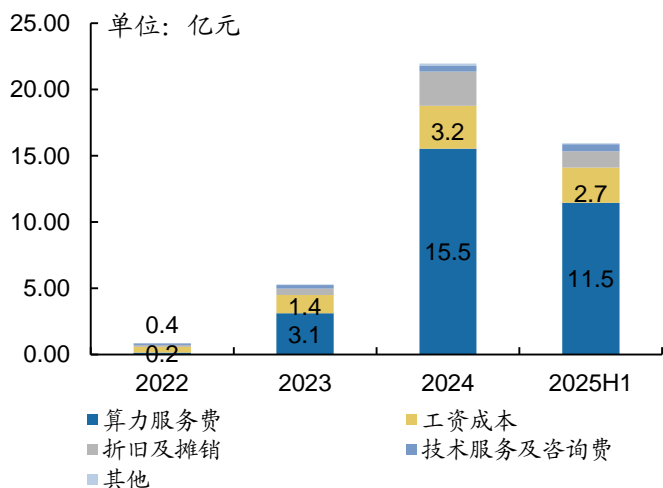
图 12: 2022-2025H1智谱各项费用率



数据来源：公司招股说明书，广发证券发展研究中心

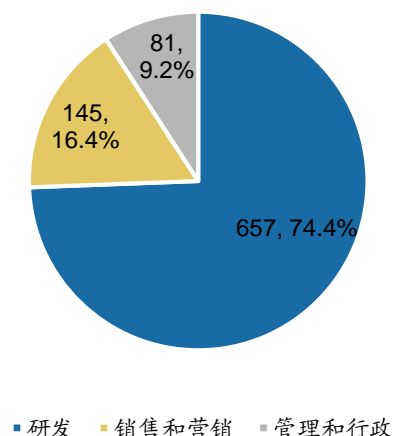
从研发费用结构看，公司研发投入的快速增长主要由算力服务费驱动：算力服务费由2022年的0.2亿元大幅提升至2025H1的11.5亿元，反映出公司在大模型训练与迭代过程中需持续向第三方基础设施供应商采购更多算力资源。与之对应，研发费用结构也发生明显迁移：早期以人力成本为主，但随着模型规模扩大与训练频次提升，研发成本重心逐步转向算力与相关硬件消耗，整体呈现从人力驱动向算力驱动的结构变化。截至2025年6月30日，公司研发部门的员工数量为657名，占74.4%，销售和营销占16.4%，管理和行政占9.2%，体现出公司对核心技术人员的高度重视和对模型训练与迭代的巨大投入。

图 13: 2022-2025H1智谱研发费用结构



数据来源：公司招股说明书，广发证券发展研究中心

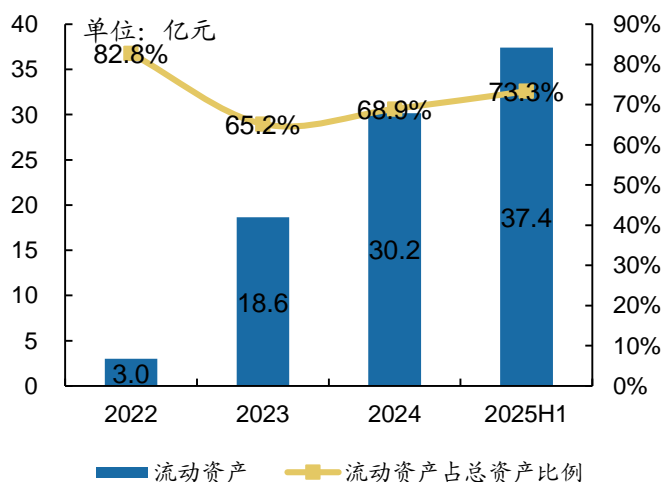
图 14: 智谱员工结构 (截至2025年6月30日)



数据来源：公司招股说明书，广发证券发展研究中心

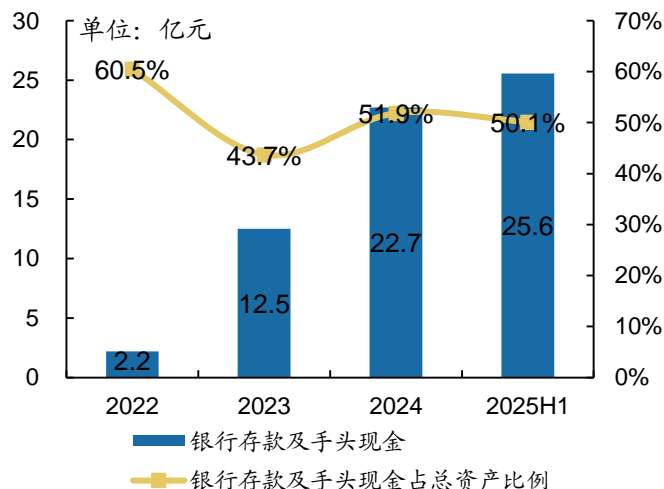
从2022年至2025H1，公司流动资产占总资产比例分别为82.8%、65.2%、68.9%和73.3%，体现出较强的资金流动性和轻资产运作模式；同时，银行存款及手头现金占总资产比例分别为60.5%、43.7%、51.9%和50.1%，为公司的资金流动性安全提供了可靠保障。

图 15: 2022-2025H1智谱流动资产



数据来源：公司招股说明书，广发证券发展研究中心

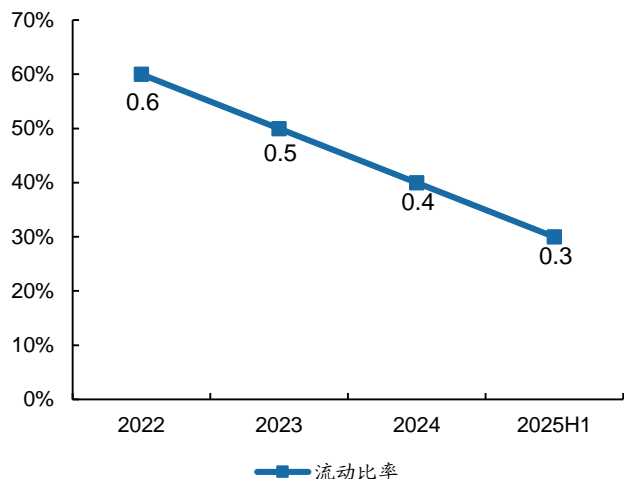
图 16: 2022-2025H1智谱银行存款及手头现金



数据来源：公司招股说明书，广发证券发展研究中心

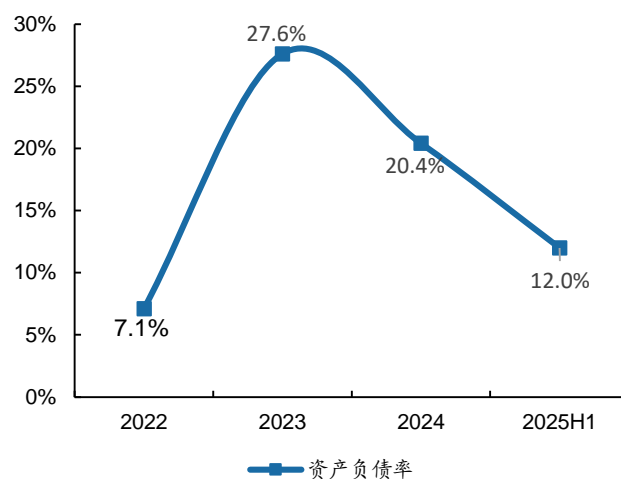
从偿债能力看，公司流动比率从2022年的0.6逐年下降至2025H1的0.3，主要系公司向投资者发行的金融工具增加流动负债净额所致；可见公司具有一定的短期偿债压力。资产负债率从2022年到2025H1分别为7.1%、27.6%、20.4%和12.0%，整体较低且呈回落趋势，说明公司总体杠杆水平可控，长期偿债压力较低。

图 17: 2022-2025H1智谱流动比率



数据来源：公司招股说明书，广发证券发展研究中心

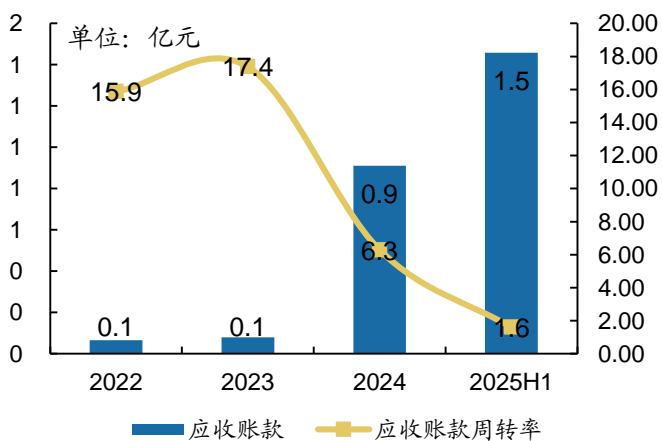
图 18: 2022-2025H1智谱资产负债率



数据来源：公司招股说明书，广发证券发展研究中心

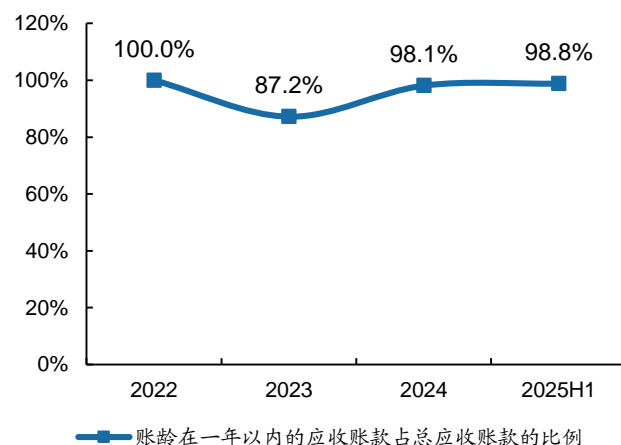
从2022年至2025H1,公司应收账款余额分别为0.1亿元、0.1亿元、0.9亿元和1.5亿元，应收账款周转率分别为15.9、17.4、6.3和1.6。应收账款的大幅上升与收入规模扩大相关。应收账款周转率在2022-2023年保持稳定，2023年起，公司的客户群和销量有显著增长，由于新客户的付款差异较大，公司也对应扩大了服务组合，尤其是采用里程碑式付款的本地化部署业务。因此，客户的付款周期也有所拉长，故周转率在2024年和2025H1逐渐下降。另一方面，公司应收账款结构健康，从2022年到2025H1分别为100.0%、87.2%、98.1%和98.8%，体现出公司对应收账款进行了有效管理。

图 19: 2022-2025H1智谱应收账款和应收账款周转率



数据来源：公司招股说明书，广发证券发展研究中心

图 20: 2022-2025H1智谱账龄在一年以内的应收账款占总应收账款的比例

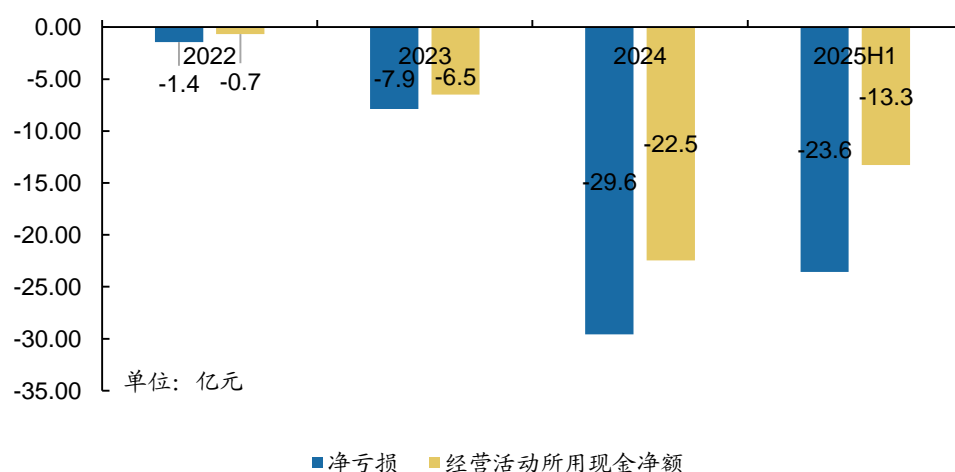


数据来源：公司招股说明书，广发证券发展研究中心

公司亏损扩大，经营活动现金流流出增加。公司净亏损从2022年到2025H1分别为-1.4亿元、-7.9亿元、-29.6亿元和-23.6亿元，呈逐年扩大趋势。其主要原因为：（1）公司随着业务扩张，需要扩大研发团队，并从第三方及相关供应商采购更多计算服务，用于模型研发与迭代，因此研发费用大幅度增加；（2）为迅速把握新兴市场机遇，公司需要扩大销售和营销团队，并增加广告开支；（3）公司目前处于商业化初期，尚未形成规模效应，短期难以依靠经营端业务“自我造血”覆盖大量投入。

公司经营活动现金净额与净亏损基本保持同比例变化。从2022年到2025H1，经营活动现金净额分别为-0.7亿元、-6.5亿元、-22.5亿元和-13.3亿元，持续为负且流出扩大，这与公司处于研发初期与算力投入高峰期相匹配；意味着当前业务现金回收难以覆盖持续扩张的研发和算力投入，经营端自我造血能力欠缺，对外部融资与现金储备的依赖度较高，故需对公司的偿债能力和流动性指标进行持续关注。随着后续收入持续增长、技术和模型的迭代以及商业化程度加深，公司有望实现亏损收窄，并改善经营活动现金流情况。

图 21: 2022-2025H1智谱净亏损和经营活动所用现金



数据来源：公司招股说明书，广发证券发展研究中心

## 二、AI 大模型技术升级驱动应用场景拓展

### （一）AI 技术持续升级，应用场景不断拓展

人工智能技术在蓬勃发展中正经历从传统判别式AI到大语言模型的重要转变。传统判别式AI侧重于识别和判断任务，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。大语言模型则是非常庞大的深度学习模型，其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，这使得其能够理解及生成自然语言以及其他类型的内容，以执行广泛的业务。因此，其能够处理传统判别式AI方法难以应对的任务。

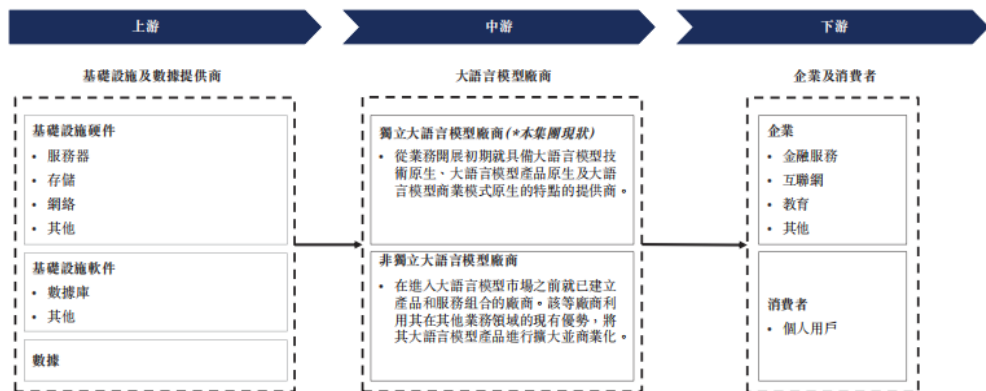
表 2: 大语言模型与判别式AI主要区别

区别	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。核心在于对上下文的理解输出符合语音逻辑的新内容。可根据外部环境的变化和设定目标自主行动。	侧重于将输入数据归入预设的类别，或基于标注数据进行预测。强调模型对已有知识的识别与归纳能力。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、语音识别及媒体内容推荐
算力	需要高性能的 GPU 或专用 AI 芯片，并需要优化硬件架构以支持大规模训练和推理。尤其是在训练阶段，需要大规模分布式计算机群。	通常参数量较小，可以使用通用 CPU 或 GPU 进行训练和推理。通常不需要大规模分布式计算机群。
算法	主要基于拥有数十亿或更多参数量的变换器架构以及强化学习技能	通常应逻辑回归、支持向量机、决策树、随机森林以及较小的神经网络等技术。
通用 2B 服务	主要依靠大规模非结构化数据（尤其预训练阶段），对标注数据依赖较低（主要在后期微调阶段）	高度依赖结构化、高质量的标注数据

数据来源：公司招股说明书，广发证券发展研究中心

当前人工智能正处于从狭义人工智能（Artificial Narrow Intelligence, ANI）迈向通用人工智能（Artificial General Intelligence, AGI）的过渡阶段，大语言模型正作为这一转变核心，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，打破了传统判别式AI应用场景割裂的局限，吸引众多行业参与到大语言模型应用市场。在大语言模型行业价值链中的主要参与者包括上游参与者（基础设施及数据提供商）、中游参与者（以独立大语言模型厂商以及非独立大语言模型厂商为代表）及下游参与者（包括企业及个人消费者）。

图 22: 大语言模型行业价值链示意图



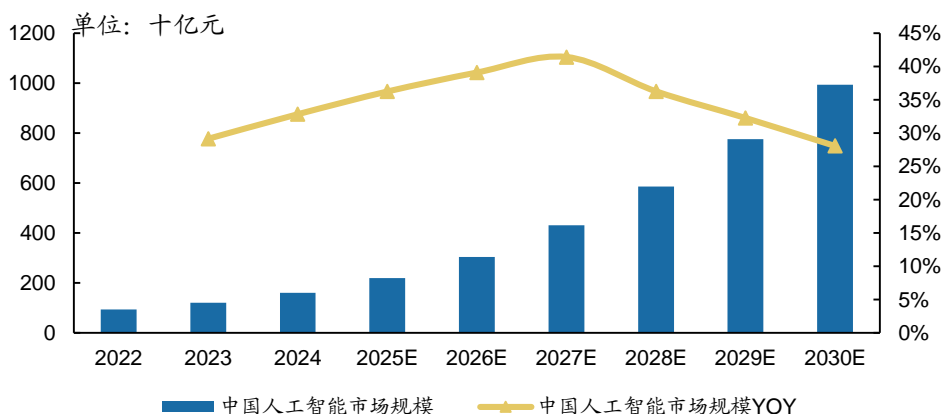
数据来源: 公司招股书, 广发证券发展研究中心

## (二) AI 大模型应用场景广阔, 市场规模快速增长

**人工智能市场快速增长。**人工智能在赋能其他数字技术、实现智能化和自动化、创造新的商业模式和产业、物理与数字世界融合方面扮演了不可替代的角色。人工智能的发展正深刻地改变经济增长、企业运行以及人类生活, 预计至2030年, 人工智能将赋能全球至少20%的日常商业决策, 为全球至少80%的消费者主流智能设备提供支持, 创造超过20万亿美元的AI影响的经济。

根据弗若斯特沙利文统计, 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, CAGR达到31.0%。2030年中国人工智能市场规模将增至9930亿元, CAGR达到35.5%。

图 23: 全球大模型市场规模预测



数据来源: 灼识咨询, 广发证券发展研究中心

大语言模型的商业化设计面向企业客户提供模型能力、工具链支持及训练或调优服务等, 以及面向个人用户提供AIGC服务。商业化进展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，但用户对大模型的付费意愿仍处于较低水平。相比之下，企业级场景是大语言模型市场增长的主要驱动因素。企业级用户对大语言模型部署有明确且个性化需求，表现出更强支付意愿与能力。对不同机构用户的服务场景主要体现在：

(1) 客户服务：基于大语言模型的聊天机器人及虚拟智能体可根据客户档案及交互历史已提供个性化沟通；

(2) 营销及销售活动：实现营销内容自动生成及个性化信息推送，提升参与度及转化率；

(3) 生产力及工作流程自动化：大语言模型嵌入生产力软件及工具，为邮件起草、会议记录生成及文件加工等任务提供支持，从而减少日常工作耗时。

**企业级大语言模型市场商业化路径呈现多元化格局。**目前企业级大语言模型的商业化模式主要包括：云端部署以及本地化部署两种模式，分别满足企业在成本、数据安全与业务实时性方面的差异化需求。云端部署解决方案提供由大语言模型厂商通过云向企业客户提供的大语言模型工具链、应用开发平台及模型服务。本地化部署解决方案则是将企业客户的定制化软件与服务安装在企业自身IT环境中，使机构能在自有基础设施内构建、训练以及使用针对自身需求定制的大语言模型。

**表 3：云端部署与本地化部署主要区别**

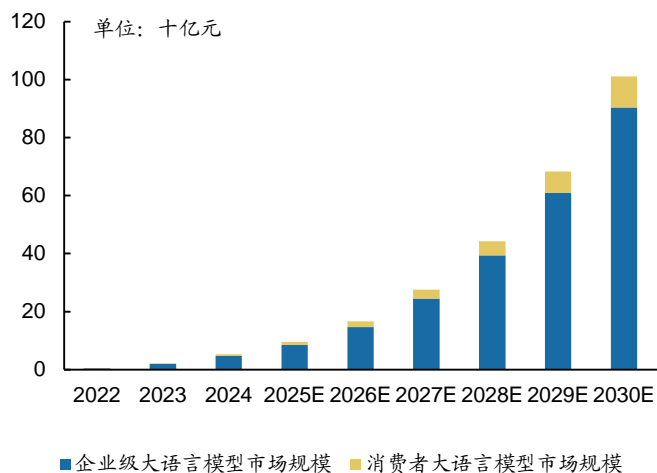
区别	大语言模型	判别式 AI
定价模式	订阅制/按调用量计费 嵌入软件授权/硬件绑定	基于向客户提供的具体产品及服务
典型客户类型	软件应用服务商、智能硬件厂商及中小企业	大型企业
核心特点	标准化、借入便捷、按需计费及高度灵活性 适合轻量化部署和快速试用	可根据具体需求进行定制、数据可控性好及安全性高 适合对数据隐私和系统稳定性要求高的行业
应用举例	其企业通过 API 接口接入通用大语言模型能力	企业私有化部署大语言模型以构建专有模型系统

数据来源：公司招股说明书，广发证券发展研究中心

根据弗若斯特沙利文统计，以收入计，2024年中国大语言模型市场规模已达到53亿元，其中企业客户贡献47亿元，个人客户贡献6亿元。预计2030年该市场规模将增至1011亿元，CAGR达到63.5%，企业客户仍将作为核心增长驱动力。

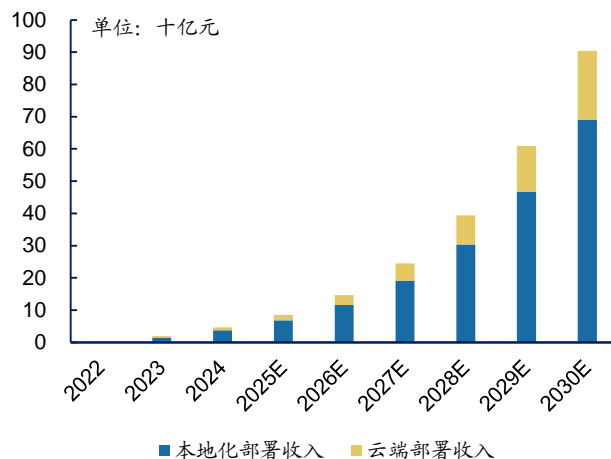
根据公司招股说明书，具体到企业级，2024年中国企业级大语言模型市场规模达到47亿元，其中云端部署市场规模9亿元，本地化部署市场规模38亿元。到2030年，中国企业级大语言模型市场规模预计将增至904亿元，其中云端部署市场规模预计为213亿元，本地化部署市场规模预计为691亿元。同期内云端部署市场规模CAGR达到69.4%，超过整体大语言模型复合年增长率。

图 24: 中国大语言市场模型规模



数据来源: 弗若斯特沙利文, 广发证券发展研究中心

图 25: 中国企业级大语言市场模型规模



数据来源: 弗若斯特沙利文, 广发证券发展研究中心

### (三) AI 大模型行业竞争激烈, 公司模型功能全面且具有中立性特点

中国大语言模型市场参与者可分为独立提供商及非独立提供商。独立提供商从业务开展初期就具备大语言欧兴技能原生、大语言模型产品原生及大语言模型商业模式原生的特点, 其自身业务在产业生态中通常不与其客户所运营的业务存在竞争关系。相较而言, 非独立提供商在进入大语言模型市场之前就已建立产品和服务组合, 并利用其其他业务领域的现有优势, 将其大语言模型产品进行扩大并商业化。以2024年收入计, 中国大模型厂商前五名中有四家为非独立提供商, 仅智谱一家独立提供商, 市占率位居整体第二名。

表 4: 中国头部大语言模型厂商排名 (2024年)

公司	排名	类型	收入 (十亿元)	市占率
公司 A	1	非独立	0.44	9.4%
智谱	2	独立	0.31	6.6%
公司 B	3	非独立	0.30	6.4%
公司 C	4	非独立	0.29	6.1%
公司 D	5	非独立	0.22	4.7%

数据来源: 公司招股说明书, 广发证券发展研究中心

在服务平台方面, 行业厂商提供的主要包括PaaS以及MaaS平台。PaaS为开发人员提供包括中间件、数据库及API的环境, 主要用户为开发人员或企业内部开发团队。MaaS平台在更高抽象层级上运作, 尤其专注于提供预训练大模型及模型开发工具套件服务, 使企业及开发人员得以调用、微调及集成先进的AI模型。

当前国内头部公司中, 大部分厂商无法支持语言、代码、图像、视频、音频多模态的MaaS平台服务, 美国头部厂商公司E与公司F可以提供全模态MaaS平台服务支持,

但在进一步产品中的平台推广存在缺陷，智谱是唯一一家将MaaS平台同时运用于多模态模型训练以及GUI智能体、手机应用与计算机应用产品推广的头部玩家。

表 5: 中国头部大语言模型厂商的模型能力对比

公司	语言	代码生成	图像生成	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	计算机应用
智谱	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
公司 A	✓		✓		✓		✓				
公司 B	✓	✓	✓	✓	✓		✓				
公司 C	✓		✓			✓	✓				
公司 D	✓		✓		✓		✓				
公司 E	✓	✓	✓	✓	✓	✓	✓				
公司 F	✓	✓	✓	✓	✓	✓	✓				

数据来源：公司招股说明书，广发证券发展研究中心

### 三、公司模型布局全面，基座模型能力突出

智谱的模型产品矩阵完整，功能全面。智谱围绕深度思考、认知世界和工具使用三方面能力构建了完整的模型产品矩阵，涵盖基座模型、反思及沉思模型、多模态模型、AI agent 与代码模型五大类。具体来看：

(1) 基座模型：定位为高性能通用大语言模型。其旗舰产品GLM-4.7不仅具备强大的通用处理能力，同时还承担着各类专业模型的底层开发底座功能。

(2) 反思及沉思模型：基于基座模型进行搭建，其核心是通过生成答案前的“深度思考”机制，以此适配各类复杂推理类任务，代表产品包含 GLM-Z1系列。

(3) 多模态模型：具备处理和整合文本、图片、音频、视频等多类型信息的能力，已覆盖 GLM-Image（图片生成）、GLM-4.6V（视觉理解及推理）等模型。

(4) AI agent：以 AutoGLM 为核心基座产品，集成推理、规划与工具调用能力，可无需持续人工输入自主完成多步骤任务，延伸出企业级智能体CoCo等。

(5) 代码模型：以 CodeGeeX 为代表，主打自动生成代码功能，助力用户提升编程效率。

表 6：智谱的AI模型及智能体

人类能力	模型及智能体	
深度思考	反思及沉思模型	GLM-Z1(反思)
		GLM-Z1-Rumination (沉思)
认知世界	多模态模型	GLM-Image (图片生成)
		GLM-4.6V(视觉理解及推理)
		CogVideoX(视频生成)
		GLM-Realtime (实时视频通话)
		GLM-4-Voice (端到端语音模型)
工具使用	AI agent	AutoGLM (“从对话到执行” — 自 主任务完成智能体)
		AutoGLM — Rumination(“边想边 干” — 具有深度思考能力的自主任 务完成智能体)
	代码模型	CoCo (企业级智能体)
		CodeGeeX (代码编写)

数据来源：公司招股说明书，广发证券发展研究中心

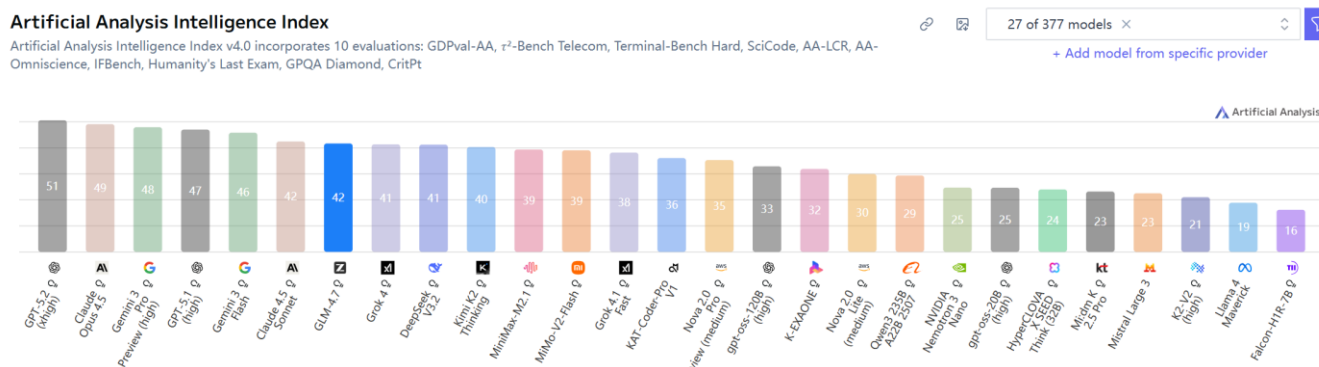
智谱全面的模型矩阵不仅可拓宽用户覆盖与商业落地空间，还能凭借统一技术底座实现性能协同提升。一方面，全面覆盖基座、推理、多模态、智能体及代码领域的模型布局，可满足B端金融、政务、教育等多元行业的定制化需求，同时可覆盖C端用户的日常交互、内容创作、编程辅助等场景，大幅拓宽用户覆盖边界与商业落地空间。另一方面，所有模型均以GLM系列旗舰基座模型为统一技术底座，能够实现底层技术能力的高效复用，不仅降低了各细分模型的适配难度，更可通过基座模型

的迭代升级，同步带动全矩阵模型性能的协同提升，进一步巩固公司在大模型赛道的核心竞争力。

### （一）GLM 基座模型各维度能力处于行业领先水平

2025 年 12 月智谱发布新一代旗舰模型 GLM-4.7，该模型作为新一代 AI 编程搭档实现编程、工具调用、思考模式等多维度能力升级。模型核心编程能力较 GLM-4.6 显著提升，主流基准测试中代码能力对齐 Claude Sonnet 4.5，同时优化思考模式。模型通过交错式、保留式、轮级思考提升指令遵循度、减少信息偏差。此外，模型的 UI 输出、工具调用及数学推理能力也同步进阶。截至 2026 年 1 月 13 日，GLM-4.7 在 Artificial Analysis 智能指数中位列全球开源模型第一、全球模型第七。

图 26：全球语言模型排行榜（截至2026年1月13号）



数据来源：Artificial Analysis，广发证券发展研究中心

**GLM-4.7相较前代模型性价比显著提升。**在内部定价体系上，GLM-4.7 与 GLM-4.6、GLM-4.5 保持完全一致，均按输入输出 token 长度分档计费。核心优势在于 GLM-4.7 与 GLM-4.6 同步将长文本输入支持上限提升至 200K tokens，相较 GLM-4.5 的 128K tokens 显著扩容，在同等使用成本下可适配更复杂的长文本处理需求。与海外旗舰模型相比，GLM-4.7较 Claude Opus 4.5 等产品成本大幅降低，凸显国内大模型的价格优势。与国内同档位模型对比，其定价高于字节跳动 Doubao-Seed-1.8，与 MiniMax M2 基本持平，低于阿里 Qwen3-Max 及月之暗面 Kimi K2 Thinking。

表 7：主流大语言模型 API 定价对比表

公司	模型	发布日期	上下文长度 (tokens)	API 价格 (每百万 tokens)
智谱华章	GLM-4.7	12 月 22 日	200K	输入 ≥ 2 元 输出 ≥ 8 元
	GLM-4.6	9 月 30 日	200K	输入 ≥ 2 元 输出 ≥ 8 元
	GLM-4.5	7 月 28 日	128k	输入 ≥ 2 元 输出 ≥ 8 元

MiniMax	MiniMax M2	12月23日	205K	输入 0.3 美金/2.1 元人民币 输出 1.2 美金/8.4 元
字节跳动	Doubao-Seed-1.8	12月18日	256K	输入 ≥ 0.8 元 输出 ≥ 2.0 元
阿里	Qwen3-Max	9月24日	256K	输入 ≥ 2.5 元 输出 ≥ 10.0 元
DeepSeek	DeepSeek-V3.2	12月1日	128K	输入 2 元 输出 3 元
月之暗面	Kimi K2 Thinking	11月6日	256K	输入 4 元 输出 16 元
Anthropic	Claude Opus 4.5	11月25日	200K	输入 5 美金 输出 25 美金
OpenAI	GPT-5.2	12月11日	400K	输入 1.75 美金 输出 14 美金
Google	Gemini 3 Pro	11月18日	1M	输入 ≥ 2 美金 输出 ≥ 12 美金
xAI	Grok 4.1	11月17日	2M	输入 0.2 美金 输出 0.5 美金

数据来源：各公司官网，广发证券发展研究中心

**GLM Coding Plan有望成为智谱云端部署业务的核心增长点。**该 AI 编码订阅套餐于 2025 年 9 月 1 日上线，面向全类型个人开发者，以每月最低 20 元的普惠价即可在主流 AI 编码工具中提供 GLM-4.7 等旗舰模型的在线使用服务。其定价仅为 Claude Code 的 1/7 且用量达 3 倍。上线四个月内，其主力模型在 Artificial Analysis Index、Code Arena 等多个榜单斩获开源、国产及全球第一，全球付费用户超 15 万、覆盖 184 个国家，用量最高用户发起 0.2 亿次请求，消耗 token 价值相当于 70 万元的宝马 X5，同时吸引超 3.7 万用户参与“拼好模”活动。该产品通过云端提供模型能力，未来有望成为带动云端业务增长的核心引擎。

表 8: GLM Coding Plan 订阅套餐详情

	GLM Coding Lite	GLM Coding Pro	GLM Coding Max
原价	40 元/月	200 元/月	400 元/月
首月优惠价	20 元/月	100 元/月	200 元/月
用量标准	120 prompts/5 小时, Claude Pro 套餐的 3 倍用量	600 prompts/5 小时, Claude Pro 套餐的 15 倍用量	2400 prompts/5 小时, Claude Pro 套餐的 60 倍用量
使用用户	处理轻量级工作负载的个人开发者	处理复杂工作负载的个人开发者	处理海量工作负载的个人开发者
核心权益	1. 模型支持 GLM-4.7、GLM-4.6、GLM-4.5 和 GLM-4.5-Air; 2. 适用于 Claude Code 等 20+ 编程工具; 3. 赠“智谱 AI 输入法”，畅享 Voice Coding; 4. 订阅期内，享受同等级最新模型更新	1. 享受 Lite 套餐所有权益; 2. 生成速度高于 Lite; 3. 视觉理解、联网搜索/读取、开源仓库 MCP	1. 享受 Pro 套餐所有权益; 2. 用量高峰优先保障; 3. 新功能抢先体验

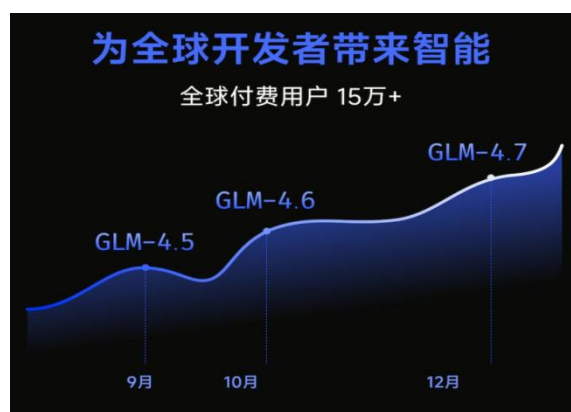
数据来源：公司官网，广发证券发展研究中心

图 27: GLM Coding Plan 全球用户使用数据展示



数据来源：公司官方微信公众号，广发证券发展研究中心

图 28: GLM Coding Plan 付费用户增长趋势



数据来源：公司官方微信公众号，广发证券发展研究中心

GLM 基座模型已深度渗透办公、招聘、快消、金融物流、学术研究、汽车制造和代码等多个领域，为各行业客户提供智能内容生成、任务协作与业务流程优化的核心技术支撑。例如，金山办公将 GLM 模型集成至 WPS AI，强化内容生成质量与跨产品适配，实现大纲、幻灯片全流程自主生成并降低运维成本。智联招聘依托该模型打造 AI 招聘助手，高效完成招聘需求理解、简历筛选及简历润色，提升人岗匹配精准度。

表 9: GLM 基座模型商业化案例

公司	所在行业	使用案例
金山办公	办公软件	基于 GLM 旗舰基座模型赋能 WPS AI，提升内容生成质量与跨产品集成能力，支持主题到大纲、幻灯片的全流程自主生成及格式优化，解决原有智能化不足、可扩展性弱的问题，降低运维成本
智联招聘	人力资源	依托 GLM 旗舰基座模型打造 AI 招聘助手，通过对话式交互精准理解雇主招聘需求，高效筛选简历、提供针对性建议及面试后评估；同时为求职者生成、润色专业简历，提升人岗匹配精准度与招聘效率
蒙牛乳业	快消	以 GLM 旗舰基座模型为核心开发 AI 营养师“蒙蒙”，支持自然语言交互，提供专家级个性化营养咨询服务，搭载 AI 健康规划师功能，可生成定制化健康营养计划，实现进度追踪、动态调整及共情激励
综合性企业	金融物流	凭借 GLM 旗舰基座模型提供多场景集成解决方案：财务端支撑金融新规问答、数据查询分析与数据治理，自动化率超 50%；物流端赋能订单管理、智能客服、图像识别等，提升全链路运营效能；同步升级办公协同系统，构建“知识库 + 工作流程 + 智能体”一体化平台
AMiner	学术研究	由 GLM 旗舰基座模型驱动 AI 研究助手，依托海量学术论文数据库，提供自然语言学术检索、学者画像解析、文献自动摘要、交互式问答及综述撰写辅助服务，支持学术趋势个性化订阅与语言润色
高端纯电汽车制造商	汽车制造	基于 GLM 旗舰基座模型升级智能座舱系统，通过多轮引导式对话提升口语化指令理解精度，具备情感智能与对话风格动态切换能力，可生成即兴娱乐内容、支持语音控制游戏，实现从指令式交互到人性化交流的转型

数据来源：公司招股说明书，广发证券发展研究中心

## （二）公司的多模态大模型覆盖语音、图像、视频等多个领域

基于GLM系列基座模型，公司开发了多款多模态模型，具体包括：

（1）**GLM-4.6V**：具备原生视觉推理能力的模型，支持长文档、多页图像及连续视频理解，新增工具调用与多模态任务自动化功能，在复杂任务中性能稳定。

（2）**GLM-Image**：与华为联合开源的图像生成模型，采用独创的“自回归 + 扩散解码器”混合架构，是首个在国产芯片上完成全程训练的 SOTA 多模态模型。

（3）**CogVideoX**：高性能视频生成模型，支持 4K 超高清 60 帧输出，可精准解析复杂提示，兼容多元艺术风格，支持多通道同步生成视频。

（4）**GLM-Realtime**：实时音视频交互模型，提供流畅 AI 音视频通话，具备两分钟长上下文记忆、歌唱及工具调用能力，适配智能硬件集成。

（5）**GLM-4-Voice**：端到端语音模型，无需文本中转，支持多语言交互，可按需调整语音情感、语调等属性，交互性能优异。

表 10：智谱多模态模型功能与应用场景

产品名	功能	应用场景案例
GLM-4.6V	高效视觉理解与推理	跨页表格重建与长文档分析、电商场景看图比价与导购、网页前端精准复刻与交互修改。
GLM-Image	高质量图像生成	创意设计领域：辅助生成插画、漫画等视觉内容；基于用户需求生成配图，助力社交媒体内容创作
CogVideoX	高清晰度视频生成	短视频创作：赋能捏 Ta 平台实现文本到完整短视频的自动生成，涵盖视觉、旁白、配乐
GLM-Realtime	实时音视频交互	智能硬件：赋能消费电子厂商打造移动 AI Agent，实现音视频通话实时解读、自然语言操控设备
GLM-4-Voice	端到端多语音处理与交互	1. 智能座舱：帮助某高端智能纯电汽车的汽车制造商实现高情商语音交互，对话风格个性化定制等 2. 健康服务：为蒙牛“蒙蒙”AI 营养师提供自然流畅的语音交互能力

数据来源：公司招股说明书，广发证券发展研究中心

**GLM-4.6V 是一款具备同尺寸模型领先水平的视觉推理模型**，可高效处理长文档、多页图像及连续视频等多元视觉输入任务。2025年12月9日智谱正式推出并开源 GLM-4.6V，该系列在 GLM-4.5V 基础上全面升级，延续 106B/12B MOE 架构和 SOTA 视觉理解能力，新增 Function Call、多模态任务自动化，并将上下文扩展至 128k，支持长文档、多页图像和连续视频，兼顾推理深度与效率，同时在多项视觉 Benchmark 与真实业务场景中表现出更高稳定性与实用价值。与 Qwen3-VL 相比，GLM-4.6V 在中文视觉理解、复杂推理与 OCR 任务上表现优异，在实际部署中具备一定实用价值。

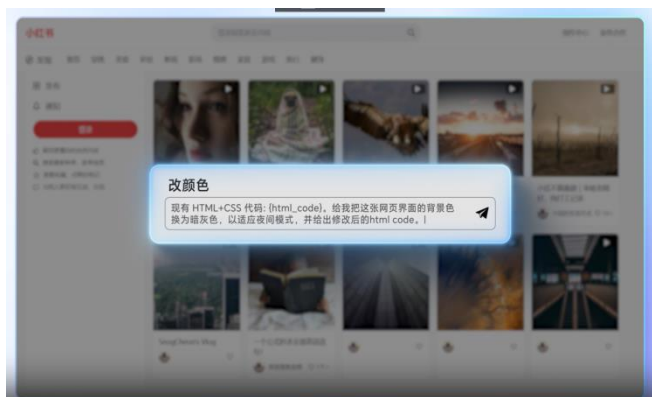
图 29: GLM-4.6V 与同类模型的性能对比

Open-source LLMs Benchmarks		GLM-4.6V	GLM-4.6V-Flash	GLM-4.5V	GLM-4.1V-9B-Thinking	Owen3-VL-8B-Thinking	Owen3-VL-235B-A22B-Thinking	Kimi-VL-A3B-Thinking-2506	Step3 32B
		104B (A22B)	9B	104B (A22B)	9B	8B	235B (A22B)	16B (A3B)	327B (A3B)
General VQA	MMBench V1.1	88.8	86.9	88.2	85.8	84.3*	90.6	84.4	81.1*
	MMBench V1.1 (CN)	88.2	85.9	88.3	84.7	83.3*	87.2*	80.7*	81.5*
	MMStar	75.9	74.7	75.3	72.9	75.3	78.7	70.4	69.0*
	BLINK (Val)	65.5	65.5	65.3	65.1	64.7	67.1	53.5*	62.7*
Multimodal Reasoning	MURBENCH	77.1	75.7	75.3	74.7	76.8	80.1	63.8*	75.0*
	MMMU (Val)	76.0	71.1	75.4	68.0	74.1	80.6	64.0	74.2
	MMMU_Pro	66.0	60.6	65.2	57.1	60.4	69.3	46.3	58.6
	VideoMMU	74.7	70.1	72.4	61.0	72.8	80.0	65.2	/
	MathVista	85.2	82.7	84.6	80.7	81.4	85.8	80.1	79.2*
	AIZD	88.8	89.2	88.1	87.9	84.9	89.2	81.9*	83.7*
	DynaMath	54.5	43.7	53.9	42.5	41.1*	56.5*	28.1*	50.1
	WebMath	69.8	60.0	68.8	63.8	66.5*	74.5*	42.0*	59.8
Multimodal Agentic	ZeroBench (sub)	25.8	22.5	23.4	19.2	-	27.7	16.2*	23.0
	MMBrowseComp	7.6	7.1	/	/	6.9*	6.7*	/	/
	Design2Code	88.6	69.8	82.2	64.7	56.6*	93.4	38.8*	34.1*
	Flame-React-Eval	86.3	78.8	82.5	72.5	56.3*	73.8*	36.3*	63.8*
	OSWorld	37.2	21.1	35.8	14.9	33.9	38.1	8.2	/
	AndroidWorld	57.0	42.7	57.0	41.7	50.0	/	/	/
	WebNoyager	81.0	71.8	84.4	69.0	47.7*	30.4*	/	/
	Webquest-SingleQA	79.5	75.1	76.9	72.1	70.7*	76.4*	35.6*	58.7*
	Webquest-MultiQA	59.0	53.4	60.6	54.7	60.3*	17.3*	11.1*	52.8*
	MMLongBench-Doc	54.9	53.0	44.7	42.4	48.0	56.2	42.1	31.8*
Multimodal Long Context	MMLongBench-128K	64.1	63.4	/	/	49.5*	61.8*	/	/
	LVBench	59.5	49.5	53.8	44.0	55.8	63.6	47.6*	/
OCR & Chart	OCRBench	86.5	84.7	86.5	84.2	81.9	87.5	86.9	83.7*
	OCR-Bench_v2 (EN)	65.1	63.5	60.8	57.4	63.9	66.8	/	/
	OCR-Bench_v2 (CN)	59.6	59.5	59.0	54.6	59.2	63.5	/	/
	ChartQAPro	65.5	62.6	64.0	59.5	58.4*	63.6*	23.7*	56.4*
	ChartMuseum	58.4	49.8	55.3	48.8	46.7*	/	33.6*	40.0*
	ChartViz_Vis-Reasoning	63.2	59.6	58.4	55.0	53.0	66.1	39.6*	/
Spatial & Grounding	OmniSpatial	52.0	50.6	51.0	47.7	51.3*	-	37.3*	47.0*
	RefCOCO-avg (val)	88.6	85.6	91.3	85.3	89.3*	92.4	33.6*	20.2*
	TreeBench	51.4	45.7	50.1	37.5	34.3*	50.9	41.5*	41.3*
	Ref-L4-test	88.9	87.7	89.5	86.8	88.6*	90.4	51.3*	12.2*

数据来源: 公司官网, 广发证券发展研究中心

GLM-4.6V不仅具备原生多模态工具调用能力的核心优势,其性价比也实现了大幅提升。GLM-4.6V无需中间转换,可直接以图像、截图、文档等作为工具参数输入并理解其视觉返回结果,形成“感知-理解-执行”闭环。同时,其前端复刻能力可实现像素级识别并生成高保真代码,支持如精准复刻小红书前端等自然语言交互式修改。网页搜索方面,模型擅长复杂文档处理及电商场景下“看图-比价-生成导购清单”的全链路任务。在实现上述功能跃升的同时, GLM-4.6V输入价格为1元/百万tokens,输出价格为3元/百万tokens,较前代下降50%,API调用成本显著降低,性价比大幅提升。此外,同时推出的GLM-4.6V-Flash轻量版也兼顾了成本与便捷性。

图 30: GLM-4.6V 前端复刻能力案例: 小红书界面像素级复刻



数据来源: 公司官网, 广发证券发展研究中心

图 31: GLM-4.6V 视觉网页搜索案例: 电商以图搜款及导购清单生成



数据来源: 公司官网, 广发证券发展研究中心

GLM-Image是智谱联合华为于2026年1月发布并开源的新一代图像生成模型，是首个在国产芯片上完成全流程训练的SOTA多模态模型。模型采用创新的“自回归+扩散解码器”混合架构，兼顾整体语义理解与细节刻画，在复杂知识型场景和文字渲染方面优势突出，在CVTG-2K、LongText-Bench等权威测评中达到开源模型最优水平，尤其擅长汉字生成等高要求任务。模型还基于昇腾NPU和昇思MindSpore AI框架完成了从数据预处理到大规模训练的全流程构建，验证了在国产全栈算力底座上训练高性能多模态生成模型的可行性。

GLM-Image在图像综合美学表现、复杂构图精准控制及长文本生成连贯性等维度上，相比Nano Banana Pro仍有较大成长空间。Nano Banana Pro凭借更成熟的训练体系，在视觉质感、艺术表现力与多元素协同生成等方面仍具优势，尤其在需要高度风格一致性与复杂语义理解的场景中体现出更强的综合生成能力。尽管GLM-Image已在文字渲染与知识型场景中表现突出，但在整体图像质感、构图合理性及长提示词下的生成稳定性方面，仍具备持续优化的空间。

图 32: GLM-Image 与同类模型的综合性能对比

Model	open source	OneIG-Bench-EN (overall)	OneIG-Bench-ZH (overall)	TIIF-Bench		DPG-Bench (overall)
				short (overall)	long (overall)	
Seedream 4.5	X	0.576	0.551	90.49	88.52	88.63
Seedream 4.0	X	0.576	0.553	90.45	88.08	88.54
Nano Banana 2.0	X	0.578	0.567	91.00	88.26	87.16
GPT Image 1 [High]	X	0.533	0.474	89.15	88.29	85.15
DALL-E 3	X	-	-	74.96	70.81	83.50
Qwen-Image	✓	0.539	0.548	86.14	86.83	88.32
Qwen-Image-2512	✓	0.530	0.515	83.24	84.93	87.20
Z-Image	✓	0.546	0.535	80.20	83.01	88.14
Z-Image-Turbo	✓	0.528	0.507	77.73	80.05	84.86
FLUX.1 [Dev]	✓	0.434	-	71.09	71.78	83.52
SD3 Medium	✓	-	-	67.46	66.09	84.08
SD XL	✓	0.316	-	54.96	42.13	74.65
BAGEL	✓	0.361	0.370	71.50	71.70	-
Janus-Pro	✓	0.267	0.240	66.50	65.01	84.19
Show-o2	✓	0.308	-	59.72	58.86	-
GLM-Image	✓	0.528	0.511	81.01	81.02	84.78

数据来源：公司官网，广发证券发展研究中心

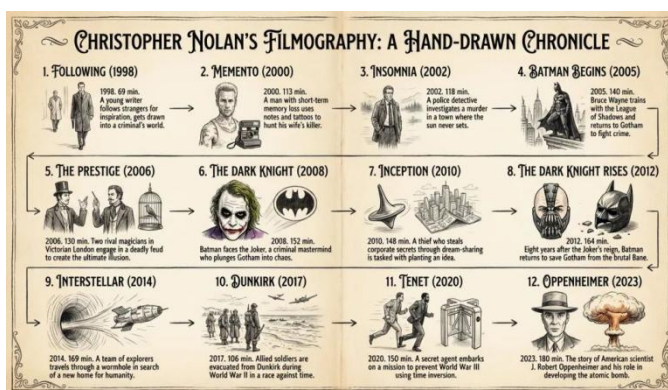
Nano Banana Pro相比GLM-Image在文字准确性、信息完整性与风格排版美观度上具有明显优势。以“文生图-影视作品年表”测评为例，Nano Banana Pro生成的图片在信息准确性、艺术风格与排版布局上均表现突出，能够完整、美观地呈现克里斯托弗诺兰全部12部电影的上映时间、片长与剧情简介；GLM-Image生成的图片则存在明显的文字错误、信息混乱与排版失调问题，如电影名称、年份、剧情介绍均出现多处不匹配和杂乱排列，未能形成清晰可读的作品年表。

图 33: GLM-Image 绘制导演电影作品年表效果图



数据来源: GLM-Image 产品, 广发证券发展研究中心

图 34: Nano Banana Pro 绘制导演电影作品年表效果图



数据来源: Nano Banana Pro 产品, 广发证券发展研究中心

在定价策略上, GLM-Image 相比 Nano Banana Pro 展现出显著的成本优势。GLM-Image 采用简单透明的按次计费, 每次生成仅需 0.1 元, 而 Nano Banana Pro 则需根据输入 Token 和输出图片分辨率进行混合计费, 其单张图片生成成本 (1K/2K 图片 0.134 美元) 远高于 GLM-Image。这种差异使得 GLM-Image 延续了中国大模型一贯的低成本优势。

表 11: Nano Banana Pro 与 GLM-Image 大模型价格对比表

	GLM-Image	Nano Banana Pro
定价模式	按次计费	按 Tokens+图片数计费
标准价格	0.1 元 / 次 (0.014 美元/次)	输入: 文字 2 美元/百万Tokens; 图片 0.0011 美元/百万Tokens 输出: 1K/2K 图片 0.134 美元/张; 4K 图片 0.24 美元/张

数据来源: 谷歌官网, 智谱官网, 广发证券发展研究中心

公司布局多模态模型, 核心源于向 AGI 终极目标的推进与商业客户拓展的双重考量。一方面, 正如智谱 CEO 张鹏在采访中所言, 公司的唯一目标就是 AGI (通用人工智能), 而多模态能力是构成完整智能的必备要素, 因此公司会在该领域持续深耕。另一方面, 从客户与市场拓展角度, 多模态模型的布局既能拓展影视、传媒、创意设计等 B 端行业客户, 精准匹配其智能化转型需求, 也能在 B 端业务根基稳固后, 依托模型对消费级场景的适配能力切入 C 端市场, 将 C 端业务打造为第二成长曲线。

多模态模型凭借文本、图像、音视频跨模态融合处理与生成能力, 赋能多行业内容创作与交互体验升级。例如, 全球领先消费电子生产商依托 GLM-Realtime 模型实现端侧 AI 音视频通话智能交互, 同时结合多模态的图像理解和生成能力打造移动 AI Agent, 完成社交媒体内容自动化创作。

**表 12: 智谱多模态模型商业化案例**

公司或产品	所在行业	使用案例
全球领先消费电子生产商	消费电子	依托智谱 GLM-Realtime 模型实现端侧 AI 音视频通话功能, 可实时解读音视频画面、识别屏幕文本与视觉细节, 通过流式推理保障超低延迟响应, 实现流畅自然的互动沟通; 同时基于多模态模型的图像理解与文本生成能力打造移动 AI Agent, 支持选取相册图片自动创作社交媒体文案
捏 Ta	AIGC 内容创作	依托智谱多模态模型的文本 - 视觉 - 音频融合生成能力推出智能短视频生成智能体, 支持自然语言输入转化为含视觉效果、旁白及背景音乐的完整短视频, 降低普通用户的创作技术门槛
智谱清言	通用智能交互	由智谱多模态模型的图像生成、视频理解与语音解说能力驱动, 支持图像、视频生成, 覆盖旅行场景景点实时解说与环境识别、艺术设计领域原创作品创作等多元需求

数据来源: 公司招股说明书, 广发证券发展研究中心

### (三) 面向 B 端和 C 端, 公司 AI agent 产品赋能各场景

AI Agent 作为智能交互与自主执行的核心载体, 是人工智能技术落地的重要形态。智谱针对不同市场需求, 围绕 To C、To B 双端场景布局两款核心 AI Agent 产品, 分别是面向消费端的 AutoGLM 与面向企业端的 CoCo。前者聚焦个人生活与办公场景的智能任务处理, 后者专攻企业业务流程的自动化提效, 共同形成覆盖个人与企业的全场景智能服务矩阵。

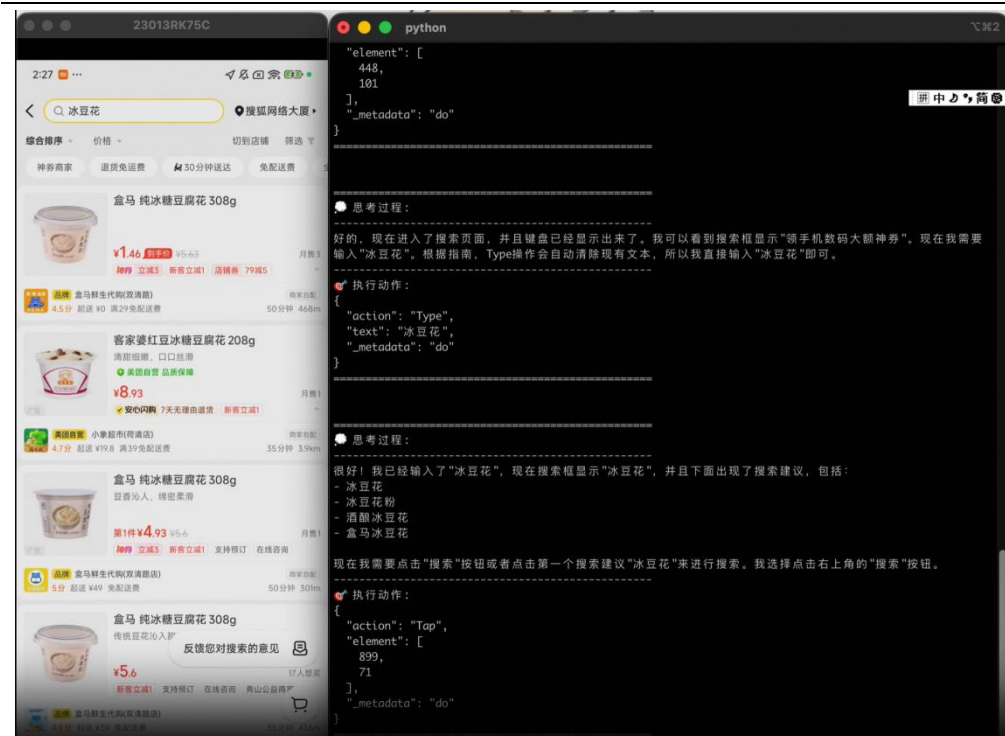
**表 13: 智谱 Agent 产品一览**

Agent 产品	功能	应用场景	案例
AutoGLM	模拟人类操作多类移动应用及网站, 自主完成指定任务	消费端生活与办公场景	全球领先消费电子生产商基于其打造移动 AI Agent, 实现自然语言操控设备、社交媒体内容自动化生成
CoCo	安全调度企业资源, 一键自动化复杂工作流, 适配企业系统集成	企业端流程提效与办公协同	中国综合性企业依托 CoCo 构建“知识库 + 工作流程 + 智能体”一体化办公系统, 提升内部管理效能

数据来源: 公司招股说明书, 广发证券发展研究中心

**AutoGLM 已实现开源并展现突出技术与应用价值。**该智能体 2024 年 10 月 25 日发布, 2025 年 12 月 9 日以 Open-AutoGLM 开源, 提供完整技术底座及 50+ 场景 Demo, Hugging Face 开放轻量级模型支持多端部署, 截至 2026 年 1 月初 GitHub 仓库获 20.9K Star, 跻身 2025 年增速领先的大模型开源项目。其依托 GLM 系列模型, 结合端到端异步强化学习在 ComputerRL、MobileRL、AgentRL 方向实现技术突破, Device Use 基准测试表现优于多款主流 Agent, 具备精准理解规划、全场景执行等核心能力, 可覆盖点外卖、订机票和飞书操作等生活与办公场景。

图 35: AutoGLM 外卖选购样例展示



数据来源：公司官网，广发证券发展研究中心

CoCo 作为一款专为企业智能自动化设计的先进企业级 AI Agent，可深度赋能企业业务流程提效与价值创造。它具备安全访问并自主编排企业软件、知识库、数据库等关键资源的能力，依托先进个性化记忆系统实现复杂工作流一键式自动化，核心优势涵盖结果导向型交付、智能个性化支持、与企业架构无缝集成三大方面，既能确保任务产出可衡量成果，也能动态适配员工需求提供定制服务，还能融入现有系统整合企业数据与运营工具，助力企业提升业务绩效。

图 36: CoCo 政务解读示例



数据来源：CoCo 产品，广发证券发展研究中心

AI Agent 凭借自主任务执行、系统集成与个性化交互能力，深度落地消费端与企业端场景，以 AutoGLM、CoCo 等核心产品为技术载体，为各领域提供全链路智能解决方案。例如，全球领先消费电子生产商基于 AutoGLM 基座智能体打造移动 AI Agent，实现系统级函数调用与长上下文记忆；中国综合性企业依托 CoCo 企业级智能体升级办公协同，构建一体化智能工作平台。

表 14: 智谱AI Agent 商业化案例

公司或产品	所在行业	使用案例
全球领先消费电子生产商	消费电子	基于 AutoGLM 基座智能体开发移动 AI Agent，集成系统级函数调用、长上下文记忆能力，支持音视频实时交互与社交媒体内容自动化生成，实现自然语言操控设备功能
中国综合性企业	金融物流 / 办公协同	依托智谱 CoCo 企业级智能体构建“知识库 + 工作流程 + 智能体”一体化办公系统，搭配个人知识库工具，实现文档理解、创作与工作流程智能化，提升内部管理效能
智谱清言	通用智能服务	依托 GLM 系列模型支持用户自主创建专属智能体，提供无代码 AI Agent 定制平台。截至最后实际可行日期，平台活跃自定义 AI Agent 数量已突破 80 万，覆盖多元场景需求

数据来源：公司招股说明书，广发证券发展研究中心

2026 年1月15日，千问App上线400多项办事功能，正式从“聊天对话”迈入“AI 办事时代”。该产品依托Qwen旗舰模型与阿里生态，深度接入淘宝、支付宝、飞猪、高德等业务，全球首发点外卖、购物、订机票酒店等 AI 购物功能并开放全用户测试。其核心优势在于全流程闭环服务，支持在对话界面内完成下单支付，无需跳转应用，同时覆盖50项政务民生服务与“找卷子、讲难题”等学习功能，定向邀测的“任务助理”可处理应用开发、办公调研等复杂任务。根据千问App新品发布会，千问App上线两个月，其C端月度活跃用户突破1亿，未来有望继续上升。

图 37: 千问APP宣布全面接入阿里生态场景



数据来源：千问 APP 发布会，广发证券发展研究中心

**AutoGLM 与千问的核心差异在于技术路线与生态模式。**AutoGLM 采用基于视觉识别 (GUI) 的 Agent 路线，无需应用接口支持，可快速适配微信、淘宝、美团等

50+ 款主流中文应用，但对模型要求能力较高，存在流畅度不足，且易触发如同豆包手机被各大厂商围剿的风控问题。千问 APP 则采用基于 MCP 和 A2A 的混合协议架构，依托阿里生态实现系统级打通，解决了决策信任与支付断点痛点，但服务范围高度依赖自有生态，仅能接入淘宝、支付宝等5 个阿里系应用，对外第三方应用覆盖能力较弱。AutoGLM 未来破局需突破合规瓶颈，通过与平台达成官方合作、探索商业分成模式，平衡自身优势与生态适配。

**表 15: 智谱AI Agent 和千问APP的功能和生态对比**

Agent 产品	技术路线	可调用 App
AutoGLM	基于视觉识别 (GUI) 的 Agent 路线	微信、QQ、微博、淘宝、京东、拼多多、美团、饿了么、肯德基、携程、12306、滴滴出行、bilibili、抖音、爱奇艺、网易云音乐、QQ 音乐、喜马拉雅、大众点评、高德地图、百度地图、小红书、知乎、豆瓣等 50+ 款主流中文应用
千问 App	基于 MCP 和 A2A 的混合协议架构	接入淘宝、支付宝、淘宝闪购、飞猪、高德共 5 个阿里生态 App

数据来源：智谱公司官网，千问 APP 微信公众号，广发证券发展研究中心

以点外卖这一高频场景为例，AutoGLM 的平台模拟操作与千问 APP 的生态内高效闭环形成鲜明对比。AutoGLM 会直接进入美团等外卖 APP，通过模拟人类的搜索、点击等 GUI 操作完成下单，全程耗时约 1 分半至 3 分钟，其 GUI 操作模式不仅流程与真实用户行为完全一致，也支持多平台比价选择的可能性；千问 APP 则在自身独立界面内，依托 MCP 和 A2A 混合协议架构直接接入淘宝闪购外卖平台，无需跳转至外部 APP 即可完成下单，全程仅需约 1 分钟，速度更快且体验更流畅，但受限于阿里生态，仅能在淘宝闪购单一渠道下单。这种体验差异清晰印证了 GUI 路线的普适性与生态风险，以及协议对接路线的流畅性与生态局限性，也揭示了两类 Agent 在商业化落地的核心矛盾。

图 38: AutoGLM 模拟人类操作外卖 APP 界面



图 39: 千问 APP 独立界面点外卖交互



数据来源: AutoGLM 产品, 广发证券发展研究中心

数据来源: 千问 APP 发布会, 广发证券发展研究中心

#### (四) 公司重点布局 AI 辅助编程赛道, 模型和工具齐备

智谱于2024年7月推出的 CodeGeeX4-ALL-9B 模型, 是公司新一代开源代码生成模型的核心代表。该模型基于智谱自有的GLM-4-9B 通用大语言模型深度优化, 在百亿参数规模下展现出顶尖的性能水平。模型能够支持代码补全与生成、代码解释、联网搜索、函数调用、仓库级代码问答等多种功能。此外, 模型支持超过300种编程语言, 并能处理超长代码文本, 可有效应用于复杂项目开发。在实际应用中, 截至2025年6月30日, CodeGeeX4日均生成代码量已超过1亿行, 在开发效率提升方面获得开发者广泛认可。

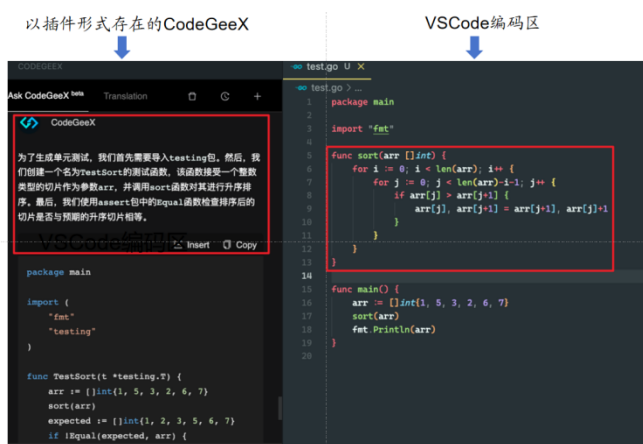
图 40: CodeGeeX4-ALL-9B 模型核心能力示意图



数据来源: CodeGeeX 官方使用手册, 广发证券发展研究中心

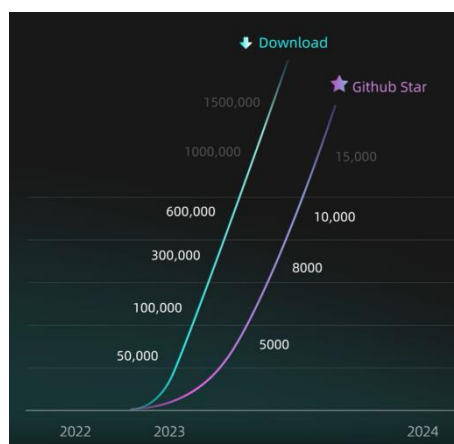
CodeGeeX是智谱于2022年9月推出的基于大模型的全能智能编程助手,以插件形式嵌入VS Code等主流IDE,通过侧边栏与开发者交互,并可外接GLM等多种主流大模型为编程全流程提供赋能,支持包括代码生成与补全、自动添加注释、代码翻译及智能问答等核心功能,兼容Python、Java、C++等上百种编程语言。CodeGeeX工具对个人用户完全免费,上线一年下载量已超60万次,开源模型在GitHub累计获得超过1万星标。

图 41: CodeGeeX在VS Code中辅助生成单元测试代码示例



数据来源: CodeGeeX 产品演示, 广发证券发展研究中心

图 42: 2022-2024 年 CodeGeeX 插件下载量与 GitHub Star 数增长趋势 (单位: 个)



数据来源: 公司官网, 广发证券发展研究中心

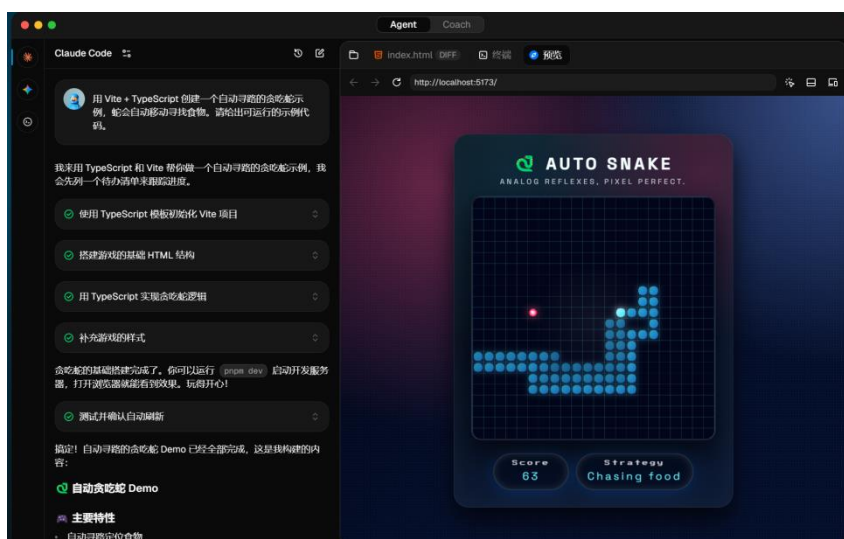
CodeGeeX 的功能在金融、医疗、汽车等行业广泛应用。以金融行业为例,某 500 人以上客户团队主打 Java、Python、JavaScript 等语言,需在编码生成、存量代码维护、研发问题检索等场景提效,CodeGeeX 凭借代码补全、生成能力提升编码效率,通过自动生成注释与代码解释优化存量代码维护效率,依托智能问答强化信息检索利用效率,还能通过代码修复、优化建议提升代码质量,提高单元测试及注

释覆盖率，助力代码贴合编码规范。

支持企业私有化部署解决方案：为了满足企业用户对代码数据安全和隐私的需求，CodeGeeX支持企业私有化部署的解决方案。企业可以将大模型部署在自己的服务器或私有云上，根据自身需求进行调整和优化。

**Z Code**是智谱于2025年12月推出的AI原生集成开发环境（IDE），旨在通过自然语言指令驱动AI Agent完成从编码、调试到预览的全流程开发任务。其具有一键切换 Gemini 3、 Claude Code等多个Agent等功能。核心特点在于完全通过Agent驱动执行，不支持手动代码编辑。其定位正从集成主流CLI工具的图形化界面，向支持多Agent协同的Vibe Coding平台演进。Z Code 客户端本身完全免费，但需用户自行配置智谱 GLM Coding Plan、 Claude Code、 OpenAI 等平台的 API 密钥或订阅来获取 AI 服务能力。

图 43: Z Code 自动生成贪吃蛇游戏演示



数据来源：Z Code 官网，广发证券发展研究中心

在AI辅助编程领域，智谱正构建覆盖“插件辅助”到“Agent驱动IDE”的全链路产品体系。其中，CodeGeeX作为开源编程插件，已积累超60万次下载量并支持GLM、GPT、Claude等多类模型；Z Code则定位为自然语言驱动的多Agent原生IDE，面向全流程自动化编程任务。与国内TRAE、通义灵码等工具相似，智谱旗下产品同样强调对国产模型的适配；相较于GitHub Copilot、Cursor等海外工具，其在中文场景适配与国产模型生态支持方面具备一定的差异基础。未来，智谱若能进一步发挥其在多模型支持与工具链完整性方面的特点，有望在国产AI编程生态中形成自身的商业化路径。

表 16: AI辅助编程工具对比

AI 辅助编程工具	产品功能	应用场景	开发公司	调用 AI 大模型
<b>CodeGeeX</b>	1. 智能编程助手插件, 支持代码生成/补全、注释添加、代码翻译、智能问答等 2. 支持超百种编程语言 (Python/Java/C++ 等) 3. 以侧边栏形式嵌入 VS Code 等主流 IDE 4. 对个人用户完全免费, 上线一年下载量超 60 万次	1. IDE 内的实时代码辅助 2. 多语言项目快速开发 3. 金融、医疗、汽车等企业私有化部署	智谱	GLM 系列
<b>ZCODE</b>	1. AI 原生 IDE, 通过自然语言驱动 Agent 完成编码、调试、预览全流程 2. 支持一键切换 Gemini 3、Claude Code 等多个 Agent 3. 完全 Agent 驱动, 不支持手动代码编辑 4. 客户端免费, 需自备 API 密钥	1. 全流程自动化编程任务 2. 多 Agent 协同复杂工程 3. 低代码/无代码开发场景	智谱	支持 GLM 系列、Claude Code、Gemini 系列等
<b>GitHub Copilot</b>	1. 基于 GPT 大模型的代码补全工具, 支持多语言 (如 Python、JavaScript 等) 2. 提供行级/函数级代码建议, 但中文理解能力较弱 3. 依赖插件形式, 无法自动解析整个项目代码库 4. 支持教育版免费试用吸引早期用户	1. IDE 内的实时代码辅助 2. 简单功能模块的快速实现 3. 依赖开源代码库的场景	微软	GPT 系列、Gemini 系列、Claude 系列、Grok 系列
<b>Cursor</b>	1. 基于 VS Code 的 AI 原生 IDE, 支持项目级代码库理解 (自动解析整个项目) 2. 多模态输入: 支持上传图片生成代码 3. 集成主流闭源大模型 API, 无需单独配置 4. 支持中文回复定制及小语种 5. 提供 Composer 功能实现跨文件代码修改	1. 复杂项目开发 (如全栈应用) 2. 多模态需求 (如设计图转代码) 3. 团队协作中的高效代码生成与调试	Cursor	GPT 系列、Gemini 系列、Claude 系列、DeepSeek 系列、Grok 系列、Composer 系列、o3、o4
<b>TRAE</b>	1. 基于 VS Code 二次开发的 AI 编程工具, 支持中文指令和上下文感知 2. 智能问答、代码补全、自动编程, 可自定义智能体 (如数据分析专用智能体) 3. 支持 #Web 实时搜索和 #Doc 文档集功能 4. 提供代码审查、错误检测及优化建议, 响应速度快	1. 端到端全栈开发 2. 复杂代码分析 3. 中文环境下的高效开发	字节跳动	GPT 系列、Gemini 系列、Kimi K2 系列、DeepSeek 系列、Grok 系列、o3
<b>通义灵码</b>	1. 基于通义千问大模型, 支持中文自然语言生成代码 2. 提供单元测试生成、异常报错排查、研发智能问答 3. 深度集成阿里云服务 4. 支持离线使用和开源模型微调	1. 阿里云生态内的开发 (如云函数、容器服务) 2. 需要与阿里云产品深度集成的项目 3. 企业级代码质量管控 (如单元测试覆盖率提升)	阿里巴巴	DeepSeek 系列、Qwen 系列
<b>SnapDevelop</b>	1. 低代码平台, 支持可视化设计 2. 自动生成 C#/Vue 代码, 符合云原生架构标准 3. 集成 AI 辅助 (如代码注释生成、错误诊断) 4. 支持信创需求	1. 云原生应用快速开发 (如微服务架构) 2. 非技术人员参与的业务流程数字化 3. 企业级信创项目 (如政府、金融行业)	艾普阳科技	GPT 系列、DeepSeek 系列、豆包系列、Qwen 系列

数据来源: 各公司官网, 广发证券发展研究中心

## 四、盈利预测和投资建议

智谱依托全栈式模型产品矩阵构建本地化部署 + 云端服务双业务架构。公司GLM基座模型持续迭代，多模态、AI Agent、代码模型等产品实现多行业商业化落地。凭借较强的模型能力，B端行业客户渗透和C端个人开发者市场拓展均呈加速态势。随着算力成本下行、模型研发效率提升及计费模式优化，公司商业化规模与盈利效率协同提升，亏损增幅将进一步收窄。基于产品竞争力与业务增长逻辑，预计2025-2027年营收保持高速增长，规模效应逐步显现，长期盈利具备较强确定性。

假设条件：

1. 本地化部署业务以标准化模型授权交付为核心，依托统一技术底座实现低边际成本扩张，客户自主承担算力、存储等硬件成本，业务持续保持高毛利率水平。结合行业数字化转型需求提升及公司海外业务落地节奏，客户规模稳步扩大推动收入持续增长，且随基数提升增速逐步放缓，预计2025-2027年该业务收入同比增长90%/85%/80%；同时产品标准化程度持续提升，规模效应进一步摊薄成本，毛利率随业务成熟度小幅抬升，预计分别为67%/68%/69%。

2. 云端部署业务涵盖通用API、编码订阅套餐Coding Plan、云端私有化部署等形式，是公司营收核心增长极。核心编码订阅产品全球化快速起量，叠加阶梯式计费模式落地、海外高定价市场持续拓展，收入实现爆发式增长，随基数扩大增速逐步回落，预计2025-2027年该业务收入同比增长400%/150%/100%；同时算力硬件成本下行、产品运营效率提升，毛利率呈稳步攀升趋势，预计分别为15%/20%/25%。

表 17：2023-2027年智谱营收拆分（单位：百万元）

	2023	2024	2025E	2026E	2027E
<b>本地化部署</b>					
收入	112.61	263.93	501.47	927.71	1669.89
增长率	105.44%	134.37%	90.00%	85.00%	80.00%
成本	35.83	89.67	165.48	296.87	517.66
毛利	76.78	174.26	335.98	630.85	1152.22
毛利率(%)	68.18%	66.02%	67.00%	68.00%	69.00%
<b>云端部署</b>					
收入	11.92	48.48	242.42	606.05	1212.10
增长率	359.68%	306.61%	400.00%	150.00%	100.00%
成本	8.22	46.85	206.06	484.84	909.08
毛利	3.70	1.63	36.36	121.21	303.03
毛利率(%)	31.04%	3.37%	15.00%	20.00%	25.00%
<b>合计</b>					
收入	124.54	312.41	743.89	1533.76	2881.99
增长率	116.93%	150.86%	138.11%	106.18%	87.90%
成本	44.06	136.53	371.54	781.71	1426.74
毛利	80.48	175.89	372.35	752.06	1455.25
毛利率	64.62%	56.30%	50.05%	49.03%	50.49%

数据来源：公司招股书，Wind，广发证券发展研究中心

公司2025-2027年收入预计为7.4亿元、15.3亿元和28.8亿元，营收增速预计为138.1%、106.2%和87.9%。预计2025-2027年EPS分别为-7.79元/股、-8.14元/股、-7.26元/股。2025-2027年，随着公司营收规模的扩大，各项费用率有望下降，具体来看：

**1. 研发效率提升：**研发费用为公司核心费用项，主要用于模型迭代与算力投入，2025年在模型持续优化下，费用率仍处于高位，后续随着算力成本下行、模型架构优化带来的研发效率提升，叠加现有模型产品矩阵成熟度提高，研发投入增速预计低于营收增速，费用率快速下降，预计2025-2027年分别为420%/230%/130%。

**2. 销售规模效应显现：**销售团队规模预计保持稳定增长态势，核心依托产品竞争力与行业口碑实现客户拓展，叠加 B端客户粘性持续提升、C端产品获客成本随规模下降，销售投入增速趋于平缓，费用率持续回落，预计2025-2027年分别为60%/35%/20%。

**3. 管理效率优化：**管理费用随公司经营规模扩大呈现稳步优化趋势，内部管理效率持续提升，叠加股份支付等非经常性费用影响逐步减弱，费用率稳步下降，预计2025-2027年分别为30%/17%/10%。

表 18：2023-2027年智谱盈利预测（单位：百万元）

	2023	2024	2025E	2026E	2027E
营业收入	124.54	312.41	743.89	1,533.76	2,881.99
毛利润	80.48	175.89	372.35	752.06	1,455.25
毛利率	64.62%	56.30%	50.05%	49.03%	50.49%
研发费用	528.9	2195.4	3,124.33	3,527.66	3,746.58
研发费用率	424.68%	702.73%	420.00%	230.00%	130.00%
销售费用	101.2	387.5	446.33	536.82	576.40
销售费用率	81.26%	124.03%	60.00%	35.00%	20.00%
管理费用	66.3	133.6	223.17	260.74	288.20
管理费用率	53.24%	42.76%	30.00%	17.00%	10.00%
归母净利润	-787.96	-2,956.49	-3,430.02	-3,583.15	-3,195.95

数据来源：公司招股书，Wind，广发证券发展研究中心

从行业特性来看，高成长阶段的 AI 模型企业核心以营收规模与增长潜力为估值核心，因此采用市销率（PS）对公司进行估值。综合考虑公司业务领域、经营模式，我们选取商汤和MiniMax为智谱的可比公司，二者均聚焦 AI 大模型领域与智谱相近，其 PS 估值具备参考意义。相较商汤，智谱营收来自AI大模型商业化落地的比例更高；相较MiniMax的C端主导模式，智谱收入更多来自于B端。综合考虑公司AI基础能力、技术优势和行业地位，给予智谱2026年70倍PS估值，合理总价值为1073.63亿元人民币，对应每股合理价值273.22港元/股（参考汇率1港元=0.89人民币），给予“增持”评级。

表 19: 智谱可比公司估值情况 (市值统计截止2026.2.3收盘)

公司名称	公司代码	业务类型	市值 (亿港元)	PS (2026)	营收 (百万元)			EPS (元/股)		
					2024A	2025E	2026E	2024A	2025E	2026E
商汤	00020.HK	生成式AI、视觉AI	974.2	14.1	3772.1	4876.8	6157.7	-0.13	-0.07	-0.02
MiniMax	00100.HK	AI大模型	1822.2	115.7	219.4	561.2	1405.1	-30.8	-11.4	-8.2

数据来源: Wind, 广发证券发展研究中心

备注: 商汤和 MiniMax 的预测来自 Wind 一致预测。

## 五、风险提示

### （一）下游企业客户对AI大模型付费意愿和能力不高的风险

一方面，部分垂直领域客户对于AI大模型付费能力受限于企业自身经营状况；另一方面，AI大模型处于快速迭代阶段，企业对于AI大模型选择范围较广，在业务赋能带来经济效益未充分显现之前，往往选择开源或免费的AI大模型服务。

### （二）研发投入巨大和成果转化的风险

公司AI大模型产品类型较多，相应研发投入较大。若公司不能及时把握行业发展趋势和市场需求，无法研发出具有商业价值、符合市场需求的产品，公司将面临研发投入难以收回的风险。

### （三）科技巨头在AI大模型领域的投入及行业竞争加剧的风险

在AI大模型领域，阿里巴巴、字节跳动等大型科技公司具有较强的资金优势和算力基础设施的优势。在与大型科技公司竞争的过程中，公司的模型迭代和商业拓展均面临较大挑战。

**资产负债表**

单位:人民币百万元

	2023A	2024A	2025E	2026E	2027E
<b>流动资产合计</b>	<b>1,863</b>	<b>3,016</b>	<b>3,456</b>	<b>1,075</b>	<b>966</b>
现金及现金等价物	1,249	2,268	2,786	500	500
应收账款及票据	8	91	83	85	80
存货	29	32	41	43	40
其他	578	624	546	446	346
<b>非流动资产合计</b>	<b>997</b>	<b>1,360</b>	<b>1,206</b>	<b>1,206</b>	<b>1,206</b>
固定资产净值	530	243	243	243	243
长期投资	115	307	200	200	200
商誉及无形资产	352	714	714	714	714
其他	0	97	50	50	50
<b>资产总额</b>	<b>2,860</b>	<b>4,376</b>	<b>4,662</b>	<b>2,281</b>	<b>2,172</b>
<b>流动负债合计</b>	<b>3,609</b>	<b>7,838</b>	<b>7,372</b>	<b>8,474</b>	<b>11,461</b>
短期借款	0	269	300	366	2,302
应付账款及票据	7	58	72	109	159
其他	3,602	7,510	7,000	8,000	9,000
<b>非流动负债合计</b>	<b>234</b>	<b>493</b>	<b>500</b>	<b>600</b>	<b>700</b>
长期借款	0	0	0	0	0
其他非流动负债	234	493	500	600	700
<b>总负债</b>	<b>3,843</b>	<b>8,331</b>	<b>7,872</b>	<b>9,074</b>	<b>12,161</b>
普通股股本	28	36	36	36	36
储备	-1,011	-3,993	-3,248	-6,831	-10,026
其他	0	0	0	0	0
归母权益总额	-983	-3,957	-3,212	-6,795	-9,990
少数股东权益	0	1	1	1	1
<b>股东权益合计</b>	<b>-983</b>	<b>-3,955</b>	<b>-3,210</b>	<b>-6,793</b>	<b>-9,989</b>
<b>负债及股东权益合计</b>	<b>2,860</b>	<b>4,376</b>	<b>4,662</b>	<b>2,281</b>	<b>2,172</b>

**利润表**

单位:人民币百万元

	2023A	2024A	2025E	2026E	2027E
<b>主营收入</b>	<b>125</b>	<b>312</b>	<b>744</b>	<b>1,534</b>	<b>2,882</b>
营业成本	44	137	372	782	1,427
毛利	80	176	372	752	1,455
其他收入	0	0	0	0	0
营销费用	101	387	446	537	576
行政管理费用	66	134	223	261	288
研发费用	529	2,195	3,124	3,528	3,747
其他营业费用合计	0	0	0	0	0
<b>营业利润</b>	<b>-616</b>	<b>-2,541</b>	<b>-3,421</b>	<b>-3,573</b>	<b>-3,156</b>
利息收入	0	0	0	0	0
利息支出	26	38	9	10	40
权益性投资损益	0	21	0	0	0
其他非经营性损益	-10	2	0	0	0
除税前利润	-788	-2,958	-3,430	-3,583	-3,196
所得税	0	0	0	0	0
<b>合并净利润</b>	<b>-788</b>	<b>-2,958</b>	<b>-3,430</b>	<b>-3,583</b>	<b>-3,196</b>
少数股东损益	0	-2	0	0	0
<b>归母净利润</b>	<b>-788</b>	<b>-2,956</b>	<b>-3,430</b>	<b>-3,583</b>	<b>-3,196</b>

**现金流量表**

单位:人民币百万元

	2023A	2024A	2025E	2026E	2027E
<b>经营活动现金流净额</b>	<b>-648</b>	<b>-2,245</b>	<b>-4,166</b>	<b>-3,442</b>	<b>-2,997</b>
合并净利润	-788	-2,958	-3,430	-3,583	-3,196
折旧与摊销	0	0	0	0	0
营运资本变动	-96	0	-744	132	159
其他非经营性调整	236	712	9	10	40
<b>投资活动现金流净额</b>	<b>-785</b>	<b>-49</b>	<b>487</b>	<b>1,100</b>	<b>1,100</b>
处置固定资产收益	0	115	0	0	0
资本性支出	-509	-133	0	0	0
投资资产支出	-235	-27	432	1,000	1,000
其他	-41	-3	54	100	100
<b>融资活动现金流净额</b>	<b>2,463</b>	<b>3,312</b>	<b>4,196</b>	<b>56</b>	<b>1,897</b>
长期债权融资	-32	330	0	0	0
股权融资	98	0	4,174	0	0
支付股利	0	0	0	0	0
其他	2,397	2,982	22	56	1,897
<b>现金净增加额</b>	<b>1,030</b>	<b>1,019</b>	<b>518</b>	<b>-2,285</b>	<b>0</b>
期初现金余额	219	1,249	2,268	2,786	500
期末现金余额	1,249	2,268	2,786	500	500

**主要财务比率**

	2023A	2024A	2025E	2026E	2027E
<b>成长能力</b>					
主营收入增长率	116.9%	150.9%	138.1%	106.2%	87.9%
营业利润增长率	-	-	-	-	-
归母净利润增长率	-	-	-	-	-
<b>获利能力</b>					
毛利率	64.6%	56.3%	50.1%	49.0%	50.5%
净利率	-632.7%	-946.8%	-461.1%	-233.6%	-110.9%
ROE	-	-	-	-	-
<b>偿债能力</b>					
资产负债率	134.4%	190.4%	168.9%	397.8%	559.8%
有息负债率	0.0%	6.2%	6.4%	16.0%	106.0%
流动比率	0.5	0.4	0.5	0.1	0.1
利息保障倍数	-23.8	-65.7	-400.6	-357.8	-78.9
<b>营运能力</b>					
应收账款周转天数	21.0	57.1	40.0	20.0	10.0
存货周转天数	159.8	80.8	40.0	20.0	10.0
应付账款周转天数	30.3	85.8	70.0	50.0	40.0
<b>每股指标</b>					
每股收益	-	-	-7.79	-8.14	-7.26
每股净资产	-	-	-7.29	-15.43	-22.69
每股经营现金流	-	-	-9.46	-7.82	-6.81
<b>估值比率</b>					
PE	-	-	-	-	-
PB	-	-	-	-	-
EV/EBITDA	-	-	-	-	-

## 广发计算机行业研究小组

- 刘雪峰：首席分析师，东南大学工学士，中国人民大学经济学硕士，1997年起先后在数家IT行业跨国公司从事技术、运营与全球项目管理工作。2010年就职于招商证券研究发展中心负责计算机组行业研究工作，2014年加入广发证券发展研究中心。
- 吴祖鹏：资深分析师，中南大学材料工程学士，复旦大学经济学硕士，曾先后任职于华泰证券、华西证券，2021年加入广发证券发展研究中心。
- 周源：资深分析师，慕尼黑工业大学硕士，2021年加入广发证券，曾任职于TUMCREATE自动驾驶科技公司，负责大数据相关工作。
- 王钰翔：高级研究员，哥伦比亚大学运筹学硕士，2024年加入广发证券发展研究中心。
- 戴亚敏：高级研究员，北京大学金融硕士，2024年加入广发证券发展研究中心。

## 广发证券—行业投资评级说明

- 买入：预期未来12个月内，股价表现强于大盘10%以上。
- 持有：预期未来12个月内，股价相对大盘的变动幅度介于-10%~+10%。
- 卖出：预期未来12个月内，股价表现弱于大盘10%以上。

## 广发证券—公司投资评级说明

- 买入：预期未来12个月内，股价表现强于大盘15%以上。
- 增持：预期未来12个月内，股价表现强于大盘5%-15%。
- 持有：预期未来12个月内，股价相对大盘的变动幅度介于-5%~+5%。
- 卖出：预期未来12个月内，股价表现弱于大盘5%以上。

## 联系我们

	广州市	深圳市	北京市	上海市	香港
地址	广州市天河区马场路26号广发证券大厦47楼	深圳市福田区益田路6001号太平金融大厦31层	北京市西城区月坛南街2号月坛大厦18层	上海市浦东新区南泉北路429号泰康保险大厦37楼	香港湾仔骆克道81号广发大厦27楼
邮政编码	510627	518026	100045	200120	-
客服邮箱	gfzqyf@gf.com.cn				

## 法律主体声明

本报告由广发证券股份有限公司或其关联机构制作，广发证券股份有限公司及其关联机构以下统称为“广发证券”。本报告的分销依据不同国家、地区的法律、法规和监管要求由广发证券于该国家或地区的具有相关合法合规经营资质的子公司/经营机构完成。

广发证券股份有限公司具备中国证监会批复的证券投资咨询业务资格，接受中国证监会监管，负责本报告于中国（港澳台地区除外）的分销。广发证券（香港）经纪有限公司具备香港证监会批复的就证券提供意见（4号牌照）的牌照，接受香港证监会监管，负责本报告于中国香港地区的分销。

本报告署名研究人员所持中国证券业协会注册分析师资质信息和香港证监会批复的牌照信息已于署名研究人员姓名处披露。

## 重要声明

广发证券股份有限公司及其关联机构可能与本报告中提及的公司寻求或正在建立业务关系，因此，投资者应当考虑广发证券股份有限公司及其关联机构因可能存在的潜在利益冲突而对本报告的独立性产生影响。投资者不应仅依据本报告内容作出任何投资决策。投资者应自主作出投资决策并自行承担投资风险，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者口头承诺均为无效。

本报告署名研究人员、联系人（以下均简称“研究人员”）针对本报告中相关公司或证券的研究分析内容，在此声明：（1）本报告的全部分析结论、研究观点均精确反映研究人员于本报告发出当日的关于相关公司或证券的所有个人观点，并不代表广发证券的立场；（2）研究人员的部

分或全部的报酬无论在过去、现在还是将来均不会与本报告所述特定分析结论、研究观点具有直接或间接的联系。

研究人员制作本报告的报酬标准依据研究质量、客户评价、工作量等多种因素确定，其影响因素亦包括广发证券的整体经营收入，该等经营收入部分来源于广发证券的投资银行类业务。

本报告仅面向经广发证券授权使用的客户/特定合作机构发送，不对外公开发布，只有接收人才可以使用，且对于接收人而言具有保密义务。广发证券并不因相关人员通过其他途径收到或阅读本报告而视其为广发证券的客户。在特定国家或地区传播或者发布本报告可能违反当地法律，广发证券并未采取任何行动以允许于该等国家或地区传播或者分销本报告。

本报告所提及证券可能不被允许在某些国家或地区内出售。请注意，投资涉及风险，证券价格可能会波动，因此投资回报可能会有所变化，过去的业绩并不保证未来的表现。本报告的内容、观点或建议并未考虑任何个别客户的具体投资目标、财务状况和特殊需求，不应被视为对特定客户关于特定证券或金融工具的投资建议。本报告发送给某客户是基于该客户被认为有能力独立评估投资风险、独立行使投资决策并独立承担相应风险。

本报告所载资料的来源及观点的出处皆被广发证券认为可靠，但广发证券不对其准确性、完整性做出任何保证。报告内容仅供参考，报告中的信息或所表达观点不构成所涉证券买卖的出价或询价。广发证券不对因使用本报告的内容而引致的损失承担任何责任，除非法律法规有明确规定。客户不应以本报告取代其独立判断或仅根据本报告做出决策，如有需要，应先咨询专业意见。

广发证券可发出其它与本报告所载信息不一致及有不同结论的报告。本报告反映研究人员的不同观点、见解及分析方法，并不代表广发证券的立场。广发证券的销售人员、交易员或其他专业人士可能以书面或口头形式，向其客户或自营交易部门提供与本报告观点相反的市场评论或交易策略，广发证券的自营交易部门亦可能会有与本报告观点不一致，甚至相反的投资策略。报告所载资料、意见及推测仅反映研究人员于发出本报告当日的判断，可随时更改且无需另行通告。广发证券或其证券研究报告业务的相关董事、高级职员、分析师和员工可能拥有本报告所提及证券的权益。在阅读本报告时，收件人应了解相关的权益披露（若有）。

本研究报告可能包括和/或描述/呈列期货合约价格的事实历史信息（“信息”）。请注意此信息仅供用作组成我们的研究方法/分析中的部分论点/依据/证据，以支持我们对所述相关行业/公司的观点的结论。在任何情况下，它并不（明示或暗示）与香港证监会第5类受规管活动（就期货合约提供意见）有关联或构成此活动。

## 权益披露

(1) 广发证券（香港）跟本研究报告所述公司在过去12个月内并没有任何投资银行业务的关系。

## 版权声明

未经广发证券事先书面许可，任何机构或个人不得以任何形式翻版、复制、刊登、转载和引用，否则由此造成的一切不良后果及法律责任由私自翻版、复制、刊登、转载和引用者承担。