

【AI金融新纪元】系列报告（二）

——AI+金融大模型的两条技术路线

证券分析师：胡 翔

执业证书编号：S0600516110001 联系邮箱：hux@dwzq.com.cn

二零二四年二月二十三日

- **当前AI与金融的结合主要有两条技术路径：①通用模型+金融语料训练金融大模型，②金融垂类大模型。1) 双方优劣具有相对性。通用大模型优势：**泛用性强、灵活性和利用率高、可迁移性强。**劣势：**特定领域深度较浅、模型复杂、训练时间长；**金融垂类模型优势：**领域专业性、针对性的解决方案、高精度和合规性。**劣势：**适应性限制、更新和维护复杂度、数据利用率低。2) 通用大模型通过金融语料训练超越金融垂类模型可能性较小。通用大模型在行业数据量，性价比，精确性、适用性、实时性、推理速度，合规性和风险控制等方面表现欠佳。
- **通用大模型“百模大战”，头部模型国外领先较大，平均水平国内外差距较小，中文上国内表现更优。1) 国外通用GPT4-Turbo遥遥领先。**OpenAI震撼发布GPT4-Turbo，开启新一代人工智能模型的大门；谷歌将在谷歌云上部署 Claude，并于推出自研的大模型LaMDA 的聊天机器人 Bard；AWS 推出自有基础模型 Titan 和 AIGC 服务 Bedrock，以及 AI 编程助手 Amazon CodeWhisperer。Anthropic推出Claude，是最接近ChatGPT的商业竞品；xAI发布其首个AI大模型产品Grok，模型通过X平台实时了解世界，GrokV1.5或于2024年3月发布。2) 国内通用百度先行，多家企业推出相关产品。百度推出“文心大模型”，是目前国内预训练大模型应用端生态最好的大模型之一；阿里发布通义千问2.0，专业维度能力较强；vivo发布BlueLM大模型，应用的场景广泛；月之暗面发布Moonshot大模型，目前位于第一梯队。3) 在金融领域中，通用模型应用表现各有差异。其中GPT系列、文心一言、通义千问、腾讯混元以及科大讯飞表现较好。
- **金融垂类模型国外发展先行，国内成品问世。1) 国外彭博BloombergGPT率先登场。**BloombergGPT的混合训练方法使其模型在金融任务上的表现大大超过了现有的大语言模型，而在通用场景上的表现则与之相当，甚至优于现有模型。AI4Finance Foundation开发FinGPT，为金融大型语言模型提供互联网规模的数据，以此推动金融领域的开源发展。2) 国内金融垂类模型百花齐放。奇富科技率先宣布推出自研的金融行业通用大模型“奇富GPT”；度小满推出国内首个千亿级中文金融大模型“轩辕”；腾讯云公布腾讯云金融行业大模型，TI-OCR大模型帮助银行解决日常业务问题；恒生电子进一步升级金融大模型LightGPT，并发布多款光子系列大模型应用产品；蚂蚁集团公布蚂蚁金融大模型，在多项金融专属任务中表现突出；东方财富、同花顺加大AI研发技术投入，筹建人工智能事业部，重点推进金融垂直大模型研发应用。
- **投资建议：**我们预计2024年金融垂类模型产品落地较多，建议关注具备AI模型技术领先优势、较大金融交易数据基础、较好应用场景入口、积极推进AI模型构建的金融科技企业，推荐【东方财富】、【同花顺】，建议关注【恒生电子】。
- **风险提示：**监管环境趋严抑制行业创新；行业竞争加剧；权益市场大幅波动。



■ 1. 通用+金融VS金融垂类哪方更强

■ 2. 国内外通用大模型在金融领域应用表现

■ 3. 国内外金融垂类模型发展进程

■ 4. 投资建议

■ 5. 风险提示

通用+金融VS金融垂类哪方更强

1.1. 通用+金融VS金融垂类：优劣势对比

- 当前AI与金融的结合主要有两条技术路径：①通用模型+金融语料训练金融大模型，②金融垂类大模型。
- 由于设计和训练目的不同，通用语言大模型与金融垂类模型在优劣上具有相对性。通用语言大模型在泛用性、灵活性、数据利用率、迁移性上相比金融垂类更有优势，而在专业性、针对性、高精度和合规性上，金融垂类模型更胜一筹；在复杂度问题上，通用语言大模型在结构上更加复杂，影响模型效率，而金融垂类模型则是在更新维护上具有复杂性。

图表：通用与金融垂类大模型优劣对比

	通用语言大模型		金融垂类大模型	
优势	泛用性强	由于在多样化的数据集上进行了训练，通用模型能够处理各种话题和领域的问题	领域专业性	在金融领域具有专业的理解能力，更精熟于金融术语和概念
	灵活性和利用率高	大量数据训练提高了模型的准确率，并可以应用于多种任务中	针对性的解决方案	更适合解决金融行业的具体问题，与行业需求和合规性要求相匹配
	可迁移性强	在数据量较少的特定任务上，通用模型也可以通过微调进行有效的迁移学习，大大减少模型的训练时间和计算资源	高精度和合规性	提供更精确、可信赖的金融信息和建议，同时符合行业规定
劣势	特定领域深度较浅	可能不具备特定领域（如金融）的深入理解和专业知识	适应性限制	由于主要针对金融领域，因此在其他领域的适用性可能有限
	模型复杂	通用大模型的结构非常复杂，使得模型的解释性变得困难，导致模型的计算量增加，影响模型的效率	更新和维护复杂度	金融政策和法规的变化可能要求模型频繁更新以保持准确性和合规性
	训练时间长	通用大模型的训练需要大量的计算资源和时间，这对于一些小型企业来说是一个挑战	数据利用率低	垂直领域模型的训练数据相对较少，模型的准确率可能会受影响

1.2. 通用+金融VS金融垂类：通用金融训练超越金融垂类可能较小

- **使用金融数据对通用大模型进行训练，数据欠缺，成本过高。**金融领域的语料应该充分覆盖各种金融产品、市场情况和业务流程等。但金融领域的数据分散在各个金融机构，通用大模型缺少金融数据进行训练，金融专业知识不足；另外一方面，如果从底层开始训练大模型，所需要投入的算力成本非常高，比如千亿级别的通用大模型，训练一次需要付出几千万的成本
- **在特定任务上，精确性与适用性欠缺，需要更多优化与定制。**金融垂类大模型通常会投入大量的时间和资源来收集和整理金融领域的专业知识。这些专业知识包括金融术语、金融工具和金融法规等。通用语言模型虽然可以通过金融语料的训练来提高在这方面的理解能力，但是否能达到金融垂类大模型的专业性仍有待验证。
- **金融领域要求实时性和高效的推理速度。**金融领域的决策和分析通常要求实时的响应和快速的推理速度。金融垂类大模型可能会针对这一需求进行了优化，以提供更快的响应时间。通用语言大模型在处理金融领域的实时应用时可能需要进一步的优化。
- **金融领域对合规性和风险控制要求极高。**金融领域对于数据保护、隐私和风险控制具有严格的要求，专门训练的金融垂类大模型可能会更好地满足这些合规性需求。

“通用模型难以胜任金融领域任务，金融大模型是大模型落地金融行业的必由之路。”

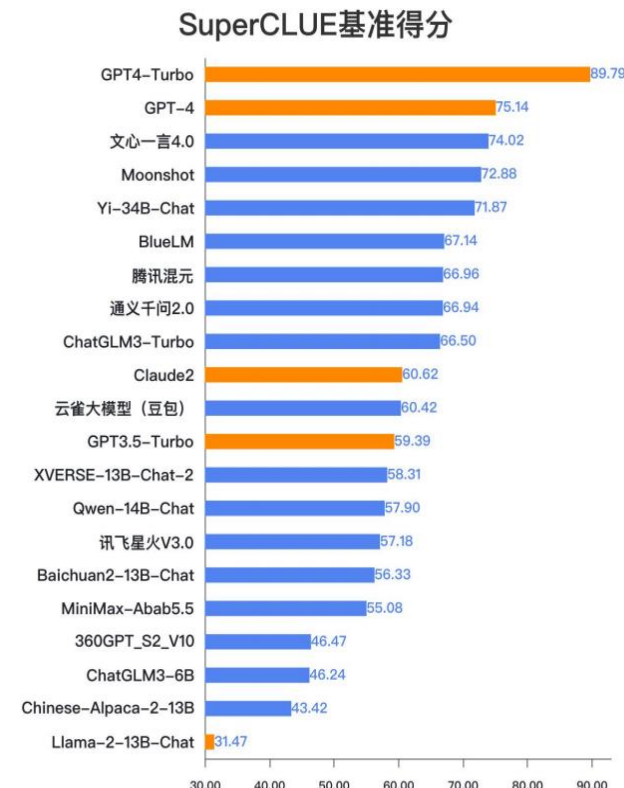
——度小满CEO 许东亮

国内外通用大模型在金融领域应用表现

2.1. 国内外通用AI大模型发展历程：国外领先，国内紧追

- 2023年6月国内AI大模型迎来爆发式增长，技术和应用不断发展，但与国外顶尖AI大模型尚有差距。自2022年OpenAI发布ChatGPT以来，国内迅速形成大模型共识，开始追赶国外。目前各行各业开闭源大模型不断出新，竞争形势越发激烈。
- 综合表现上，头部模型国外领先，平均水平国内外差距较小。在所有模型中，GPT4-Turbo遥遥领先，国内最好的大模型为百度文心一言，但仍有15.77分的差距。国内方面，虽仍有差距，但在过去一年发展迅速，平均水平上与国外差距并不明显。此外，国内开源大模型在中文上的表现要优于国外开源大模型。

图表：国内外大模型综合表现 (2023年11月28日)



图表：AI大模型2023年关键进展



国内外头部模型差距依然明显

GPT4-Turbo总分89.79分遥遥领先

2.2. 国外AI通用大模型案例，OpenAI&微软引领业界

OpenAI&微软

OpenAI携手微软 ChatGPT4-Turbo业界领先

微软和OpenAI是目前大模型技术水平、产品化落地最为前沿的领军者。2023年3月，OpenAI发布工程化的多模态GPT-4，并与各个领域的软件开展合作。2023年11月7日，OpenAI开发者大会重磅发布GPT-4 Turbo，这项新模型带来了六大升级，包括更长的上下文长度、更强的控制、模型的知识升级、多模态、模型微定制以及更高的速率限制。



谷歌

谷歌框架领先 模型丰富，应用偏弱

谷歌于2023年2月宣布将在谷歌云上部署ChatGPT的有力竞品Claude，并于同月推出自研的基于1270亿参数大模型LaMDA的聊天机器人Bard。应用方面，谷歌在过去更注重发表论文，未能及时将成果产品化。同时，旗下DeepMind研发的基础语言模型Gopher、Chinchilla和对话模型Sparrow也尚未产品化。



AWS

AWS领跑AIGC 推出Titan大语言模型

2023年4月，AWS正式入局AIGC，推出自有基础模型Titan和AIGC服务Bedrock，以及AI编程助手AmazonCodeWhisperer。Titan系列模型分为用于内容生成的文本模型Titan text和可创建矢量嵌入的嵌入模型Titan Embeddings。此外，基于自研推理和训练AI芯片的最新实例AmazonEC2Trn1n和AmazonEC2 Inf2正式可用。



Anthropic

Anthropic推出Claude 最接近ChatGPT的商业竞品

2023年7月，Anthropic宣布Claude2正式开始上架。Claude是基于transformer架构的大语言模型，被认为是最接近ChatGPT的商业产品。相比V1.3，Claude2编码能力提升巨大，具备更强大的逻辑能力，同时对训练数据进行更新。此外，Claude2聊天工具完全免费。



xAI

xAI推出Grok GrokV1.5或于2024年3月发布

2023年11月5日，马斯克旗下xAI团队发布其首个AI大模型产品Grok。Grok通过X平台实时了解世界，还能回答被大多数其他AI系统拒绝的辛辣问题。马斯克表示，Grok使用来自公开数据的数十亿个数据点进行训练。2024年2月22日消息，马斯克在社交媒体平台“X”上表示，xAI的Grok V1.5于2周后发布。



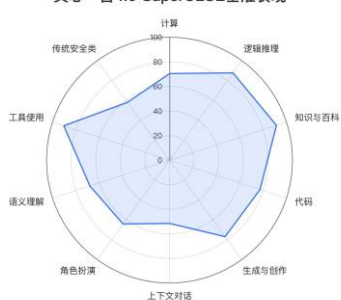
2.3. 国内AI通用大模型案例，多家企业推出相关产品

文心一言

百度具备先发优势，文心大模型国内领先

文心一言是百度全新一代知识增强大语言模型，于2023年3月16日正式发布，已进行多个版本迭代，10月17日发布V4.0版本。据百度官方介绍，文心一言目前已有7000万用户。文心一言4.0的能力栈较为广泛，可应用的场景较多。其在查询搜索知识应用、任务拆解规划Agent、文案写作以及代码编写及纠错等方面的应用表现不俗。

文心一言4.0 SuperCLUE基准表现

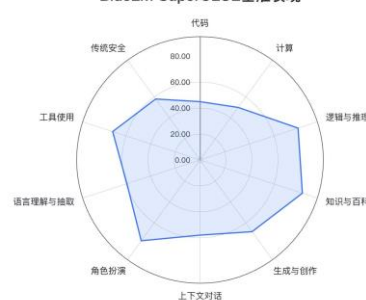


vivo BlueLM

vivo发布BlueLM大模型，应用的场景相对广泛

BlueLM是vivo自主训练的大语言模型，出自于vivo AI全球研究院。vivo于11月1日开发者大会上正式发布自研AI大模型矩阵，包括十亿、百亿、千亿三个不同参数量级的5个自研大模型。BlueLM主要可以应用在手机智能应用中，进行基础手机指令操作、实时语音助手、查询信息以及一些基于手机端的办公应用。

BlueLM SuperCLUE基准表现

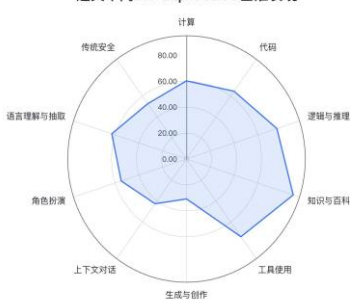


通义千问

阿里发布通义千问2.0，专业维度能力较强

通义千问，是阿里云推出的大语言模型，于2023年4月11日在阿里云峰会上正式发布1.0。9月13日，阿里云宣布通义千问大模型已首批通过备案。10月31日，阿里云正式发布千亿级参数大模型通义千问2.0，8大行业模型同步上线。通义千问2.0专业维度上的能力较强，可应用于相对专业复杂场景，例如金融、医疗、汽车等垂直专业场景。

通义千问2.0 SuperCLUE基准表现

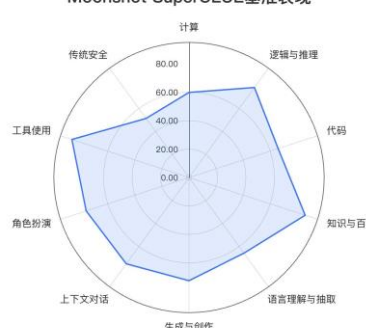


Moonshot AI

月之暗面发布Moonshot大模型，位于第一梯队

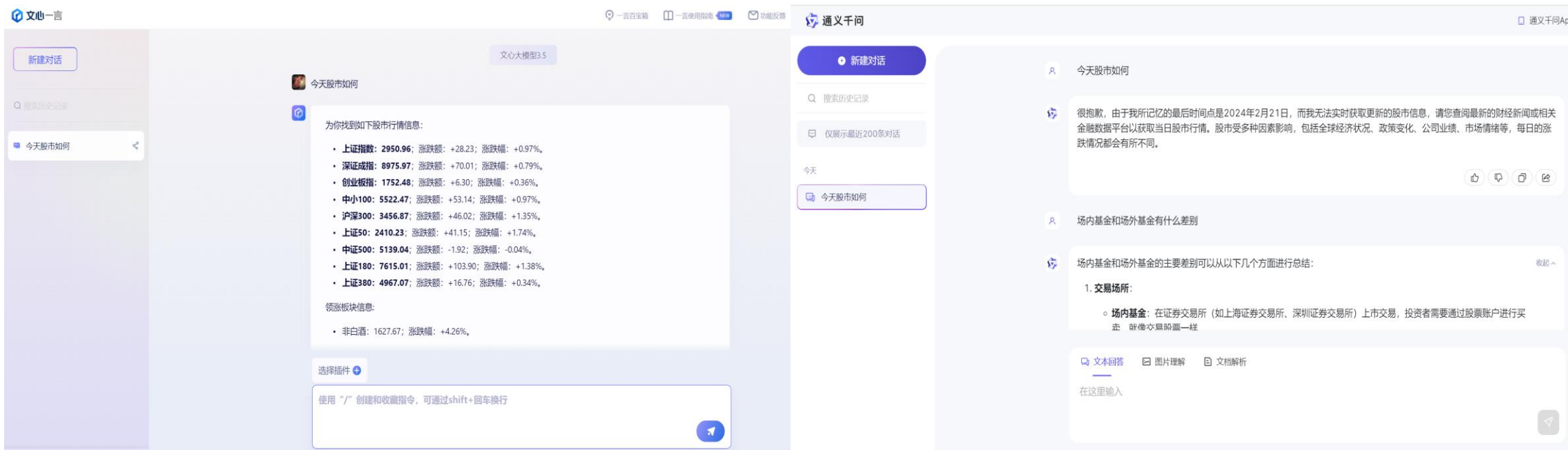
Moonshot是月之暗面自主训练的大语言模型，于10月9日正式发布。该模型具备多语言能力，支持约20万汉字上下文，通过创新的网络结构和工程优化，实现了无损的长程注意力机制，有较强的文档理解、归纳和处理能力。长程对话、长文本阅读、AI智能体等方面的应用表现强劲。另外，在数学运算、内容创作、虚拟数字人等场景也有不错的表现。

Moonshot SuperCLUE基准表现



2.4. 通用大模型在金融领域应用表现各有差异

- **国外GPT系列在金融领域表现较好。** GPT系列能较好理解金融术语，解答金融相关问题，且具备一定的实时性，在金融领域能够自动生成金融报告、做市场研究、客服机器人等。
- **国内百度文心、阿里通义千问、腾讯混元以及科大讯飞在金融领域表现较好。** 百度文心一言具备较高的实时性，且金融术语的理解较好，准确度高，能够在金融新闻分类、问答系统和智能写作中发挥作用。阿里通义千问能够较好的解释金融概念、提供一般性的金融投资分析指导，但存在时效性限制。腾讯混元可以优化金融服务的个性化体验、市场情绪的追踪等。科大讯飞火星则将语音识别和处理技术用于交易系统、智能财经助手等。



国内外金融垂类模型发展进程

3.1. 国内外金融垂类模型发展历程：国外发展先行，国内成品问世

- **金融大模型始于2023年3月BloombergGPT，通过应用金融大模型，金融效率将得到大幅提升。** 彭博推出了为金融领域量身定BloombergGPT模型，吸引了行业关注，大模型有了新的发展方向。金融行业沉淀了大量高质量数据。各金融平台的用户数以亿计，各种用户画像数据、交易数据浩如烟海。利用大模型对上述数据的分析处理，可大幅提高金融效率。比如，金融机构可以预测用户行为偏好，更高效、准确评估客户风险；AI还可以实时监测交易和市场波动，及时制定策略。
- **当前国内外金融行业都在主动拥抱大模型。** IDC（国际数据公司）一项调研显示，超半数的金融机构计划在2023年投资生成式人工智能技术，只有10%的金融机构表示没有试验计划。国外自BloombergGPT后，也出现了如FinGPT等一系列金融大模型。而国产金融大模型也已分出了明显的两个“流派”。一派来自于传统金融机构，另一派来自于金融系科技企业或互联网企业。

图表：国内外金融垂类AI模型发布时间&发布机构

时间	发布机构	金融垂类模型
2023年3月	彭博	BloombergGPT
2023年6月	AI4Finance Foundation	FinGPT
	奇富科技	奇富GPT
	度小满	轩辕大模型
	恒生电子	LightGPT
2023年9月	蚂蚁集团	蚂蚁金融大模型AntFinGLM
	腾讯云	腾讯云金融行业大模型
2023年10月	恒生电子	LightGPT (升级) +WarrenQ+光子
2024年1月	同花顺	HithinkGPT+AI iFinD
2024年1月	东方财富	妙想大模型

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/458042013011006040>