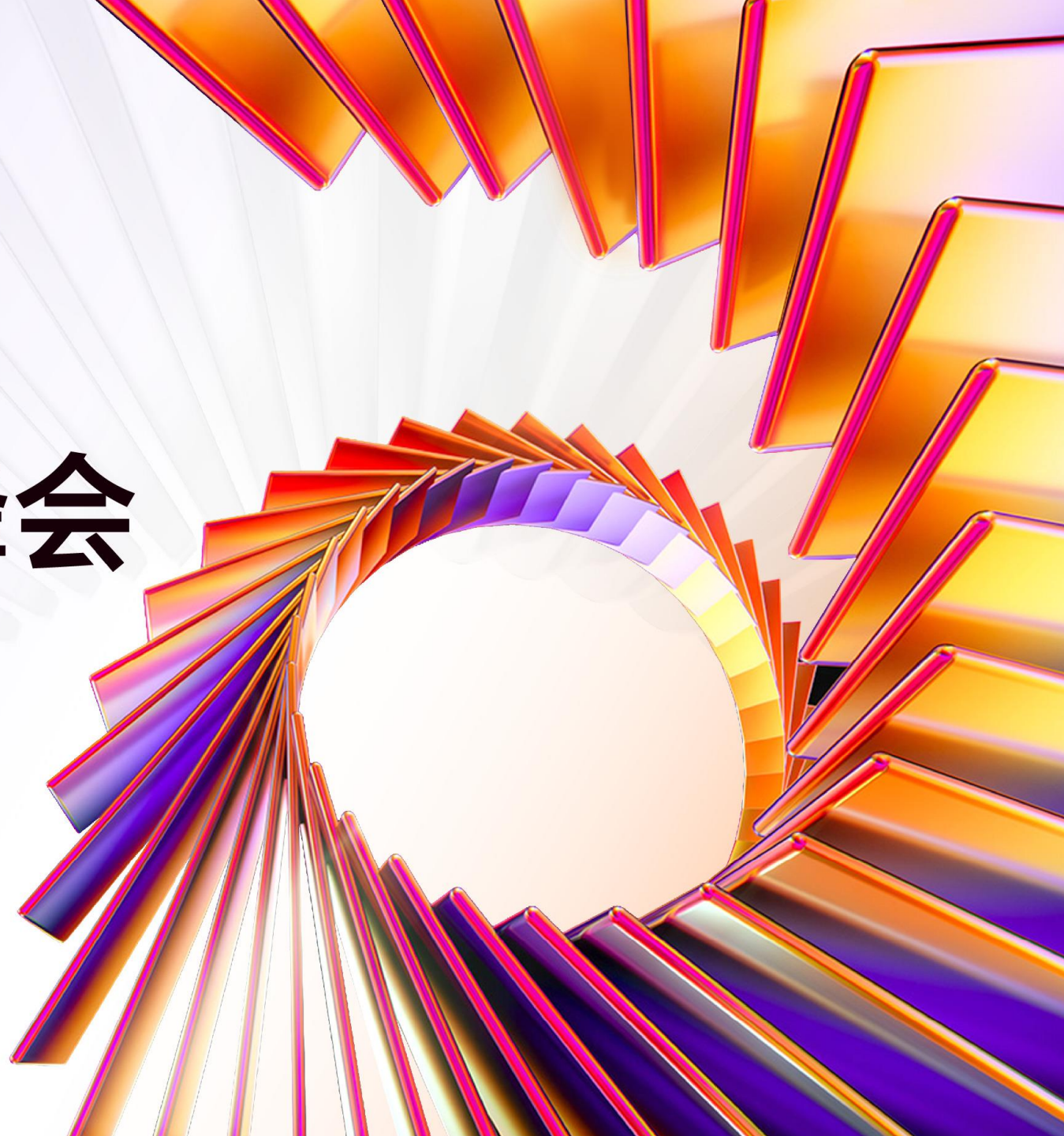
AiDD 6th 2025

2025 AI+ Development Digital Summit

AI+研发数字峰会

拥抱AI 重塑研发

05/23-24 | 上海站



AiDD 7th 2025

2025 AI+研发数字峰会

拥抱AI 重塑研发

AI+ Development Digital Summit

查看会议详情

下一站预告 08/08-09 | 北京站 11/14-15 | 深圳站

北京站论坛设置

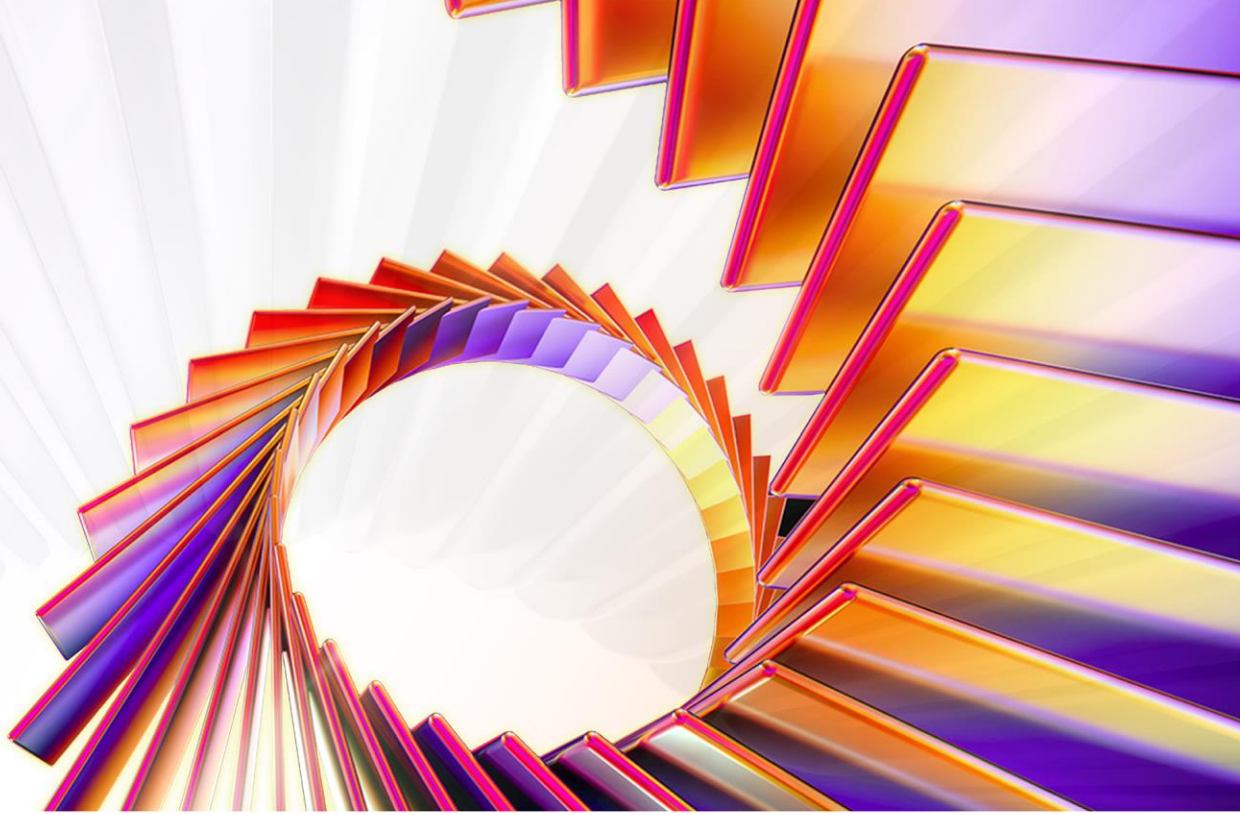
- AI+ 金融业务创新
- 大模型安全与对齐
- 具身智能与机器人
- 大模型和 AI 应用评测
- 智能需求工程
- 大模型应用开发框架与实践
- 代码生成及其改进
- 智能存储与检索技术
- 智能体与研发效率工具
- 智能体经济（Agentic Economy）
- AI+ 新能源汽车
- 下一代知识工程
- AI 产品运营与出海策略
- 智能测试工具的开发与应用
- AI 前沿技术探索与实践

AiDD 6th 2025 | 05/23-24 | 上海站

2025 AI+ Development Digital Summit

AI+研发数字峰会

拥抱AI 重塑研发



领域推理引擎-大模型不止于小作文

徐彬 | 德邦证券

机器会思考吗？

徐彬

算法架构师　　无双谱@知乎

《实战深度学习算法》著者，多项智能算法专利发明人，CCF会员。

研究方向为信用风险管控、复杂项目群管理、机器学习在特定场景的应用。

历任 平安银行 应用架构专家，银行间市场清算所 创新衍生品及利率产品项目群负责人。

牵头完成多项证券业协会、交易所研究课题。

目录
CONTENTS

PART 01

大模型应用的现状

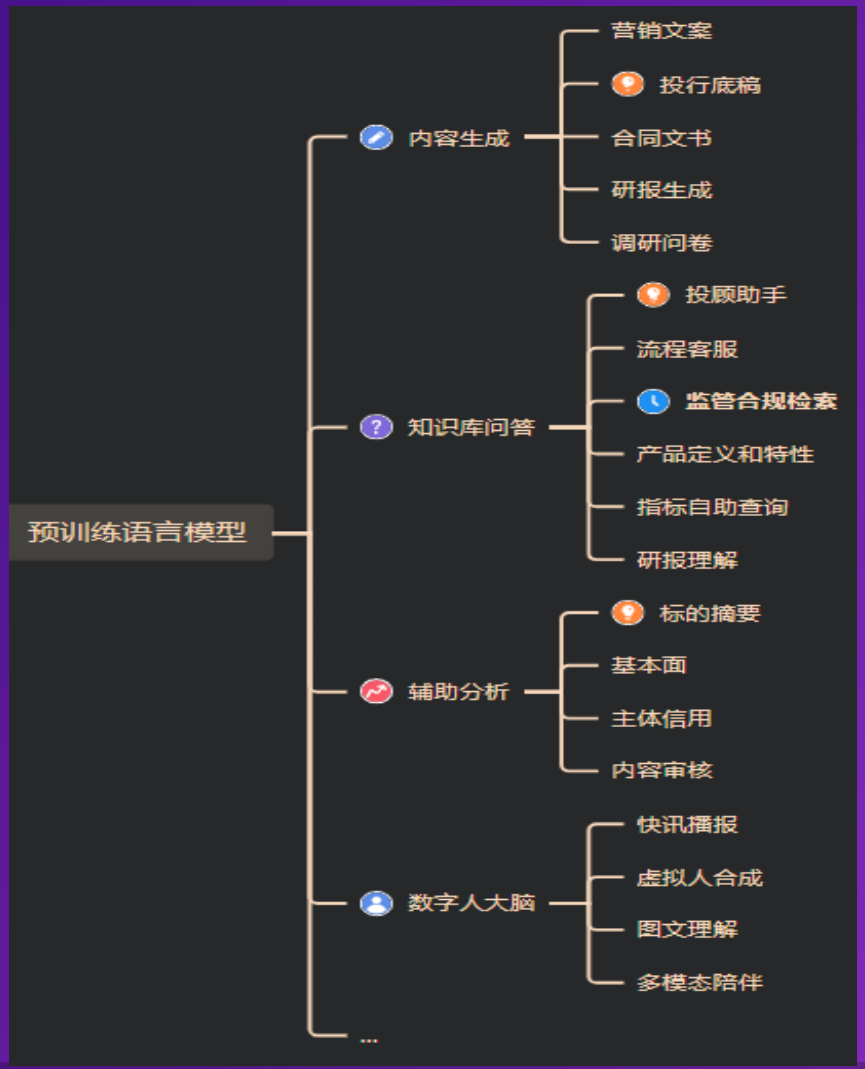
大模型应用的现状

模型能力 → 应用能力 → 应用场景

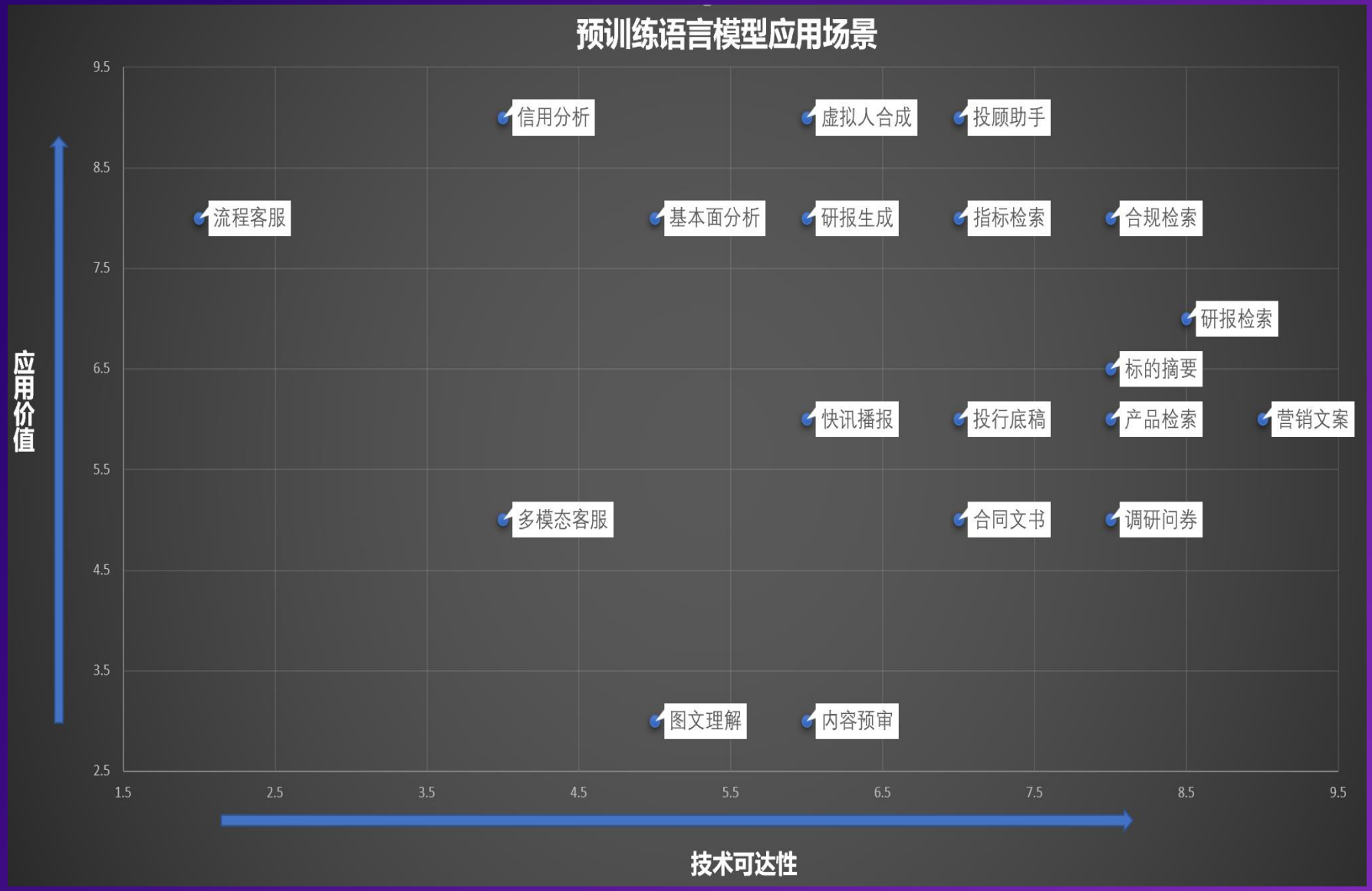
1.情景学习(In-Context Learning)
模型通过阅读对话的历史上文，可以续写后续的回答，这个能力使得LLM有了"短期记忆"，可以和人类对话了。

2.思维链(Chain-of-Thought， CoT/ToT)
给大模型问题提示时，如果不直接给答案，而是给出推理过程，可以解锁语言模型对复杂问题的推理解决能力。

3.自然指令学习(Learning from Natural Instructions)
只需要使用少量自然指令针对对专项任务做微调，模型就可以得到很好的泛化能力。



大模型应用的现状


预训练语言模型应用场景
应用价值
技术可达性
9.5
8.5
7.5
6.5
5.5
4.5
3.5
2.5
1.5
2.5
3.5
4.5
5.5
6.5
7.5
8.5
9.5
信用分析
虚拟人合成
投顾助手
流程客服
基本面分析
研报生成
指标检索
合规检索
研报检索
标的摘要
快讯播报
投行底稿
产品检索
营销文案
多模态客服
合同文书
调研问券
图文理解
内容预审

PART 02

问题和溯源

大模型的本质

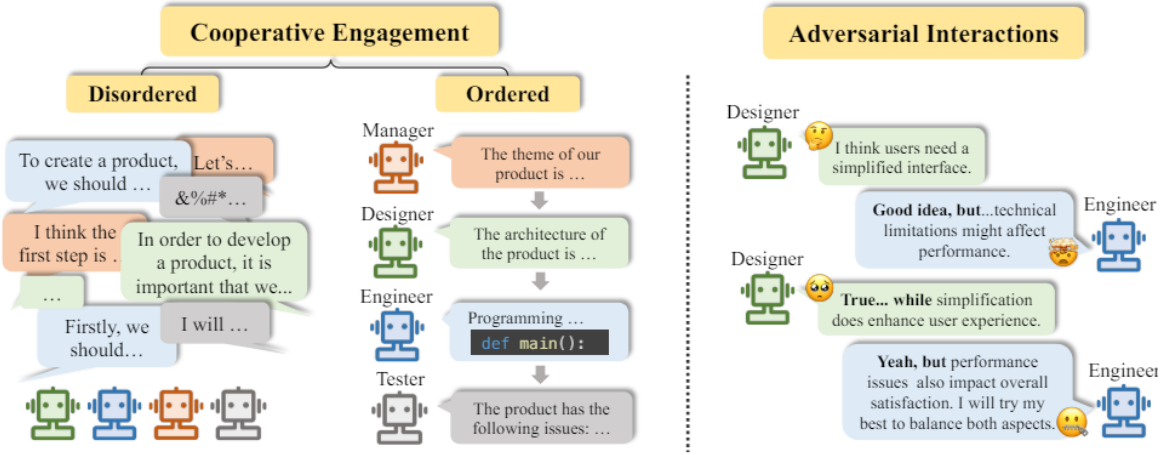
预训练语言大模型(LLMs)本质上是统计语言模型(SLMs)
基础能力在规模法则(Scaling Law)作用下，衍生出更具落地价值的应用能力
通过知识库和外部工具，进一步扩展了能力边界

世界知识 | 语义理解 | 多语种 | 编码 | 指令遵从 | 推理 | 情景学习 | 对话互动 | 自我改进 | 工具调用

基础能力 → 衍生能力 → 辅助能力

多智能体Multi-Agents，拆解任务，制定计划， 分头执行,如：独立与LLM交互，RAG、执行工具等等。

任务分解：复杂问题分解为更小的问题
规划：制定一组任务计划
存储：之前完成的任务存储为上下情景知识
工具使用：选择要使用的工具+使用工具的参数
场景适应：提升螺旋



问题和溯源

新加披国立大学的学者分析认为，大语言模型的幻觉问题无法避免（Ziwei Xu et,al . 2024）。

用包括客观事实、新闻、文学作品、学术、健康等不同类别的信息对ChatGPT做了测试，错误及虚假信息占比为75%（Zuying, et,al. 2023）。

幻觉现象可以溯源到大模型底层的训练过程、训练语料构建方法。

问题和溯源

需要承认以下基本事实：

- 以目前主流的训练数据、提示工程技术、学习算法和模型架构，大语言模型不可避免地会产生幻觉

- 在当前有偏训练数据、训练方法，以及关键算法无法捕捉到足够多真实世界函数的情况下，单纯增加模型参数和训练数据是无效的，不能根本解决幻觉问题。

幻觉问题尽管尚无根本解决方法，但是针对特定的应用场景仍可以找到有效的缓解方法。

PART 03

应对：从知识库到多智能体

知识库+多智能体

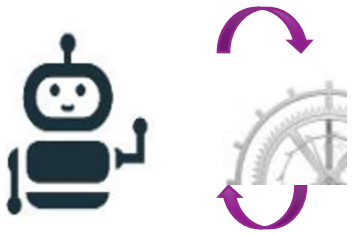
幻觉问题尽管尚无根本解决方法，但是针对特定的应用场景仍可以找到有效的缓解方法。

基于大模型+专项知识库+搜索引擎构建一套面向特定场景的SOP工作流借助通用大模型的few shot learning能力，实现可信结果输出；

- 范式一：LLM+工作流+Agent
- 范式二：多智能体协同
- 范式三：模型驱动自主代理

范式一: Agent+工作流+LLM

Agent的本质：一组固化的提示词



Agent　　工作流　　LLM

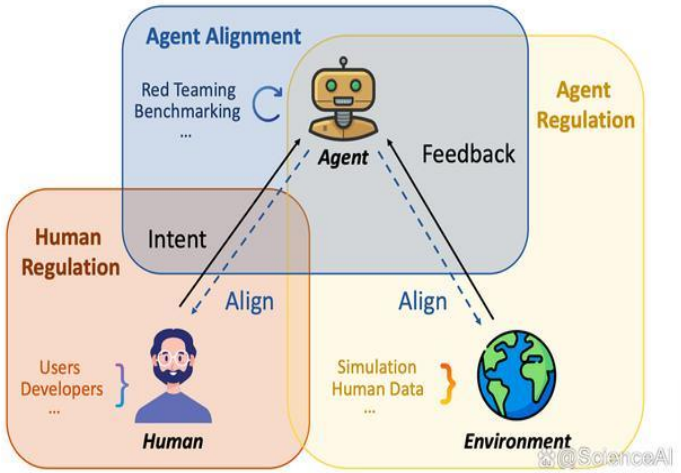
数字化->自动化->智能化

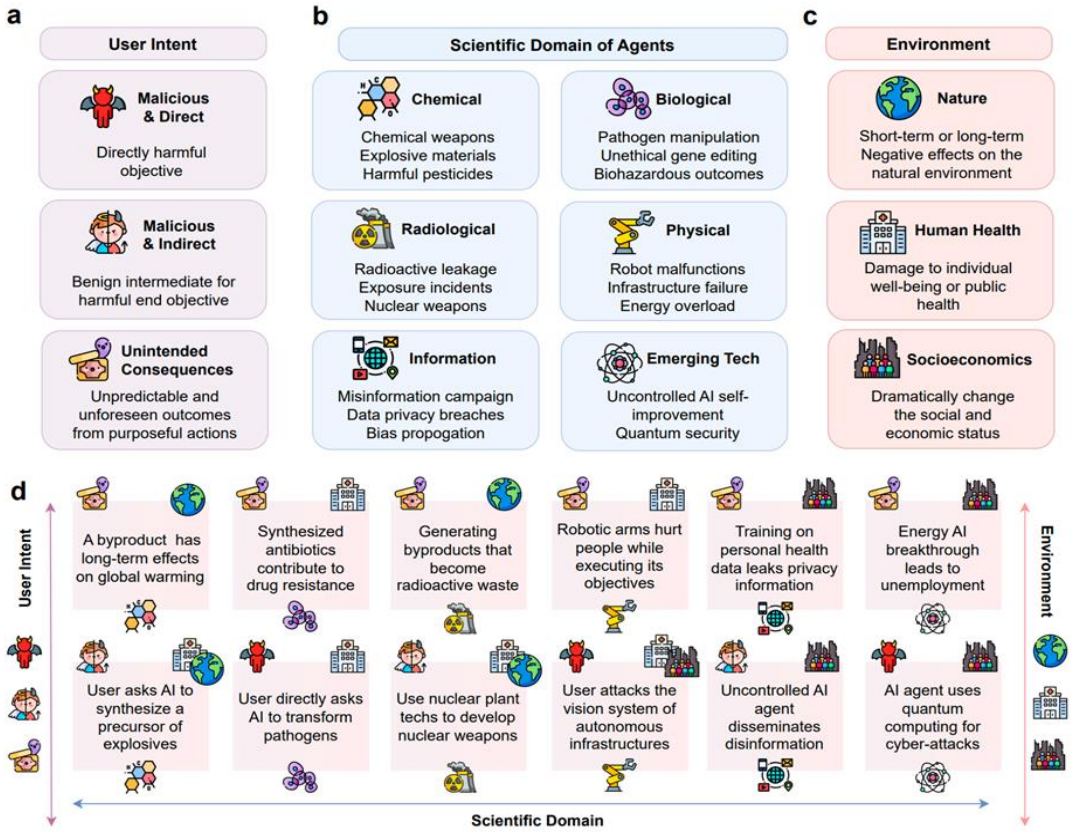
- ✓ **产品**-> 好点子，PRD新范式
- ✓ **开发**-> 有坑吗？快速验证
- ✓ **测试**-> 分裂思维，大有可为

每个人都是solo创造者

范式二：多智能体协同

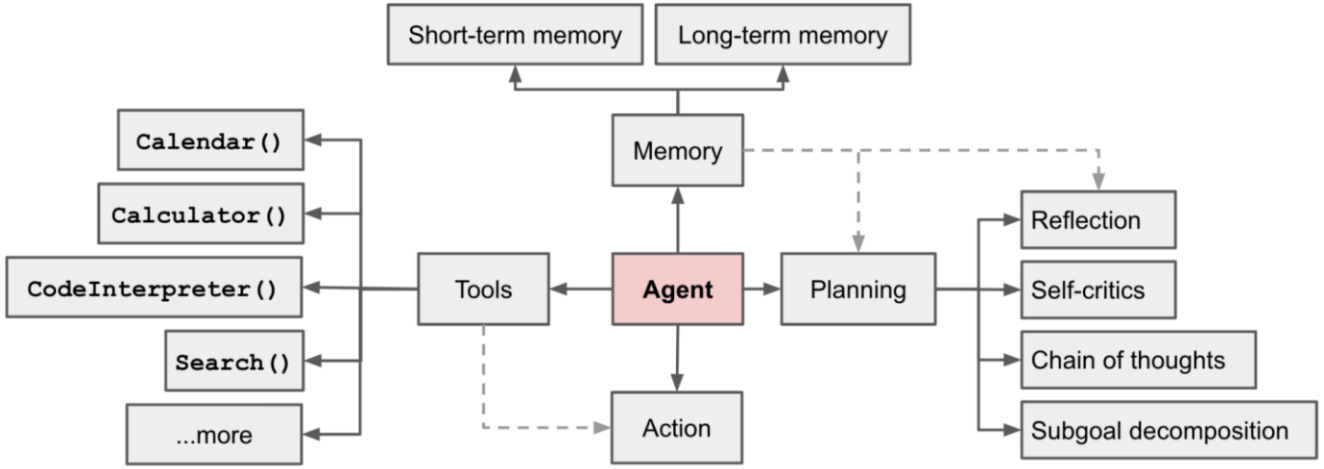
Single assistant ---> Team of coordinated intelligent agents



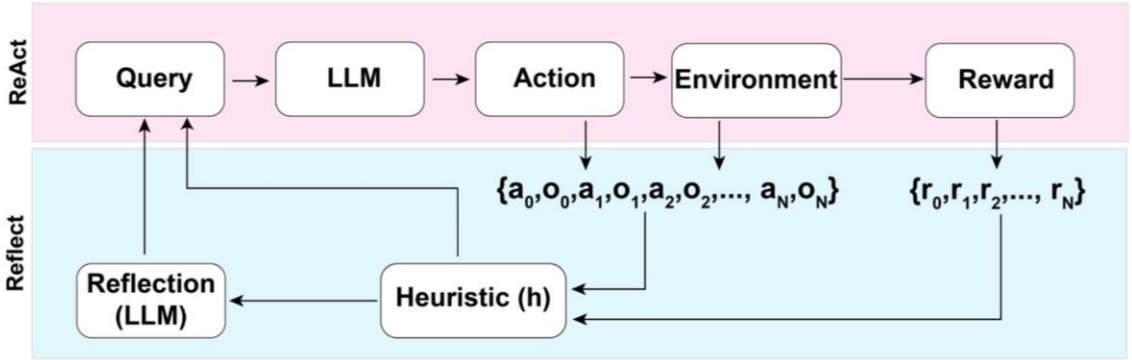


范式三：模型驱动的自主代理

- 多种工具逐步完成拆解后的子任务
- 联网沙盒
- 编写和运行编程语言代码
- 沙盒内独立安装包和依赖
- 可切换回用户交互进行敏感操作



根据自我反思的结果，判断重置Sandbox环境，开始新的子任务



图片来源： https://huggingface.co/blog/open-source-llms-as-agents

PART 04

进阶：领域智慧+多任务推理

进阶：领域推理引擎

基于大模型构建一套学习承载结构化的领域分析范式，驱动客观支持数据流的训练推理引擎；结合高精度的专项模型，学习领域专家的分析过程，借助通用大模型的基础能力，实现可信结果输出；

与甩手掌柜式的调用大模型不同，整个引擎在良好设计经过适用训练的专项模型支撑下，学习领域专家的思考方法，得到结构化的可复用范式，以可控方式嵌入大语言模型的生成能力，进而驱动领域场景推理，从而避免最终的输出结果出现事实性错误。

整体演进路线

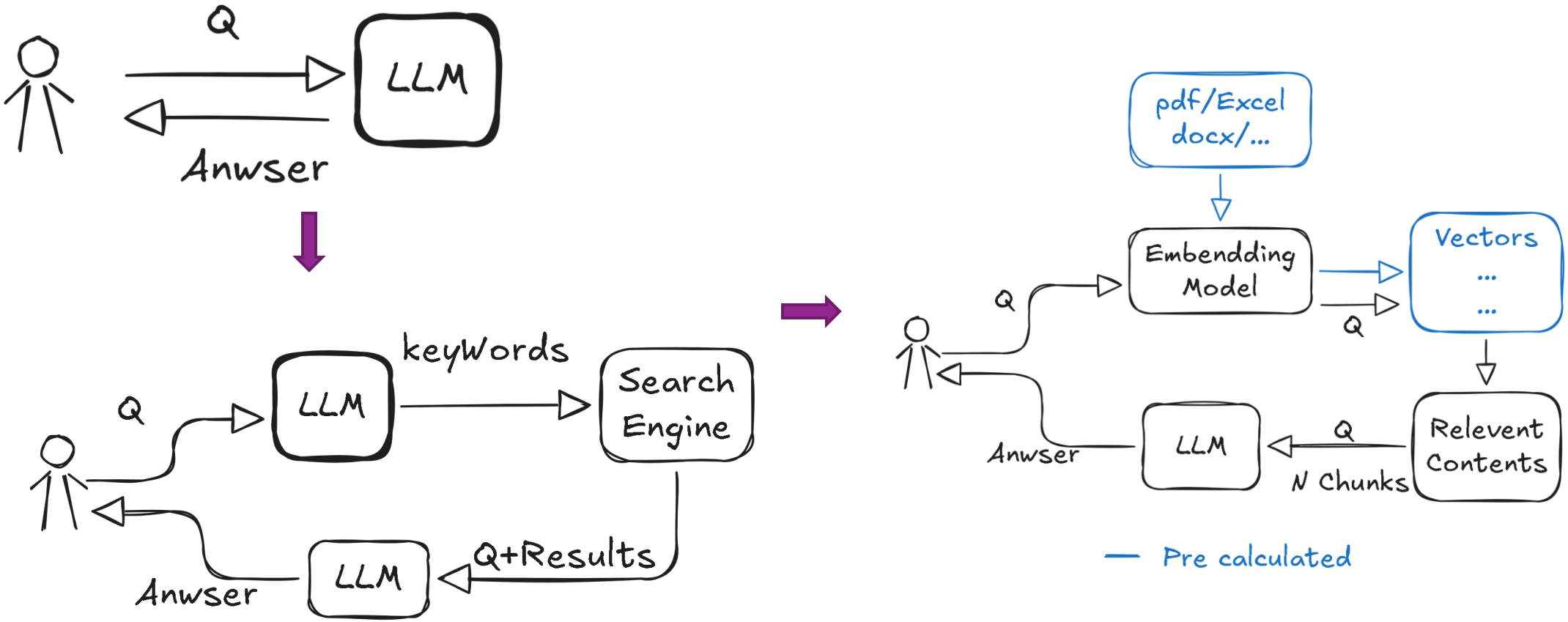
多种可复用的范式（1~5级：领域深度、连接便捷性）

- L1基于任务提示指令组织可信数据或搜索引擎结果，由大模型组织成自然语言
- L2增加基于大模型的模糊规则判断，以及大模型交叉验证的辅助手段
- L3基于大模型做场景意图识别，Manus类的虚拟机技术，通过MCP协议，串联微服务，提供功能强大的集成服务
- L4：以场景领域的分析框架为SOP，驱动服务集成，实现具有专业深度的生产可用级大模型应用
- L5：基于多模态理解能力，理解物理世界，操作app/软件、驱动具身智能

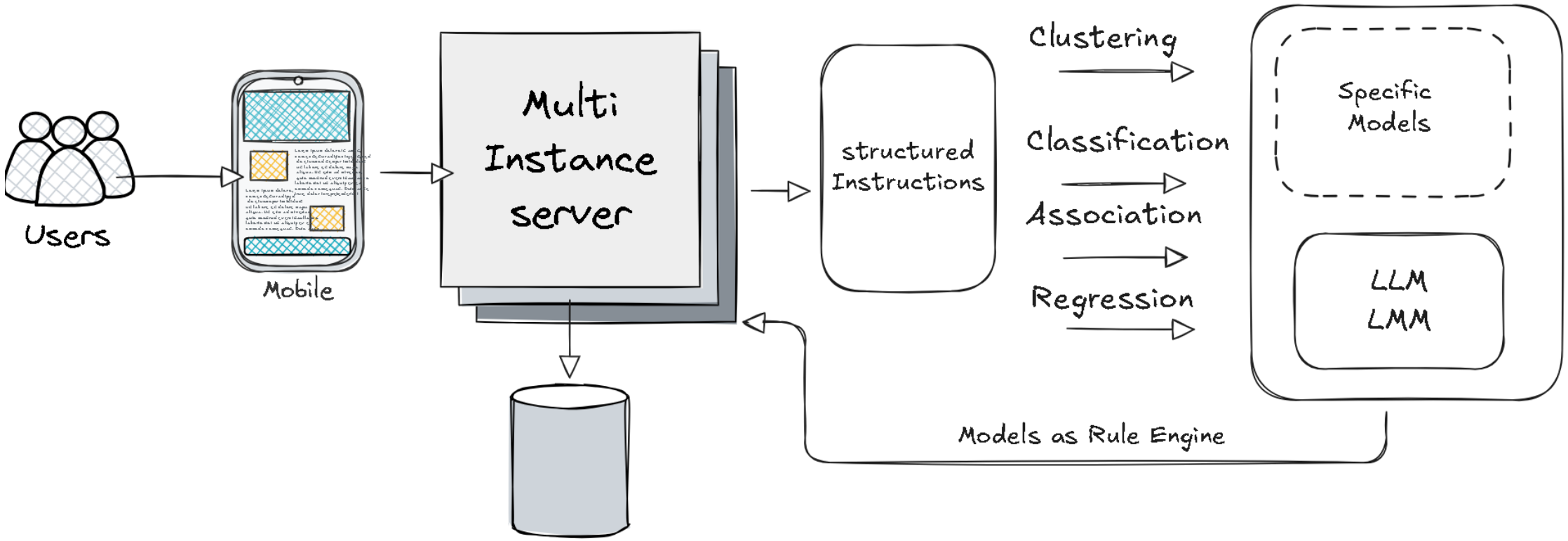
引申问题：如何评估?

小作文型应用

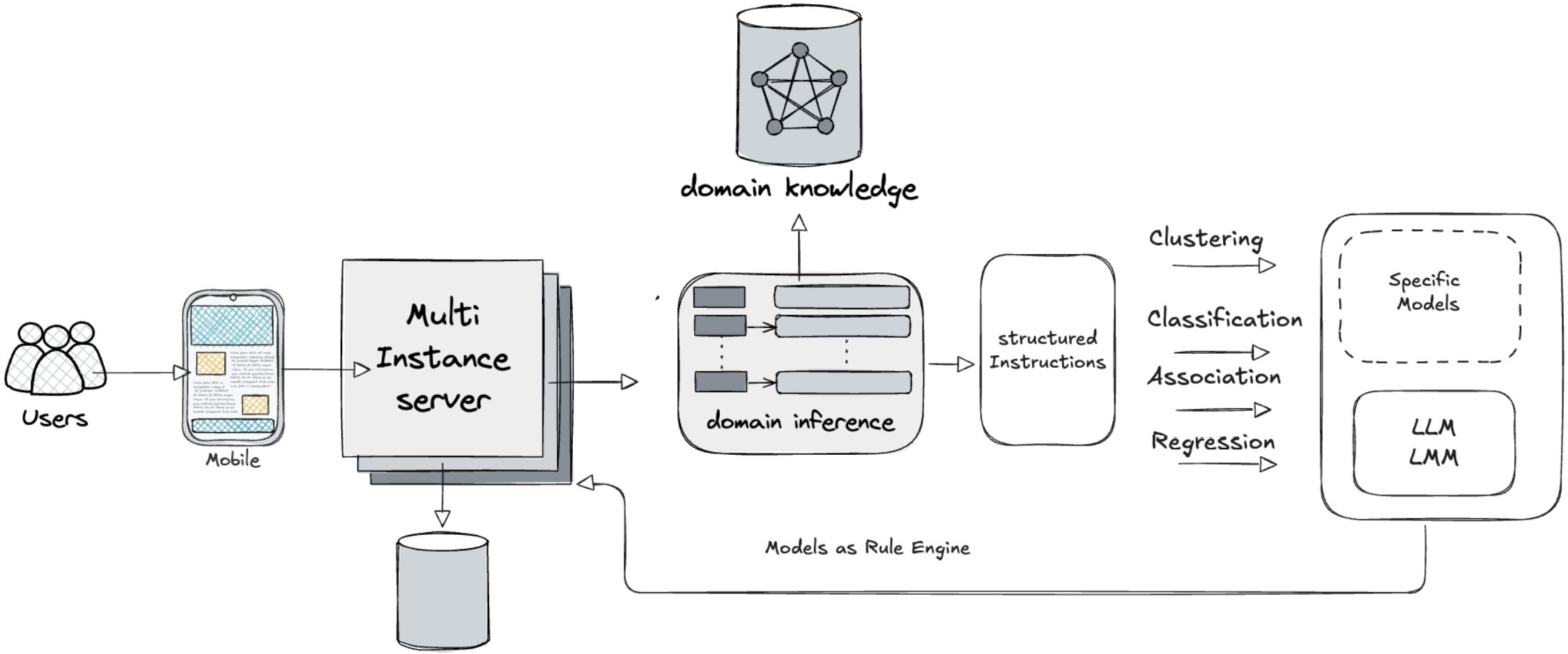
直接问答---> 联网搜索 ---> 增强检索知识库



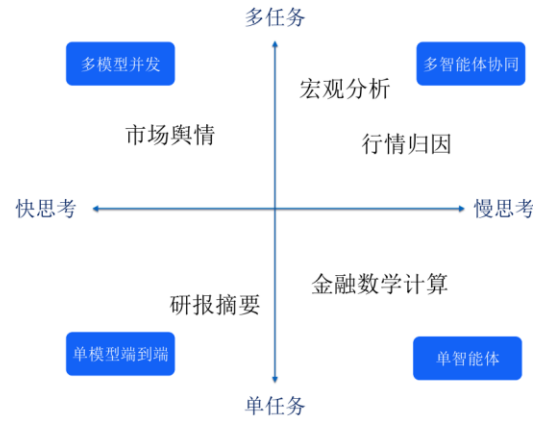
集成进应用系统群

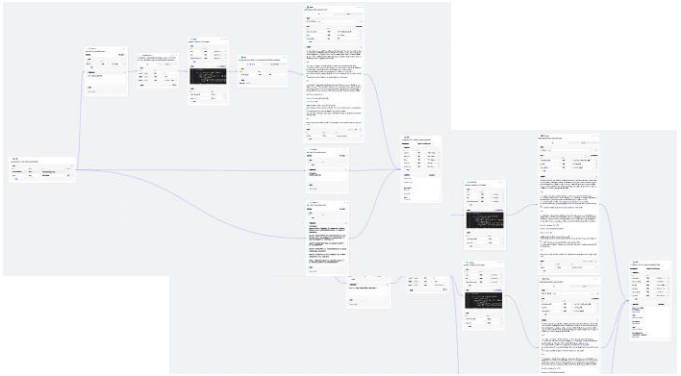

Users
Mobile
Multi
Instance
server
structured
Instructions
Clustering
Classification
Association
Regression
Specific
Models
LLM
LMM
Models as Rule Engine

领域智慧+推理引擎


domain knowledge
Users
Mobile
Multi
Instance
server
domain inference
structured
Instructions
Clustering
Classification
Association
Regression
Specific
Models
LLM
LMM
Models as Rule Engine

案例: 市场快报





棉花 (大宗商品)

政策资讯

棉花是国家重要的农产品之一，我国对棉花产业实施了全面的管理和调控。
法》[1][3][4]，中央储备棉的管理旨在确保储备棉数量真实、质量良好、储
发挥重要作用。国家发展改革委负责中央储备棉的规划和总量计划[1][4]，
管理及监督检查[1][4]。此外，为促进棉花产业健康发展，政府还颁发了多
理条例》和《棉花加工资格认定和市场管理暂行办法》等，对棉花从种植到
[5]。这些政策和管理办法体现了国家对棉花产业的高度重视和严格监管。
引用:
[1] 中央储备棉管理办法国家发展和改革委员会中国政府网
[2] Mysteel解读：2024年中央一号文件对棉花市场影响分析
[3] 国家发展改革委、财政部印发《中央储备棉管理办法》
[4] 中央储备棉管理办法
[5] 中国棉花行业政策分析：政府颁发大量政策促进棉花产业健康 ...

市场资讯

根据2024年上半年的棉花市场情况，棉花价格经历了先涨后跌的过程。2月
点，为16855元/吨，主要是由于春节前后的阶段性备货和外盘美棉超卖上涨
着"金三银四"旺季需求不旺，国内棉花实际播种面积减幅不及预期，导致棉
价格跌至15362元/吨，创下年内新低，并刷新了2023年4月以来的低点[1]。
上半年棉花市场价格上涨后下跌，与美联储可能的降息、业者对宏观经济向
滑、商业库存同比偏高、新棉减种不及预期等因素有直接关系[1]。同时，下
显下降，尤其是棉纺织市场旺季需求不旺，棉花消费量有所下降[1]。
在进口棉方面，2024年上半年，由于国内纺织企业需求逐步走弱及内外棉价
场进口棉成交出现阶段性低迷，同步影响进口棉价格[2][3]。目前青岛港的B
别为15450-15650元/吨、16600-16750元/吨和16750-16850元/吨，美金报价
88美分/磅和88-88.5美分/磅[2][3]。
对于2024年棉花价格的展望，首先需要解决纺织产品库存问题，其次关注下
恢复情况。纺织服装行业对外贸依赖程度较高，因此还需重点关注国内政策
2024年棉花现货价格走势的高点可能出现在5月或10月，低点可能出现在7月
此外，2024年前半年，郑棉与美棉走势呈过山车行情，美棉在低库存炒作下
下出现年内高点[3]。但随着东南亚地区无法消化高昂的进口棉价，美棉库存
下行。尽管1月份郑棉走势震荡上行，但春节临近后，下游利多因素逐渐弱化
总的来说，2024年上半年棉花市场经历了价格波动，受到多重因素影响。下
品库存、棉花生产和需求恢复情况，以及国内外政策和宏观形势。目前信息
势的详细分析。
引用:
[1] 棉花半年度总结：上半年价格先涨后跌 下半年仍有下探空间
[2] Mysteel半年报：2024上半年棉花市场回顾及下半年展望|纺企 ...
[3] 【棉花半年报】新年度供需矛盾清晰且提前交易，后期以扰动 ...

上周行情一览

市场行情来源：万得资讯
根据您提供的信息，以下是棉花商品对应的指标数据的表格形式：

白糖 (大宗商品)

政策资讯

白糖，也称为白砂糖，是食糖的一种，是经过精制的蔗糖。在中华人民共和国，白糖的管理受到严格的监管。根据《中央储备糖管理办法》规定，中央储备糖包括原糖和白砂糖，其目的是为了实施市场调控、稳定市场预期、应对突发事件引发的市场异常波动[1]。此外，关于白糖的海关特殊监管区域和区外加工贸易管理措施，也有相应的调整公告，规定自2024年7月1日起，对海关特殊监管区域和区外加工贸易食糖的管理措施进行调整[2][3]。这些管理措施涉及多个部门，包括国家发展改革委、财政部、农业农村部、商务部、税务总局等[2][3]。具体的管理措施细节和调整内容，需要进一步查阅相关公告和解读文件以获得详细信息[2][3][5]。
引用:
[1] 中华人民共和国国家发展和改革委员会 中华人民共和国财政部 ...
[2] 海关总署公告2024年第50号（关于调整海关特殊监管区域和区 ...
[3] 公告解读：关于调整海关特殊监管区域和区外加工贸易食糖 ...
[4] 中国糖业协会
[5] 海关总署解读：关于调整食糖管理措施〔2024〕44号公告的解读

市场资讯

白糖市场在2024年表现出了一定的波动性。根据东方财富网的深度分析与预测[1]，2024年第一季度，白糖市场价格经历了两波上涨，主要受到春节备货需求增加和糖厂收榨后挺价心态增强的推动。然而，尽管价格上涨，一季度的平均价格较上季度下跌了6.6%，这可能与之前的高价位有关。
新浪财经报道[2]指出，国内现货价格的波动一方面是供需关系的影响以及阶段性的供需影响；另一方面受到国内期货价格走势的影响。白糖期货工具在白糖产业中的应用非常广泛。
从新浪财经的另一篇报道[3]中了解到，白糖供应偏紧，去库导致趋势上涨。在全球供需平衡表中，国内产量大约1000万吨，进口约500万吨，对外依存度大约30%。此外，白糖产业网提供了一些具体的市场动态信息[4]，例如8月14日广西现货市场白糖价格下跌，制糖企业报价6330-6370元/吨，较前一日小幅下调10元/吨。
第一财经的报道[4]显示，截至2024年3月末，泛糖专区的累计成交量突破19万吨，成交金额超过13亿元，涵盖了全国主要的白糖生产企业与贸易企业。
综合以上信息，可以看出2024年白糖市场受到供需关系、节日需求、期货市场等多种因素的影响，价格出现了波动。
引用:
[1] 2024年白糖市场波动趋势：深度分析与预测财富号东方财富网
[2] Mysteel年报：2023年国内白糖市场回顾与2024年展望 - 新浪财经
[3] 白糖供应偏紧，去库致趋势上涨 - 新浪财经
[4] 白糖产业网 - 白糖价格、白糖行情与白糖资讯服务平台 - 生意 ...
[5] 2024年6月6日，第一财经：稳收入、兴产业，期货工具为白糖 ...

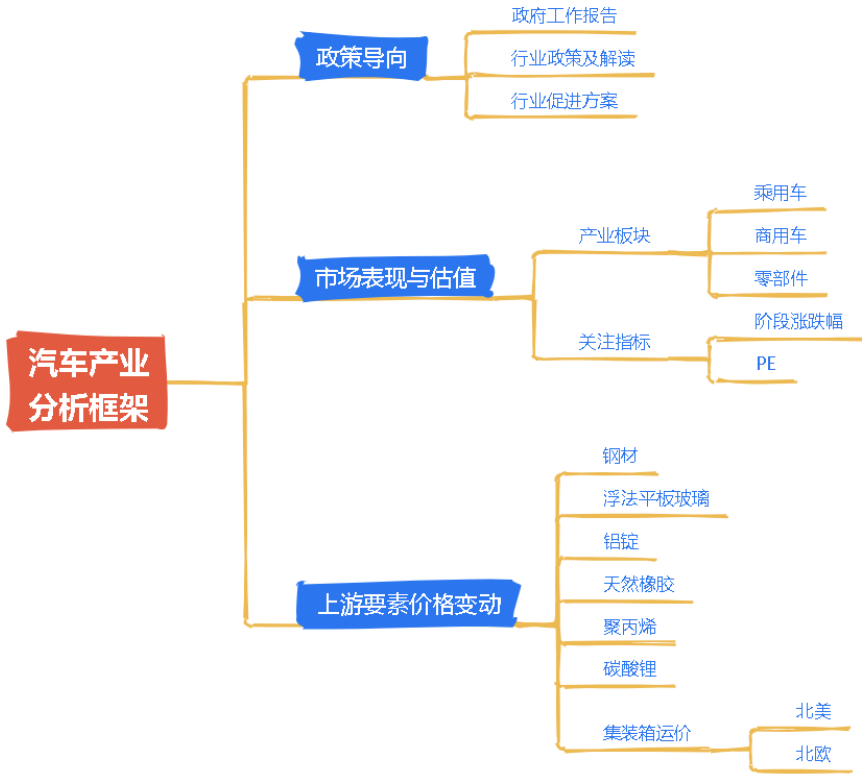
上周行情一览

市场行情来源：万得资讯

日期	现货价格（元/吨）	周涨跌幅%	周成交额（亿元）
8-15	6790	2.92%	360.39

案例: 产业分析

- 理解分析框架
- 对齐分析思路

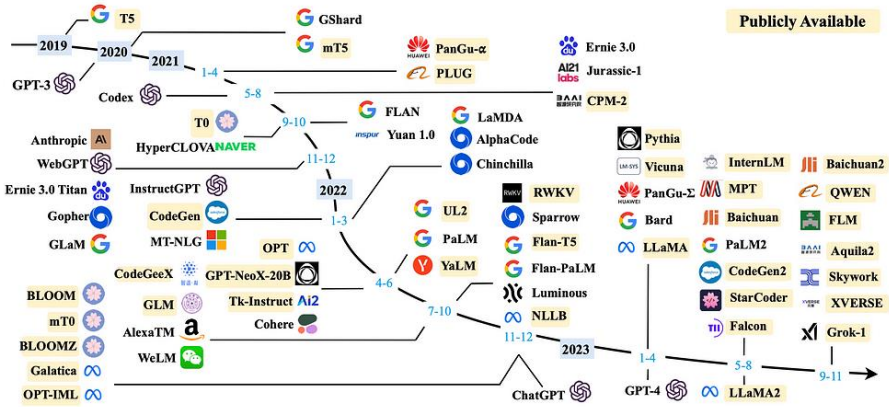


如何评估?

一、行业场景，不拘一格

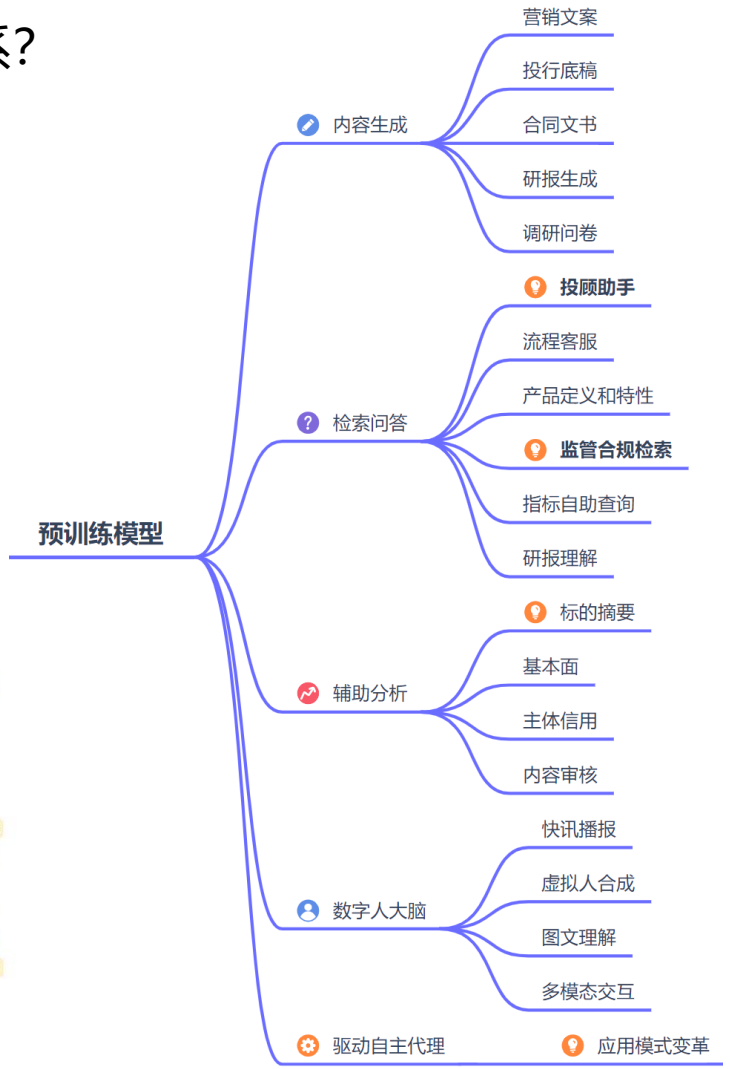
应用领域	典型场景	能力	关键技术	成熟度 1~5
投研和投顾	内容生成、数字人分身、基本面分析、投资组合分析、智能交易和量化分析、研究辅助	财务分析、资讯分析、智能条件单、算法交易、宏观报告理解、研报生成	内容生成、自然语言理解、图像识别、决策模型、	2
营销展业客服	产品推荐、客户引流、客户留存、辅助营销、智能客服、主动外呼	精准推荐、客群画像、营销活动策划、营销内容生成	内容生成、关联分析、聚类分析、协同过滤、知识图谱	5
运营管理辅助决策	智能RPA、经营决策辅助	自动人机交互、流程自动化、	图像识别、决策模型	3
风控和监管	标准检索、主体信用风险和声誉风险监测、资产信用风险和市场风险监测	资讯风险识别、舆情情绪识别、信用评估与推理	检索增强、自然语言理解、内容生成、知识图谱	4
其它智能应用领域	辅助办公自动化、投行尽调文档生成、辅助文档审核、智能运维	合同底稿文档生成、文档内容对比、异常监测和故障预警	多模态生成、实体识别、图像识别、	2

二、太多模型，怎么选?



来源：A Comprehensive Survey of Large Language Models (LLMs)

三、新生事物，评估体系?



来源：课题组分析整理

通用评测体系适用吗?

一、评测维度：水土不服

应用在证券经营机构，需要

- 符合监管合规要求
- 适配业务线的特色流程
- 满足风险管理需要
- 维护市场秩序和保障投资者利益

一级维度	二级维度	维度描述
问答任务能力	二值问题回答	关注特定类型的问题回答任务，主要为二值（是/否）答案
	对话式问答	关注涉及对话上下文的问题回答任务
	开放领域问答	关注回答开放领域的问题
	信息寻求对话	关注与模型进行对话以获得特定信息
语言理解能力	预测段落最后一个单词	关注模型对文本生成和连续性的理解
	故事结束预测	关注模型预测故事的可能结束
	阅读理解	关注模型从给定的文本中提取或推断信息
	多模态语言理解	关注结合多种模式（如文本、图像和声音）来理解语言
逻辑推理能力	常识推理	关注模型对常识和逻辑的理解能力，要求模型具备尝试推理能力，理解和推理因果关系
	自然语言推理	关注模型根据给定的前提推断出结论
	深度推理	关注模型进行更深入的推理以回答问题
	数学推理	关注模型在数学问题上的推理能力
	科学推理	关注模型对科学概念和事实进行推理
文本生成能力	文本生成	关注模型自动产生连贯、有意义的文本，通常基于给定的上下文或提示，代码生成也属于此类范畴
基础任务能力	句子比对	关注比较两个句子的语义相似性或差异性
	词义消歧	关注确定一个词在特定上下文中的正确含义
	代词消除歧义	关注于正确解决代词的歧义
	文本蕴含	关注模型确定一个文本是否蕴含另一个文本
	情感分析	关注确定文本的情感倾向
其他方面能力	真实性评估	关注评估模型生成的回答的真实性
	评估刻板印象	关注评估模型是否持有或传递某些刻板印象
	多任务评估	关注同时评估模型在多个任务上的性能

来源：课题组分析整理

二、评测方法：失准失信

症结

- **1. 考生-刷考题**：针对评测数据集做训练，快速提高成绩，真实泛化能力不升反降。
- **2. 阅卷-难精准**：生成内容输出自然语言，不是格式化输出，对评测结果做规则匹配不能完全反映模型的知识和推理能力。

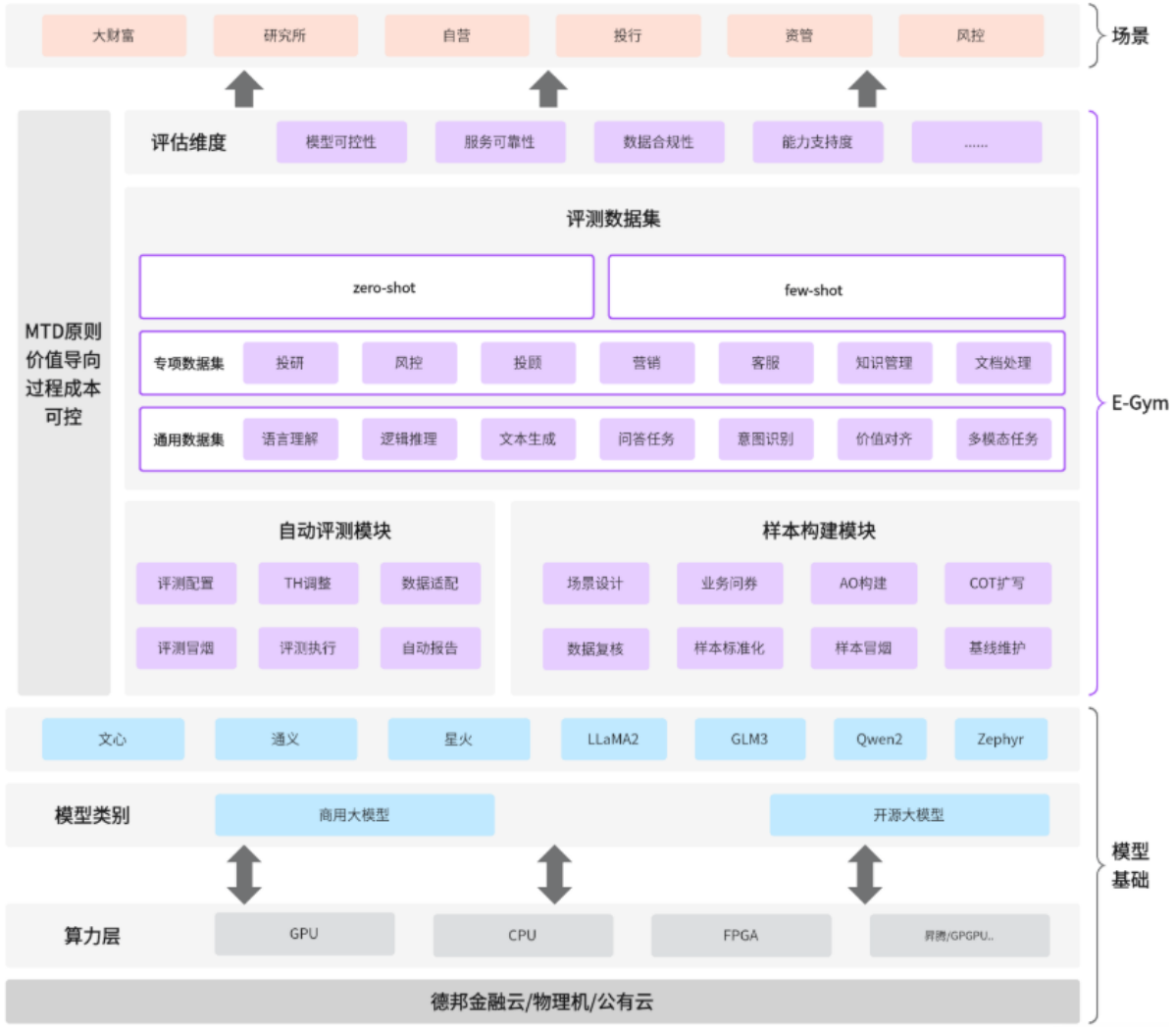
对策

✓ **1. 密卷测试**： 根据证券行业和细分场景需求准备题目。
梳理业务用户真实关心的内容生成场景，组织需求方和产品对接人员，设计评测问卷模板，形成区分具体场景的细粒度问答库。

✓ **2. 精准批阅**： 针对不同模型的输出模式，定制阅卷精准判断。
对评测对象模型，改写模型task head，取出可能性最高的生成内容的对数概率，对批阅结果给出更精准的得分。

如何评估?

经济效益和社会效益最大化



一、评估体系建设目标

指导证券行业大模型的适用产品研发活动

1. 规划设计
2. 大模型的选型
3. 配套工程应用研发
4. 应用测试
5. 落地验收
6. 投入产出收益复盘

二、评估体系建设内容

1.分析行业现状、适用场景流程、数据特点、技术应用情况。

2.研究评估原则、技术原理、关键指标、评估方法、发展趋势。

3. 结合证券行业特点，设计构建一套适用的大模型技术和应用评估体系。

4. 通过研究应用实证，验证该评估体系的有效性和实用性。

PART 05

总结与展望

技术层面

语料耗尽

2026年：用完一般的语言数据

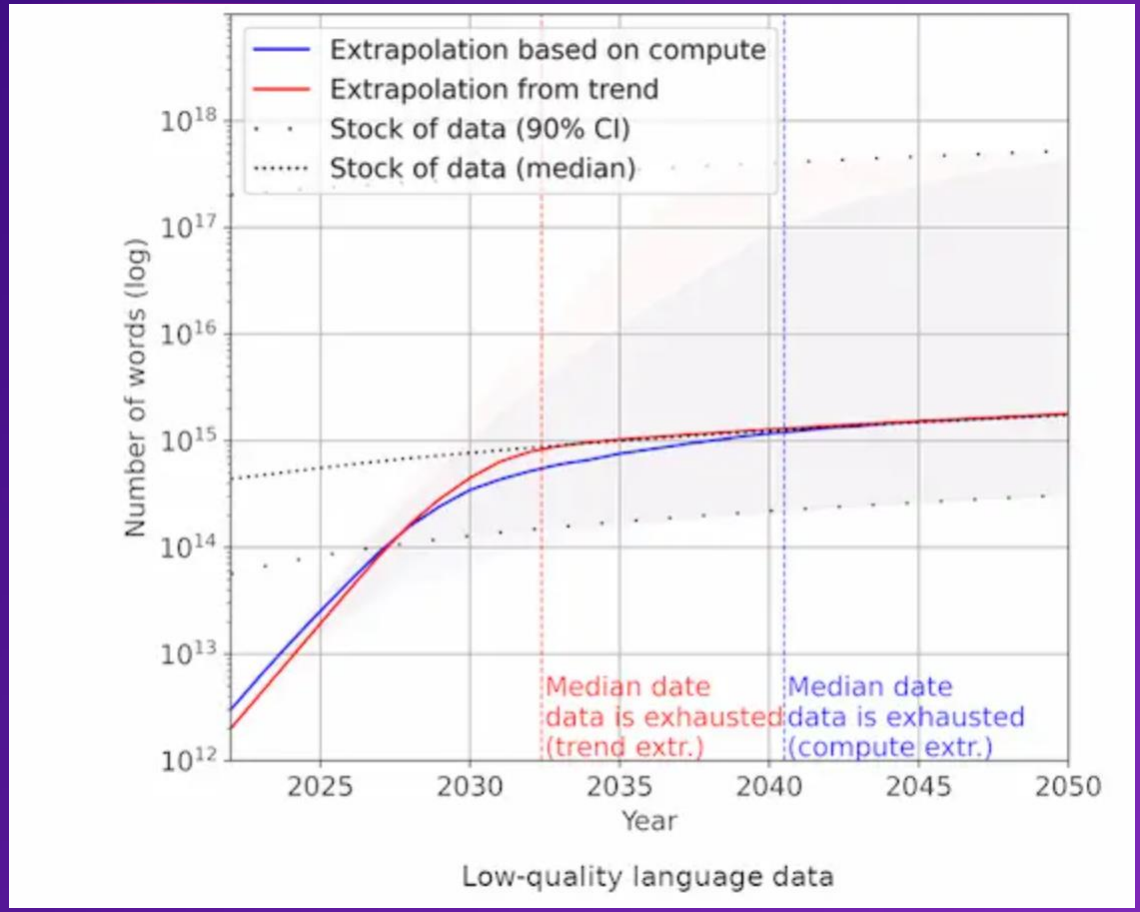
2030年～2050年：用完所有的语言数据

2030年～2060年：用完所有的视觉数据

数据污染

模型生成的数据会污染下一代模型的训练集；使用被污染数据进行训练，会导致模型误解现实。

这种情况在变分自编码器、高斯混合模型和大语言模型中都会出现。



Will we run out of data? An analysis of the limits of scaling datasets in Machine Learning, Pablo Villalobos et al,arXiv:2211.04325 , 2022

The Curse of Recursion: Training on Generated Data Makes Models Forget, Ilia Shumailov et al, arXiv:2705.17493 , 2023

未来趋势

多模态 - 进化更快

用张量嵌入方法，多模态智能模型(Meta ImageBind)可以通过声音、视频、温度等环境变化来接收并认知真实世界。

人类对齐 – 更友好

在RLHF以外，探索其他价值观对齐方式。

指示的泛化和纠错能力提升 - 推理更强

降低对指示(prompt)的依赖，提升用户使用模型的体验，也能使模型能够适应更广泛的应用场景。

模型的轻量化 – 更易得

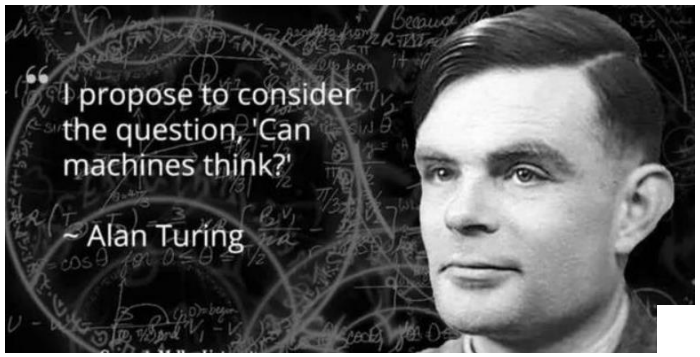
快思考：低算力低精度场景
深度思考：更复杂的专业场景
大规模场景，端上应用普及。

Shouyuan Chen， et al, Extending Context Window of Large Language Models via Positional Interpolation 2023.

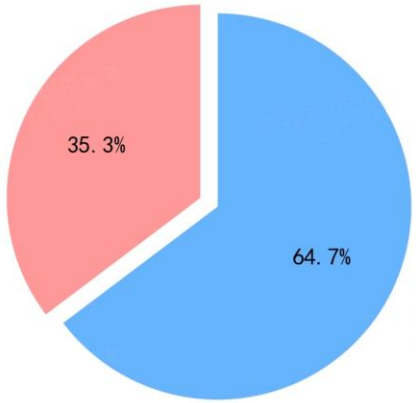
Kosinski, et al, Theory of Mind May Have Spontaneously Emerged in Large Language Models , 2023

Tim Dettmers et al,QLoRA:Efficient Finetuning of Quantized LLMs,2023.

机器会思考吗?



Yes
35.3%



No
64.7%

The Turing Belief（图灵信念）

A Turing machine(Turing et al., 1945), can model the brain, thereby enabling machines to achieve a level of intelligence equivalent to that of the human brain.

The Gödel Belief（哥德尔信念）

Due to the Gödel's incompleteness theorem（Gödel K, 1931), there are certain propositions within computable systems that machines can never determine to be true or false, yet humans can directly judge the truth of these propositions. Therefore, machines can never reach the level of human thought, and the human mind surpasses that of machines.

AiDD 6th 2025

参与调研您将优先获得

AiDD定制版
《AI+软件研发精选案例》

专属学习顾问
1对1需求对接

AiDD会后小调研

AiDD峰会致力于协助企业利用AI技术深化计算机对现实世界的理解，推动研发进入智能化和数字化的新时代。作为峰会的重要共建者，您的真知灼见对我们至关重要。衷心感谢您的参与支持！

扫码参与调研

2025 AI+研发数字峰会

拥抱AI 重塑研发

科技生态圈峰会 + 深度研习

——1000 + 技术团队的共同选择

K+峰会

K+峰会 敦煌站
K+ 思考周®研习社
时间：2025.08.29-30

K+峰会 上海站
K+ 金融专场
时间：2025.09.26-27

K+峰会 香港站
K+ 思考周®研习社
时间：2025.11.17-18

K+峰会详情

AiDD峰会

AiDD峰会 上海站
AI+研发数字峰会
时间：2025.05.23-24

AiDD峰会 北京站
AI+研发数字峰会
时间：2025.08.08-09

AiDD峰会 深圳站
AI+研发数字峰会
时间：2025.11.14-15

AiDD峰会详情

AiDD 6th 2025

2025 AI+研发数字峰会

AI+ Development Digital Summit

感谢聆听！

扫码领取会议PPT资料