

新一批国产大模型通过备案，垂直领域大模型商业化加速

—行业跟踪点评

核心观点：

- 事件：**近日，第四批国产大模型通过备案，此批共有 13 家企业共计 14 款国产大模型入列。大模型产品通过备案的企业中，既有一批人工智能企业，如第四范式、零一万物、衍远科技、识因智能等，又有一批深耕垂直领域的厂商，如小米、智联招聘、Boss 直聘、脉脉、什么值得买、微步在线、新壹科技、创思远达、掌阅等。
- 国产大模型由通用走向垂直领域，商业化加速。**2023 年 8 月，首批 8 家企业的大模型通过备案，其中，百度、抖音、百川智能、中科院、商汤、MiniMax、上海人工智能实验室 7 家厂商的大模型参数量均在千亿级，均为通用大模型。2023 年 11 月，第二批 10 家企业通过了大模型备案，既有月之暗面（MoonShot AI）、昆仑万维（天工大模型）、知乎（知海图 AI 大模型）、面壁智能（Luca 大模型）这样的千亿级大模型，也出现了出门问问（序列猴子大模型）、网易有道（子曰大模型）这样的百亿级大模型。而第三批、第四批通过备案的大模型，大部分的参数量均在百亿级以下，聚焦垂直领域。比如什么值得买的“AI 问答机器人”，是一款消费领域的垂类产品，将作为客户的私人购物助手，通过对话的形式为客户搜寻全球好物链接。新壹科技推出的新壹视频大模型，则是目前国内第一款聚焦视频领域的人工智能大模型，其作为多模态视频模型，可以实现脚本生成、素材匹配、AI 视频生成、智能剪辑配音、数字人播报等功能。

“2023 年的主旋律是如何做大模型，2024 年的主旋律是如何用大模型。”千亿级通用大模型拥有多种技能，但是将垂直大模型进行专业化的训练后，其性能在单项技能上不输千亿大模型，且百亿参数大模型的成本更低，性价比更高，更适合企业使用。在解决具体行业的特定问题上，垂直领域的大模型更加具有优势，且企业也逐渐认识到了垂直大模型的应用潜力，积极投入研发与应用，垂直大模型商业化加速。
- 伴随着第四批大模型备案通过，大模型的监管也在持续常态化推进。**《生成式人工智能服务管理暂行办法》是中国首份生成式人工智能（AIGC）监管文件，也是全球第一部 AIGC 管理办法。该文件中提出坚持发展和安全并重，促进创新和依法治理相结合的原则，采取有效措施鼓励生成式人工智能创新发展，对生成式人工智能服务实行包容审慎和分类分级监管。这既展现了监管部门对大模型发展的关注与支持，又表明了对大模型管理和规范的严格要求。这份文件赋予了大模型开发和应用明确的法律依据，有助于在保护用户权益的同时防止技术滥用。截至目前，国内核心企业的大模型产品几乎都获得了备案，理论上讲，获得大模型备案有助于各个厂商大模型产品形成数据飞轮效应，对国产大模型的发展起到了促进作用。
- 展望 2024 年，多模态成为行业热点，AIGC 商业化进程不断推进。**根据工信部赛迪研究院最新数据，2023 年我国大模型市场规模实现较快提升，应用场景不断丰富，预计 2023 年我国大模型市场规模达到 132.3 亿元，增长率达 110%。目前，行业愈发关注多模态大模型。多模态大模型对于 AI 的发展具有重要意义，因为要实现通用人工智能（AGI）需要分析多样化的数据，甚至包

计算机行业

推荐

维持评级

分析师

吴砚靖

☎：(8610) 66568589

✉：wuyanqing@chinastock.com.cn

分析师登记编码：S0130519070001

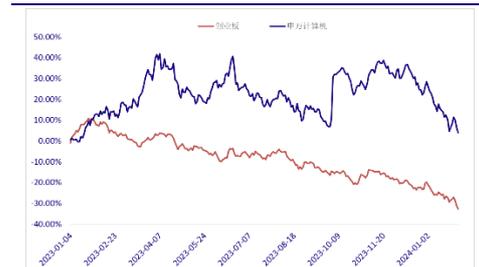
邹文倩

☎：(8610) 86359293

✉：zouwenqian@chinastock.com.cn

分析师登记编码：S0130519060003

相对指数表现图



资料来源：中国银河证券研究院

相关研究

【银河计算机】计算机行业 2024 年度策略：AI 三要素共振，AIGC 云到端加速推进

括触觉。谷歌于 2023 年 12 月发布大模型 Gemini，其在多模态测试中的表现优于 GPT-4V。Gemini 直接用多模态数据（包括文字、音频、图片、视频、PDF 文件等）进行训练，可以无缝地理解和操作上述不同类型的信息，因此是“原生”多模态；而 GPT-4V 是分别训练了纯文本、纯视觉和纯音频模型再进行拼接，因此多模态表现不如 Gemini。而 OpenAI 也将于 2024 年初发布 GPT-5 多模态大模型。多模态大模型的涌现将更加丰富应用场景，使 AI 不再局限于文本和图像的理解，而是将拓展到视频领域，推进 AIGC 商业化进程。

- **投资建议：**我们认为 2024 年大模型应用将进入落地期，一方面垂直领域大模型的商业化应用正在加速，另一方面多模态大模型涌现，应用场景将更加丰富。我们看好 AIGC 的商业化发展，建议关注科大讯飞（002230.SZ）、金山办公（688111.SH）、同花顺（300033.SZ）、萤石网络（688475.SH）、卫宁健康（300253.SZ）、中科创达（300496.SZ）、拓尔思（300229.SZ）、虹软科技（688088.SH）、万兴科技（300624.SZ）。
- **风险提示：**全球宏观经济低迷的风险；下游需求萎靡的风险；监管加强的风险；AIGC 商业化进度不及预期的风险。

附录：

表 1：四批通过备案的大模型基本情况

批次（通过日期）	模型名称	参数量	公司	应用场景
第一批（2023 年 8 月）	文心一言	2600 亿	百度	应用于智能问答场景，服务于对话问答类企业端客户，根据用户提出的问题，结合上下文，生成相应的文本或多模态回答。
	云雀大模型	1300 亿	抖音	应用于智能对话、智能写作场景，根据用户输入的文本信息，生成相对应的答复。可用于日常闲聊，具有一定创造力，能根据聊天内容创作段子、小诗等。
	GLM 大模型	60 亿	智谱 AI	应用于对话生成场景，根据用户输入的文本内容，应用对话模型，生成对话文本回复。“智谱清言”已具备通用问答、多轮对话、创意写作、代码生成以及虚拟对话等丰富能力，未来还将开放多模态等生成能力。
	百川大模型	超千亿	百川智能	应用于游戏、影视等诸多领域数字角色代码集成和构建，同时将支持医疗应用。搜索增强知识库和 192K 超长上下文窗口的结合，让模型可以连接全部企业知识库以及全网信息。
	日日新大模型	千亿级	商汤	适用于逻辑推理、文本续写、文案生成、内容摘要、代码生成、知识检索等大多数文本处理相关的场景，并且还提供声音大模型能力。
	ABAB 大模型	千亿级	MiniMax	适用于大多数文本处理相关的场景，如逻辑推理、文本续写、文案生成、文本扩写、文本改写、内容摘要、代码生成、知识检索等。此外，MiniMax 开放平台还提供声音大模型能力，可应用于配音和音频交互等场景中。目前，MiniMax 开放平台已接入超百家付费客户，提供办公协作、互动娱乐、客服、搜索、教育等十余个行业场景。
	紫东太初大模型	1000 亿	中科院	支持知识问答和开放领域闲聊。紫东太初理解能力强，可进行多轮交互。拥有视觉、文本、语音多种模态，基于全栈国产化基础软硬件平台建立。目前，“紫东太初”已在文本创作、医疗、物流、政务等多个行业实现应用落地。
	书生通用大模型	千亿级	上海人工智能实验室	支持开放域对话，具备一定的理解和创新能力。
	第二批（2023 年 11 月）	未对外公布名称		美团
百灵大模型		万亿级	蚂蚁集团	全线应用到蚂蚁的各个业务场景，并在各垂直领域创新研发产品，更多应用产品即将面向社会开放。
知海图 AI		千亿级	知乎	具备大语言模型能力的“知海图 AI”还具备强大的多模态处理能力，助力知乎实现会员故事智能配图，赋能会员业务发展。同时还运用到了知乎教育业务中，开展了基于大模型的产品创新，通过 AI 智能申论批改、智能作业批改、AI 教务领航员，极大提升了教学效率和教学质量。
序列猴子大模型		百亿级	出门问问	应用于智能问答、智能写作场景，服务于企业端客户，提供根据输入的文本诉求，生成与之相关的文本内容的功能。支持文字、语音、图片、3D 内容生成等多个任务。核心能力覆盖“知识、对话、数学、逻辑、推理、规划”六个维度。
天工大模型		双千亿级	昆仑万维	应用于智能对话场景，根据输入的文本诉求，通过大模型，生成输入相关的文本回复。拥有强大的自然语言处理和智能交互能力，能够实现个性化 AI 搜索、智能问答、聊天互动、文本生成、编写代码、语言翻译等多种应用场景。
Moonshot AI 大模型		千亿级	月之暗面	具备较强的多语言、长文本分析能力。其在中文上具备显著优势，实际使用效果能够支持约 20 万汉字的上下文，其能够清晰地梳理一部长篇小说的主要情节、主要人物和人物的命运，也可以根据用户上传的财报进行关键数据和公司发展情况的分析等。
面壁 Luca 大模型		千亿级	面壁智能	多模态智能对话助手。经过多次迭代更新，Luca 不仅在中英文语言对话方面表现出色，还具备强大的代码、知识、逻辑及图片理解能力。
子曰大模型		百亿级	网易有道	应用于智能对话场景，根据用户输入的文本数据，结合场景需求，通过大语言模型，生成符合用户意图的回答。
九章大模型(MathGPT)		千亿级	好未来	应用于在线教育场景，服务于企业端客户，根据输入的语音，结合语音识别模型与打分模型，输出语音得分。结合认知规律、题库资源、解题算法等核心优势，展示了该模型在

				多个场景中的应用成果，覆盖口语训练、作文批改、习题解答等六大领域。
	WPS AI	超千亿级	金山办公	应用于文本生成场景，根据用户输入的文本信息，结合用户意图，实现文档写作、阅读理解、概括总结等功能。
	360 奇元大模型	千亿级	360 公司	生成与创作、多轮对话、代码能力、逻辑与推理、知识问答、阅读理解、文本分类、翻译、文本改写、多模态。
第三批（2023 年 12 月）	言犀大模型	千亿级	京东	在线咨询机器人、智能语音外呼、智能语音应答
	福禄瓜视觉大模型		抖音	
	快意大模型	多种参数规模	快手	内容创作、信息咨询、数学逻辑、代码编写、多轮对话等广泛任务
	小冰大模型	混合模型(大中小模型有机组合)	红棉小冰科技	小冰克隆人、歌手克隆人分支、小冰数字员工等
	CharacterGLM	百亿级	聆心智能	定制中文 AI 角色
	孟子 GPT	百亿级	澜舟科技	智能客服、智慧营销、企业知识问答、机器翻译、智能写作等。
	雅意大模型	百亿级	中科闻歌	实时联网问答、领域知识问答、多语言内容理解、复杂场景信息抽取、多模态内容生成
	语鲸大模型	80 亿	深言科技	具备强大的语义理解与文本生成、代码理解与生成、逻辑推理等能力，提供文本创作、数学计算、代码补全等服务
	山海大模型	千亿级	云知声	智慧医疗、智慧教育、知识管理、智慧营销、智能客服
第四批（2024 年 1 月）	零一万物大模型	60 亿、340 亿	零一万物	处理知识库信息、生成文本、文档处理
	式说大模型	千亿级	第四范式	营销/销售、产品创新、业务助手
	品商大模型	百亿级		市场解读、产品创新、内容设计、营销策划、智能导购、知识小秘书
	慕小仙大模型	35 亿	衍远科技	专为服装、美妆、食品饮料等行业设计。生成高度可控的商品形态、光影效果，以及模特形象，为用户提供模特定制、人台变超模等创新服务。
	一叶轻舟大模型	70 亿、130 亿、330 亿	识因智能	在企业场景、政务场景和消费场景中成为数字化转型与智能化应用的核心技术引擎
	新壹视频大模型	文本数据百亿级，图片视频数据 20 亿左右	新壹科技	脚本生成、素材匹配、AI 视频生成、智能剪辑配音、数字人播报
	魔方大模型	百亿级	中科创达	在智能汽车和机器人领域将操作系统与人工智能技术相结合，以实现更高效、更智能的数据处理和决策
	微步情报智脑大模型		微步在线	将网络安全大模型与微步 99.99%高精度威胁情报能力及实战化安全检测能力相结合，具备对常见告警、日志等数据的分析能力，可达到中级安全分析师水平
	南北阁大模型	百亿级	BOSS 直聘	以 BOSS 直聘平台官方网站、官方移动 App、微信小程序等相关产品为载体，向用户提供的生成式 AI 服务
	“AI 改简历”新功能		智联招聘	可以根据用户初步填写的简历版本进行再加工，自动生成较规范的简历内容。
	“智能问答”新功能		脉脉	拥有 AI 问答能力，可以在社区提问中做出实时的答复，可帮助用户求职以及获取职场知识
	“小爱同学 AI 助手”新功能		小米	基于小米的硬件生态和海量数据，为用户提供语音识别、NLP 等完整的 AI 技术服务，可以实现答疑解惑等功能
	AI 问答机器人		什么值得买	提供商品百科、品牌导航等场景应用
	“阅爱聊”微信小程序		掌阅科技	提供“聊书”、“聊人”两大场景，可与系统中的形象对话，用该形式阅读名著或小说。

资料来源：各企业官网、网信办、数据观，中国银河证券研究院

分析师承诺及简介

本人承诺以勤勉的执业态度，独立、客观地出具本报告，本报告清晰准确地反映本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与、未来也将不会与本报告的具体推荐或观点直接或间接相关。

吴砚靖 TMT/科创板研究负责人

北京大学软件项目管理硕士，10年证券分析从业经验，历任中银国际证券首席分析师，国内大型知名PE机构研究部执行总经理。具备一二级市场经验，长期专注科技公司研究。

邹文倩 计算机/科创板团队分析师

复旦大学金融硕士，复旦大学理学学士；2016年加入中国银河证券研究院；2016年新财富入围团队成员。

免责声明

本报告由中国银河证券股份有限公司（以下简称银河证券）向其客户提供。银河证券无需因接收人收到本报告而视其为客户。若您并非银河证券客户中的专业投资者，为保证服务质量、控制投资风险、应首先联系银河证券机构销售部门或客户经理，完成投资者适当性匹配，并充分了解该项服务的性质、特点、使用的注意事项以及若不当使用可能带来的风险或损失。

本报告所载的全部内容只提供给客户做参考之用，并不构成对客户投资咨询建议，并非作为买卖、认购证券或其它金融工具的邀请或保证。客户不应单纯依靠本报告而取代自我独立判断。银河证券认为本报告资料来源是可靠的，所载内容及观点客观公正，但不担保其准确性或完整性。本报告所载内容反映的是银河证券在最初发表本报告日期当日的判断，银河证券可发出其它与本报告所载内容不一致或有不同结论的报告，但银河证券没有义务和责任去及时更新本报告涉及的内容并通知客户。银河证券不对因客户使用本报告而导致的损失负任何责任。

本报告可能附带其它网站的地址或超级链接，对于可能涉及的银河证券网站以外的地址或超级链接，银河证券不对其内容负责。链接网站的内容不构成本报告的任何部分，客户需自行承担浏览这些网站的费用或风险。

银河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。银河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。

银河证券已具备中国证监会批复的证券投资咨询业务资格。除非另有说明，所有本报告的版权属于银河证券。未经银河证券书面授权许可，任何机构或个人不得以任何形式转发、转载、翻版或传播本报告。特提醒公众投资者慎重使用未经授权刊载或者转发的本公司证券研究报告。

本报告版权归银河证券所有并保留最终解释权。

评级标准

评级标准	评级	说明
评级标准为报告发布日后的6到12个月行业指数（或公司股价）相对市场表现，其中：A股市场以沪深300指数为基准，新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准，北交所市场以北证50指数为基准，香港市场以摩根士丹利中国指数为基准。	行业评级	推荐：相对基准指数涨幅10%以上
		中性：相对基准指数涨幅在-5%~10%之间
		回避：相对基准指数跌幅5%以上
公司评级	推荐：相对基准指数涨幅20%以上	
	谨慎推荐：相对基准指数涨幅在5%~20%之间	
	中性：相对基准指数涨幅在-5%~5%之间	
	回避：相对基准指数跌幅5%以上	

联系

中国银河证券股份有限公司 研究院

深圳市福田区金田路3088号中洲大厦20层

上海浦东新区富城路99号震旦大厦31层

北京市丰台区西营街8号院1号楼青海金融大厦

公司网址：www.chinastock.com.cn

机构请致电：

深广地区：程曦 0755-83471683 chengxi_yj@chinastock.com.cn

苏一耘 0755-83479312 suyiyun_yj@chinastock.com.cn

上海地区：陆韵如 021-60387901 luyunru_yj@chinastock.com.cn

李洋洋 021-20252671 liyangyang_yj@chinastock.com.cn

北京地区：田薇 010-80927721 tianwei@chinastock.com.cn

唐嫚玲 010-80927722 tangmanling_bj@chinastock.com.cn