仅供内部参考，请勿外传

2026 年 06 月 19 日

计算机行业研究

买入（维持评级）

行业周报

证券研究报告

计算机组
分析师：李可夫（执业 S1130525120009） 分析师：刘高畅（执业 S1130525120005）
likefu@gjzq.com.cn liugaochang@gjzq.com.cn

国产模型&国产算力交相辉映

GLM-5.2 正式发布，国产大模型长任务、代码能力大幅追赶海外顶尖模型

智谱于 2026 年 6 月 16 日推出旗舰模型 GLM-5.2，依托 IndexShare 稀疏注意力架构实现稳定 100 万 Token 超长上下文，大幅降低长文本计算开销，投机解码接收长度同步提升 20%。在长周期任务评测中，模型多项指标仅次于 Claude Opus 4.8，领先 GPT-5.5；代码与智能体场景提升显著，Imarena WebDev 代码榜单位列全球第二，智能体榜单稳居上游。马斯克预判中国大模型 2027 年一季度可达

大模型技术进步已获得全球产业关注。

头部互联网大厂批量采购国产 GPU，软硬件 Day0 协同验证国产算力生态成熟度

字节跳动洽谈采购至少 5 万颗天数智芯国产推理芯片，叠加原有寒武纪、华为供应商，国产芯片正式进入头部大厂核心供应链；寒武纪实现 GLM-5 系列模型 Day0 同步适配，通过算子、

片算力效率。需求端，豆包大模型日均 Token 调用量突破 120 万亿，近三月规模翻倍、两年暴涨千倍，海量推理算力缺口持续打开国产芯片商业化落地空间，头部厂商落地落地充分验证国产芯片可承接大规模线上大模型业务。

模型迭代与国产算力相互赋能，产业形成正向循环，持续看好国产 AI 产业链发展

GLM-5.2 代表国产大模型在长周期、代码、智能体核心场景持续缩小与海外差距，国产 GPU 快速适配头部模型、接入互联网大厂真实业务需求，二者形成双向驱动的产业闭环：国产算力凭借软硬件协同优势承接增量负载，落地实践又反向推动芯片与模型持续优化，国产模型与国产算力交相辉映，长期成长逻辑清晰

投资建议

相关标的：

大模型：智谱、MiniMax、阿里巴巴、腾讯控股。

国内算力：寒武纪、海光信息、东阳光、摩尔线程、天数智芯、壁仞科技、利通电子、协创数据、京基智农、浪潮信息、利扬芯片、禾盛新材、沐曦股份、华勤技术、杰华特、中国长城、晶科科技、罗曼股份、网宿科技、盈峰环境、芯原股份、华丰科技、晶科科技、亿田智能、豫能控股、星环科技、鸿日达、盛视科技、首都在线、神州数码、百度集团、中芯国际、华虹半导体、中科曙光、润泽科技、大位科技、润建股份、奥飞数据、云赛智联、瑞晟智能、科华数据、潍柴重机、金山云、欧陆通、杰创智能、奥尼电子。

风险提示

行业竞争加剧的风险；技术研发进度不及预期的风险；特定行业下游资本开支周期性波动的风险。

内容目录

一、GLM-5.2 发布，国产大模型长周期任务能力进阶 3
1.1 100 万 Token 上下文与架构优化共同支撑长周期任务 3
1.2 长周期与代码能力较前代明显提升 3
1.3 马斯克预计中国模型或于 2027 年一季度达到
二、字节采购国产 GPU，Day 0 适配体现国产软硬件生态协同 6
2.1 字节被报道洽谈采购至少 5 万颗国产 AI 芯片 6
2.2 寒武纪 Day 0 适配 GLM-5，国产软硬件协同提速 7
风险提示 7

图表目录

图表 1： GLM-5.2 长周期任务评测 3
图表 2： GLM-5.2 标准代码与智能体评测 4
图表 3： lmarena WebDev 排行 4
图表 4： lmarena Agent 排行 5
图表 5： X 平台关于中国模型追赶
图表 6： 豆包大模型日均 Token 调用量 7

一、GLM-5.2 发布，国产大模型长周期任务能力进阶

1.1 100 万 Token 上下文与架构优化共同支撑长周期任务

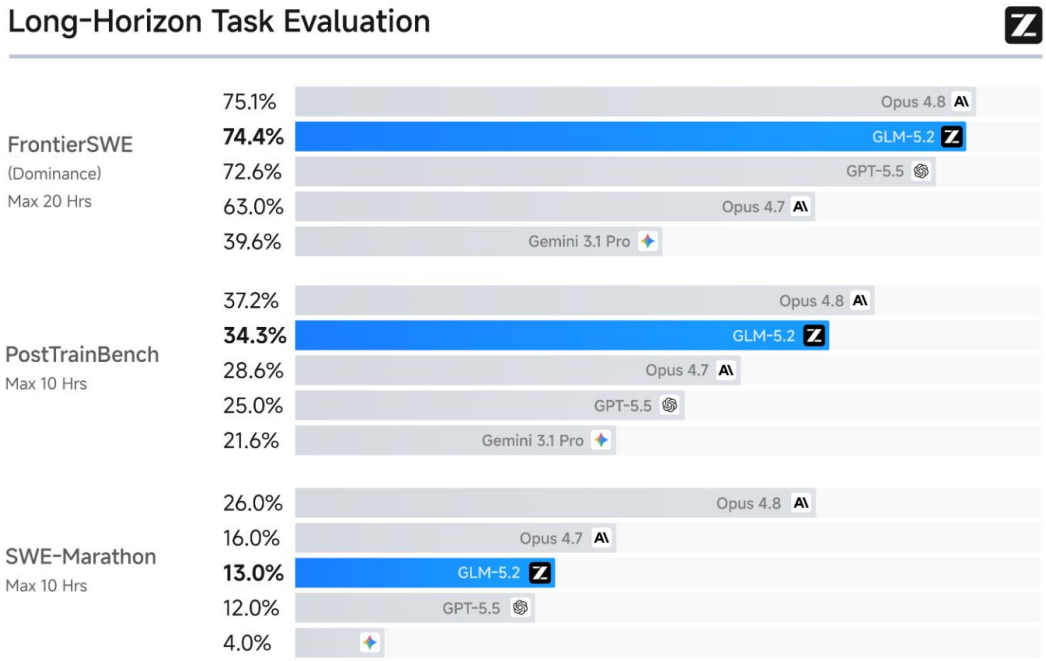
智谱于 2026 年 6 月 16 日发布 GLM-5.2，并将其称为面向长周期任务的最新旗舰模型。相较 GLM-5.1，GLM-5.2 首次在可稳定使用的 100 万 Token 上下文上交付长周期任务能力，同时提供不同思考强度，以平衡模型表现、执行速度与计算成本。

架构方面，智谱提出 IndexShare 方案，每四层稀疏注意力层共享同一索引器；官网称，该方案在 100 万 Token 上下文下将单 Token
于投机解码，官网披露接受长度最高提升 20%。模型权重采用 MIT 开源许可。

1.2 长周期与代码能力较前代明显提升

智谱官方评测显示 GLM-5.2 长周期任务表现优秀，在
74.4%，低于 Claude Opus 4.8 的 75.1%，高于 GPT-5.5 的 72.6%；在 PostTrainBench 中，GLM-5.2 得分为 34.3%，低于 Claude Opus 4.8 的 37.2%，高于 GPT-5.5 的 28.4%；在 SWE-Marathon 中，GLM-5.2 得分为 13.0%，低于 Claude Opus 4.8 的 26.0%，略高于 GPT-5.5 的 12.0%。

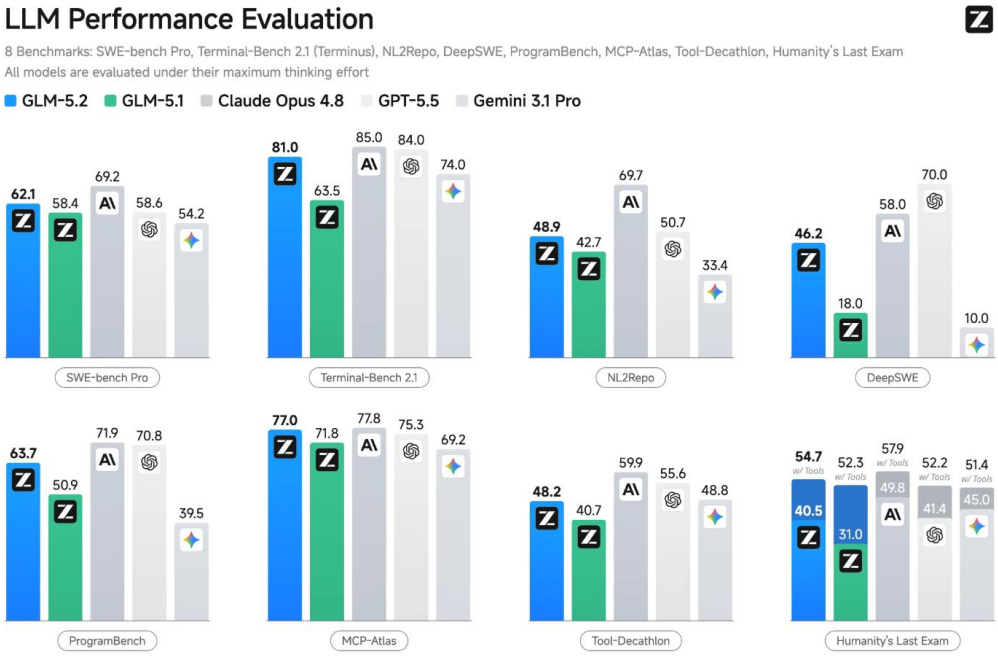
图表 1：GLM-5.2 长周期任务评测



来源：智谱官网，国金证券研究所

在标准代码评测中，智谱官网披露 GLM-5.2 在 Terminal-Bench 2.1 上取得 81.0 分，GLM-5.1 为 63.5 分；在 SWE-bench Pro 上取得 62.1 分，GLM-5.1 为 58.4 分。官网同时指出，GLM-5.2 在 Terminal-Bench 2.1 上的 81.0 分与 Claude Opus 4.8 的 85.0 分仍存在差距。

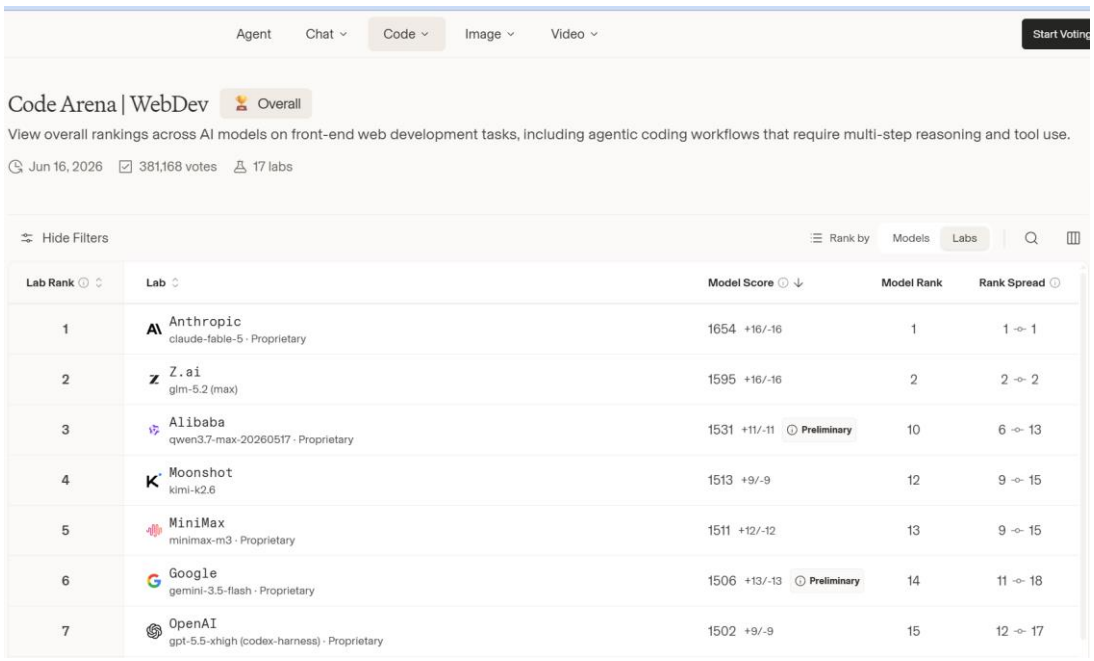
图表2：GLM-5.2 标准代码与智能体评测



来源：智谱官网，国金证券研究所

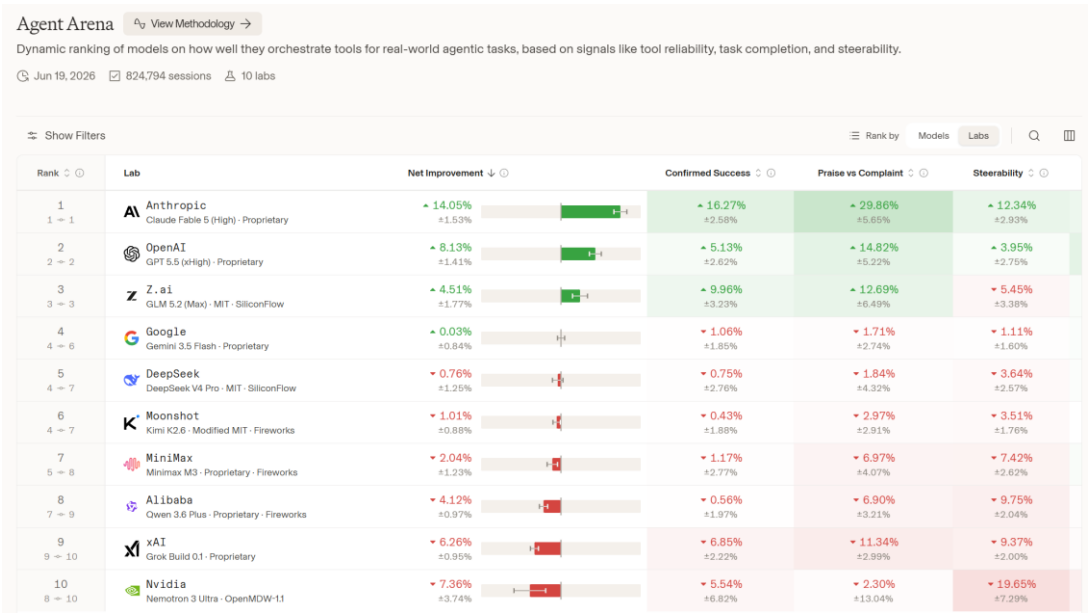
在第三方评测 lmarena 榜单上，智谱的 GLM-5.2 在 Coding 能力榜单 WebDev 排行第二，仅次于海外前沿模型 Claude Anthropic 的 Claude 系列模型。

图表3：lmarena WebDev 排行



来源：lmarena，国金证券研究所

图表4：lmarena Agent 排行



来源：lmarena，国金证券研究所

1.3 马斯克预计中国模型或于 2027 年一季度达到

据新浪财经，6 月 19 日有网友于 X 上提问称：“中国大模型预计何时能够达到 Anthropic 的
在 2027 年一季度”。智谱创始人唐杰很快回应称，“不需要那么久”。

图表5：X 平台关于中国模型追赶



来源：新浪财经，国金证券研究所

我们认为，海外产业人士的评论不能代替标准化评测，但反映出国产开源模型的能力进展正在获得更高的全球关注。相较笼统讨论参数规模，代码、智能体和长周期任务表现更接近模型在真实生产环境中的交付能力。

二、字节采购国产 GPU，Day 0 适配体现国产软硬件生态协同

2.1 字节被报道洽谈采购至少 5 万颗国产 AI 芯片

财联社 2026 年 6 月 18 日报道，字节跳动正与天数智芯讨论采购至少 5 万颗 AI 芯片，主要用于推理工作。报道进一步称，本次洽谈供货芯片主要用于大模型推理负载，对应天数智芯智铠系列云端推理 GPU，训练场景使用天垓系列。

该报道同时指出，若交易达成，天数智芯将成为华为和寒武纪之后字节跳动的第三家 GPU 供应商；截至发稿，字节跳动与天数智芯暂未回应。因此，相关信息应理解为采购洽谈而非已经确认的订单或收入。

Token 调用量爆发带来海量算力刚需，显著利好国产芯片规模化落地。4 月 2 日武汉巡展上，火山引擎披露豆包大模型最新日均 Token 调用量已突破 120 万亿，近三月规模直接翻倍；对比两年前仅 1200 亿的基数，整体调用量实现千倍增长。海量推理算力缺口下，字节跳动洽谈采购国产 GPU、将国产芯片纳入核心供应链，正是国产芯片产品力获得头部互联网大厂真实业务验证的核心信号，大模型流量爆发与国产算力导入形成双向利好，持续打开国产芯片商业化成长天花板。

图表6：豆包大模型日均 Token 调用量



来源：澎湃新闻，国金证券研究所

2.2 寒武纪 Day 0 适配 GLM-5，国产软硬件协同提速

寒武纪官方微信公众号此前发布 GLM-5 Day 0 适配相关文章。寒武纪一直重视芯片和算法的联合创新，致力于以软硬件协同的方式，优化大模型部署性能，降低部署成本。此前，我们对 GLM 系列模型进行了深入的软硬件协同性能优化，达成了业界领先的算力利用率。针对本次 GLM-5 模型，寒武纪依托深度集成的 Triton 的兼容能力完成快速适配，利用 BangC 和原生
优化，再次达成了业界领先的计算效率。依托 GLM-5 强大的模型能力，叠加寒武纪极致的计算效率，可大幅降低训推成。

我们认为，Day 0 适配意味着芯片厂商的软件栈需要在模型发布初期完成框架、算子和推理部署支持，显示了国产芯片与模型之间的深度协同合作。

同时我们认为，相较单颗芯片峰值性能，新模型能否快速运行、性能能否稳定释放以及开发者迁移成本是否可控，正在成为国产算力可用性的关键评价维度。

模型迭代、生态适配与需求导入形成正向循环。我们认为，GLM-5.2 验证国产模型在长周期任务和代码智能体方向继续进步，寒武纪 GLM-5 Day 0 适配体现国产算力生态的响应速度，字节洽谈采购国产 GPU 则指向互联网大厂的真实需求。模型能力提升带来推理调用和任务时长增长，国产芯片通过快速适配承接新增负载，真实业务部署再反哺软硬件优化，国产大模型与国产算力有望形成相互促进的产业循环。

风险提示

行业竞争加剧的风险：

众多新兴玩家参与到市场竞争之中，若市场竞争进一步加剧，竞争优势偏弱的企业或面临出清，某些中低端品类的毛利率或受到一定程度影响。

技术研发进度不及预期的风险：

技术开发需投入大量资源，如果相关厂商新品研发进程不及预期，表观层面将呈现出投入产出在较长时期的滞后特征。

特定行业下游资本开支周期性波动的风险：

部分公司下游资本开支波动与行业周期性相关性较强，或在个别年份对于上游厂商的营收表现产生扰动。

行业投资评级的说明：

买入：预期未来 3—6 个月内该行业上涨幅度超过大盘在 15%以上；

增持：预期未来 3—6 个月内该行业上涨幅度超过大盘在 5%—15%；

中性：预期未来 3—6 个月内该行业变动幅度相对大盘在 -5%—5%；

减持：预期未来 3—6 个月内该行业下跌幅度超过大盘在 5%以上。

特别声明：

国金证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本报告版权归“国金证券股份有限公司”（以下简称“国金证券”）所有，未经事先书面授权，任何机构和个人均不得以任何方式对本报告的任何部分制作任何形式的复制、转发、转载、引用、修改、仿制、刊发，或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发，需注明出处为“国金证券股份有限公司”，且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料，但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证。本报告反映撰写研究人员的不同设想、见解及分析方法，故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致，国金证券不对使用本报告所包含的材料产生的任何直接或间接损失或与此有关的其他任何损失承担任何责任。且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，在不作事先通知的情况下，可能会随时调整，亦可因使用不同假设和标准、采用不同观点和分析方法而与国金证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。

本报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可能会受汇率影响而波动。过往的业绩并不能代表未来的表现。

客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突，而不应视本报告为作出投资决策的唯一因素。证券研究报告是用于服务具备专业知识的投资者和投资顾问的专业产品，使用时必须经专业人士进行解读。国金证券建议获取报告人员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。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议，国金证券不就报告中的内容对最终操作建议做出任何担保，在任何时候均不构成对任何人的个人推荐。

在法律允许的情况下，国金证券的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供多种金融服务。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。国金证券并不因收件人收到本报告而视其为国金证券的客户。本报告对于收件人而言属高度机密，只有符合条件的收件人才能使用。根据《证券期货投资者适当性管理办法》，本报告仅供国金证券股份有限公司客户中风险评级高于 C3 级（含 C3 级）的投资者使用；本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断。使用国金证券研究报告进行投资，遭受任何损失，国金证券不承担相关法律责任。

若国金证券以外的任何机构或个人发送本报告，则由该机构或个人为此发送行为承担全部责任。本报告不构成国金证券向发送本报告机构或个人的收件人提供投资建议，国金证券不为此承担任何责任。

此报告仅限于中国境内使用。国金证券版权所有，保留一切权利。

上海	北京	深圳
电话：021-80234211	电话：010-85950438	电话：0755-86695353
邮箱：researchsh@gjzq.com.cn	邮箱：researchbj@gjzq.com.cn	邮箱：researchsz@gjzq.com.cn
邮编：201204	邮编：100005	邮编：518000
地址：上海浦东新区芳甸路 1088 号紫竹国际大厦 5 楼	地址：北京市东城区建内大街 26 号新闻大厦 8 层南侧	地址：深圳市福田区金田路 2028 号皇岗商务中心 18 楼 1806

【小程序】
国金证券研究服务

【公众号】
国金证券研究