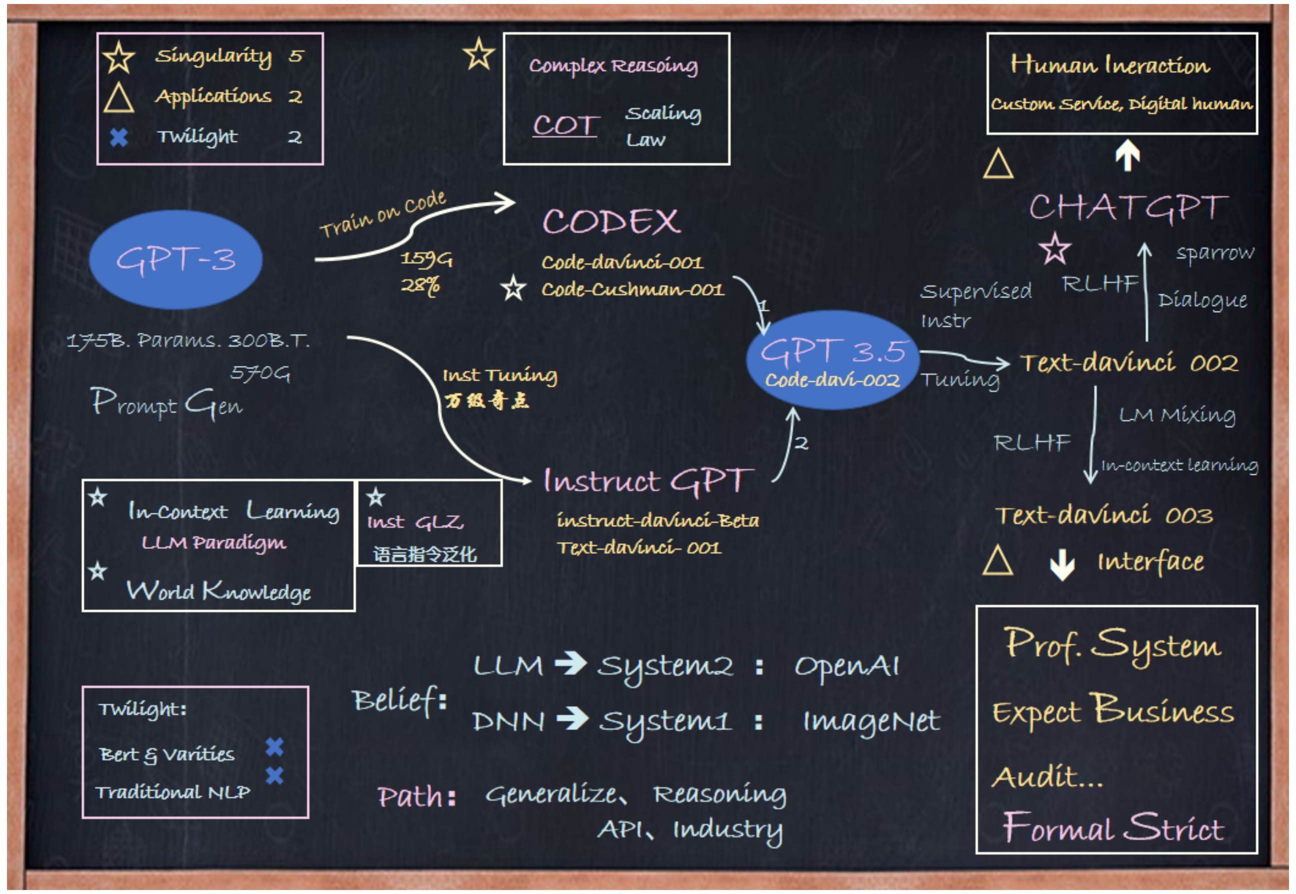
金融大模型的研究与应用

2023年8月

LLM研究路径的发展脉络及测评基准

2022年11月 OpenAI公司发布了语言类大模型ChatGPT，其展示出的通用化智能潜力，引发了新一轮的技术竞赛、产业革命。也成为各界研究对象，我行重点总结了其打造历程和相关技术原理，总结出**5个能力奇点sigularity、2类应用Application、2项末端Twilight**。



Leaderboard - C-Eval

Results for different subjects and the average test results are shown below. The results are from either zero-shot or few-shot prompting ---- note that few-shot is not necessarily better than zero-shot, for example, zero-shot is better for many instruction-tuned models in our own runs. In cases we tested the models in both zero- and few-shot settings, we report the setting with higher overall average accuracy. (Model details including prompting format can be viewed by clicking into each model)

You are welcome to submit your model's test results to C-Eval at any time (either zero-shot or few-shot eval is fine). Click here to submit your results (your results will not be public on the leaderboard unless you request to do so).

(Note: * indicates that the model was evaluated by the C-Eval team, while other results are obtained through user submissions.)

#	Model	Creator	Submission Date	Avg ▾	Avg(Hard)	STEM	Social Science	Humanities	Others
0	ChatGLM2	Tsinghua & Zhipu.AI	2023/6/25	71.1	50	64.4	81.6	73.7	71.3
1	GPT-4*	OpenAI	2023/5/15	68.7	54.9	67.1	77.6	64.5	67.8
2	AiLMe-100B v2	APUS	2023/7/25	67.7	55.3	65.4	72.3	71.2	64
3	CW-MLM-13B	CloudWalk	2023/8/21	66.7	47	58.1	81.4	70.8	64.9
4	GS-LLM-Beta	共生矩阵科技（深圳）有限公司	2023/8/20	66.7	43.2	57.4	79.7	73	65.5
5	SageGPT-V0.2	4Paradigm	2023/7/25	66.6	61.1	67.9	76.6	66.9	54.9
6	SenseChat	SenseTime	2023/6/20	66.1	45.1	58	78.4	67.2	68.8
7	Mengzi-7B	澜舟科技	2023/8/16	64.9	44.4	56	78.6	70.1	63.6
8	vivo_Agent_LM_7B	vivo	2023/8/9	64.4	43	57.4	75.9	66.9	64.4
9	赤兔	北京容联易通信息技术有限公司	2023/8/8	64.1	43.2	58.5	76.6	66.9	60.3
10	InternLM	SenseTime & Shanghai AI Laboratory (equal contribution)	2023/6/1	62.7	46	58.1	76.7	64.6	56.4

	人类	gpt-4	文心一言(v2.2.0)	Claude-2	gpt-3.5-turbo	ChatGLM-130B	讯飞星火(v1.5)	Claude instant-v1	360智脑(4.0)	internlm-chat-7b
总分	83.66	70.89	62.00	60.94	59.79	59.35	58.02	56.31	55.04	53.91
基础能力	85.03	70.04	61.11	62.01	64.40	53.78	63.32	58.85	56.68	54.85
中文特性	82.29	72.67	71.38	61.18	63.19	71.39	65.72	55.91	62.54	61.35
专业与学术能力		69.96	53.50	59.63	51.78	52.89	45.03	54.16	45.88	45.53

	ChatGLM2-6B	MiniMax-abab5.5	通义千问(v1.0.3)	Baichuan-13B-Chat	BELLE-13B	IDEA姜子牙-13B-v1.1	Phoenix-7B	MOSS-16B	Llama-2-13B-chat	Vicuna-13B	RWKV-7B World-CHN tuned
总分	53.85	53.06	51.52	49.35	46.60	43.80	41.57	35.36	34.26	31.70	27.83
基础能力	55.60	53.61	52.84	50.46	48.71	47.55	45.39	37.01	35.85	34.61	30.71
中文特性	63.59	62.79	61.73	55.38	52.99	48.61	44.62	38.01	37.37	33.71	28.13
专业与学术能力	42.37	42.77	39.98	42.21	38.10	35.26	34.70	31.07	29.57	26.80	24.66

2023年8月全球已发布数百个GPT大模型，我国年内已发布80多个大模型。针对GPT大模型的测评榜陆续出台，对大模型的评测体系也在逐步完善。

C-Eval作为全球大模型综合性考试测试榜，覆盖了人文、社科、理工、其他专业四个方向，能够更加全面地对模型的语言处理能力进行评估。

SuperCLUE着眼于综合评价大模型的能力，从3个维度评价模型能力：基础能力、专业能力和中文特性能力。

国外大模型调研情况

产品/模型名称	关联企业/高校实验室	是否开源	基模型	参数量	发布时间
ChatGPT	OpenAI	否	GPT3.5/4.0	-	2022年12月
Claude	Anthropics	否	-	-	2023年4月
LLaMA2	Meta	是	-	7B、13B、70B	2023年7月
LLaMA	Meta	是	-	7B、13B、33B和65B	2023年3月
BLOOMZ	BigScience	是	BLOOM	7B、176B	2022年11月
Falcon	echnology Innovation Institute (TII)	是	-	7B、40B	2023年5月
Bard	Google Deepmind	否	PALM	-	2023年2月
Alpaca	Stanford	是	LLaMA	7B	2023年3月
Vicuna	伯克利、卡内基、斯坦福	是	LLaMA	7B、13B	2023年3月
OPT-IML	Meta	是	-	30B	2022年12月
WizardLM	微软	是	LLaMA	7B、13B、30B、70B	2023年5月

大模型技术的进展

领域化

- 在模型发展方向上，业界和学术界也在讨论，未来是**一个通用的大模型**，还是**数个领域精通的模型并存**；
- 近期出现了一批垂直领域的应用模型，包括法律领域ChatLaw，交通领域TransGPT，科技领域Mozi等。在当前的模型结构、算力成本、数据情况下，单一通用大模型是否可以在全领域获得较好的效果，仍存在疑问；
- 在近期采访中贾扬清认为，之后的模型训练，不会再走单一模型路线，而是逐渐转为专业化模型。

开源化

- 自Meta开源LLaMA以来，特别是7月LLaMA2开源且有条件地开放了商业使用后，**社区迎来了众多的模型及数据的开源工作**；
- 开源模型，相比闭源大模型的黑箱，由于公布了训练数据、方法、标注等细节，**透明化程度更高**，企业用户用起来更放心。

MoE化

- MoE（Mixture-of-Experts, 混合专家模型）架构，**包含多个专家模型，每个专家模型负责处理输入数据的不同方面或子任务**；
- MoE的优势包括：
  - **适合复杂性和多样性任务**：通过将输入数据分配给不同的专家模型，每个模型可以专注于自己擅长的领域，从而提高整体模型的效果；
  - **可扩展性**：MoE可以通过添加更多的专家模型来进一步增加模型的容量和表达能力；
  - **提升性能**：在具体场景上，可以对专家模型进行删减部署，从而提升场景任务上的推理效率。

生态化

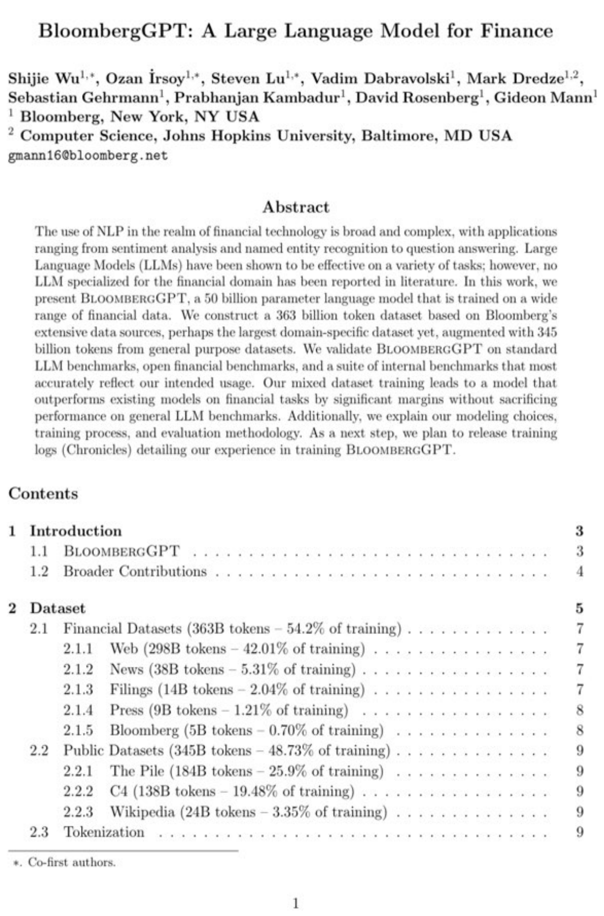
- 类似Bert、LLaMA这些强大的基座模型被放到开源社区后，都会有一大波寒武纪生命大爆发的现象；
- 近期，**很多业界领域模型的建设及学术界算法优化的工作都在基于LLaMA2展开，逐渐形成了应用生态圈**。

金融领域大模型整理

BloombergGPT：**彭博社**基于自建金融数据集训练的500亿参数量LLM，没有开源，仅可申请试用。

摩根士丹利将GPT4用于财富管理知识库搜索（投资策略、市场研究和分析报告），赋能投资顾问。

哥伦比亚大学推出开源金融语言模型FinGPT,提供自动化数据管理和类似机器人顾问和算法交易等应用的潜力。


BloombergGPT: A Large Language Model for Finance

Shijie Wu[1,*], Ozan İrsoy[1,*], Steven Lu[1,*], Vadim Dabravolski[1], Mark Dredze[1,2], Sebastian Gehrmann[1], Prabhanjan Kambadur[1], David Rosenberg[1], Gideon Mann[1]
[1] Bloomberg, New York, NY USA
[2] Computer Science, Johns Hopkins University, Baltimore, MD USA
gmann16@bloomberg.net

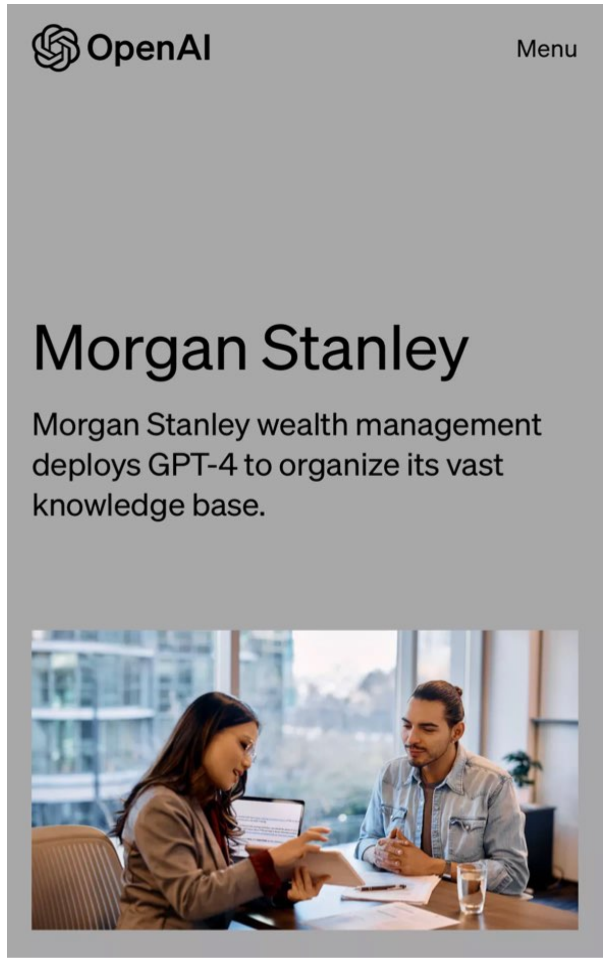
Abstract

The use of NLP in the realm of financial technology is broad and complex, with applications ranging from sentiment analysis and named entity recognition to question answering. Large Language Models (LLMs) have been shown to be effective on a variety of tasks; however, no LLM specialized for the financial domain has been reported in literature. In this work, we present BloombergGPT, a 50 billion parameter language model that is trained on a wide range of financial data. We construct a 363 billion token dataset based on Bloomberg's extensive data sources, perhaps the largest domain-specific dataset yet, augmented with 345 billion tokens from general purpose datasets. We validate BloombergGPT on standard LLM benchmarks, open financial benchmarks, and a suite of internal benchmarks that most accurately reflect our intended usage. Our mixed dataset training leads to a model that outperforms existing models on financial tasks by significant margins without sacrificing performance on general LLM benchmarks. Additionally, we explain our modeling choices, training process, and evaluation methodology. As a next step, we plan to release training logs (Chronicles) detailing our experience in training BloombergGPT.

Contents

*. Co-first authors.

1



AI应用概览

服务

- 小招客服
- 手机银行转账核身
- 坐席助手
- AI公积金客服
- 薪福通实名认证
- E餐通
- 95555语音导航
- 企业APP实名认证
- ……

营销

- 智能外呼
- 招行智网
- 人机协呼
- 交易链分析
- 天玑舆情专区
- 天玑画像平台
- 扶持有招
- ……
- AI双录机器人
- 企业新知

风控

- 贷后风险预警
- 合作机构关联风险预警
- 境外舆情预警
- 法规部可疑交易识别
- 同业风险预警
- 企业分层
- ……

运营

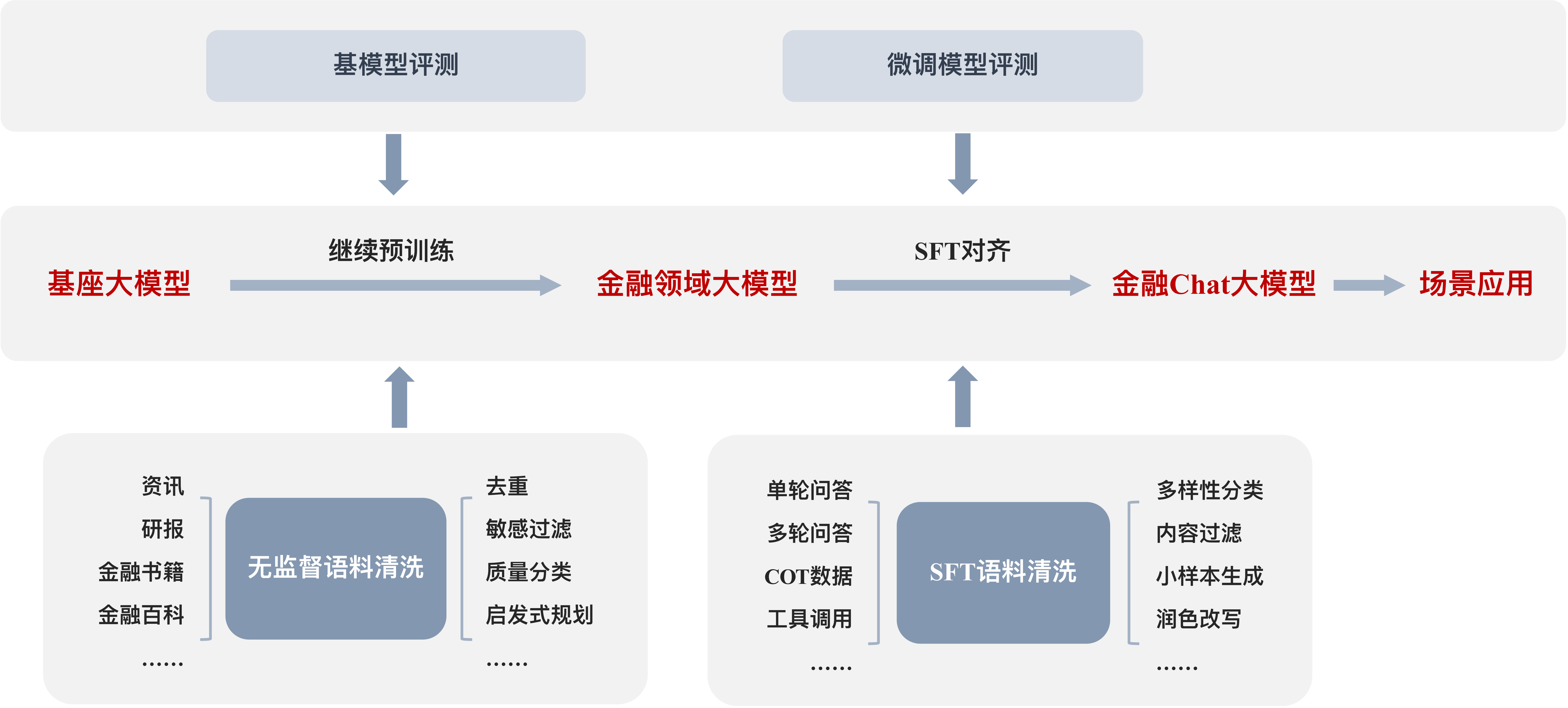
- AI对练机器人
- 同业存单融资规划
- 资债图中台
- 店周客流量统计
- 智能头寸预测
- 智能作业匹配与调度
- 实时指挥系统
- 智能审核
- 外汇预审
- 客服音频转录
- 智恒资产负债管理系统
- 大额现金智能预约
- 报单易
- 开户卡证识别
- 人脸识别
- 网点服务质量监测
- 支付结算审核
- 法规保函审核
- 合同审查
- ……

大财富

- AI小招
- 私募基金解读
- 行情云
- 天仪R+门户
- 行业研究中心
- 研究图谱
- CMBR
- 初心计划
- 托管投研一体化辅助平台
- 资产配置可视化平台
- ……

知识门户 - 知道

金融大模型的阶段性训练



AIGC的诞生及产业图谱

AIGC（AI Generated Content）是指通过人工智能技术和算法自动生成文本、图片、音频、视频等多种类型的内容。大语言模型促进了AIGC的快速发展。

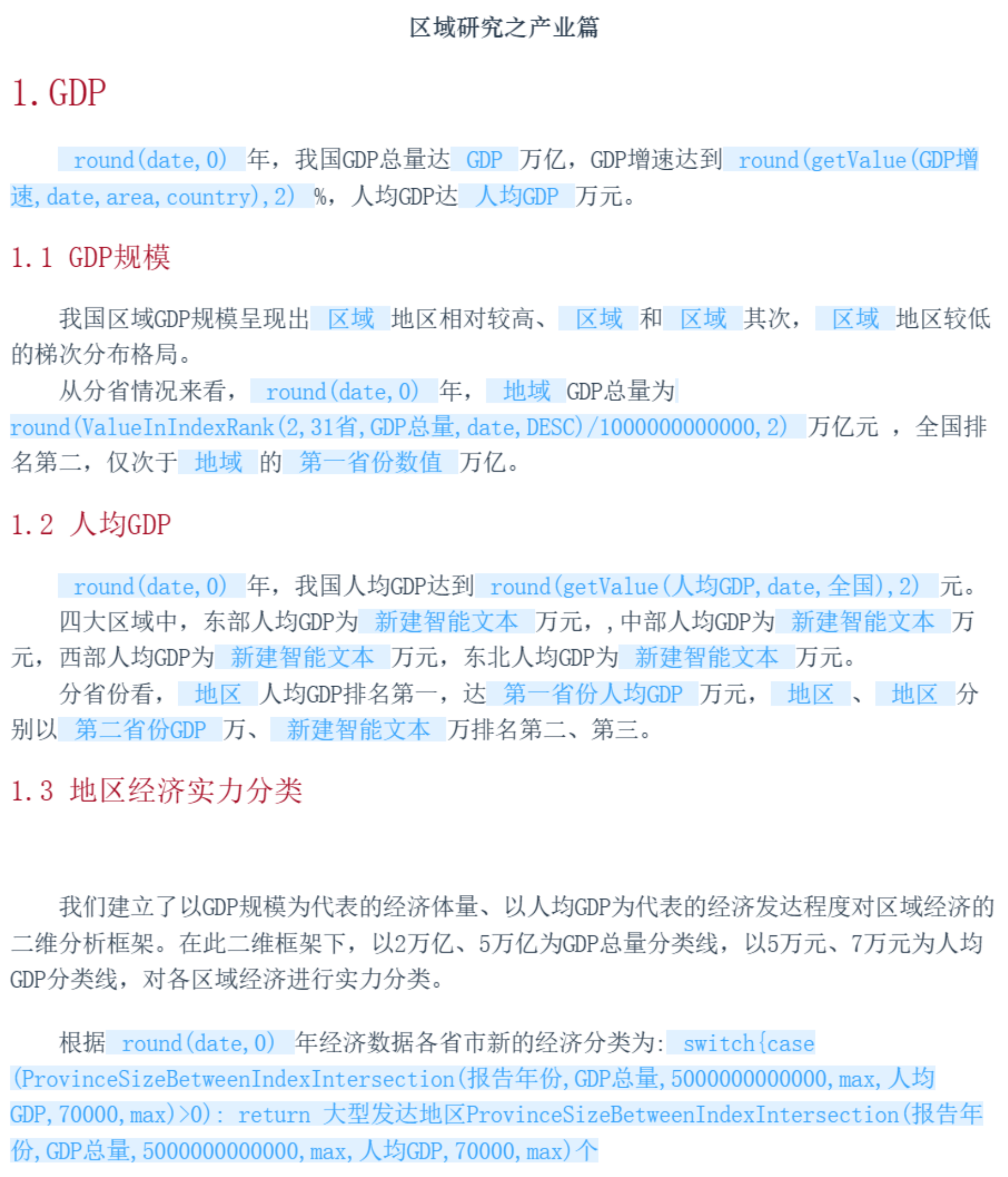
场景层	AIGC+工业、AIGC+医疗、AIGC+金融、AIGC+传媒、AIGC+教育、AIGC+娱乐、AIGC+电商、……
应用层	文本情感分析、新闻撰写、剧情续写、虚拟客服、视频插帧、电影风格转换、AI换脸、搜索引擎、Copilot、摘要生成、问答系统、短视频配音、地图语音导航
功能层	文本理解、结构化写作、非结构化写作、交互性文本；语音机器人、音乐合成、语音克隆；图像编辑与融合、2D图像生成3D模型；切换视频风格、画质增强修复、视频内容创作、动态面部编辑；文本生成图像、文本生成代码、文本生成音频、文本生成视频、图像/视频生成文本
模型层	自然语言理解NLU、自然语言生成NLG；Transformer、Flow-based models、BERT、VAE、CLIP、NeRF、CAN、GLM、Diffusion、DALLE、……
算力层	CPU、GPU、DPU、TPU、FPGA、ASIC

财富管理的业务视图

产品端
理财
公募基金
私募
FOF/MOM
收益/风险
发行人分析
FICC
权益
REITs
现金类
资产端
β被动
主动α
指数
风格
个体
事件
客户端
财富管理
资金供给
收益诉求
风险承受
流动性
期限
有效连接资金与资产
成长估值
风险预警
业务模式
实体经济
企业发展
产业趋势
授信管理
财务规划
并购
债券承销
经济走势
货币政策
研究端
宏观环境
金融环境
GDP
PMI
CPI
总投资
总产能

投研分析--研报辅助生成

- 探索大模型+投研能力，在研究员沉淀的投研框架之上，从传统规则模板式的辅助研报编写，向结合大模型的多样化生成转变，表达更灵活，题材更多样；
- 基于研究员总结的投研逻辑和大模型NL2SQL的技术，解决研报生成中大模型的幻觉问题。



区域研究之产业篇

1. GDP

round(date,0) 年，我国GDP总量达 GDP 万亿，GDP增速达到 round(getValue(GDP增速,date,area,country),2) %，人均GDP达 人均GDP 万元。

1.1 GDP规模

我国区域GDP规模呈现出 区域 地区相对较高、 区域 和 区域 其次， 区域 地区较低的梯次分布格局。

从分省情况来看， round(date,0) 年， 地域 GDP总量为 round(ValueInIndexRank(2,31省,GDP总量,date,DESC)/1000000000000,2) 万亿元 ，全国排名第二，仅次于 地域 的 第一省份数值 万亿。

1.2 人均GDP

round(date,0) 年，我国人均GDP达到 round(getValue(人均GDP,date,全国),2) 元。

四大区域中，东部人均GDP为 新建智能文本 万元，,中部人均GDP为 新建智能文本 万元，西部人均GDP为 新建智能文本 万元，东北人均GDP为 新建智能文本 万元。

分省份看， 地区 人均GDP排名第一，达 第一省份人均GDP 万元， 地区 、 地区 分别以 第二省份GDP 万、 新建智能文本 万排名第二、第三。

1.3 地区经济实力分类

我们建立了以GDP规模为代表的经济体量、以人均GDP为代表的经济发达程度对区域经济的二维分析框架。在此二维框架下，以2万亿、5万亿为GDP总量分类线，以5万元、7万元为人均GDP分类线，对各区域经济进行实力分类。

根据 round(date,0) 年经济数据各省市新的经济分类为: switch{case (ProvinceSizeBetweenIndexIntersection(报告年份,GDP总量,5000000000000,max,人均GDP,70000,max)>0): return 大型发达地区ProvinceSizeBetweenIndexIntersection(报告年份,GDP总量,5000000000000,max,人均GDP,70000,max)个

大模型+数字人

行内数字人技术与大模型技术有机结合，支持数字人主播与粉丝高效互动，并根据直播主题，智能生成台词。

■ 数字人直播

- **人机协同 无缝衔接**
  - 针对直播间的活动实况，后台人员可随时通过**语音驱动**数字人唇形开合，无缝切入，完成直播间粉丝互动。
- **完美主播 有问必答**
  - 直播间文本机器人可及时**理解用户意图**，积极与直播间粉丝互动，引导粉丝完成必要的购买或引流操作。
- **智能讲稿 AIGC自备台词**
  - 根据直播主题，大模型可**智能生成**主题讲解词，可支持7*24h不间断直播内容投放。

■ 数字人短视频

- **低成本制作 真假难辨**
  - 基于3-5分钟真人录像，即可完成数字人IP定制；
  - 播报效果真实自然，难辨真假，使客户有很好的观感体验。
- **多场景多渠道适配 低成本生产并投放优质内容**
  - 大模型可**基于关键词/核心语句描述生成指定的主题文本内容**，并通过数字人对此项内容完成播报和讲解；
  - 优质的短视频内容可在**多渠道**进行投放和传播，以扩充客户经理管户半径受限，实际服务不到位等问题。如，数字人替代客户经理，在手机端为私行客户解读资配建议书。

《生成式人工智能服务管理暂行办法》

当前，生成式人工智能正在加速发展，不断催生新场景、新业态、新模式和新市场，赋能千行百业。2023年7月国家互联网信息办公室联合多部委发布《生成式人工智能服务管理暂行办法》，界定了生成式人工智能技术的基本概念，规定了生成式人工智能服务提供者的制度要求，为生成式人工智能的健康发展指明了方向。

国家互联网信息办公室
中华人民共和国国家发展和改革委员会
中华人民共和国教育部
中华人民共和国科学技术部
中华人民共和国工业和信息化部
中华人民共和国公安部
国家广播电视总局
令

第15号

《生成式人工智能服务管理暂行办法》已经2023年5月23日国家互联网信息办公室2023年第12次室务会会议审议通过，并经国家发展和改革委员会、教育部、科学技术部、工业和信息化部、公安部、国家广播电视总局同意，现予公布，自2023年8月15日起施行。

国家互联网信息办公室主任 庄荣文
国家发展和改革委员会主任 郑栅洁
教育部部长 怀进鹏
科学技术部部长 王志刚
工业和信息化部部长 金壮龙
公安部部长 王小洪
国家广播电视总局局长 曹淑敏
2023年7月10日

第一章 总 则

第一条 为了促进生成式人工智能健康发展和规范应用，维护国家安全和社会公共利益，保护公民、法人和其他组织的合法权益，根据《中华人民共和国网络安全法》、《中华人民共和国数据安全法》、《中华人民共和国个人信息保护法》、《中华人民共和国科学技术进步法》等法律、行政法规，制定本办法。

第二条 利用生成式人工智能技术向中华人民共和国境内公众提供生成文本、图片、音频、视频等内容的服务（以下称生成式人工智能服务），适用本办法。

国家对利用生成式人工智能服务从事新闻出版、影视制作、文艺创作等活动另有规定的，从其规定。

行业组织、企业、教育和科研机构、公共文化机构、有关专业机构等研发、应用生成式人工智能技术，未向境内公众提供生成式人工智能服务的，不适用本办法的规定。

第三条 国家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。

第四条 提供和使用生成式人工智能服务，应当遵守法律、行政法规，尊重社会公德和伦理道德，遵守以下规定：

（一）坚持社会主义核心价值观，不得生成煽动颠覆国家政权、推翻社会主义制度，危害国家安全和利益、损害国家形象，煽动分裂国家、破坏国家统一和社会稳定，宣扬恐怖主义、极端主义，宣扬民族仇恨、民族歧视，暴力、淫秽色情，以及虚假有害信息等法律、行政法规禁止的内容；

（二）在算法设计、训练数据选择、模型生成和优化、提供服务等过程中，采取有效措施防止产生民族、信仰、国别、地域、性别、年龄、职业、健康等歧视；

（三）尊重知识产权、商业道德，保守商业秘密，不得利用算法、数据、平台等优势，实施垄断和不正当竞争行为；

（四）尊重他人合法权益，不得危害他人身心健康，不得侵害他人肖像权、名誉权、荣誉权、隐私权和个人信息权益；

（五）基于服务类型特点，采取有效措施，提升生成式人工智能服务的透明度，提高生成内容的准确性和可靠性。

Thank you