

芯跑科技 | 月报资讯

2024 年 4 月，第 051 期：国内大模型面面观



国内大模型面面观

前言：

自 ChatGPT 横空出世以来，AGI（通用人工智能）的梦想对于这一代的创业者来说，似乎不再是空中楼阁，而是可望又可及的远方。国内大模型创业企业蜂拥而上，如雨后春笋般，不停冒出，以至于 2023 年被称之为“百模大战”的一年。声量很大、噪声很多，可以说是国内大模型行业的一个形象的描述。为了方便读者对国内大模型的行业现状及未来有个基本认知，本文尝试从三个角度来对国内大模型产业进行介绍，包括国内大模型的能力评测、商业能力和各厂商对 AGI 的愿景进行分析介绍。

一、国内大模型能力测评

1、国内大模型厂商一览

2023 年 4 月 11 日，国家互联网信息办公室发布的关于《生成式人工智能服务管理办法（征求意见稿）》提出，利用生成式人工智能产品向公众提供服务前，应当按照《具有舆论属性或社会动员能力的互联网信息服务安全评估规定》向国家网信部门申报安全评估，并按照《互联网信息服务算法推荐管理规定》履行算法备案和变更、注销备案手续。同时，利用生成式人工智能生成的内容应当真实准确，采取措施防止生成虚假信息。

2023 年 8 月 31 日，首批通过备案的超过 8 个 AI 大模型对外公布。第一批大模型产品获批名单中，既有巨头企业，也包含了创业公司与科研院校。具体的企业机构包括：

产品名称	所属公司	发布时间	产品特点	价格	其他信息
文心一言	百度	2023 年 3 月 16 日	首个生成式 AI 产品，包含基础版和专业版，4.0 版本是最强大的文心大模型，与 GPT-4 相比毫不逊色	59.9 元/月（专业版单月购买），49.9 元/月（连续包月优惠价），基础版免费	百度 CEO 李彦宏评价
百川	百川智能	2023 年	发布了多款开源和闭源大模型，每 28 天发布一款新的大模型，Baichuan-7B/13B 两款开源大模型在多个权威评测榜单名列前茅，累计下载量超过六百万次	免费	创始人王小川
日日新 SenseNova	商汤科技	2023 年 4 月	提供自然语言、内容生成、自动化数据标注、自定义模型训练等多种能力，包括商量 SenseChat、秒画 SenseMirage、如影 SenseAvatar 等功能强大且易用的内容生成平台产品	未提及	最新版大模型“商量 SenseChat 2.0 版本”已发布，参数达千亿级规模

云雀	抖音集团	2023 年 8 月 17 日	基于云雀模型开发的 AI 工具，提供聊天机器人、写作助手以及英语学习助手等功能，支持网页 Web 平台，iOS 以及安卓平台	未提及	iOS 需要使用 TestFlight 安装
智谱清言	智谱 AI	2023 年	包括双语千亿级超大规模预训练模型 GLM-130B，中英双语对话模型 ChatGLM、认知大模型平台 Bigmodel.ai，包括 CodeGeeX 和 CogView 等产品	未提及	与多家国内互联网巨头，政企达成合作
紫东太初	中科院自动化研究所	2021 年	定位为跨模态通用人工智能平台，参数已达千亿级别，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	未提及	2023 年 6 月发布 2.0 版本，加入了视频、传感信号、3D 点云等新的模态数据
ABAB	Minimax	未提及	文本到视觉（text-to-visual）、文本到语音（text-to-audio）、文本到文本（text-to-text）三个模态的基础模型架构	未提及	自研通用大模型
书生（INTERN）	上海人工智能实验室	2021 年	从国内首个覆盖多种视觉任务的通用大模型，逐步升级为涵括视觉、语言和三维等在内的大模型体系	未提及	提供了首个面向大模型研发与应用的全链条开源体系
通义千问	阿里云	2023 年 4 月	2.0 版本在复杂指令理解、文学创作、通用数学、知识记忆、幻觉抵御等能力上均有显著提升，综合性能已经超过 GPT-3.5，加速追赶 GPT-4	未提及	2023 年 10 月 31 日发布 2.0 版本，通义千问 APP 在各大手机应用市场正式上线

监管部门对大模型服务提供商颁布的许可备案以地域划分。除了以上几家外，腾讯混元、华为盘古属于广东备案。科大讯飞星火大模型和阿里云通义千问大模型也宣称完成备案。

2023 年 11 月 4 日，第二批通过备案的 11 款 AI 大模型产品对外公布。2023 年 12 月 28 日，第三批 AI 大模型备案落地。2024-02-01，第四批 AI 大模型备案落地。截止目前，国内已经有超过 40 款 AI 大模型产品获得了备案审批。

产品名称	所属公司	特点	其他信息
子曰	网易有道	教育垂类大模型，专业预训练语料，提供对话式大语言模型，快速迭代升级	搭载在智能硬件产品及 App 中，为全年龄段学习者提供高效学习体验
面壁露卡 Luca	面壁智能	多模态智能对话助手，基于自研千亿参数基座模型 CPM，具备代码、知识、逻辑及图片理解能力	经过多次迭代更新
序列猴子	出门问问	能力体系涵盖“知识、对话、数学、逻辑、推理、规划”，AI 写作应用「奇妙文」，支持多种 AI 写作功能	用户可在写作过程中一键唤醒 AI 帮助创作，支持图文并茂
天工	昆仑万维	双千亿级大语言模型，AI 搜索引擎，对话式 AI 助手，强大的自然语言处理和智能交互能力	应用场景广泛，涵盖科学、技术、文化、艺术、历史等领域
AI 大模型	美团	扩张算法团队，筹划平台部门，探索视觉、语音、智能交互、自然语言处理能力	今年 5 月发布招聘对话大模型应用产品岗位

知海图 AI	知乎	千亿级参数，基于成熟大模型 CPM-Bee，具备更强的逻辑推理能力，更快的训练和推理速度	业务提效显著，降低人工标注量，提升业务准召效果
Moonshot	月之暗面	2023 年 10 月推出，支持输入 20 万汉字的智能助手产品 KimiChat，多语言能力强	在中文上具备显著优势，支持约 20 万汉字的上下文
WPS AI	金山办公	类 ChatGPT 式应用，接入 WPS 多款产品，解决日常办公难题，提供简洁的交互框架和 AI 功能	构建稳定的 Office 基建服务
MathGPT	好未来	数学计算能力覆盖中小学数学题，支持题目追问，提供对话式解答反馈	用户可用文字或图片上传数学题，支持中文、英文版本
奇元	360	大模型定位灵活，以 B 端用户为主，聚焦商业化应用和垂直领域，提升工作效率	360 智脑大模型已全面接入 360 搜索、360 浏览器等
百灵	蚂蚁集团	采用 Transfromer 架构，基于万亿级 Token 语料训练，支持长窗口推理，全栈布局	在算力、安全力和知识力上表现突出
言犀	京东	产业研发，融合通用数据与原生数据，高产业属性，强泛化能力，多安全保障	解决真实产业问题，已在多个供应链场景中试点接入
山海	云知声	十大能力，包括语言生成、理解、知识问答等，支持多语种、多文体、多风格的语言生成任务	支持可控文本生成，满足多种约束条件
CharacterGLM	聆心智能	超拟人大模型，具备“六边形能力”，使 AI 拥有个性和情感，立体化人格	基于 CharacterGLM 研发了 AiU 兴趣互动社区
孟子 GPT	澜舟科技	发布多款大模型，生成式可控大语言模型，多轮帮助用户完成工作任务，支持迁移学习和定制垂直模型	满足专业领域任务，支持融入行业数据和知识图谱
雅意	中科闻歌	企业级专属大模型，准确回答日常和专业领域问题，多种特色技能，支持实时在线和离线私有部署	为媒体行业用户提供专属领域大模型应用服务
语鲸	深言科技	孵化自清华大学自然语言处理实验室，强大语义理解与文本生成、代码理解与生成、逻辑推理能力	更懂中文，更懂中文使用场景，C-EVAL 评测榜单第二名
福禄瓜	抖音	视觉大模型，文生图和图生图能力，理解语言、感知美感，提供图像风格化、素材生产等服务	广泛应用于内容创作、广告营销、教育和培训等领域
快意	快手	系列大模型产品，预训练模型和对话模型，出色的语言理解和生成能力，支持广泛任务	快手 App 内测“AI 对话”功能，提供多种基于短视频应用场景的服务
小冰	红棉小冰科技	多模态大模型，理解人类用户交互中的情感表达，实现情商与智商兼备的交互体验	广泛应用于 B 端数字员工和 C 端泛娱乐类情感陪伴产品
小爱同学 AI 助手	小米	基于小米硬件生态和海量数据，提供语音识别、NLP 等 AI 技术服务	2023 年 8 月大模型加持，内测内置交互模型
式说	第四范式	生成语言能力，解决 GPT 技术应用时内容可信、数据安全和成本问题	式说 1.0-3.0 不断迭代，增强多模态输入输出和企业级 Copilot 能力
Yi-34B 和 Yi-6B	零一万物	开源大模型，双语（英文/中文），支持长上下文窗口	Yi-6B 适合个人及研究用途，Yi-34B 具备大模型涌现能力
魔方	创思远达	物联网上市公司中科创达旗下，覆盖边缘端、语言大模型、多模态、机器人等	与车厂客户合作，应用于智能座舱的语音助手、3D 图形图像
微步情报智脑 XGPT	步刻科技	网络安全大模型与微步威胁情报能力结合	具备安全知识、威胁分析、安全运营三大核心能力
南北阁	BOSS 直聘	以 BOSS 直聘平台为载体	平台仍在内测阶段，需要激活才能服务
品商	衍远科技	数百亿级参数规模，理解人与商品关系	提供从市场洞察到营销创新的全链路支持
摹小仙	衍远科技	多模态 AI 模型，专为服装、美妆、食品饮料等行业设计	引入自研的 Protected-SD 技术，提供版权授权、生成与追溯保护

新壹视频	新壹科技	国内首款聚焦视频领域的人工智能大模型	具备视频脚本生成、素材匹配原生 AI 视频生成等全流程 AIGC 视频生产能力
一叶轻舟	识因智能	生成式大语言模型产品	330 亿、130 亿和 70 亿参数量级商用推理模型，成为北京市石景山区通用人工智能大模型产业集聚区首批落地的重点项目
AI 问答机器人	什么值得买	对话形式与用户进行交互，提供智能的消费辅助决策服务	帮助用户快速甄选全网优质好物链接
智能问答	脉脉	职场招聘平台，在 ChatGPT 爆火后推出 AI 功能产品	社区提问中做出实时答复，帮助求职及获取职场知识
AI 改简历	智联招聘	根据用户初步填写的简历版本进行再加工	求职者写一份不到 100 个字的简历，后台系统可以把它变成 400 多个匹配字段
阅爱聊	掌阅	AI 对话机器人式产品	提供“聊书”、“聊人”两大场景，可和系统中既定角色形象对话，支持多种语言情感风格，支持角色和读书双重模式阅读

可以看到，大模型赛道已成为云厂商、互联网厂商、人工智能企业、创业企业等各路选手的“兵家必争之地”，各家各凭本事，没有一家甘居人后，大模型刷榜层出不穷。

2、不同的测试标准

目前，主流的大模型测试策略有以下几种：

测试策略类别	具体测试方法
基准测试	通用语言理解评估基准（GLUE）：包括多种语言理解任务，如句子相似性、语义角色标注等。
	斯坦福问答数据集（SQuAD）：评估模型在阅读理解和问答任务中的表现。
	机器阅读理解挑战赛（MRC）：考察模型在阅读理解任务中的表现。
	自然语言推理任务（SNLI）：评估模型在自然语言推理任务中的表现。
	语言模型评估基准（BERT-Score）：评估模型生成的文本与人类生成的文本之间的相似度。
特定任务测试	文本摘要：评估模型在生成简洁、准确的摘要方面的能力。
	情感分析：评估模型在判断文本情感倾向方面的能力。
	对话系统：评估模型在多轮对话中的表现，包括生成回复、理解用户意图等。
	代码理解和生成：评估模型在理解和生成代码方面的能力。
人类评估	专家评估：邀请领域专家对模型的输出进行评价。
	用户调查：通过问卷调查或直接与用户互动来评估模型的性能。
自定义评估	业务场景模拟：在特定的业务场景中评估模型的性能，如电商推荐、金融分析等。
	定制化测试集：根据特定需求设计测试集，以评估模型在该场景下的表现。
持续迭代	实时反馈：收集用户使用过程中的实时反馈，不断优化模型。
	迭代优化：根据评估结果对模型进行迭代优化，提高性能。
跨领域评估	多任务学习：评估模型在不同任务上的表现，以考察其泛化能力。
	迁移学习：在多个领域上评估模型的迁移学习能力。
安全性和合规性评估	内容过滤：评估模型在过滤不当内容方面的能力。
	隐私保护：评估模型在处理敏感信息时的安全性。
成本效益分析	计算资源消耗：评估模型在生成或理解文本时所需的计算资源。
	经济效益：评估模型在实际应用中的经济效益。

在评价大模型能力时，通常需要结合多种策略，并根据具体的应用场景和需求来选择最合适的评价方法。

估方法。同时，评估结果也需要与实际业务目标相结合，以全面评价大模型的性能。

在具体测试中，会考虑以下几方面的能力：

能力类别	具体能力
文本生成能力	流畅性：模型生成的文本是否通顺、连贯。
	多样性：模型生成的文本是否多样化，能否生成不同风格、体裁的文本。
	创造性：模型能否生成新颖、有创意的文本。
语言理解能力	准确性：模型对输入文本的理解是否准确。
	上下文理解：模型是否能理解文本的上下文，并据此作出正确反应。
	推理能力：模型是否能进行逻辑推理和抽象思维。
多语言能力	多语言支持：模型是否支持多种语言。
	语言翻译质量：模型在不同语言之间的翻译质量。
多模态能力	多模态融合：模型是否能处理文本以外的其他模态数据，如图像、音频等。
	模态理解：模型对不同模态数据的处理和理解能力。
安全性与可控性	内容安全性：模型生成的内容是否符合社会主义核心价值观，是否过滤不良信息。
	可解释性：模型决策过程是否可被理解和解释。
	鲁棒性：模型对错误输入的容忍度和鲁棒性。
效率与可扩展性	计算效率：模型在生成或理解文本时的计算资源消耗。
	可扩展性：模型是否能够在大规模数据集上训练和优化。
用户体验	交互友好性：模型与用户的交互是否自然、友好。
	个性化服务：模型是否能够根据用户需求提供个性化服务。
实际应用场景	应用适应性：模型在特定行业或应用场景中的适应能力。
	商业价值：模型在实际应用中的商业价值和市场潜力。

不同的测评标准可能会根据具体应用场景和用户需求有所侧重。在实际操作中，大语言模型的评估往往需要结合多种标准，并且需要考虑到模型的实际表现与预期的业务目标之间的匹配度。

3、测试结果对比

有人的地方，就有江湖，大模型赛道亦然。自 ChatGPT 崛起以来，评测成为大模型发展的关键。尽管国内外存在多个模型能力评测榜单，但它们的质量参差不齐，排名差异显著。国内众多模型，家家声称「第一」，但实际应用时却现原形。但综合各家榜单来看，也能管中窥豹，了解国内各大模型的基础实力。

本文选取三个评测结果，分别是评测平台 SuperBench、C-Eval 和 HuggingFace 的开源大模型排行榜 (Open LLM Leaderboard)。

1) SuperBench 结果

为提供客观、科学的评测标准，清华大学基础模型研究中心联合中关村实验室研制了 SuperBench 大模型综合能力评测框架，旨在推动大模型技术、应用和生态的健康发展。

SuperBench 提出了五项大模型原生评测基准，包括：模型安全能力评测 SafetyBench、模型智能体性能评测 AgentBench、模型长文本性能评测 LongBench、模型对齐性能评测 AlignBench 以及模型工具调用能力评测 ToolBench 等。

5 项大模型原生评测基准

评测基准	简介	特色
AlignBench	首个评估人类对齐能力的基准数据集	<ul style="list-style-type: none"> ✓ 来自智谱清言的真实使用场景 ✓ 直接检验LLM的生成能力 ✓ 覆盖8大能力，100+子类别
LongBench	首个评估长文理解能力的双语基准数据集	<ul style="list-style-type: none"> ✓ 6大长文本理解场景 ✓ 数据长度长达5K-15K ✓ 中英双语，题量近5K
ToolBench	首个大规模LLM工具学习能力基准数据集	<ul style="list-style-type: none"> ✓ 覆盖3.5K个工具，16,000+个真实 API ✓ 支持单工具和多工具场景
AgentBench	首个评估LLM作为Agent的能力基准数据集	<ul style="list-style-type: none"> ✓ 8大典型的智能体使用场景 ✓ 评估决策和执行能力
SafetyBench	首个中英文双语大模型安全基准数据集	<ul style="list-style-type: none"> ✓ 7种典型安全场景 ✓ 选择题形式，支持快速评估 ✓ 中英双语，题量达到11K

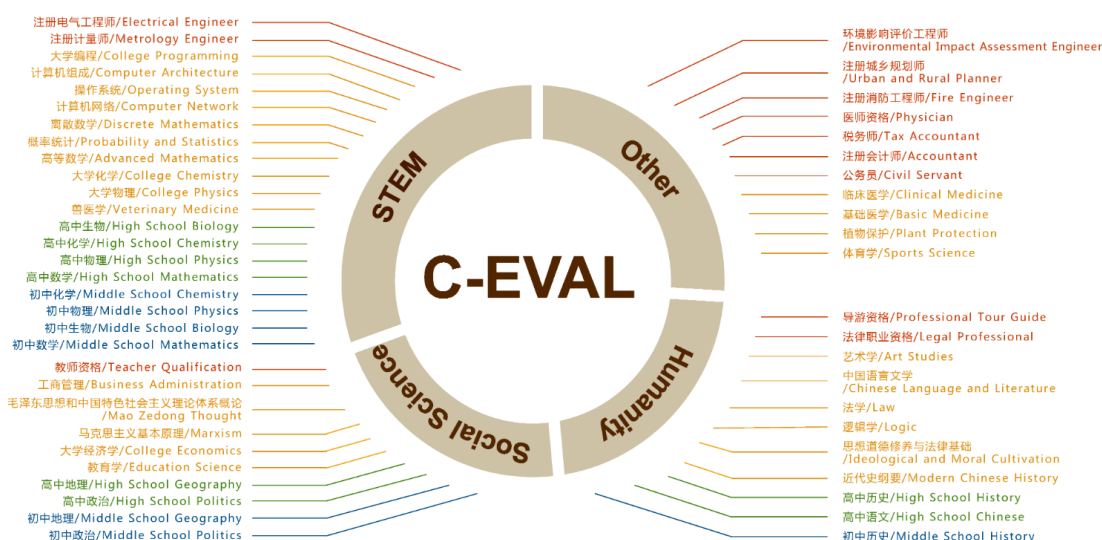
根据 2024 年 3 月的《SuperBench 大模型综合能力评测报告》，SuperBench 得出以下结论：

整体来说，GPT-4 系列模型和 Claude-3 等国外模型在多个能力上依然处于领先地位，国内头部大模型 GLM-4 和文心一言 4.0 表现亮眼，与国际一流模型水平接近，且差距已经逐渐缩小。

国内大模型中，GLM-4 和文心一言 4.0 在本次评测中表现最好，为国内头部模型；通义千问 2.1、Abab6、moonshot 网页版以及 qwen1.5-72b-chat 紧随其后，在部分能力评测中亦有不俗表现；但是国内大模型对比国际一流模型在代码编写、作为智能体两个能力上依然有较大差距，国内模型仍需努力。

2) C-Eval 评测结果

C-Eval 榜单是由清华大学、上海交通大学和爱丁堡大学合作构建的面向中文语言模型的综合性考试评测集，涵盖 52 个不同学科，共有 13948 道多项选择题，是目前较为权威的中文 AI 大模型评测榜单。CMMLU 数据集则是一个综合性的中文评估基准，由 MBZUAI、上海交通大学、微软亚洲研究院共同推出，在评估语言模型在中文语境下的知识和推理能力方面极具权威性。C-Eval 是一个全面的中文基础模型评估套件。它包含了 13948 个多项选择题，涵盖了 52 个不同的学科和四个难度级别，如下所示：



C-Eval 从知识型任务出发，通过不同数量集，将模型能力分类以进行准确衡量。其中，C-Eval Hard 自己从强推理的学科专门抽出来，用来衡量模型的推理能力，这部分内容要求模型首先拥有数学知识，然后有逐步的解题思路，并具有各种复杂计算的能力，并用一定格式表达，这部分题非常难。C-Eval HARD 基准也是中国第一个提供复杂推理问题的基准。

C-Eval 研发团队介绍，希望 C-Eval 可以在整体上对标 MMLU (这个数据集被用于 GPT-3.5, GPT-4, PaLM, PaLM-2, Gopher, Chinchilla 的研发)，希望在 Hard 的部分对标 MATH (这个数据集被用于 GPT-4, PaLM-2, Minerva, Galactica 的研发)。**根据官网榜单，可以看到国内大模型在特定测试集上已经可胜过 GPT-4。**榜单本身只测模型潜力，不测真实用户感受。要模型真的被用户喜好，还是需要大量的人工评价的。

公开访问的模型

#	模型名称	发布机构	访问方式	提交时间	平均 v	平均(Hard)	STEM	社会科学	人文科学	其他
0	Qwen-72B	Alibaba Cloud	Weight	2023/11/30	82.8	64.7	77.1	91.7	84.7	82.9
1	Yi-34B	零一万物	Weight	2023/11/2	81.4	58.7	73.7	89.6	84.6	84.9
2	APUS-xDAN 大模型 4.0 (MoE) 136B	新巨智能 & APUS	Weight	2024/4/22	81.3	63.1	76.3	87.5	81.8	84.3
3	TuringMM-34B-Chat	北京光年无限科技有限公司	Weight	2024/2/27	80.7	60.2	73.8	89.3	82.3	83.7
4	Linly-Chinese-LLaMA2-70B	深圳大学大数据系统计算技术国家工程实验室 & APUS	Weight	2024/2/3	80.6	63	76	87.2	80	83.4
5	PCI-TransGPT	佳都科技	API, Web	2024/1/4	80.4	62.5	75.4	89.2	81.7	80.3
6	Taichu-70B	紫东太初	Weight	2024/1/12	80.1	59.8	73.8	89.5	82.9	80.4
7	OrionStar-Yi-34B-Chat	OrionStarAI	Weight	2023/11/22	78.1	55.8	70.1	88	80.7	80.9
8	XuanYuan-13B	度小满AI-Lab	Weight	2024/2/2	76.8	59	71.3	86.5	80.1	74.9
9	YAYI2-30B	中科闻歌	Weight	2023/12/18	75.3	53.1	67.2	83.8	80.6	76.8
10	XuanYuan-6B	度小满AI-Lab	Weight	2024/2/2	74.4	58	69.5	84.5	76.8	71.9
11	xDAN-L2-Chat-lite-v1.0	xDAN-AI	API, Private	2023/12/17	74.3	50.7	66.5	84.8	78.1	75.3
12	BlueLM-7B	vivo	Weight	2023/11/7	73.3	48.9	64.3	83.3	76.5	77.1
13	XuanYuan2-70B	度小满AI-Lab	Weight	2024/2/2	72.7	53.1	67.2	84.2	75.8	69
14	XVERSE-65B-2	XVERSE Technology	Weight	2023/12/8	72.4	50.8	65.7	85	74	71.8
15	Qwen-14B	Alibaba Cloud	Weight	2023/9/22	72.1	53.7	65.7	85.4	75.3	68.4
16	Yi-6B	零一万物	Weight	2023/11/2	72	46.6	62.3	83.9	76.3	74.6
17	XuanYuan-70B	度小满AI-Lab	Weight	2023/9/21	71.9	53.6	67.7	83.3	73.9	67.4
18	ChatGLM3-6B-base	Tsinghua & Zhipu AI	Weight	2023/10/26	69	46.8	61	82.4	73.4	66.9
19	GPT-4*	OpenAI	API, Web	2023/5/15	68.7	54.9	67.1	77.6	64.5	67.8
20	XVERSE-65B	XVERSE Technology	Weight	2023/11/5	68.6	46.2	61.3	81.4	71	67.8
21	Aquila2-70B-Expr	北京智源人工智能研究院	Weight	2023/11/27	66.8	47.2	61.6	79.7	69.4	62
22	Nanbeige-16B-Base	Nanbeige LLM Lab	Weight	2023/11/8	63.8	43.5	57.8	77.2	66.9	59.4
23	LingoWhale-8B	深言科技(DeepLangAI)	Weight	2023/11/3	63.6	46.4	57	73.7	68.5	61.5
24	Qwen-7B v1.1	Alibaba Cloud	Weight	2023/9/12	63.5	46.4	57.7	78.1	66.6	57.8
25	XVERSE-13B-2	XVERSE Technology	Weight	2023/11/4	63.5	41.6	55.7	77.3	66	62.7
26	TeleChat	中国电信人工智能科技有限公司	Weight	2024/1/8	63.1	45.5	59.7	76.1	63.4	57.3
27	Alaya-7B-Base	北京九章云极科技有限公司 DataCanvas Limited	Weight	2023/12/1	62.8	32.5	54.1	75.9	63.2	66.3
28	Erlangshen-UniMC-1.3B	IDEA研究院	Weight	2023/8/4	61	36.7	49.6	74.9	70.7	59.4
29	Qwen-7B	Alibaba Cloud	Weight	2023/7/29	59.6	41	52.8	74.1	63.1	55.2
30	BatGPT-15b-sirius-v2	SJTU & WHU	Weight	2023/8/4	57.4	36.9	50.5	72.1	60.7	53.3

3) HuggingFace 的开源大模型排行榜 (Open LLM Leaderboard)

Open IM Leaderboard 是最大的大模型和数据集社区 Huggingface 推出的开源大模型排行榜单，基于 Eleuther AI Language ModelEvaluation Hanes (Eleuther AI 语言模型评估框架)封装。

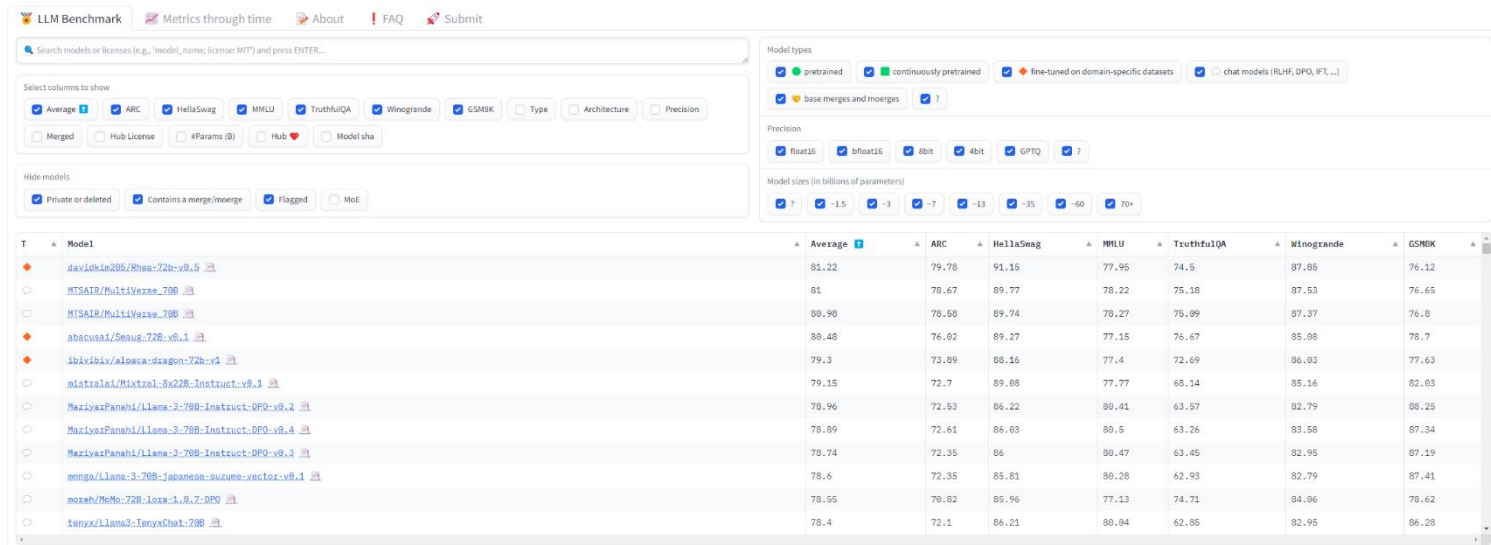
由于社区在发布了大量的大型语言道型(LM)和即天机器人之后，往往伴随看对其件能的夸大宣传，很难过源出开源社区取得的真正讲展以及目前 89 最先进模型。因此，Hugging face 使用 leuther A)语言模型评估框架对模型进行四个关键基准测试评估。这是一个统一的框架，用于在大量不同的评估任务上测试生成式语言模型。

Open LLM Leaderboard 的评估基准：

- A12 推理挑战(25-shot):一组小学科学问题
- HellaSwag(10-shot):一个测试常识推理的任务，对人类来说很容易(大约 95%)，但对 SOTA 模型来说具有挑战性。
- MML(5-shot)-用于测量文本模型的多任务准确性。测试涵盖 57 个任务，包括基本数学、美国历史、计算机科学、法律等等。
- TruthfulQA(0-shot)-用于测量模型复制在在线常见虚假信息中的倾向性

其最新公布榜单如下：

👉 Open LLM Leaderboard



T	Model	Average	ARC	HellaSwag	MMLU	TruthfulQA	Winogrande	GSMBK
1	davieskil205/8hse-72b-v0.5	81.22	79.78	91.15	77.95	74.5	87.85	76.12
2	MISAIR/MultiVerse_70B	81	78.67	89.77	78.22	75.18	87.53	76.65
3	MISAIR/MultiVerse_70B	80.90	78.58	89.74	78.27	75.09	87.37	76.8
4	ahucuan1/Smaug-72B-v0.1	80.48	76.02	89.27	77.15	76.67	85.08	78.7
5	ibivibiv/loose-dragon-72b-v4	79.3	73.09	88.16	77.4	72.69	86.03	77.63
6	mistralai/Mistral-8x22B-Instruct-v0.3	79.15	72.7	89.88	77.77	68.14	85.16	82.03
7	MariyaPanahi/Llama-3-70B-Instruct-RPQ-v0.2	78.96	72.53	86.22	80.41	63.57	82.79	88.25
8	MariyaPanahi/Llama-3-70B-Instruct-RPQ-v0.4	78.89	72.61	86.03	80.5	63.26	83.58	87.34
9	MariyaPanahi/Llama-3-70B-Instruct-RPQ-v0.3	78.74	72.35	86	80.47	63.45	82.95	87.19
10	mona/Llama-3-70B-japanese-suzume-vector-v0.1	78.6	72.35	85.81	80.28	62.93	82.79	87.41
11	moezh/MoMo-72B-LoRA-1.8.7-RPQ	78.55	70.82	85.96	77.13	74.71	84.06	78.62
12	tenyx/Llama3-TenyxChat-70B	78.4	72.1	86.21	80.04	62.85	82.95	86.28

其中，排名第一和第四的都是基于 Smaug-72B-v0.1，而 Smaug-72B-v0.1 是基于 Qwen 72B 微调而来，排名第二、第三的 MultiVerse_70B 就是基于阿里云的 Qwen 72B 调整而来的，可以说阿里云的 Qwen 72B 在 Huggingface 上已经超过了开源标杆 Llama 2-70B 和大部分商用闭源模型。

二、国内大模型厂商商业能力对比

1、团队背景及融资情况梳理

虽然国内大模型厂商数量大，但从测试榜单、用户反馈及行业应用来看，百度、智谱、阿里、MiniMax、月之暗面和百川应该都属于第一梯队，以下是根据公开信息整理的团队背景：

公司名称	产品名称	核心人员	教育背景	工作经历
百度	文心一言	王海峰	百度首席技术官	人工智能领域取得显著成就，获得“吴文俊人工智能科技进步奖”特等奖
智谱	智谱清言	唐杰	清华大学计算机系教授，入选 2023 AAAI Fellow, ACM / IEEE Fellow	研究方向包括人工智能、数据挖掘、社交网络、机器学习和知识图谱等，曾获 ACM SIGKDD Test-of-Time Award
阿里	通义千问	周靖人	中国科学技术大学毕业，获哥伦比亚大学计算机博士学位，前微软研发合伙人	达摩院智能计算实验室、大数据智能计算和搜索推荐平台负责人，入选 IEEE Fellow

MiniMax	Adab6	闫俊杰	前商汤科技副总裁、通用智能技术负责人，商汤研究院副院长	深度学习和计算机视觉领域发表顶级会议和期刊论文 100 余篇，Google Scholar 引用超过 10000 次
月之暗面	Kiwi	杨植麟	清华大学计算机系本科毕业，卡内基梅隆大学语言技术研究所 (LTI) 博士	师从苹果 AI 研究负责人 Ruslan Salakhutdinov 和谷歌首席科学家 William Cohen
百川	baichuan	王小川	1996 年获得国际奥林匹克信息学竞赛金牌，进入清华大学	搜狐前高管，推出搜狗搜索，以“自然交互 + 知识计算”为核心推出 AI 翻译机等产品

在融资上，根据公开信息整理，各创业公司融资历史如下：

1) 北京智谱华章科技有限公司

序号	融资日期	融资轮次	融资金额	企业估值	投资方	所属机构	来源	↑
1	2024-03-14	B+轮	-	-	北京市人工智能产业投资基金 (有限合伙)	顺德基金	相关新闻	
2	2024-01-17	B+轮	-	-	天津合远优择壹号创业投资合伙企业 (有限合伙) 杭州光合贰期创业投资合伙企业 (有限合伙)	光速光合 光速光合	-	
3	2023-09-19	B+轮	-	约10亿美元	好未来 小米集团 红杉基金 美国 腾讯投资 金山软件 阿里巴巴 顺为资本 高瓴资本 社保基金中关村自主创新投资基金 (北京) 合伙企业 (有限合伙) 上海云研企业管理咨询有限公司 北京华品博睿网络技术有限公司	好未来 小米集团 红杉基金 美国 腾讯投资 金山软件 阿里巴巴 顺为资本 高瓴资本 君联资本 - -	相关新闻	
4	2023-07-17	B+轮	数亿元人民币	约5亿美元	天津三快科技有限公司	美国	相关新闻	
5	2022-09-26	B轮	数亿元人民币	-	昆山市启明融凯股权投资合伙企业 (有限合伙) 领投 苏州君联相道股权投资合伙企业 (有限合伙) 领投 苏州启明融乾股权投资合伙企业 (有限合伙) 领投 青岛华控成长股权投资合伙企业 (有限合伙)	启明创投 君联资本 启明创投 华控基金	相关新闻	
6	2022-01-30	股权融资	-	-	北京创新智源科技有限公司	-	-	
7	2021-09-14	A轮	超亿元人民币	-	图灵创投 凌云光技术股份有限公司 北京华控产业投资基金 (有限合伙) 北京将门创业投资中心 (有限合伙) 北京星连峰基金企业管理合伙企业 (有限合伙) 南京图灵一期创业投资合伙企业 (有限合伙) 枣庄通智股权投资合伙企业 (有限合伙) 深圳市财智创赢私募股权投资企业 (有限合伙) 深圳市达晨创鸿私募股权投资企业 (有限合伙) 北京莱品投资管理有限公司	图灵创投 凌云光 华控基金 将门创投 海南智桥 IMO Ventures 海南智桥 达晨财智 达晨财智 -	相关新闻	
8	2019-08-13	Pre-A轮	-	-	中科创星 华控技术转移有限公司	中科创星 -	-	

据悉，智谱最新一轮估值为 200 亿元，员工人数近 1000 人。

2) 北京月之暗面科技有限公司

序号	融资日期	融资轮次	融资金额	企业估值	投资方	所属机构	来源
1	2024-02-19	A轮	超10亿美元	约25亿美元	阿里巴巴 领投	阿里巴巴	相关新闻
					Monolith砺思资本	Monolith砺思资本	
					小红书	小红书	
					红杉中国	红杉中国	
2	2023-06-12	天使轮	超2亿美元	约3亿美元	美团	美团	-
					真格基金	真格基金	
					红杉中国	红杉中国	

据悉，月之暗面最新一轮估值为 25 亿美元，员工人数 200+。

3) 北京百川智能科技有限公司

序号	融资日期	融资轮次	融资金额	企业估值	投资方	所属机构	来源
1	2023-10-17	A轮	3亿美元	-	卓源亚洲	卓源亚洲	相关新闻
					卓源资本	卓源资本	
					基石资本	基石资本	
					小米集团	小米集团	
					心资本	心资本	
					深创投	深创投	
					红点中国	红点中国	
					联想之星	联想之星	
					腾讯投资	腾讯投资	
					阿里巴巴	阿里巴巴	
2	2023-04-10	天使轮	5000万美元	-	顺为资本	顺为资本	相关新闻
					王小川	-	

投资界独家获悉，由王小川创立的百川智能进行新一轮数亿美元融资，将成为国内 2024 年最大 AI 领域融资之一。来自参与的投资人证实，百川智能今年以来估值已超 18 亿美元（约合人民币 130 亿）。目前，百川员工人数超过 170 人。

4) 上海稀宇极智科技有限公司 (MiniMax)

融资轮次	融资时间	融资金额	投资方
A轮	2023-06	超2.5亿美元	腾讯投资
Pre-A轮	2022-07	未透露	明势资本, 米哈游, 高瓴创投, IDG资本
天使轮	2022-01	未透露	云启资本

今年 3 月，彭博爆出 MiniMax 新一轮融资估值将超 25 亿美元。据悉阿里和红杉中国已承诺将参与，其余跟投方还在洽谈中，相关条款可能会有所调整。目前，MiniMax 员工人数近 300+。

以上大模型公司中，除了百度、阿里之外，融资金额都超过 10 亿人民币，可谓是新时代的“AI 四小龙”。投资机构除了传统的 VC、PE 之外，各大互联网厂商如阿里、小米、腾讯、美团、米哈游等都有参与投资，这场大模型的“完美风暴”将市场的所有参与者都裹挟进去，即使前路有太多的不确定性。

2、商业模式选择及落地情况

1) 百度-文心一言商业模式及现状

2024 年 2 月 28 日，百度发布 2023 年第四季度及全年财报时披露，自发布以来，百度不断降低文心大模型的推理成本，目前已降低至去年 3 月版本的 1%。2024 年生成式人工智能与基础模型业务将为百度带来数十亿元人民币的增量收入。

李彦宏在电话会上透露，百度智能云四季度总营收 84 亿元，其中大模型为云业务带来约 6.6 亿元增量收入。李彦宏称，越来越多的企业选择在百度云上构建模型、开发应用程序。

百度利用生成式 AI 重构百度所有的 To C 产品，对于搜索而言，生成式 AI 的引入使百度能够回答更广泛范围的问题，包括更复杂、开放式和比较性的查询。在 2023 年 10 月，李彦宏展示了基于文心一言重构的百度搜索、如流、地图、网盘、文库等十余款 AI 原生应用。李彦宏称，经监测，文心大模型的日调用量已超过 5000 万次，季度环比增长 190%。

公开数据显示，截至 2023 年 12 月底，文心一言用户规模超过 1 亿。四季度内，中国三星、荣耀、汽车之家等头部企业均与百度达成合作。以三星为例，其最新旗舰手机 Galaxy S24 系列集成文心大模型多项能力，包括通话、翻译、智能摘要等。

可以看到，文心一言对于百度而言，不仅是就一个语言模型，而是将百度的所以业务进行重构，文心一言的商业模式就是百度目前的商业模式。

2) 智谱商业模式及现状

根据智谱公开披露信息的得到，智谱 2023 年商业订单收入过亿元，采用的是通用大模型+行业应用的 B 端商业模式。智谱披露公司已收获超 2000 家生态合作伙伴，实现超 1000 个大模型规模化应用，与超 200 家企业进行深度共创，多个领域的头部公司，涵盖了传媒、咨询、消费、金融、新能源、互联网、智能办公等多个细分场景，均已成为智谱 AI 的代表客户。

目前智谱 AI 探索出四种大模型商业化模式：最轻量级的 API 调用服务，基于云端的私有化方案，完全私有化的方案，以及提供国产化适配的软硬件一体化方案。

第一种，最轻量级的 API 调用模式。将大模型封装成 API 开放平台，按照 API 调用量付费。

标准版定价

通过我们提供的标准API产品，快速构建自己专属的大模型应用，您可按需使用，我们将按照模型实际处理文本的token数量计量计费。●

[充值体验](#)

[咨询标准版定价](#)

模型	说明	上下文长度	单价
GLM-4	提供了更强大的问答和文本生成能力。适合于复杂的对话交互和深度内容创作设计的场景。	128K	0.1元 / 千tokens
GLM-4V	实现了视觉语言特征的深度融合，支持视觉问答、图像字幕、视觉定位、复杂目标检测等各类图像理解任务。	2K	0.1元 / 千tokens ●
GLM-3-Turbo	适用于对知识量、推理能力、创造力要求较高的场景，比如广告文案、小说写作、知识类写作、代码生成等。	128K	0.005元 / 千tokens

第二种，云端私有化方案。部分企业希望又相对独立的数据保障措施，同时不想自己维护存储、计算等基础设施，智谱 AI 专门开辟基于云端算力的模型专区，让客户在调用模型的过程中该，保持数据与其他客户的相对隔离。

云端私有化定价

提供额外的模型微调服务，提高模型精度，快速构建自己的专属大模型。

咨询云端私有化定价

模型	套餐包含		单价 (年付)
ChatGLM-130B	推理机数量	3台	120万元/年
	训练语料额度	5,000,000 tokens	
	在线教程培训	1年	
ChatGLM-66B	推理机数量	3台	100万元/年
	训练语料额度	3,000,000 tokens	
	在线教程培训	1年	
ChatGLM-32B	推理机数量	2台	50万元/年
	训练语料额度	2,000,000 tokens	
	在线教程培训	1年	
ChatGLM-12B	推理机数量	1台	25万元/年
	训练语料额度	2,000,000 tokens	
	在线教程培训	1年	

第三种，完全私有化部署。对于央国企及其他保密性更强的行业，可以将大模型部署在企业内部，通过企业自身的算力和技术平台开发应用，提供 AI 服务。

本地私有化定价

智谱AI大模型私有化部署服务是基于大模型算法模型的云计算服务，为客户提供高效的数据处理、模型训练和部署服务。我们为您提供智谱大模型文件和相关的工具包，客户可自行训练和部署推理服务，同时智谱会提供部署应用相关的技术支持和咨询，以及模型的更新。通过私有化部署方案，实现数据的完全掌控和模型的安全运行。

咨询本地私有化定价

模型	套餐包含		单价 (年付)
ChatGLM-130B	推理实例license	不限量/年	3960万元/年
	推理&微调工具包	1年	
	咨询服务	15人天/年	
ChatGLM-66B	推理实例license	不限量/年	1680万元/年
	推理&微调工具包	1年	
	咨询服务	15人天/年	
ChatGLM-32B	推理实例license	不限量/年	680万元/年
	推理&微调工具包	1年	
	咨询服务	10人天/年	
ChatGLM-12B	推理实例license	不限量/年	180万元/年
	推理&微调工具包	1年	
	咨询服务	6人天/年	

注：ChatGLM3-6B 现已开放免费商用授权，申请授权证书即可使用 [免费申请授权](#)

第四种，软硬件结合一体机方案。把模型和一些国产化设备进行绑定，提供软硬一体服务，免去了在客户的环境当中进行部署和调试的过程。

3) 阿里通义千问商业模式及现状

阿里云面向生态开放以“通义千问”为代表的模型能力和训练底座，与产品生态伙伴展开在行业模型方向的MaaS（模型即服务）产品的集成与被集成合作，旨在为用户提供基于AIGC领域研究成果在行业中的

最新技术应用与解决方案，而共同开展的一项关于技术合作的产品合作计划。

合作模式为：

合作模式

被集成合作：通义千问开放模型API、专属模型定制等能力，伙伴基于通用模型能力及行业特性集成至自有产品。集成合作：伙伴将成熟能力API插件集成至通义千问，服务于云上用户。



合作示例



阿里云大模型的逻辑是通过开源的方式提供技术产品，降低门槛，推动技术普惠，为企业客户到个人开发者提供多元化、全方位的技术服务。在通义千问的基础上创建的大模型、小模型越丰富，AI生态就越繁荣，阿里云的前景也会越广阔。

4) MiniMax 商业模式及现状

在商业模式方面，MiniMax 基于 C 端的 Glow，以及 B 端的 ABAB 大模型，分别衍生出了两套不同的商业模式。

具体来说，在 C 端主要采取了按量付费”（Pay-As-You-Go）模式，而 B 端则采取了 PaaS（“平台即服务”Software as a Service）模式。MiniMax 推出了 AI 虚拟聊天社交软件 Glow，这是一款面向 C 端的产品，用户可以通过与智能体进行对话来获得互动体验。同时通过购买不同的“会员权限卡”来解锁相应次数的对话额度。

在 B 端，MiniMax 通过技术授权的方式，对外开放 API，允许企业客户直接接入其大模型技术。企业可以利用 MiniMax 的技术进行产品开发或服务优化，而 MiniMax 则从中获得授权费用或合作收益。

例如，通过 MiniMax 的 Assistants API，用户可以构建一个助手，该助手能够根据用户指令，智能调用工具，可以融入各类应用和服务中，满足复杂数学运算、文件处理、数据分析、文档分析等任务，为各行各业创造价值。

产品定价

MiniMax 开放平台提供基于大语言模型的丰富 API 能力，以满足不同企业客户及专业开发者的需求。计价方式简单灵活，按量计费，满足不同业务规模。

文本大模型

基于 MiniMax 端到端自研多模态大语言模型，为企业客户和个人开发者提供功能丰富的 API 能力和专区方案，适用多数文本处理场景。

ChatCompletion Pro

提供了更强大的问答和文本生成能力，包括多人对话、对话示例，支持集合搜索引擎、调用自定义函数、限制返回格式，适合于复杂的对话交互和深度内容创作设计的场景。

计费项	单价	计费说明	tokens 中文字符比 (估算)
abab6.5	0.03 元/千tokens	计费 token 数包括输入和输出	1000 tokens : 1600 中文字符
abab6.5s	0.01 元/千tokens	同上	1000 tokens : 1600 中文字符
abab6	0.1 元/千tokens	同上	1000 tokens : 1600 中文字符
abab5.5	0.015 元/千tokens	同上	1000 tokens : 750 中文字符
abab5.5s	0.005 元/千tokens	同上	1000 tokens : 1200 中文字符

注：计费项是 token 数，以 1000 个 token 为单位；tokens 字符比值根据使用场景的不同略有浮动，以实际消耗为准，字符数包括标点等。

ChatCompletion V2

该接口支持 OpenAI 参数定义和基础功能。

计费项	单价	计费说明	tokens 中文字符比 (估算)
abab6.5	0.03 元/千tokens	计费 token 数包括输入和输出	1000 tokens : 1600 中文字符
abab6.5s	0.01 元/千tokens	同上	1000 tokens : 1600 中文字符
abab6	0.1 元/千tokens	同上	1000 tokens : 1600 中文字符
abab5.5	0.015 元/千tokens	同上	1000 tokens : 750 中文字符
abab5.5s	0.005 元/千tokens	同上	1000 tokens : 1250 中文字符

注：计费项是 token 数，以 1000 个 token 为单位；tokens 字符比值根据使用场景的不同略有浮动，以实际消耗为准，字符数包括标点等。

5) 月之暗面-Kiwi 商业模式及现状

成立之初，月之暗面便定下了独特的商业模式，不做目前看来商业模式更清晰的 B 端，而是在做基础大模型研究的同时，面向 C 端发力，意图通过技术提升，攻克 C 端用户个性化服务的需求。月之暗面的特点在于对长文本的处理能力，不过目前 Kiwi 在 C 端还处于免费阶段，商业模式尚未清晰。

目前，有消息称月之暗面称年内将有商业化初步方案，针对付费用户提供更加高效的用户体验，也会将 API 开放给 B 端合作伙伴。

6) 百川商业模式及现状

2023 年 9 月 25 日，百川智能发布 Baichuan2-53B 闭源大模型，全面升级了 Baichuan1-53B 的各项能力。与此同时，作为首批通过备案的大模型企业，百川智能此次还开放了 Baichuan2-53B API 接口，正式进军 To B 领域，开启商业化进程。收费模式如下：

价格说明

通用大模型

计费项	时间（每日）	价格	备注
模型调用 Baichuan2-Turbo	00:00 ~ 24:00	0.008元/千tokens	包含输入和输出
模型调用 Baichuan2-Turbo-192k	00:00 ~ 24:00	0.016元/千tokens	包含输入和输出
模型调用 Baichuan2-53B	00:00 ~ 8:00	0.01元/千tokens	包含输入和输出
	8:00 ~ 24:00	0.02元/千tokens	
搜索增强服务	00:00 ~ 24:00	0.03元/次	开启 with_search_enhance 后，接口自动判断调用搜索增强服务的次数

* 假设您的请求包含了1500tokens的文本，您将被收取1.5次的“千tokens”费用，即此次API调用将收取1.5*调用时段的千tokens费用。

* 搜索可以为模型引入更多相关信息，有助于提高输出的质量、及时性等。如果您不需要引入搜索结果，可以通过传参 with_search_enhance 关闭。

角色大模型

计费项	时间（每日）	价格
Baichuan-NPC-Lite	00:00 ~ 24:00	0.0099元/千tokens
Baichuan-NPC-Turbo	00:00 ~ 24:00	0.015元/千tokens

知识库

知识库API接口，对Embeddings模型和文件数据存储两部分分别收费，具体收费明细如下：

计费项	时间（每日）	价格
模型调用 Baichuan-Text-Embedding	00:00 ~ 24:00	0.0005元/千tokens
文件存储	00:00 ~ 24:00	1.5元/GB/天

* 知识库API在2024年1月19日结束免费公测、正式开启收费，单个用户限制5G存储容量。

向量化模型 Embeddings

使用百川自研的向量模型Baichuan-Text-Embedding，具体价格如下：

计费项	时间（每日）	价格
模型调用 Baichuan-Text-Embedding	00:00 ~ 24:00	0.0005元/千tokens

三、国内大模型厂商对 AGI 的想象

根据 DeedMind 的创始人兼首席 AGI 科学家 Shane Legg 的定义，AGI 能够执行一般人类可完成的认知任务、甚至超越这个范围。具体而言，AGI 应能够学习广泛任务，能够执行复杂、多步骤的任务。DeepMind 根据 AI 模型性能和学习处理任务的广泛性对 AGI 水平进行分类，从 Level-0 无人工智能，到 Level-5 超越人类共 6 个等级。

图表：AGI 可以根据性能和广泛性划分为 6 个等级

等级	主要特征
Level-0 无人工智能 (Narrow Non-AI)	只能完成明确定义的任务，比如计算器软件或编译器
Level-1 初现 (Emerging AGI)	性能相当于或略优于一个不熟练的人类。比如一些前沿语言模型在某些任务上已经达到了初现 AGI 的水平
Level-2 熟练 (Competent AGI)	至少能够在大多数任务上达到熟练人类的水平。目前的前沿语言模型在某些任务上已经接近熟练 AGI 的水平
Level-3 专家 (Expert AGI)	在大多数任务上能够达到专家人类的水平
Level-4 大师 (Virtuoso AGI)	在大多数任务上能够达到顶尖人类的水平
Level-5 超越人类 (Superhuman AGI)	在所有任务上都能超过 100% 的人类

来源：《Levels of AGI: Operationalizing Progress on the Path to AGI》，国金证券研究所

按照 DeepMind 的 6 级 AGI 水平分类，目前国际顶尖大模型仍处于 Level-1 Emerging AGI 阶段。

语言大模型能力相对完备，底层技术路线大多选择 Transformer Decoder-only 架构，结合 MOE 和多模态 embedding，算法细节优化方向区别较小。以 GPT-4、Gemini 1.5、Claude 3 为例，**语言大模型在推理、长文本、代码生成领域已经能够完成初级任务，但距复杂、专业水平仍有差距**；多模态大模已经能够面向 B\C 端提供商业化产品，底层技术路线主要采用 Diffusion Transformer，但细节优化空间较大，高质量和成规模的数据集仍在发展初期；具身智能类大模还在探索阶段，底层技术路线尚不清晰，数据收集、训练方法、测评方法等都处于发展初期。在实际应用场景中准确率较低。

对于 AGI，国内厂商有着怎样的愿景？技术路径和商业路径上有什么新的可能性？

公司名称	AGI 愿景	技术路径	商业路径
百度	自然语言将成为下一代的编程语言，人人都能成为开发者。这一愿景的核心是百度的文心大模型 4.0 工具版，它不仅能够让开发者通过自然语言交互进行复杂数据和文件处理，还能生成图表或文件，快速洞察数据中的特点、分析变化趋势、为后续的决策提供高效和精准的支撑。	文字、图片、语音、视频等多模态的融合，是基础模型非常重要的长期发展方向，是通往人工通用智能（AGI）的必经之路。	智能体可能是未来最主流的大模型使用方式。基于强大的基础模型，智能体可以批量生成，应用在各种各样的场景。而百度的文心智能体平台已经有超过 3 万个智能体被创建，5 万多名开发者和上万家企业入驻。这些智能体的开发和应用，正在改变人们的工作和生活方式，提升效率和创造力。
智谱	实现超级 AGI，就是超越人的智能的智能，当超级 AGI 实现的时候，如何正确地引导和控制 AGI 系统是关键	智谱 AI 将继续坚持基础模型能力升级和商业化落地加速推进这两条路线。在核心技术方面要坚持自主创新，这是“买不来，求不来”的。	智谱将突破更多行业领域，打造更多标杆案例，向行业、市场内外展示大模型落地能力，让大模型真正落下来，接地气，是我们的重要任务。我们还要努力打造分层大模型生态圈，联手产业合作伙伴，避免“重复造轮子”问题。
阿里	阿里云是做基础设施的。阿里云的愿景是成为全球领先的云计算服务提供	阿里云创业者计划与上百家机构合作，不仅自家正	一面自研大模型、一面做投资和生态布局。

	商，为客户提供高效、安全、可靠的云计算服务。	开发通义千问等大模型产品，而且还不断投资布局	
MiniMax	AGI 不是大杀器，是普通人每天用的产品	始终基于用户需求，这样算法才能不断突破临界点，我们也才能为越来越多的客户带来实打实的价值。	我们的目标是在算法上持续保持领先，在业务上始终为用户创造实际的价值。
月之暗面	成为 AI 时代的 Super App	月之暗面的技术路线，就是不走捷径，踏实地解决算法与工程的双重挑战，在算力、存储、带宽等技术层面做了极致的优化。	开源模型无法构建产品壁垒。比如，在海外有几百个基于开源扩散模型 Stable Diffusion 的应用出现，但最后其实没有任何一个跑出来。探索智能边界，并且让智能真正有用、实现个性化价值，这就需要提高技术与产品的迭代效率。
百川	让大众能够轻松普惠地，获取知识和专业服务	搜索增强是大模型发展的必由之路，能够有效解决幻觉、时效性差、专业领域知识不足等阻碍大模型应用的核心问题。	做成 AI 时代的“超级 APP”，背后一定是拥有强大能力的 AI 大模型的支持

大模型的潜在应用潜力决定了这是一条所有人必须关注的赛道，但极高的人才壁垒、资金门槛、规模效应决定了不是所有人都有资格走上赛场。大模型的竞争一场智力、魄力与勇气的比拼，它比马拉松还漫长，更像一场充满未知、刺激、风险极高但潜在收益也极高的航行，就像大航海时代一样。对于目前市场上绝大多数企业来说，即使没有上场，也必须下注。

百度、阿里占据地利，搜索及云服务市场是其基本盘，大模型不过是土地上新冒出来的枝丫，它们进可攻退可守。而对于腾讯、金山、字节、网易等互联网厂商而言，技术重要，但应用更重要，大模型只是改善用户体验、产品模式的手段而已，不会以命相拼。独角兽如智谱、MiniMax、月之暗面和百川，还需要奋力前行，在资金消耗、技术迭代、商业落地和组织管理的风暴中稳住船舵，稍有不慎就有可能功亏一篑。而对于绝大多数其他厂商，都会面临资金及人才短缺，退而求其次，寻求应用端、服务端的二次开发市场，有幸的成为软件巨头，不幸的消失于层层风暴中。

参考资料：

1. 中国数谷- 14 款！中国第四批大模型备案获批（http://www.cbdio.com/BigData/2024-02/01/content_6176476.htm）；
2. SuperBench- 2024 年 3 月的《SuperBench 大模型综合能力评测报告》；
3. 排行榜 - C-Eval（https://cevalbenchmark.com/static/leaderboard_zh.html）；
4. hugging face（https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard）；
5. 企查查-北京智谱华章科技有限公司、上海稀宇极智科技有限公司、北京月之暗面科技有限公司、北京百川智能科技有限公司；
6. 国金证券-如何实现 AGI-大模型现状及发展路径展望；
7. 通义官网（https://tongyi.aliyun.com/?spm=5176.28195423.J_3592841540.3.246f2aacep8L39C）；
8. 百川官网（<https://platform.baichuan-ai.com/price>）；
9. 智谱官网（<https://open.bigmodel.cn/pricing>）；
10. MiniMax 官网（<https://www.minimaxi.com/document/price?id=6433f32294878d408fc8293e>）；
11. 月之暗面官网（<https://kimi.moonshot.cn/>）；