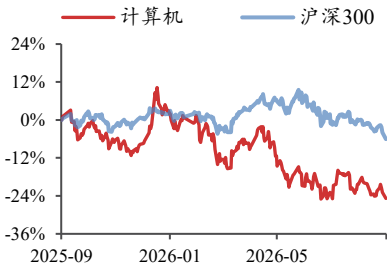


计算机

2026 年 09 月 28 日

投资评级：看好（维持）

行业走势图



数据来源：聚源

相关研究报告

《国产算力供需持续趋紧，智能体放大 CPU 增量缺口——行业点评报告》
-2026.9.23

《AI 需求驱动存储景气上行，Agent 应用加速落地——行业周报》-2026.9.20
《大模型迭代拓展能力边界，应用落地兑现产业价值——行业深度报告》
-2026.9.18

大模型系列深度（一）：国产大模型竞争重心转向效率与生态，智谱以 GLM 全栈体系卡位第一梯队

——行业深度报告

张越（分析师）

zhangyue1@kysec.cn

证书编号：S0790524090003

叶进新（联系人）

yejinxin@kysec.cn

证书编号：S0790126060053

● 国产大模型迈入商业化效率竞争，成本与生态成为分化重点

国产大模型竞争已从能力比拼转向能力、生态、成本与变现的系统性较量。截至 8 月 31 日，国内累计 1112 款生成式 AI 服务完成备案、731 款应用或功能完成登记。头部模型综合能力趋于接近，SuperCLUE 2026 年 7 月通用框架测试中：Qwen3.8-Max-0902、DeepSeek V4.1 Flash、GLM-5.3 和 Kimi K3 的总分集中在 70 至 73 分，但优势方向各异，其中 GLM-5.3 在智能体编程和任务规划两项领先。随着开放权重、长上下文、原生多模态及 Agent 能力逐步普及，行业竞争重心正由参数规模和单次榜单转向技术能力、开发者生态、调用成本与商业化效率的系统比拼。

● GLM 由 Coding 迈向长程 Agent，模型与应用矩阵协同扩张

智谱沿 GLM 技术路线持续迭代，模型能力已由语言生成和代码辅助扩展至任务规划、工具调用及工程交付。GLM-5 实现由“写代码”向“完成工程”跃迁，GLM-5.1 和 5.2 进一步强化持续工作与百万级上下文，GLM-5.3 与 GLM-5.3-Flash 分别面向高价值复杂任务和高频规模化调用，GLM-5-Turbo 则针对工具调用、持续任务和多智能体协同进行优化。应用端，公司已形成由 Z.ai、智谱清言、ZCode、AutoClaw、AutoGLM、Zread.ai 和 AMiner 组成的产品矩阵，覆盖对话创作、编程、桌面办公、手机操作和科研等场景，推动模型能力由 API 调用向实际任务执行延伸。

● API 放量推动收入结构跃迁，算力适配与开源生态支撑扩张

2026 年上半年，智谱实现营业收入 9.54 亿元，同比增长 399.74%；其中开放平台及 API 收入 8.25 亿元，同比增长 2735.7%，占总收入约 86.5%，该业务毛利率由负 0.4% 升至 24.6%，商业模式加速由一次性本地部署转向持续调用的云端服务。公司 ARR 持续增长，2026 年末指引达 30 亿美元。智谱已兼容 40 多款主流芯片，并完成与多家国产算力平台的深度适配，实现 10 万级国产芯片规模化推理，单位 Token 推理成本较年初下降 80%。截至相关统计期，公司已开源 50 多个模型、累计下载量超过 4500 万次，MaaS 平台企业及开发者用户超过 580 万。同期净亏损同比收窄 12.12%，但研发费用率仍达 223.42%，公司仍处于以高研发投入换取模型能力和商业规模扩张的阶段。

● 受益标的

大模型：智谱、MiniMax、阿里巴巴、腾讯控股。

● **风险提示：**大模型技术迭代不及预期风险、AI 技术应用不及预期风险、相关政策不及预期风险、AI Infra 建设不及预期风险等。

目 录

1、 国产头部模型能力收敛，开源与成本效率体现竞争优势.....	3
2、 智谱沿 GLM 主线构建全栈能力，由技术积累迈向商业化.....	6
2.1、 智谱完成从清华实验室到港交所上市的跨越，收入持续放量、ARR 高速增长.....	6
2.2、 智谱 MaaS 平台依托多元模型与算力适配提供完整产品矩阵.....	9
2.3、 科研底蕴、全栈技术与开源生态共同构筑智谱竞争壁垒.....	15
3、 风险提示.....	17

图表目录

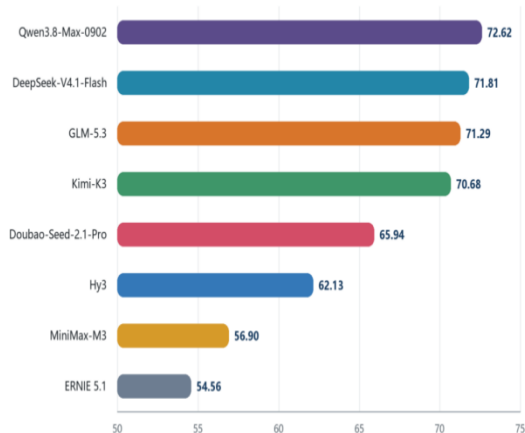
图 1： SuperCLUE 通用榜头部模型总分接近.....	3
图 2： 头部模型六维能力各有侧重.....	3
图 3： 头部开放权重模型能力接近，调用成本差异明显.....	4
图 4： 智谱沿 GLM 技术主线持续强化模型与智能体能力.....	6
图 5： 智谱营收持续高速增长，2026H1 净亏损同比收窄.....	8
图 6： 收入结构切换令毛利率阶段性承压，净利率仍处负值但明显改善.....	8
图 7： 收入放量显著摊薄销售及管理费用率，研发投入强度仍处高位.....	8
图 8： 智谱收入结构已由本地部署主导快速转向数据与云端产品主导.....	8
图 9： 智谱 ARR 持续增长，2026 年末指引达 30 亿美元.....	9
图 10： 智谱 MaaS 平台依托多元模型与算力适配提供一体化 AI 服务.....	9
图 11： GLM-5.3 和 5.3-flash 在复杂软件工程、终端操作和更广泛的真实世界 Agent 任务上均取得显著进步.....	10
图 12： DSA 通过索引器筛选 Top-k 键值条目，实现稀疏注意力计算.....	11
图 13： GLM-ASR-Nano 模型架构：“重感知、极致压缩、轻量推理”.....	12
图 14： 基于多奖励融合强化学习，实现工业级语音合成.....	12
图 15： GLM-5-Turbo 在 OpenClaw 场景中，多项关键任务领先于多家主流模型.....	12
图 16： 智谱构建覆盖多元场景的 AI 产品矩阵.....	13
图 17： 生成式 AI 产品 Z.ai 面向全球市场.....	13
图 18： 智谱清言深耕国内市场.....	13
图 19： Agent 产品均需下载到本地部署.....	14
图 20： AMiner 依托 GLM 模型覆盖“搜、读、写、做”全科研流程.....	14
图 21： GLM 模型已完成与多家国产算力平台的深度推理适配.....	15
图 22： 公司研发投入持续高增并显著超过营业收入.....	15
图 23： 公司研发人员占比超七成、人才结构高度研发化.....	15
图 24： 智谱已形成贯通架构设计、模型训练、部署优化与评估反馈的完整研发体系.....	16
图 25： 智谱依托多元基础设施与灵活交付方式满足客户差异化需求.....	16
表 1： 国产大模型厂商形成差异化竞争格局.....	4
表 2： 头部开放权重模型能力接近，调用成本差异明显.....	5
表 3： 大模型厂商商业化路径由 API 向多元场景延伸.....	5
表 4： 公司高管具备较强的科研经验、产业经验和经营管理能力.....	7
表 5： 最新模型 GLM-5.3 与 GLM-5.3-Flash 差异化定位，兼顾能力上界与推理效率.....	10
表 6： 稀疏注意力与线性注意力通过不同机制降低长上下文计算开销.....	11

1、国产头部模型能力收敛，开源与成本效率体现竞争优势

国产大模型竞争已从能力比拼转向能力、生态、成本与变现的系统性较量。大模型供给端由互联网平台、云厂商、独立模型公司和行业软件商共同参与，国家网信办公告显示，截至8月31日累计1112款生成式人工智能服务完成备案、731款应用或功能完成登记。头部厂商的竞争力沿四条证据链依次展开：榜单层面，头部模型能力接近、优势各有侧重，通用能力趋于收敛；生态层面，开源开放成为基准模型的主流选择，低门槛与高迭代效率持续强化竞争优势；成本层面，成本效率构成国产模型最具辨识度的优势；商业化层面，各厂商沿API、订阅、企业智能体与专属部署多路径推进，收入数据开始验证技术能力的变现。

国产头部模型能力接近、优势各有侧重。SuperCLUE 2026年7月通用框架含507题，8至9月为同框架补测。八款模型总分依次为 Qwen3.8-Max-0902 72.62、DeepSeek-V4.1-Flash 71.81、GLM-5.3 71.29、Kimi K3 70.68、Seed2.1 Pro 65.94、Hy3 62.13、MiniMax M3 56.90 和 ERNIE 5.1 54.56。分项上，DeepSeek 的数学 84.21 和幻觉控制 87.16 居首；Seed 科学 77.19 居首，精确指令 42.86 与 Qwen 并列；GLM 的智能体编程 76.84 和任务规划 91.15 居首，Kimi 编程 75.79 次之。Qwen 规划 89.53，Hy3 数学 77.19 和科学 75.44 突出，MiniMax 规划 77.68 强于其总分位置，ERNIE 的幻觉控制 83.02 和指令 41.90 优于编程 27.37。头部模型尚未形成六维能力全面领先者，竞争更多体现为不同能力组合。

图1：SuperCLUE 通用榜头部模型总分接近



资料来源：SuperCLUE Intelligence Index、开源证券研究所

图2：头部模型六维能力各有侧重

	数学推理	幻觉控制	科学推理	精确指令	智能体编程	任务规划
Qwen3.8-Max-0902	78.95	86.19	75.44	42.86	70.53	89.53
DeepSeek-V4.1-Flash	84.21	87.16	70.18	40.00	71.58	79.84
GLM-5.3	70.59	79.35	75.44	40.00	76.84	91.15
Kimi-K3	71.93	83.85	70.17	40.95	75.79	84.35
Doubao-Seed-2.1-Pro	75.44	78.85	77.19	42.86	51.58	84.95
Hy3	77.19	79.20	75.44	35.24	47.37	66.82
MiniMax-M3	70.18	70.44	70.17	20.00	46.32	77.68
ERNIE 5.1	64.91	83.02	59.65	41.90	27.37	72.21

资料来源：SuperCLUE Intelligence Index、开源证券研究所

开源开放已成为国产基准模型的主流选择，并凭借低门槛与高迭代效率持续强化其竞争力。八款基准模型中，DeepSeek V4.1 Flash、Hy3、Kimi K3、MiniMax M3 和 GLM-5.3 已直接发布权重，分别采用 MIT、Apache 2.0 或厂商自定义许可。Qwen 已发布 Qwen3.8-2.4T-A95B 权重，该模型与 Qwen3.8-Max 具有共同技术基础，但不能与公开权重完全等同。Seed2.1 Pro 和 ERNIE 5.1 尚未开放对应权重，百度则已按 Apache 2.0 许可开放 ERNIE 4.5 系列。上述开放模型均提供 vLLM、SGLang 部署方案。开放权重有助于企业开展本地及私有化部署、微调和性能优化，增强对数据及自身技术使用决策的控制，降低对单一供应商的依赖，并借助社区贡献和多服务商生态保持技术前沿；同时借助全球开发者“发布—反馈—迭代”闭环，社区评测、优化方案与衍生模型反哺模型进化，率先获得广泛应用者更易成为事实标准。

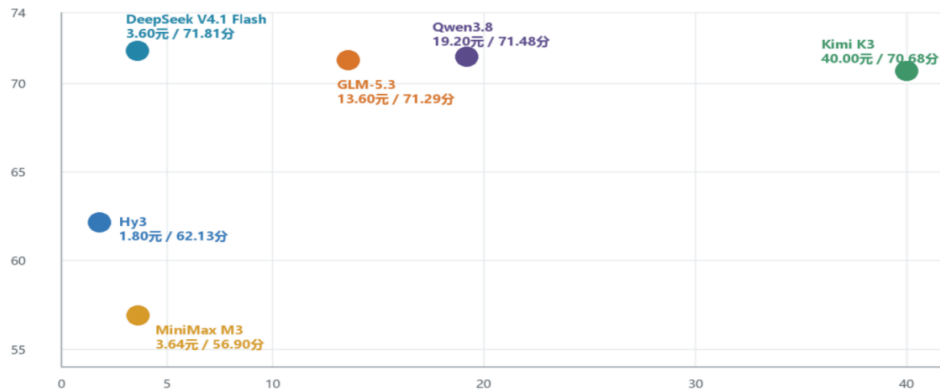
表1：国产大模型厂商形成差异化竞争格局

厂商	代表模型	公开规格与能力重点	开源情况	官方产品/API 入口
DeepSeek	V4.1 Flash	552B 主干参数 MoE；预填充/解码阶段分别激活 8B/16B；原生图文输入、最长 100 万 Token；重点优化长上下文 Agent 推理效率	MIT License	DeepSeek API；合作伙伴 WorkBuddy、OpenCode
阿里	Qwen3.8-2.4T-A95B	2.4T 总参数、95B 激活；原生 262,144 Token，可扩展至约 101 万 Token；重点覆盖 Coding、专业工作、研究与长程 Agent 任务	Qwen3.8-Max License	QwenCloud、阿里云百炼 Model Studio、Qwen Code
字节跳动	Seed2.1 Pro/Turbo	未披露参数规模；Pro/Turbo 两个版本；通用 Agent、端到端代码工程、多模态理解	未开放权重	豆包办公任务、TRAE Work/IDE、火山方舟 API
腾讯	Hy3	295B 总参数、21B 激活；256K 上下文；推理、Agent、长上下文与 Coding	Apache-2.0	元宝、WorkBuddy、Marvis、ima、腾讯云 TokenHub API
月之暗面	Kimi K3	2.8T 总参数、104B 激活；原生多模态、约 100 万 Token 上下文；重点覆盖长程 Coding、知识工作和推理任务	Kimi K3 License	Kimi、Kimi Work、Kimi Code、Kimi 开放平台 API
MiniMax	M3	约 428B 总参数、约 23B 激活；100 万 Token 上下文、原生多模态；重点覆盖 Coding、Agentic 工作与桌面操作	MiniMax Community License	MiniMax Code、MiniMax Token Plan、MiniMax 开放平台 API
智谱	GLM-5.3	沿用 GLM-5.2 基础模型，能力增量来自后训练；重点强化复杂 Coding 和长程任务，并支持低/高/最大三档思考强度	GLM-5.3 License	Z.ai、ZCode、GLM Coding Plan
百度	ERNIE 5.1	继承 ERNIE 5.0 预训练基础，总参数和激活参数分别压缩至约 1/3 和约 1/2；重点覆盖 Agent、推理与深度搜索能力	未开放权重	文心助手、飞桨 AI Studio、千帆大模型平台 API

资料来源：各公司官网、官方微博、Hugging Face、开源证券研究所

成本效率是国产模型最具辨识度的优势。国内八款模型每百万 Token 输入/输出价为：DeepSeek V4.1 Flash 闲时 1/4 元、峰时 2/8 元；Qwen3.8 Max 12/36 元；Seed2.1 Pro 6/30 元、Turbo 3/15 元；Hy3 1/4 元；Kimi K3 20/100 元；GLM-5.3 8/28 元；ERNIE 5.1 在 32K 内 4/18 元；MiniMax M3 在 512K 内按国际站现价为 0.30/1.20 美元。国外代表中，GPT-6 Astra 与 Claude Fable 5.1 均为 10/50 美元。

图3：头部开放权重模型能力接近，调用成本差异明显



口径：均按公开API最低上下文阶梯；DeepSeek取峰时；MiniMax按国际站现价并按官方汇率折算。
注：横轴是标价代理，不含缓存、重试、自部署算力及人工复核；能力分数与价格并非同一测试。

资料来源：各公司和平台官网、开源证券研究所

表2：头部开放权重模型能力接近，调用成本差异明显

厂商	模型	价格单位	输入	缓存命中	输出
DeepSeek	DeepSeek V4.1 Flash	元/百万 Token	1 闲 / 2 峰	0.02 / 0.04	4 闲/8 峰
阿里	Qwen3.8 Max	元/百万 Token	12	隐式 1.5	36
字节跳动	Seed2.1 Pro / Turbo	元/百万 Token	6 / 3	1.2 / 0.6	30 / 15
腾讯	Hy3	元/百万 Token	1	0.25	4
月之暗面	Kimi K3	元/百万 Token	20	2	100
MiniMax	MiniMax M3	美元/百万 Token	0.30	0.06	1.20
智谱	GLM-5.3	元/百万 Token	8	2	28
百度	ERNIE 5.1	元/百万 Token	4 / 6	未列示	18 / 22
OpenAI	GPT-6 Astra	美元/百万 Token	10	1	50
Anthropic	Claude Fable 5.1	美元/百万 Token	10	0.25	50

资料来源：各大模型官方定价页、开源证券研究所

国内厂商已沿 API、订阅、企业智能体和专属部署形成不同路径。阿里 2026 年二季度 AI 相关产品收入 123.76 亿元；字节在 6 月披露火山引擎日均使用 120 万亿 Token，但未拆分内外部与付费调用；腾讯未单列模型收入，披露 Hy3 持续位列 OpenRouter Token 消耗前三，WorkBuddy 出现付费意愿；百度同期 AI 业务收入 125 亿元，其中 AI 云基础设施 73 亿元。独立厂商中，MiniMax 上半年收入 1.166 亿美元，开放平台及企业服务 0.739 亿美元；智谱上半年收入 9.54 亿元，开放平台及 API 收入 8.25 亿元；月之暗面已提供 API、会员和企业席位，DeepSeek 以按量 API 和开放权重为主，两者均未披露收入或付费用户。

表3：大模型厂商商业化路径由 API 向多元场景延伸

厂商与模型	主要收费路径	公司或平台商业化数据
DeepSeek V4.1 Flash	官方 API 按量计费	已实行 API 峰谷计价；未披露独立收入
阿里 Qwen3.8	Model Studio 按量 API、Token Plan、企业 AI 云服务；QwenWork 为公测产品入口	2026Q2 AI 相关产品收入 123.76 亿元
字节 Seed2.1	火山方舟 API 及模型服务、TRAE 订阅及企业服务、豆包企业级产品	2026 年 6 月披露日均使用超 180 万亿 Token
腾讯 Hy3	TokenHub 按量 API、Hy Token Plan、WorkBuddy 及 CodeBuddy 订阅和 Token 充值、企业版或专属部署	OpenRouter Token 消耗前三；WorkBuddy 有付费意愿
月之暗面 Kimi K3	Kimi API 按量计费、个人会员、Kimi Business 企业席位	已公开 API 与席位价格；未披露收入或付费用户
MiniMax M3	AI 原生应用订阅及内购、开放平台 API、Token Plan、其他 AI 企业服务	2026H1 收入 1.166 亿美元；平台及企业 0.739 亿美元
智谱 GLM-5.3	开放平台/API 按量计费、Coding Plan、企业智能体、本地化部署	2026H1 收入 9.54 亿元；开放平台及 API 8.25 亿元
百度 ERNIE 5.1	AI 云基础设施、千帆 MaaS 及 API、AI 应用	2026Q2 AI 业务 125 亿元；AI 云基础设施 73 亿元

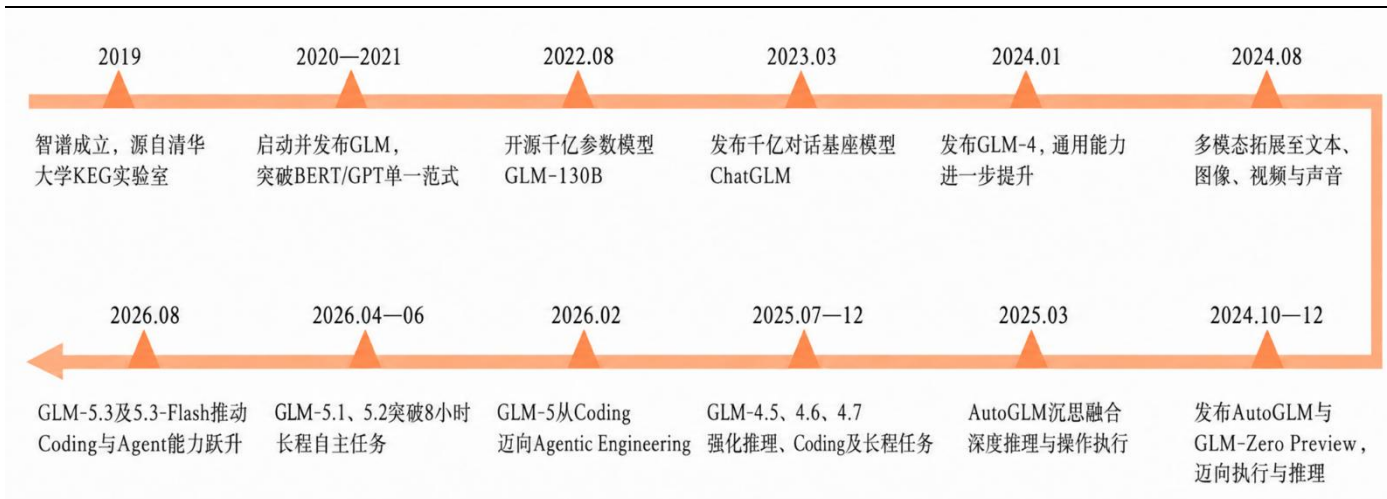
资料来源：厂商或大模型官网、开源证券研究所

2、智谱沿 GLM 主线构建全栈能力，由技术积累迈向商业化

2.1、智谱完成从清华实验室到港交所上市的跨越，收入持续放量、ARR 高速增长

智谱的技术积累可追溯至 2006 年清华大学知识工程实验室推出的科研情报挖掘平台 AMiner。2019 年 6 月，AMiner 被单独拆分，北京智谱华章科技股份有限公司由此成立，联合创始人包括清华博士张鹏、时任清华大学高级工程师刘德兵与计算机系教授唐杰。成立不到一年，OpenAI 推出 1750 亿参数的 GPT-3，此背景下团队选择走自主研发道路，投身 GLM 预训练框架研发。2022 年 8 月，智谱开源高精度千亿大模型 GLM-130B，同时发布 ChatGLM 对话模型及多模态产品矩阵；同年 11 月，GLM-130B 作为亚洲唯一入选的全球 30 个主流大模型，在鲁棒性、校准误差和无偏性方面表现出色。2023 年 8 月底，“智谱清言”上线；2024 年 1 月发布比肩 GPT-4 的基座大模型 GLM-4，7 月上线视频生成功能“清影”，8 月底清言推出国内首个面向客户端开放的视频通话功能，完成从语言文字、图片到视频的模态跨越。2026 年 1 月 8 日正式登陆港交所。上市后，智谱以约两月一次的节奏迭代 GLM-5 系列，能力从 Coding 延伸至长程自主任务与国产算力规模化推理。

图4：智谱沿 GLM 技术主线持续强化模型与智能体能力



资料来源：公司官网、招股说明书、开源证券研究所

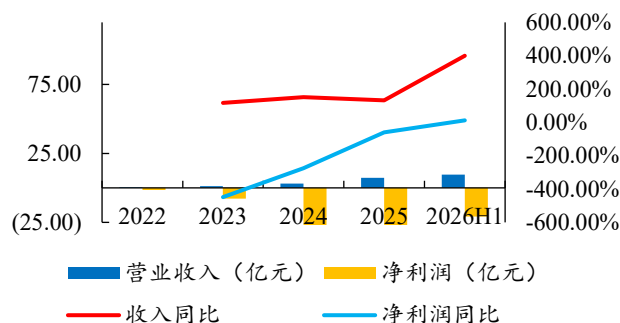
公司核心管理团队在计算机科学、人工智能及科技行业拥有丰富的从业经验，多位成员具有清华大学任职或教育背景。董事长刘德兵拥有近 18 年计算技术行业经验，曾任清华大学高级工程师，主要负责公司的战略规划、业务方向及整体管理。首席科学家唐杰长期任教于清华大学，曾任清华大学计算机系副主任，长期从事人工智能、知识图谱、数据挖掘及机器学习等领域的研究，主要负责公司整体技术战略、前沿研究以及核心模型与产品创新。首席执行官兼总经理张鹏在计算机科学领域拥有近 20 年经验，曾于 2005 年至 2020 年任职于清华大学，主要负责业务发展、研发及日常运营管理。副总经理王绍兰拥有逾 20 年科技行业经验，曾任万康世纪科技（北京）有限公司首席执行官，主要负责业务发展、融资及日常运营管理。董事会秘书肖磊具有法律、投资及财务管理经验，曾任创新奇智科技集团股份有限公司首席财务官兼副总裁等职，主要负责公司的企业管治及财务运营。

表4：公司高管具备较强的科研经验、产业经验和经营管理能力

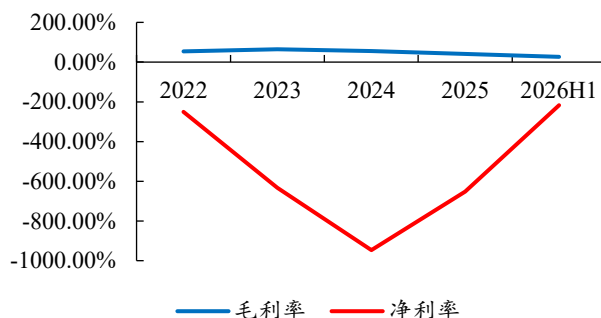
姓名	职位	任职经历
刘德兵	法定代表人、董事长	联合创始人、执行董事兼董事长，主要负责集团战略规划、业务方向及整体管理，在计算技术行业拥有近18年经验。2007年9月至2012年12月任特艺（中国）科技有限公司北京研究所研究工程师，随后在清华大学担任高级工程师。曾获中国人工智能学会科技进步奖一等奖、北京市科学技术进步奖二等奖及国家科学技术进步奖二等奖等荣誉。
唐杰	首席科学家	创始人兼首席科学家（首席 AI 官），负责统筹公司整体技术战略与前沿研究方向，推动核心模型与产品创新，健全人工智能安全与对齐体系。长期任教于清华大学，曾任清华大学计算机系副主任，现任清华大学计算机系讲席教授、人工智能研究院基础模型研究中心主任；长期从事人工智能、知识图谱、数据挖掘、社交网络与机器学习等领域研究。
张鹏	总经理、首席执行官	联合创始人、执行董事、首席执行官兼总经理，主要负责集团业务发展、研发及日常运营管理。在计算机科学领域拥有近20年经验，2005年8月至2020年12月任职于清华大学。先后取得清华大学计算机科学与技术学士、硕士及电子信息博士学位，曾获王选新闻科学技术奖、中国电子学会科技进步奖一等奖及“全国工业和信息化系统劳动模范”称号。
王绍兰	副总经理	联合创始人兼副总经理，主要负责集团业务发展、融资及日常运营管理，在科技行业拥有逾20年经验。历任北京凹凸微电子开发有限公司、星云融创（北京）科技有限公司相关职务，曾任万康世纪科技（北京）有限公司首席执行官；2018年3月至2019年6月任职于清华大学。
肖磊	董事会秘书	董事会秘书，主要负责集团企业管治及财务营运。曾任中粮置地管理有限公司法律顾问；在首泰金信（北京）股权投资基金管理股份有限公司历任法务部副总经理及总经理、董事会秘书、投资银行部总经理及副总经理；曾任创新奇智科技集团股份有限公司财务及投资总监、联席公司秘书、董事会秘书、首席财务官兼副总裁；2024年2月至2025年10月任上海黄豆网络科技有限公司副总裁。

资料来源：Wind、开源证券研究所

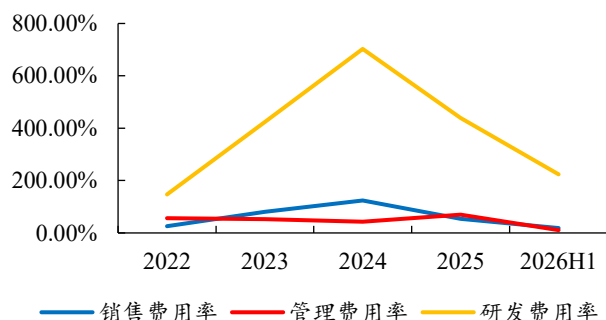
智谱近年收入规模快速扩张。2022—2025年，公司营业收入由0.57亿元增长至7.24亿元，2023—2025年分别同比增长116.93%、150.86%和131.85%，主要受模型能力持续迭代、市场需求扩张以及本地化部署和云端业务销量提升带动。2026年上半年，公司营业收入进一步增至9.54亿元，同比增长399.74%，其中数据与云端产品收入占比由2025年的26.28%升至86.51%，主要受开放平台及API业务快速放量驱动，商业模式加速由一次性本地部署向持续调用的云端服务转型。**盈利方面**，公司净亏损由2022年的1.44亿元扩大至2025年的47.18亿元，主要受基础模型迭代、研发团队扩张、股份支付及算力投入增加影响；2026年上半年净亏损为20.72亿元，同比收窄12.12%，收入和毛利快速增长以及销售、管理开支下降对亏损改善形成支撑。公司毛利率由2023年的64.62%降至2026年上半年的26.38%，主要由于云端业务占比快速提升但仍处于毛利率爬坡期，同时本地化交付业务投入增加；同期净利率虽仍为-217.21%，但已明显改善。**费用方面**，2026年上半年销售及管理费用率分别降至18.63%和10.83%，规模效应逐步显现，但研发费用率仍达223.42%，反映公司仍处于以高研发投入换取模型能力提升和商业化扩张的阶段。

图5：智谱营收持续高速增长，2026H1 净亏损同比收窄


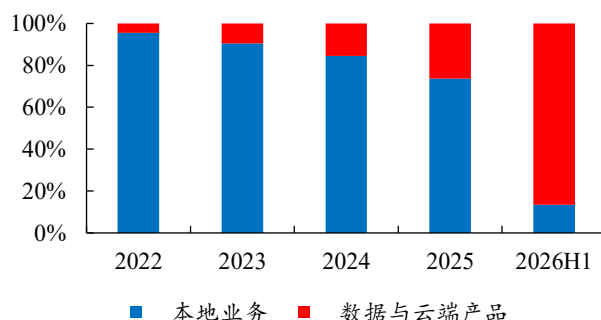
数据来源：智谱招股说明书、智谱 2025 年年报、智谱 2026 年半年报、开源证券研究所

图6：收入结构切换令毛利率阶段性承压，净利率仍处负值但明显改善


数据来源：智谱招股说明书、智谱 2025 年年报、智谱 2026 年半年报、开源证券研究所

图7：收入放量显著摊薄销售及管理费用率，研发投入强度仍处高位


数据来源：智谱招股说明书、智谱 2025 年年报、智谱 2026 年半年报、开源证券研究所

图8：智谱收入结构已由本地部署主导快速转向数据与云端产品主导


数据来源：Wind、开源证券研究所

智谱 ARR 于 2025 年下半年起进入陡峭放量通道，收入主引擎已完成由本地化部署向 MaaS 开放平台（API+订阅）的切换，商业化路径趋于清晰。起点为开发者订阅业务：公司于 2025 年 9 月推出 GLM Coding Plan，定价最低 20 元/月、约为海外同类产品的 1/7；据 CEO 张鹏披露，该业务上线短时间内 ARR 即突破 1 亿元人民币，付费开发者超 15 万，MaaS 平台累计企业及应用开发者逾 270 万。2026 年 3 月，公司进一步披露 MaaS 平台 API 业务 ARR 达 17 亿元人民币，同比增长约 60 倍，付费开发者超 24.2 万。至 2026 年 8 月 31 日中期业绩，切换基本完成：上半年 MaaS 开放平台及 API 服务收入 8.25 亿元，同比增长约 2736%，收入占比由上年同期的 15.2% 升至 86.5%。截至 8 月末，按月度收入年化口径 ARR 达 16 亿美元，较 7 月上旬的 10 亿美元提升 60%，周度年化口径达 20 亿美元。

智谱 ARR 高增长的核心支撑在于量价齐升与单位经济模型的改善，token 经济模式已初步跑通。截至 8 月末，平台 Token 调用量较年初增长超 40 倍，API 平均售价提升约 101%，打破行业价格战下“以价换量”的常规路径。开放平台及 API 业务毛利率由上年同期的 -0.4% 转正至 24.6%，验证了推理成本优化（单位 Token 推理成本较年初下降 80%）对盈利模型的实质贡献。9 月 11 日披露总额约 50 亿美元再融资后，管理层进一步将年末 ARR 目标上调至 30 亿美元。

图9：智谱 ARR 持续增长，2026 年末指引达 30 亿美元

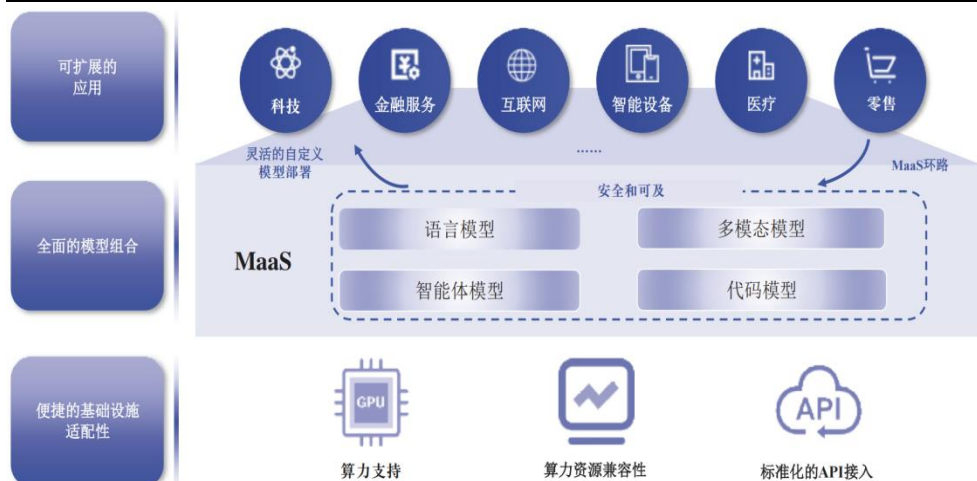


资料来源：中国证券网、网易新闻、同花顺新闻、开源证券研究所

2.2、智谱 MaaS 平台依托多元模型与算力适配提供完整产品矩阵

公司围绕一体化 MaaS 平台组织产品供应，平台由全面的模型组合、可扩展的应用和便捷的基础设施适配性三个层级构成。模型层覆盖语言、多模态、智能体及代码四类模型，并提供模型微调、部署及智能体开发工具；应用层面向科技、金融服务、互联网、智能设备、医疗、零售等场景，支持跨硬件、跨场景和跨业务流程调用，客户可借助智能体工作区，通过简化微调、增量训练和提示工程定制解决方案；基础设施层集成计算、网络、训练通信与推理加速能力，支持 15 亿至 2300 亿参数模型以及跨云、跨芯片部署，并可扩展至手机、个人电脑和智能汽车等终端。

图10：智谱 MaaS 平台依托多元模型与算力适配提供一体化 AI 服务



资料来源：智谱招股说明书

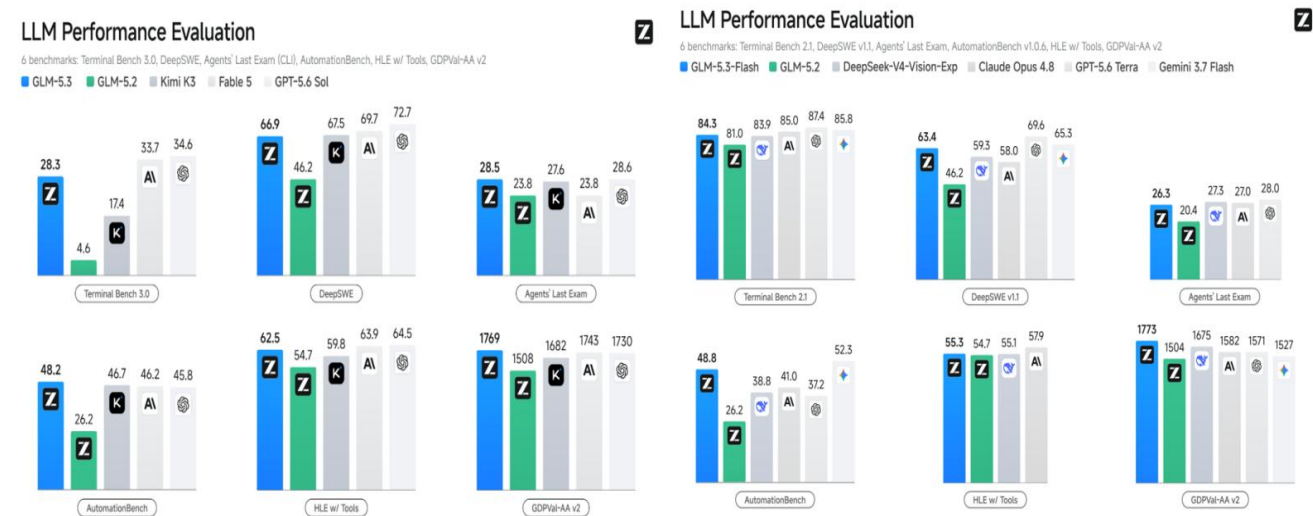
（一）全面的模型组合

智谱基座模型沿 GLM 技术路线持续演进，并于 2025 年进入以编程和智能体能力为核心的加速迭代阶段。自 2021 年推出首代 GLM 以来，公司先后发布 GLM-130B、ChatGLM 和 GLM-4，并逐步由语言理解与生成拓展至推理、代码和工具调用。2025 年，GLM-4.5 首次将推理、编程与智能体能力统一至同一基座，GLM-4.6 进一步提升长上下文、代码生成和智能体表现。2025 年 12 月发布的 GLM-4.7 重点强化编程

能力，引入保留式思考与轮级思考机制，在多项代码评测中达到当时开源模型领先水平。2026 年 2 月推出的 GLM-5 则通过扩大参数和训练数据规模、引入异步强化学习框架及稀疏注意力机制，实现由“写代码”向“完成工程”的关键跃迁。

GLM-5.1 之后，智谱将迭代重点转向长周期任务的稳定性、持续性与实际可用性。2026 年 4 月发布的 GLM-5.1 具备 8 小时级持续工作能力，并在 SWE-bench Pro 上取得领先成绩；6 月推出的 GLM-5.2 进一步支持百万级无损上下文，在综合能力与代码盲测中进入全球前列；8 月发布的 GLM-5.3 在基座不变的情况下，通过扩大后训练规模显著提升性能，并展现出网络安全能力；同期推出的 GLM-5.3-Flash 成为系列首个原生多模态模型，在保持前沿性能的同时大幅降低使用成本，体现出智谱推动先进模型普惠化的方向。

图11：GLM-5.3 和 5.3-flash 在复杂软件工程、终端操作和更广泛的真实世界 Agent 任务上均取得显著进步



资料来源：智谱官网、开源证券研究所

表5：最新模型 GLM-5.3 与 GLM-5.3-Flash 差异化定位，兼顾能力上界与推理效率

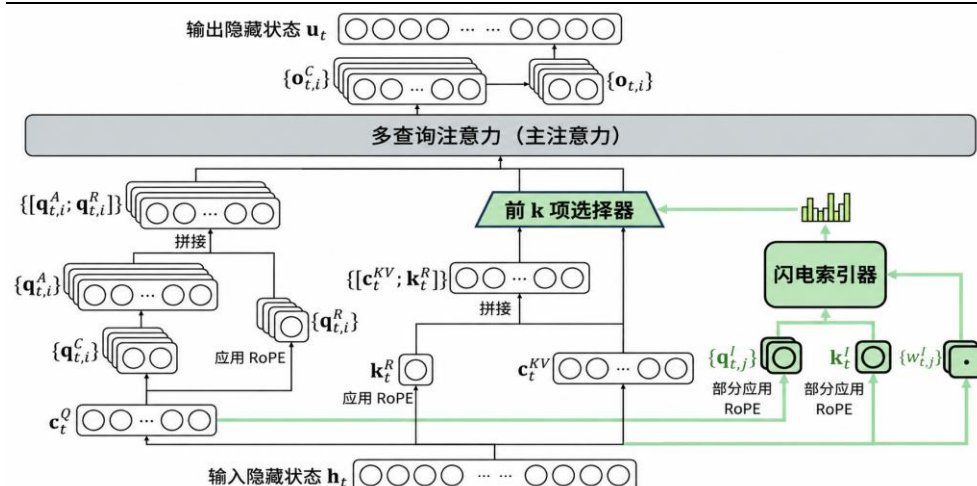
核心参数	GLM-5.3	GLM-5.3-Flash	差异
定位	旗舰模型，追求能力上界	高性价比模型，强调低成本高性能	分别覆盖高价值复杂任务与高频规模化调用
总参数量	744B	320B	Flash 总参数量约为 GLM-5.3 的 43%
激活参数量	40B	18B	Flash 单 Token 激活参数更少，利于降低推理成本
Transformer 层数	78 层	45 层	Flash 网络深度明显降低，更侧重推理效率
路由专家数	256 个	288 个	Flash 总参数更少，但专家数量更多，MoE 结构进一步细化
注意力架构	DeepSeek 稀疏注意力	线性注意力+稀疏注意力混合架构	Flash 针对长上下文及低成本推理重新设计注意力机制
基座及训练方式	沿用 GLM-5.2 基座，重点扩大后训练 Scaling	重新训练新基座，同时优化模型架构与训练体系	5.3 重在“训得更深”，Flash 重在“重新设计效率前沿”

资料来源：Z.ai、SGLang、Hugging Face、开源证券研究所

稀疏注意力与线性注意力分别通过减少计算连接和重构计算方式提升长上下文

处理效率，并可通过混合架构兼顾模型能力与推理成本。稀疏注意力由闪电索引器与细粒度词元选择机制组成。索引器为查询与此前词元计算分数，仅选取分数最高的前 k 个键值条目参与注意力计算，从而降低主注意力计算复杂度，并结合优化实现提升长上下文推理效率。线性注意力则改变计算形式，经典方法利用核特征映射和矩阵乘法结合律，先汇总键值信息，再供查询读取，避免显式构造完整的两两注意力矩阵；自回归生成时可递推更新固定大小的状态，使注意力计算量随序列长度线性增长。GLM-5.3-Flash 采用两者结合的混合架构，在保持精准长上下文能力的同时降低长上下文推理成本。

图12：DSA 通过索引器筛选 Top-k 键值条目，实现稀疏注意力计算



资料来源：《DeepSeek-V3.2: Pushing the Frontier of Open Large Language Models》、开源证券研究所

表6：稀疏注意力与线性注意力通过不同机制降低长上下文计算开销

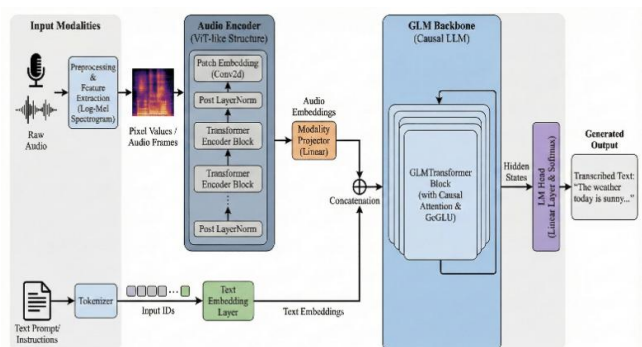
对比维度	稀疏注意力 (以 DSA 原始报告为例)	线性注意力 (以经典核函数方案为例)
核心思路	筛选参与注意力计算的词元，减少连接数量	改写计算形式，复用键值信息的聚合结果
计算过程	索引器打分、选取前 k 项，再对选中项计算注意力	通过核特征映射和矩阵乘法结合律，先聚合、后查询
序列计算量	选定 (k) 项后的主注意力计算约为 $(O(Lk))$ ，索引器仍为 $(O(L^2))$ ，另需计入筛选开销	特征维度固定时，注意力计算量为 $O(N)$
生成时的历史信息	从历史键值中选择并读取相关条目	将历史信息累积到固定大小的递推状态中
主要收益	减少长上下文中的主注意力计算	降低长序列计算开销；单步生成的状态存储不随历史长度增长

资料来源：《DeepSeek-V3.2: Pushing the Frontier of Open Large Language Models》、Katharopoulos 等《Transformers are RNNs: Fast Autoregressive Transformers with Linear Attention》(ICML 2020)、Z.ai《GLM-5.3-Flash 官方模型说明》、开源证券研究所

智谱已形成覆盖视觉理解、语音识别、语音合成与图像生成四条主线的开源多模态布局。1) 视觉侧，GLM-4.6V 支持将图像、截图和文档页直接作为工具参数，

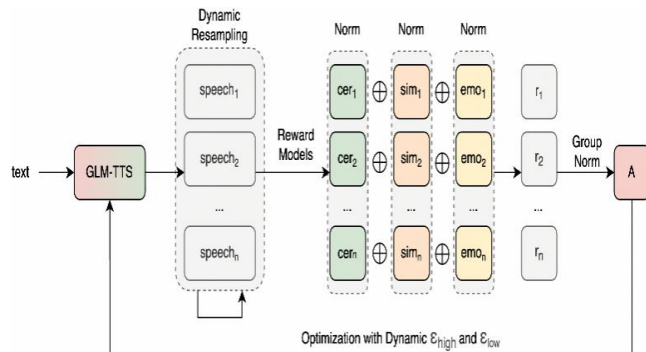
并对工具返回的图表、网页截图进行再次理解，形成“感知—理解—执行”闭环，可用于图文创作、视觉网页搜索和前端复刻。其视觉编码器与 128K 上下文对齐，单次推理约可处理 150 页文档、200 页 PPT 或一小时视频，在 20 余项主流多模态评测中达到同参数规模 SOTA。2) **语音侧**，15 亿参数的 GLM-ASR-Nano-2512 采用非对称架构，开源模型平均错误率为 4.10，方言及低音量场景表现突出；GLM-TTS 通过两阶段生成和多奖励融合强化学习，实现 3 秒音色复刻，CER 降至 0.89%。3) **图像侧**，GLM-Image 采用 9B 自回归与 7B DiT 混合架构，文字渲染在两项基准中位列开源第一，并基于昇腾 Atlas 800T A2 与昇思 MindSpore 完成全流程训练。

图13: GLM-ASR-Nano 模型架构:“重感知、极致压缩、轻量推理”



资料来源：智谱官网

图14: 基于多奖励融合强化学习，实现工业级语音合成



资料来源：智谱官网

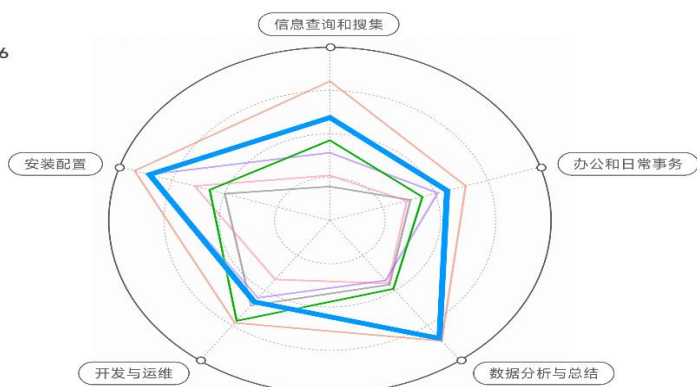
GLM-5-Turbo 是面向 OpenClaw 龙虾场景优化的基座模型，重点提升工具调用、定时与持续性任务、长链路执行等能力。模型从训练数据和优化目标入手，围绕工具调用、复杂指令拆解、高吞吐长链路执行稳定性等能力进行强化。在智谱自研的 Agent 端到端评测基准 ZClawBench 上，GLM-5-Turbo 位列国产模型第一。GLM-5-Turbo 以“Pony-Alpha-2”为代号接入 AutoClaw 澳龙开展盲测时，90% 的受访者认为其表现优于其他国产模型。围绕模型应用，智谱同步推出龙虾套餐与 Claw for Enterprise Security 体系，提供统一调度与权限编排、执行链路可视化监控、角色权限管理、完整审计日志以及关键节点人工审批等功能。

图15: GLM-5-Turbo 在 OpenClaw 场景中，多项关键任务领先于多家主流模型

ZClawBench：龙虾场景能力评测



GLM-5-Turbo
GLM-5
Claude Opus 4.6
Gemini 3.1 Pro
MiniMax M2.5
Kimi K2.5



资料来源：智谱官网

(二) 可扩展的应用

智谱以 GLM 模型体系为基础，构建了由 Z.ai、智谱清言、ZCode、AutoClaw、AutoGLM、Zread.ai 和 AMiner 组成的 AI 应用产品矩阵。根据公司官网，相关产品主要服务日常问答与创作、编程与办公、设备操作、代码理解和学术研究等需求，支持文本生成、图像理解、代码生成、多模态交互、检索增强生成及视频合成等功能。具备多任务处理能力的产品，为满足客户日益多样化的应用需求提供了支撑。

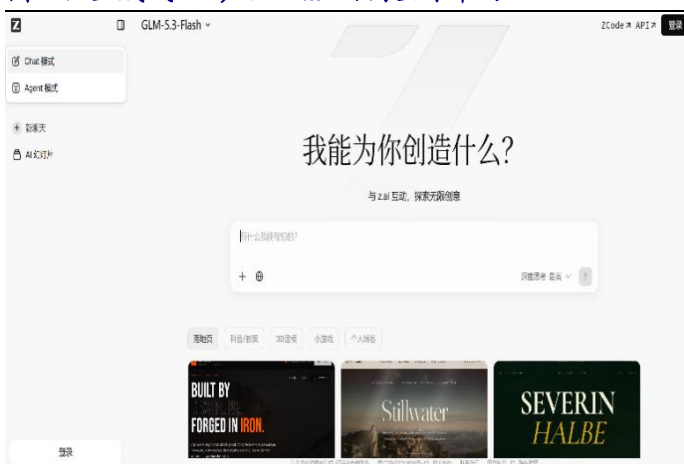
图16：智谱构建覆盖多元场景的 AI 产品矩阵



资料来源：智谱官网

Z.ai 与智谱清言均依托智谱 GLM 技术底座，为用户提供对话、内容生成、代码处理及任务执行等生成式 AI 服务。Z.ai 是智谱的国际化品牌与模型服务平台，搭载并开源 GLM 基座模型，通过在线对话和 API 提供端到端应用开发、长程 Agent 任务、多智能体编程及办公文档生成等能力，并配套 OpenClaw、Z Code 和 Excel 插件，同时完成对多类国产算力平台的推理适配。智谱清言则是智谱于 2023 年 8 月上线的生成式 AI 助手，上线初期基于 ChatGLM2，主要服务工作、学习和日常生活场景，支持智能问答、多轮对话、文案及文档创作和代码生成，并通过“灵感大全”与 Improve 功能降低使用门槛。相较而言，Z.ai 更侧重工程开发、Agent 执行和产业服务，智谱清言更强调低门槛的日常对话与个人提效，目前 GLM-5.3 均同步在两款产品上线。

图17：生成式 AI 产品 Z.ai 面向全球市场



资料来源：智谱官网

图18：智谱清言深耕国内市场



资料来源：智谱官网

ZCode、AutoClaw 和 AutoGLM 均为智谱围绕 Agent 任务执行推出的产品，分别面向软件开发、桌面工作和手机操作场景。ZCode 定位为智能体开发环境，致力于将繁琐的 CLI 交互转化为直观的图形化多 Agent 协作平台，能灵活调度 claude code、codex、gemini 等主流 CLI 工具，支持可视化配置、多设备远程开发、多任务并行、会话版本管理和内置浏览器等使用场景。AutoClaw 是深度适配 GLM 模型的桌面端 AI 工作助手，能够根据用户目标理解需求、拆解步骤、调用工具并持续执行，交付文档、报告、网页、代码或设计方案；目前已提供 50 余项 Skills，覆盖内容创作、网站开发、邮件与飞书办公、营销 SEO、代码开发、金融投研和学术研究等场景。AutoGLM 则是能够操作手机的 AI Agent，可根据文字或语音指令模拟点击、滑动和输入等操作，完成点单、购物、出行及社交等跨 App 任务；其开源内容包括核心模型、Phone Use 能力框架与工具链、覆盖 50 余款 App 的可运行 Demo 及相关文档。

图19: Agent 产品均需下载到本地部署



资料来源：智谱官网、开源证券研究所

Zread.ai 和 AMiner 均是智谱面向专业工作场景推出的大模型辅助工具，分别服务软件开发与学术科研需求。Zread.ai 是基于大模型的一站式代码理解与技术文档生成工具，旨在解决开发者接手旧项目困难、文档撰写低效、开源项目难以上手等痛点。Zread.ai 以 GLM 模型为核心底座，凭借其代码理解能力强、低幻觉、支持 Deep Research 及 Agent 能力适配好等优势，实现“读懂代码、生成知识、服务协作”。AMiner 是集学术检索、科研智能体、学术资源、个性化学术空间于一体 AI 科研助手，已覆盖 220 个国家和地区的 1000 万用户。AMiner 接入智谱 GLM 模型，依托 3.2 亿篇论文、1.7 亿件专利等亿级资源与百亿级知识图谱提供可靠输出，覆盖科研“搜、读、写”全流程：双模检索精准找文献，AI 阅读助手辅助精读，AMiner 沉思平均 7 分钟完成万字报告，另提供学术简历生成、学术空间知识沉淀及多端订阅推送服务。

图20: AMiner 依托 GLM 模型覆盖“搜、读、写、做”全科研流程



资料来源：智谱官网

（三）便捷的基础设施适配性

智谱已形成覆盖不同模型规模、云环境、国产芯片及终端设备的广泛基础设施适配能力。通过与基础设施业务合作伙伴共同设计先进的算力基础设施，其 MaaS 平台能够提供集成计算、网络、训练通信及推理加速能力，支持 15 亿至 2,300 亿参数规模的模型，并实现大规模、实时的跨云及跨芯片组部署。目前，GLM-5 已与华为昇腾、摩尔线程、寒武纪、昆仑芯、沐曦、燧原、海光等国产算力平台完成深度推理适配，并通过底层算子优化与硬件加速，在国产芯片集群上实现高吞吐、低延迟的稳定运行。该适配能力也可将模型扩展至手机、电脑及智能汽车等大众设备。

图21：GLM 模型已完成与多家国产算力平台的深度推理适配

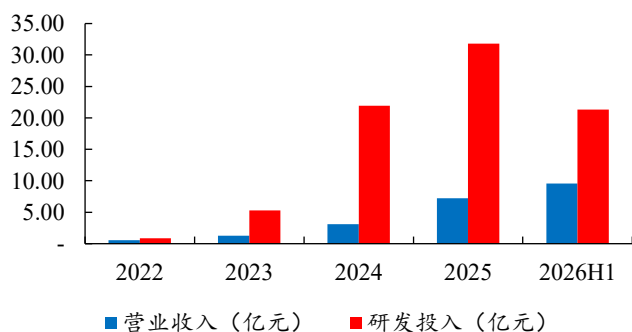


资料来源：华兴资本公众号

2.3、科研底蕴、全栈技术与开源生态共同构筑智谱竞争壁垒

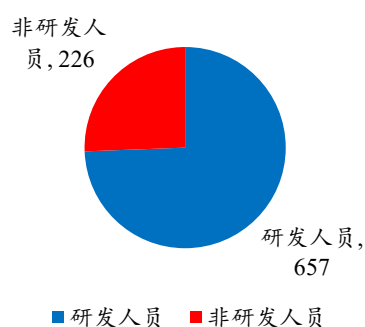
深厚的人才储备与高强度研发投入构成公司模型持续进化的底层支撑。公司核心团队与清华大学知识工程实验室保持长期深度合作，张鹏、唐杰等核心成员兼具清华学术背景与多年产业经验。截至 2025 年 6 月末，公司研发团队达 657 人，占员工总数超过 74%，技术领导平均拥有逾 12 年相关领域经验，研究方向覆盖自然语言处理、复杂系统决策及多模态语义分析；2025 年公司研发开支达 31.80 亿元，同比增长 44.9%，持续投向研发人才与模型训练基础设施。依托上述积累，公司形成覆盖预训练、中训练与后训练的完整训练体系，自研 Slime 框架实现生成与训练解耦；GLM-4.5 采用 3550 亿总参数的混合专家架构，GLM-5 进一步优化模型架构与强化学习体系，推动模型能力向任务规划、工具调用和软件工程交付延伸。人才、投入与训练体系相互衔接，持续扩展模型可处理的任务范围。

图22：公司研发投入持续高增并显著超过营业收入



资料来源：智谱招股说明书、2025 年年报、2026 年半年报、开源证券研究所

图23：公司研发人员占比超七成、人才结构高度研发化



资料来源：智谱招股说明书、开源证券研究所

图24：智谱已形成贯通架构设计、模型训练、部署优化与评估反馈的完整研发体系



资料来源：智谱招股说明书、开源证券研究所

多芯片适配与推理工程优化，将模型能力延伸为规模化部署与成本控制优势。截至 2025 年 6 月末，公司模型已兼容 40 多款主流硬件芯片，并通过自研算子库支持多样化硬件环境，GLM-5 已完成对七家国产芯片厂商平台的适配。在此基础上，公司持续推进软硬件协同设计，以定制化融合内核、通信优化及推理架构调整提升不同硬件条件下的运行效率；据 2026 年中期业绩公告，公司已实现 10 万级国产芯片的规模化推理，单位 Token 推理成本较年初下降 80%。多芯片适配拓宽可选算力来源，推理优化则为控制服务成本、扩大调用规模奠定基础，形成自模型研发贯通至实际部署的工程化能力。

图25：智谱依托多元基础设施与灵活交付方式满足客户差异化需求



资料来源：智谱招股说明书

依托 MaaS 平台、多元交付与开源生态，公司模型能力加速向商业收入与用户覆盖转化。公司通过 MaaS 平台集成模型调用、微调、智能体开发及场景模板等工具，

以云端与本地化部署并行、贯穿需求沟通至后续支持的交付流程，匹配客户在便捷调用、数据控制与定制开发等方面的差异化需求。2026 年上半年，公司开放平台及 API 收入达 8.25 亿元，同比增长 2735.7%，占总收入比重由上年同期的 15.2% 提升至 86.5%，业务毛利率由负 0.4% 升至 24.6%，标准化服务在收入结构中的重要性持续提高，模型能力的商业转化获得经营层面验证。生态侧，截至 2025 年 6 月末公司已开源 50 多个模型，累计下载量超过 4500 万次，GLM-4.5 以 MIT 许可开放并提供部署适配说明；公司同步通过“Z 计划”向合作伙伴提供模型、基础设施访问及定制化技术支持，2026 年 6 月末 MaaS 平台企业及开发者用户累计超过 580 万。开源扩大技术覆盖与开发者参与，伙伴网络补充行业需求理解与应用交付能力，共同支撑公司使用场景拓展与实践反馈积累。

3、风险提示

大模型技术迭代不及预期风险。大模型技术迭代是推动产业能力边界拓展与下游场景深化的核心驱动力，当前行业正处于多模态感知、复杂推理、智能体执行等前沿技术的攻坚突破期，技术突破节奏、能力落地效果均存在一定不确定性。若后续核心技术突破进度、模型能力提升幅度低于市场预期，无法有效支撑下游高价值场景的深度适配与功能升级，可能导致产业落地节奏整体延后，行业成长进程存在阶段性放缓的可能性。

AI 技术应用不及预期风险。下游场景规模化渗透是大模型产业价值兑现的核心支撑，当前大模型正从技术概念验证期向垂直行业规模化应用期过渡，落地进度受行业数字化基础、客户付费意愿、投入产出比验证等多重因素共同制约。若后续垂直行业适配进度、企业付费转化节奏不及市场预期，产业商业价值的兑现周期可能有所拉长，行业长期成长动能的释放节奏存在一定波动可能。

相关政策不及预期风险。大模型产业发展高度依赖政策环境的引导规范与配套支持，当前人工智能领域处于快速发展阶段，数据安全合规、技术监管标准、产业扶持政策、市场准入规则等相关制度仍处于持续完善过程中。若后续行业监管要求趋严、扶持政策落地进度慢于预期，或合规要求抬升导致企业研发与运营成本增加，可能对大模型技术迭代、商业化落地及行业竞争格局形成阶段性扰动，行业发展节奏存在相应调整的可能性。

AI Infra 建设不及预期风险。AI 基础设施是大模型技术迭代与产业规模化落地的底层算力与工程支撑，算力集群建设、高速互联部署、存储体系配套等基建环节具备投入规模大、建设周期长、产业链协同要求高的特征，落地进度受硬件供给、资金投入、工程能力等多重因素约束。若后续高端算力供给受限、基建项目建设进度慢于市场预期，可能形成算力供给瓶颈，制约大模型训练迭代与应用规模化落地，行业整体发展节奏存在阶段性承压的可能。

特别声明

《证券期货投资者适当性管理办法》、《证券经营机构投资者适当性管理实施指引（试行）》已正式实施。根据上述规定，开源证券评定此研报的风险等级为R3（中风险），因此通过公共平台推送的研报其适用的投资者类别仅限定为专业投资者及风险承受能力为C3、C4、C5的普通投资者。若您并非专业投资者及风险承受能力为C3、C4、C5的普通投资者，请取消阅读，请勿收藏、接收或使用本研报中的任何信息。

因此受限于访问权限的设置，若给您造成不便，烦请见谅！感谢您给予的理解与配合。

分析师声明

本研究报告的署名人员具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，以勤勉的职业态度，独立、客观地出具本报告，并对内容和观点负责。本报告清晰准确地反映了署名人员的研究观点，所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。本报告采用的各种估值方法及模型均有其局限性，估值结果不保证所涉及证券能够在该价格交易。本报告署名人员保证他们报酬的任何一部分不曾与，不与，也将不会与本报告中的具体的推荐意见或观点有直接或间接的联系。

股票投资评级说明

	评级	说明
证券评级	买入（Buy）	预计相对强于市场表现 20%以上；
	增持（outperform）	预计相对强于市场表现 5%~20%；
	中性（Neutral）	预计相对市场表现在-5%~+5%之间波动；
	减持（underperform）	预计相对弱于市场表现 5%以下。
行业评级	看好（overweight）	预计行业超越整体市场表现；
	中性（Neutral）	预计行业与整体市场表现基本持平；
	看淡（underperform）	预计行业弱于整体市场表现。

备注：评级标准为以报告日后的 6~12 个月内，证券相对于市场基准指数的涨跌幅表现，其中 A 股基准指数为沪深 300 指数（北交所基准指数为北证 50 指数）、港股基准指数为恒生指数、新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）、美股基准指数为标普 500 或纳斯达克综合指数。我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议；投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者应阅读整篇报告，以获取比较完整的观点与信息，不应仅仅依靠投资评级来推断结论。

法律声明

开源证券股份有限公司是经中国证监会批准设立的证券经营机构，已具备证券投资咨询业务资格。

本报告仅供开源证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。本报告是发送给开源证券客户的，属于商业秘密材料，只有开源证券客户才能参考或使用。

本报告是基于本公司认为可靠的已公开信息，但本公司不保证该等信息的准确性或完整性。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他金融工具的邀请或向人做出邀请。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动，过往的业绩表现不应作为其日后表现的预示。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。客户应当考虑到本公司可能存在可能影响本报告客观性的利益冲突，不应视本报告为做出投资决策的唯一因素。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。本公司未确保本报告充分考虑到个别客户特殊的投资目标、财务状况或需要。本公司建议客户应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。若本报告的接收人非本公司的客户，应在基于本报告做出任何投资决定或就本报告要求任何解释前咨询独立投资顾问。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口头承诺均为无效。

开源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。开源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。

本报告的版权归本公司所有。本公司对本报告保留一切权利。未经本公司事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

开源证券研究所

上海

地址：上海市浦东新区世纪大道1788号陆家嘴金控广场1号楼3层
邮编：200120
邮箱：research@kysec.cn

北京

地址：北京市西城区西直门外大街18号金贸大厦C2座9层
邮编：100044
邮箱：research@kysec.cn

深圳

地址：深圳市福田区金田路2030号卓越世纪中心1号楼45层
邮编：518000
邮箱：research@kysec.cn

西安

地址：西安市高新区锦业路1号都市之门B座5层
邮编：710065
邮箱：research@kysec.cn