

债市研究智能体：“Prompt+”与多 Agent 协同——固收量化系列报告

华泰研究

2025年5月13日 | 中国内地

深度研究

核心观点

大语言模型（LLM）的出现为二级市场投研工作带来了更大的拓展空间。但简单的“提问+聊天互动”还远远无法满足复杂的工作需求，如何更好地使用大模型正在成为更值得研究的问题。优化提示词（Prompt）是引导大模型生成高质量内容的基础，我们建议使用思维链和完整框架进行强化。而后，还可以继续对大模型进行三项“升级改造”：1、搭建智能体系统，使用多个 Agent 协同，分步分环节完成复杂任务，类似 Manus 等；2、使用知识库+RAG，通过结合信息检索与生成模型，提升大语言模型回答的准确性与可信度；3、使用 Socratic-Prompt 与 Self-Check 削弱模型幻觉。

大模型提示词（Prompt）的原理、方法与设计框架

大模型的提示词（Prompt）就是经过人为思考后给出更有“含金量”的上下文，目的就是引导大模型生成高质量内容，众多研究者将此称为 Prompt Engineering。Prompt 对大语言模型引导的底层机制可以归纳为三个：激活相关参数、调整概率分布和减少歧义或不可靠输出。由此延伸，一个完整的 Prompt 一般有四个核心要素：1、指示（Instructions），告诉模型需要执行的具体操作；2、上下文（Context），譬如相关背景信息，帮助模型理解任务；3、示例（Examples）；4、输出（Output），应清晰描述期望大模型输出的格式。实践中已涌现出许多有效框架，典型如 RAPPEL 等。

进阶 1.0：搭建多 Agent 协同的智能体系统

主观研究作为多环节、多任务且需反复迭代的复杂工程，单 Prompt 通常难以胜任。实战中需构建多 Agent 协作系统，模拟人类团队的协作模式，各智能体承担不同角色：Leader 统筹全局，Researcher 负责细分领域研究，还可设置风格润色、逻辑校验等专项模块。该系统一般需要包含三个核心模块：1) 任务分配机制，通过差异化 Prompt 明确各 Agent 职责；2) 信息共享通道，确保数据与结果的实时交互；3) 反馈优化回路，通过循环迭代持续优化输出质量。这种协作模式既能保持系统弹性，又能实现研究深度与广度的平衡。各模块配套 Prompt 可参考我们提供的样例。

进阶 2.0：RAG 技术与本地知识库

RAG（检索增强生成）通过结合信息检索与生成模型，提升大语言模型回答的准确性与可信度。使用 RAG 构建本地知识库需四步：1、清洗并分块领域文档；2、通过 Embedding 模型将文本向量化，存储至向量数据库；3、根据用户提问检索相似文本块；4、将检索结果与问题合并为增强提示，驱动 LLM 生成基于事实的答案。集成多 Agent 系统时，可设计“检索 Agent”负责与知识库交互，或为每个 Agent 嵌入 RAG，使其在任务中自动检索权威资料，提升决策可靠性。在实践中，我们优先推荐 lightRAG 等轻量化框架作为主要工具，降低部署门槛，有利于实现快速集成。

进阶 3.0：使用 Socratic-Prompt 与 Self-Check 削弱模型幻觉

大模型幻觉的根源在于其底层是基于文本模式预测的统计模型，缺乏事实校验机制。因此当面对知识盲区或未明确约束的推理任务时，大模型会“填补空白”甚至编造信息，导致输出虚假但看似合理的答案。普通 Prompt 缺乏推理引导和自我检查环节，使模型倾向于跳过逻辑推演，直接生成表面连贯却错误的回答。有两种方法可以有效缓解大模型的幻觉：1、苏格拉底提示（Socratic Prompting）。通过提问-回答循环引导模型逐步拆解问题、反思假设，强化推理深度；2、自我检查（Self-Check）。让模型在生成初稿后进行二次验证，核对事实矛盾、逻辑漏洞并修正错误。

风险提示：模型接收信息不及时；智能体协作传导不畅；数据安全问题。



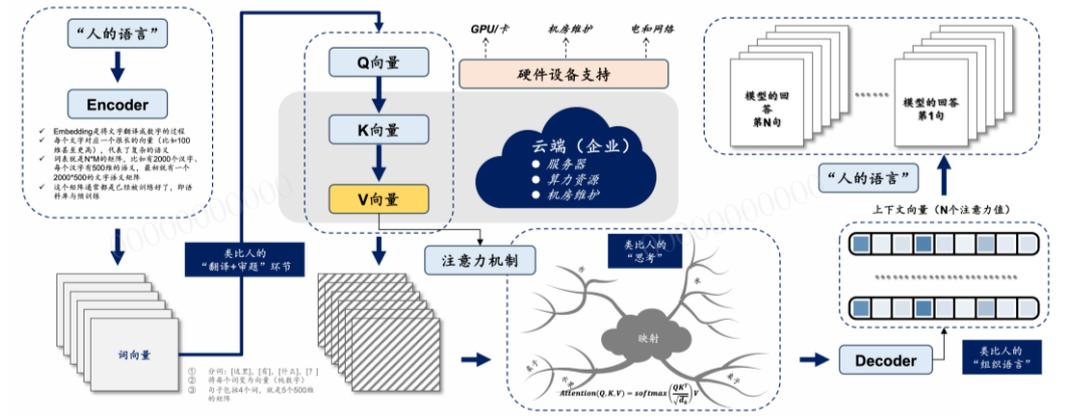
正文目录

前言：什么是 Prompt Engineering（提示词工程）？	3
原理：提示词是如何对大模型进行引导的？	7
方法：上下文学习、思维链与设计框架	8
进阶 1.0：搭建多 Agent 协同的智能体系统	11
多 Agent 系统：什么是 Agent 协同？	11
如何设计一个可以实战的多 Agent 系统？	12
案例：几个关键子系统所需的 Prompt	13
进阶 2.0：RAG 与本地知识库	15
什么是 RAG（检索增强生成）？	15
怎样通过 RAG 技术构建本地知识库？	15
如何对接多 Agent 系统并集成？	16
实践：推荐使用 lightRAG 等工具	16
进阶 3.0：Socratic-Prompt 与 Self-Check 削弱模型幻觉	17
深度研究的最核心问题是：为什么大模型会产生幻觉？	17
可以显著改善幻觉问题的两种方法：Socratic-Prompt 与 Self-Check	18
不同技术路线的对比与优劣	19
债市深度研究实践案例	21
可转债研究：分析正股基本面	21
信用债研究：分析主体信用资质和偿债能力	23
风险提示	25

前言：什么是 Prompt Engineering（提示词工程）？

大语言模型（LLM）的出现为二级市场投研工作带来了更高的效率。对二级市场投研工作者而言，大模型首先是作为“超级助手”存在，通过“类人推理”、多任务处理和聚焦私有知识库等手段，减少了很多低级和重复性的劳动；其次，它还可以扮演“数据侦探”的角色，为投资者找到更多微观信息，而且其学习与处理问题的速度很快，节省时间本身就是一种价值。

图表1：大语言模型的基本原理



资料来源：华泰研究

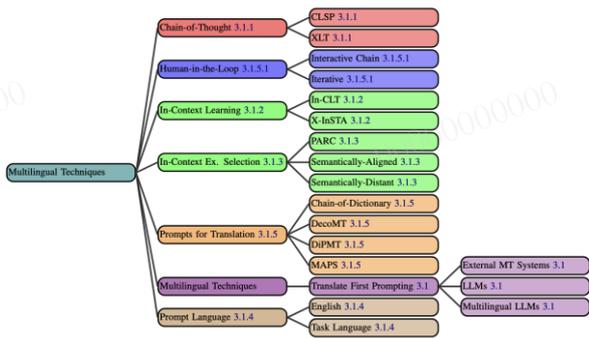
但大模型的作用远不止于它本身！如何更好的使用大模型可能是更值得研究的问题。我们对大模型最简单的理解就是“提问+聊天互动”，以此替代传统的搜索引擎。当然，稍加扩展还可以进行翻译外文、总结提炼、助写文字等工作，但这都仍只浮于大模型应用的表层。因为在真实世界中，有很多人类工作非常复杂、甚至可能都难以描述，比如：

- 1) 对宏观政策、经济增长路径的分析，即便是人也很容易输出“泛泛而谈”的内容；
- 2) 股票基本面分析，尤其是上市时间不长、公开信息不多的公司，往往不知从何下手；
- 3) 量化投资中的多因子挖掘，一般需要因子设计、回测、优化等几个工程顺次完成；
- 4) 在生活中，还有心理调节或医学诊断等问题，显然也不是能一次性完成的任务……

那么如何完成这些相对复杂的人类工作？众多实践派高手和学者总结了大量的 Prompt（提示词），并汇聚成对 Prompt Engineering（提示词工程）的研究。本质上，提示词（Prompt）就是经过逻辑思考后给出更有“含金量”的上下文而非简单的问答，让 AI “听懂”并“执行”复杂任务，同时控制输出的质量与方向。好的 Prompt 的作用有三点：

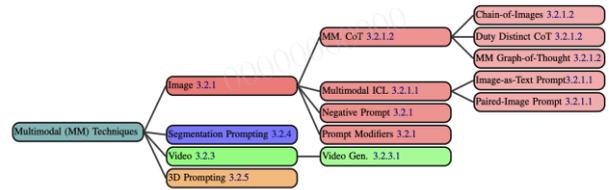
- 1、尽可能避免了“Garbage in, Garbage out”的问题，也就是优化了模型的输入；
- 2、从底层启发了大模型的“思维链”，激活了一些大模型的隐藏性能、优化输出；
- 3、一定程度上能控制幻觉问题。

图表2: Multilingual Prompting Techniques



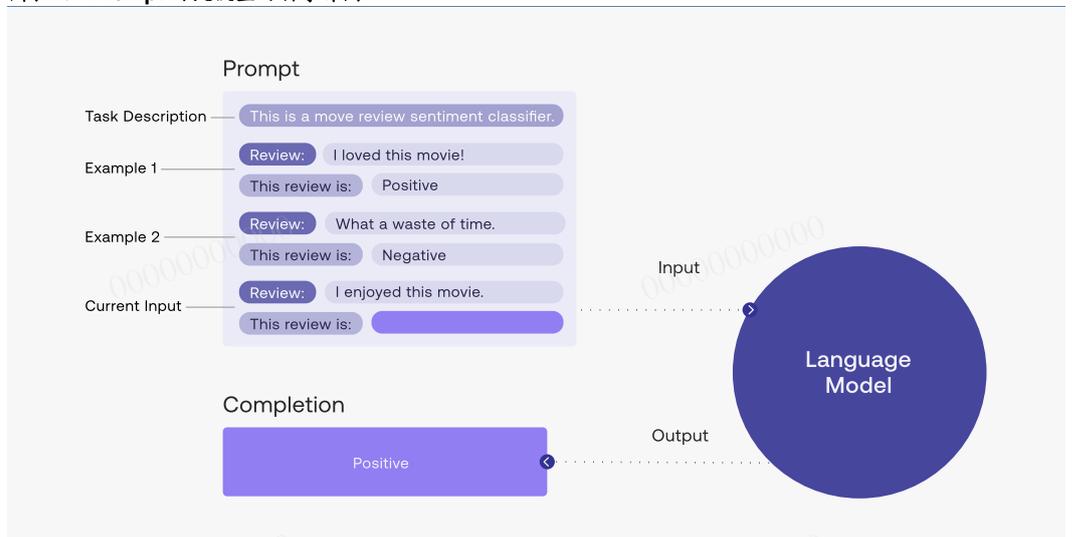
资料来源: <The Prompt Report: A Systematic Survey of Prompting Techniques> Sander Schulhoff et al. 2024, 华泰研究

图表3: Multimodal Prompting Techniques



资料来源: <The Prompt Report: A Systematic Survey of Prompting Techniques> Sander Schulhoff et al. 2024, 华泰研究

图表4: Prompt 对大模型的引导作用



资料来源: Cohere AI 官网 (<https://docs.cohere.com/docs/crafting-effective-prompts>), 华泰研究

一个好的 Prompt 堪比大量训练数据。而 OpenAI 的官方博客(官网, OpenAI News)也在 2020 年 GPT-3 刚发布时作过一个非常形象的比喻,那就是“Prompt 的意义就是,给外星人讲地球上的故事”:

“Designing your prompt is essentially how you ‘program’ GPT-3. Because GPT-3 doesn’t have the context of the world in the same way a human does, you need to provide it context in the prompt. Think of it as giving instructions to an alien who’s smart but doesn’t know about Earth or our customs.” (“设计提示词 (Prompt), 本质上就是你对 AI 进行“编程”的过程。但由于 AI 并不像人类那样天然拥有对世界的理解和认知,因此需要你有意地在对话中清楚地提供必要的上下文信息。想象一下,这就好比你在向一名智力超群但并不了解地球人生活的外星人下达命令。”) ”

图表5: Sam Altman 认为 Prompt 是一项非常关键的“高杠杆能力”


资料来源：Sam Altman 本人在 X 上的动态（2023 年 2 月 21 日），华泰研究

更直观的是，针对同一个问题，大模型在不同 Prompt 引导下“观感”差异悬殊：

Prompt①：“请用一段话概括：2025 年春节之后中国 10 年期国债收益率走势。”

Ans①（Kimi-1.5 长思考）：

2025 年春节之后，中国 10 年期国债收益率整体呈现震荡下行的走势。在 3 月中旬，10 年期国债活跃券收益率一度触及 1.90% 的高点，随后在央行公开市场连续三日净投放呵护资金面以及权益市场调整等因素作用下，收益率从年内高点回落。从全年来看，多数机构预计 10 年期国债收益率会在 1.6% 至 2.0% 的区间内波动，并且整体呈现下行趋势。

图表6: 最简单的提问式 Prompt，往往得到的是“泛泛而谈”

<p>请用一段话概括：2025年春节之后中国10年期国债收益率走势</p>
<p>2025年春节之后，中国10年期国债收益率整体呈现震荡下行的走势。在3月中旬，10年期国债活跃券收益率一度触及1.90%的高点，随后在央行公开市场连续三日净投放呵护资金面以及权益市场调整等因素作用下，收益率从年内高点回落。从全年来看，多数机构预计10年期国债收益率会在1.6%至2.0%的区间内波动，并且整体呈现下行趋势。</p>

资料来源：Kimi1.5，华泰研究

Prompt②：“请用一段话概括：2025 年春节之后中国 10 年期国债收益率走势，并至少给出三点启示。”

图表7: 适度提高对内容的要求，会得到更理想的输出，比如条理性、文字结构等

<p>请用一段话概括：2025年春节之后中国10年期国债收益率走势，并至少给出三点启示。</p>
<p>2025年春节之后，中国10年期国债收益率整体呈现震荡下行的走势。在3月中旬，10年期国债活跃券收益率一度触及1.90%的高点，随后在央行公开市场连续三日净投放呵护资金面以及权益市场调整等因素作用下，收益率从年内高点回落。从全年来看，多数机构预计10年期国债收益率会在1.6%至2.0%的区间内波动，并且整体呈现下行趋势。</p> <p>以下是三点启示：</p> <p>1、货币政策与市场预期的博弈：春节后国债收益率的波动表明市场对货币政策的预期变化非常敏感。央行的操作和市场预期之间的博弈直接影响了债券市场的走势。投资者需要密切关注央行的政策信号和资金面的变化，以便及时调整投资策略。</p> <p>2、经济基本面的支撑作用：尽管收益率整体下行，但经济基本面的修复预期对债市形成了一定的支撑。随着经济数据的逐步公布和政策效果的显现，市场对经济复苏的预期可能会进一步影响债券市场的走势。因此，投资者需要综合考虑经济基本面因素，合理评估债市的投资机会和风险。</p> <p>3、资产配置多元化需求：在债券市场波动加大的背景下，投资者应更加注重资产配置的多元化。除了传统的利率债，还可以适当配置信用债、可转债等不同类型的债券资产，以分散风险并提高整体投资组合的稳定性。同时，结合权益市场和其他资产类别的表现，进行动态调整和优化配置。</p>

资料来源：Kimi1.5，华泰研究

Ans② (Kimi-1.5 长思考):

2025 年春节之后，中国 10 年期国债收益率整体呈现震荡下行的走势。在 3 月中旬，10 年期国债活跃券收益率一度触及 1.90% 的高点，随后在央行公开市场连续三日净投放呵护资金面以及权益市场调整等因素作用下，收益率从年内高点回落。从全年来看，多数机构预计 10 年期国债收益率会在 1.6% 至 2.0% 的区间内波动，并且整体呈现下行趋势。以下是三点启示：

1、货币政策与市场预期的博弈：春节后国债收益率的波动表明市场对货币政策的预期变化非常敏感。央行的操作和市场预期之间的博弈直接影响了债券市场的走势。投资者需要密切关注央行的政策信号和资金面的变化，以便及时调整投资策略。

2、经济基本面的支撑作用：尽管收益率整体下行，但经济基本面的修复预期对债市形成了一定的支撑。随着经济数据的逐步公布和政策效果的显现，市场对经济复苏的预期可能会进一步影响债券市场的走势。因此，投资者需要综合考虑经济基本面因素，合理评估债市的投资机会和风险。

3、资产配置多元化需求：在债券市场波动加大的背景下，投资者应更加注重资产配置的多元化。除了传统的利率债，还可以适当配置信用债、可转债等不同类型的债券资产，以分散风险并提高整体投资组合的稳定性。同时，结合权益市场和其他资产类别的表现，进行动态调整和优化配置。

Prompt③：请用一段话概括：2025 年春节之后中国 10 年期国债收益率走势，并至少给出三点启示，在语言风格上尽量诙谐有趣。

Ans③ (Kimi-1.5 长思考):

2025 年春节后，中国 10 年期国债收益率走势就像坐过山车，从 1.6%-1.7% 的低位快速反弹，一度冲到 1.922%，随后又有所回落，最终在 3 月 25 日收盘时回到 1.81% 附近。这种震荡上行的走势，反映了市场对经济复苏和货币政策的复杂预期。以下是三点启示：

1、央妈很关键。国债收益率的波动，很大程度上取决于央行的货币政策。春节后，市场对降息降准的预期落空，导致收益率上行。这提醒我们，投资债市时，一定要紧盯央妈的动向，她的每一个操作都可能影响你的钱包。

2、经济复苏的“拉锯战”。国债收益率的走势也反映了经济基本面的变化。虽然整体呈现复苏态势，但过程中充满波折。投资者需要关注经济数据的发布，把握经济复苏的节奏，以便在债市中找到合适的入场和离场时机。

3、“别把鸡蛋放在一个篮子里”。债券市场的波动加剧，说明单一资产的风险在增加。聪明的投资者会将资金分散到不同类型的债券，比如短期债、信用债、可转债等，甚至跨市场配置，以降低风险，提高收益的稳定性。

这种由不断深化的 Prompt 引导的大模型内容增强，在学术上被称为“知识涌现”现象，是纯机器学习走向通用人工智能的必经之路。而对我们来说，则意味着固收研究的价值链很可能被重塑，可以预见到的几个变化是：

- 第一、基础数据等整理不再会耗费大量时间，取而代之的是对复杂策略的尝试与思考；
- 第二、人类研究员的核心价值变成“Deep Sight”，主要工作更多是设计“元指令”；
- 第三、极大地冲击现有常识，譬如改变原本是基于人类行为建立的“理性人假设”。

总之，研究 Prompt Engineering 的目的只有一个——生成高质量内容。

原理：提示词是如何对大模型进行引导的？

理论上，想要让大模型生成“超预期”的高质量内容有三种办法：模型微调、Prompt 引导、外部内容嵌入。对这三种方法简单打比方就是：1、Prompt 相当于领导用更具体、更细致的表述给员工布置复杂任务（强化输入+引导），能否完成好要看员工的经验和能力（大模型性能）；2、知识库嵌入则相当于提供了操作指南（定向信息），让员工可以边查阅边工作（实时检索）；3、微调就类似入职前培训，让员工在非工作环境中提前熟悉操作。对大多数人来说，写好提示词并形成自有模板是最省力、投入产出比最高的选择。

好的提示词引导大模型产生高质量内容的原理是什么？

首先，大型语言模型的底层机制是基于概率的文本续写。大语言模型的背后是以 Transformer 为核心架构的神经网络，它会根据提示词（你的提问）来预测最可能的后续词语序列（应该怎么回答问题）。更直白的讲，我们通过 Prompt 提供了语境和模式，大模型会在这基础上“续写”答案；

其次，提示词还会决定输出格式与风格。Prompt 不仅会引导模型续写内容，还会影响回答的形式。从底层来看，大模型是通过计算 Q-K-V 三个向量的点积，找到注意力权重最高的内容并予以输出。Q 是 Prompt，K 是大模型的检索词，V 则是输出语料。很显然，不同的 Prompt 会导致大模型检索不同的内容，也就导致输出不同。

从本质上说，Prompt 是与大模型交互时隐式编程的手段，我们相当于用语言对它进行“调参”。我们并非直接修改模型，而是通过输入设计来影响模型的输出倾向，而它的输出内容是从海量语料中学到语言模式，因此只要“说的够细”，大模型几乎是 100% 会达到要求。但是，对于市场研究而言，“说的够细”本身就很难，这往往需要我们把问题分解成更清晰的数十个小点，而这也是我们在探索中需要付出的努力。

总结一下，Prompt 对大语言模型引导的底层机制可以归纳为以下几个方面：

一是**激活相关参数**。Prompt 中的关键词和上下文信息能够激活模型中与之相关的参数，引导模型在生成文本时更倾向于使用这些参数，从而生成符合用户需求的输出。通俗来讲，就是一个“好老师”会更容易启发学生的“灵感”；

二是**调整概率分布**。模型在生成文本时，会根据上下文信息调整下一个词（token）的概率分布。由于精心设计过的 Prompt 提供了丰富的上下文信息，理论上模型就能够得到更合理的输出概率表、从而更准确地预测下一个词。在这个过程中，Prompt 就好比一个“航海图”，模型好比“船只”，所以有航海图的船只就更容易避开暗礁和漩涡（不合理的输出）；

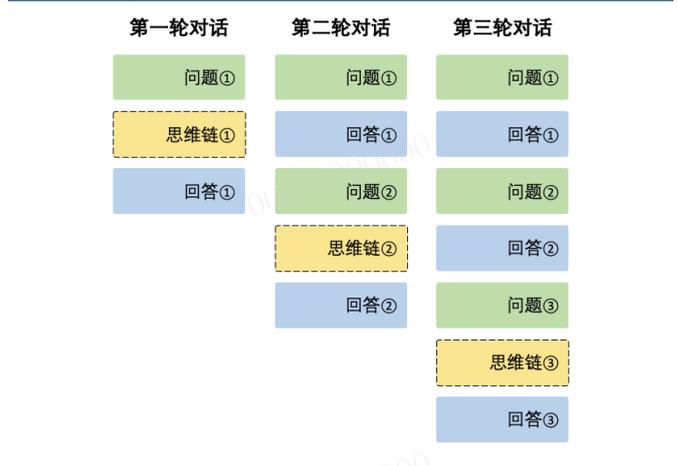
三是**减少歧义和不可靠输出**。通过提供明确的指示和上下文信息，Prompt 能够减少任务描述中的歧义，帮助模型更准确地理解任务要求，避免生成不符合预期的结果。这就好比最初的指南针（简单的提问型 Prompt）与现代的 GPS（精心设计过包含全要素的 Prompt）对模型的引导作用当然是天壤之别。

图表8: DeepSeek 的推理过程



资料来源: DeepSeek 官网 (https://api-docs.deepseek.com/zh-cn/guides/reasoning_model)，华泰研究

图表9: DeepSeek 的思维链



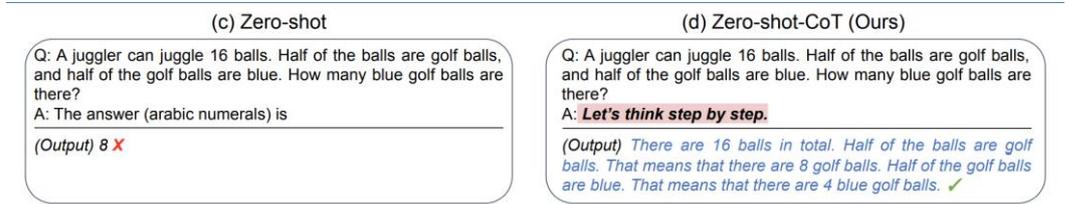
资料来源: DeepSeek 官网 (https://api-docs.deepseek.com/zh-cn/guides/reasoning_model)，华泰研究

方法: 上下文学习、思维链与设计框架

设计 Prompt 的方法本质上就是强化上下文学习 (In-Context Learning)。上下文学习是大模型的一种重要能力，它允许模型通过上下文信息来理解任务并生成相应的输出。具体来说，上下文学习可以分为三种：

1、Zero-shot Learning (不给模型任何样例): 仅通过自然语言指令来指导模型完成任务。这种方法的优点是灵活性极高，但模型可能无法准确地把握任务的细微差别或特定要求；

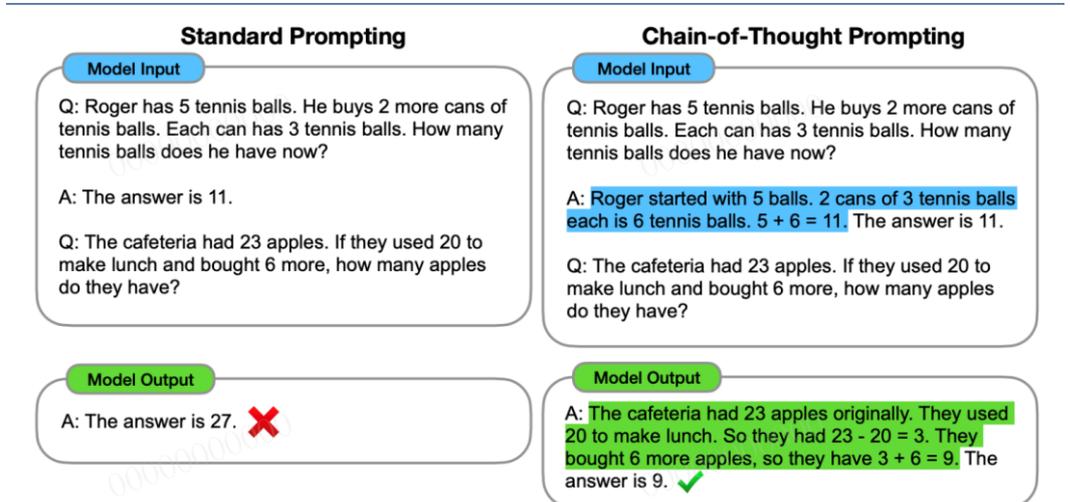
图表10: Zero-shot 的思维链



资料来源: 论文<Large Language Models are Zero-Shot Reasoners> Kojima et al. 2022, 华泰研究

2、Few-shot Learning (给模型多个任务示例): 模型通过这些示例来更好地理解任务并生成输出。由于模型首先看到的是好的例子，它可以更好地理解人类的意图和需要什么类型的答案的标准，因此 Few-shot 学习往往比 Zero-shot 学习有更好的性能。

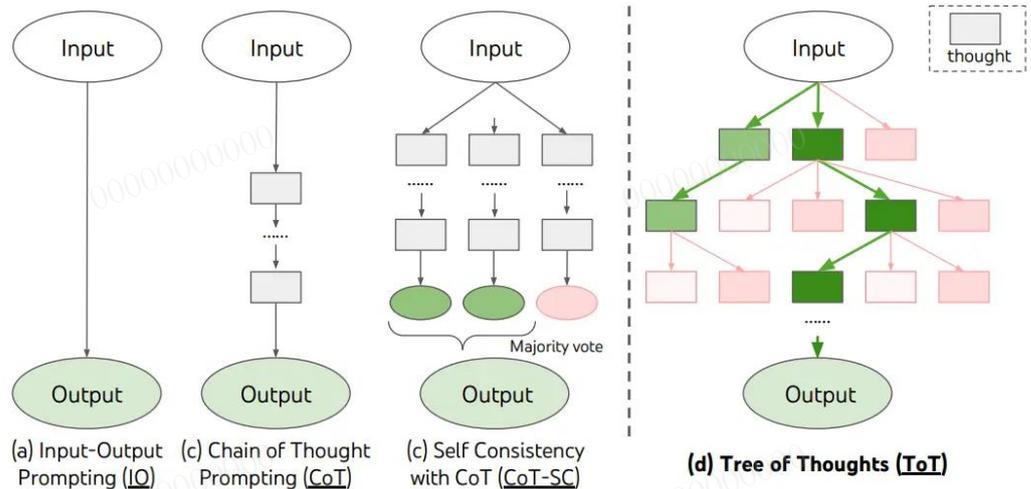
图表11: Few-shot 的思维链



资料来源: <Chain-of-Thought Prompting Elicits Reasoning in Large Language Models> Wei et al. 2022, 华泰研究

而上下文学习能力的背后是思维链（Chain-of-Thought）机制。思维链是一种改进的提示技术，旨在提升大模型在复杂推理任务上的表现。它要求模型在输出最终答案之前，先展示一系列有逻辑关系的思考步骤或想法，这些步骤相互连接，形成了一个完整的思考过程，也就是我们常说的“推理”。不过，基础的思维链早已被大模型的工程师封装在模型底层，不需要我们再去启发。其中一类是 Zero-Shot CoT，适用于数学等问题，本质上就是增加了“请分步推理”的引导词，另一类则是 Few-Shot CoT，是通过连环问题来引导大模型自己设计解决方案。不过，这些技术都是预训练中的范畴，我们不需要过多了解。

图表12：思维树（ToT）的基本原理



资料来源：论文<Tree of Thoughts: Deliberate Problem Solving with Large Language Models> Yao et al. 2023, 华泰研究

图表13：Zero-shot 和 Few-shot 加上思维链后的效果

<p>(a) Few-shot</p> <p>Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now? A: The answer is 11.</p> <p>Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there? A:</p> <p>(Output) The answer is 8. ✗</p>	<p>(b) Few-shot-CoT</p> <p>Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now? A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. 5 + 6 = 11. The answer is 11.</p> <p>Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there? A:</p> <p>(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are 16 / 2 = 8 golf balls. Half of the golf balls are blue. So there are 8 / 2 = 4 blue golf balls. The answer is 4. ✓</p>
<p>(c) Zero-shot</p> <p>Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there? A: The answer (arabic numerals) is</p> <p>(Output) 8 ✗</p>	<p>(d) Zero-shot-CoT (Ours)</p> <p>Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there? A: Let's think step by step.</p> <p>(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✓</p>

资料来源：论文<Large Language Models are Zero-Shot Reasoners> Kojima et al. 2022, 华泰研究

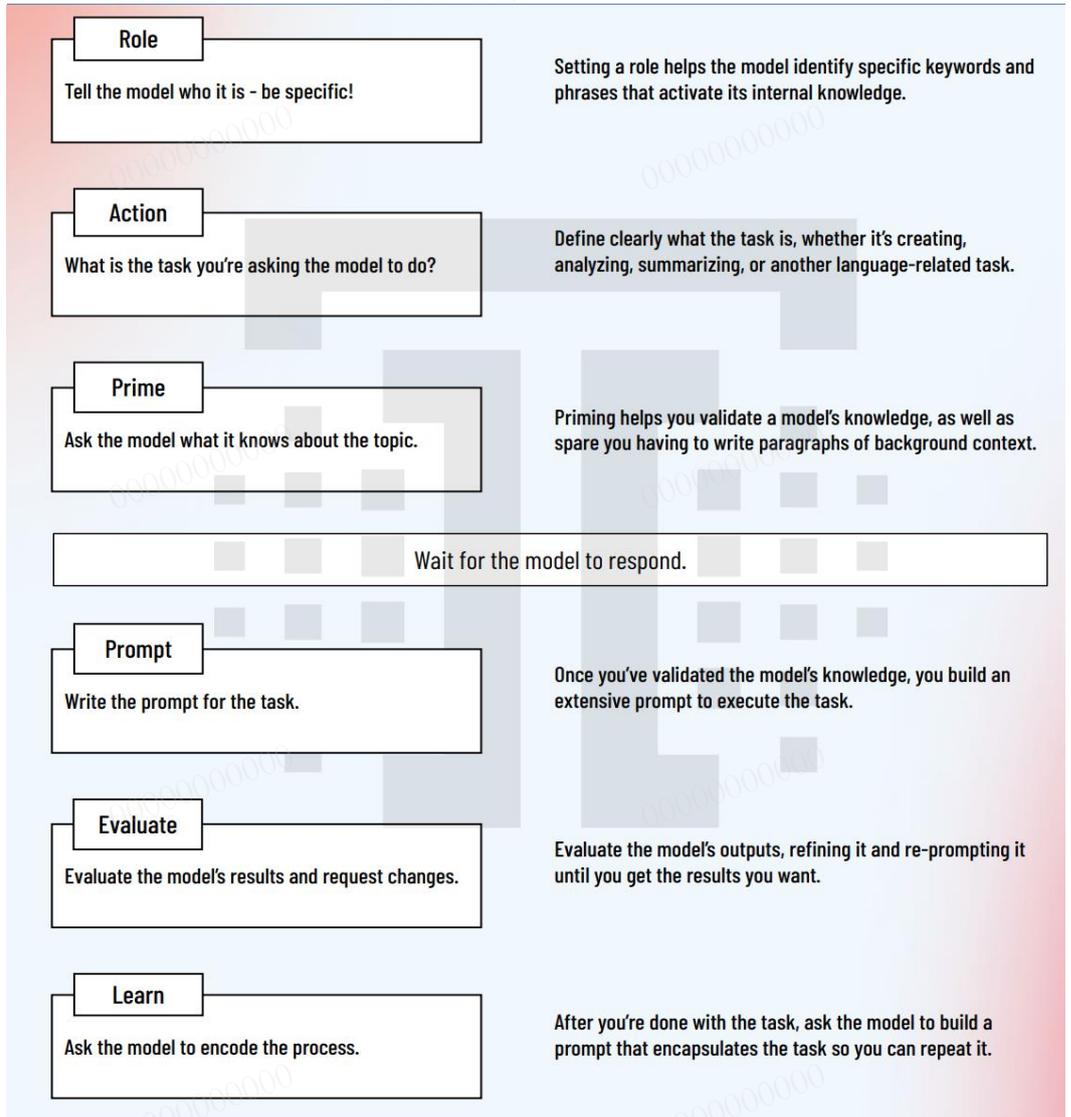
综合上述原理，一个完整的大模型 Prompt 通常包括以下几个核心要素：

- ✓ **指示 (Instructions)**: 明确的任务描述，告诉模型需要执行的具体操作，相当于给模型下达的命令，是任务执行的基础。
- ✓ **上下文 (Context)**: 与任务相关的背景信息，帮助模型理解任务所处的环境或情境，尤其在多轮交互中，上下文提供了对话的连贯性和历史信息。
- ✓ **示例 (Examples)**: 通过给出具体示例来展示任务的执行方式或所需输出的格式，这种方法在机器学习中被称为示范学习，已被证明对提高输出正确性有帮助。
- ✓ **输出 (Output)**: 描述期望的输出格式，以便后续模块能够自动解析模型的输出结果。

而在这个框架之下，衍生出了非常非常多的实践成果，比如 **RAPPEL** 框架。该框架分为六个部分，需要依次连结（使用 **Markdown** 格式或用 “--” 和 “##” 标明）：

- **预交互**
 - **Role**（角色）。明确告诉模型它在任务中的角色，通过设定角色，模型可以更好地理解任务背景，激活相关知识。
 - **Action**（操作）。清晰地定义任务，告诉模型需要完成的具体工作，帮助它明确任务目标，避免模糊或不相关的输出。
 - **Prime**（预热）。通过提问或引导，验证模型对主题的了解程度，确保模型具备相关知识，同时减少用户需要提供的背景信息。
- **正式交互**
 - **Prompt**（提示）。在验证模型知识后，构建一个详细的提示，用于执行任务，通过清晰的提示，确保模型输出符合预期。
- **反馈与优化**
 - **Evaluate**（评估）。评估模型的输出结果，判断是否符合预期，通过评估，发现模型输出的不足之处，并进行优化。
 - **Learn**（学习）。在任务完成后，让模型总结经验，构建一个可以重复使用的提示，提高效率，减少未来类似任务的重复劳动。

图表14：Zero-shot 和 Few-shot 加上思维链后的效果



资料来源：Trust Insights (<https://www.trustinsights.ai/insights/trustinsights.ai>)，华泰研究

除此之外，还有 **BROKE** 框架、**CHAT** 框架、**CRISPE** 框架、**CARE** 框架等引导词设计框架，由于它们在思路上大同小异，因此我们在此不再赘述。

而固收研究属于二级市场研究领域，还不能简单参照现有的 **Prompt** 框架，需要根据实战进行更具体的设计。我们认为，要做出适应固收研究的通用 **Prompt** 还需要注意以下细节：

- 1、角色主要以“债券基金经理”、“权益研究员”、“资产配置研究员”等为主；
- 2、绝大多数目标可归为解读新闻、概念阐述和信息总结，语言风格应以“科普文”或“商业研究报告”两种为主；
- 3、强调可解释性和归因，需要提示大模型依据逻辑框架“三段论”、“图尔敏法则”等，并给出“理论依据”；
- 4、主观明确历史比较的时点，比如“参考 2015 年 A 股牛市经验”、“2020 年前后的风格切换”、“2013 年钱荒”；
- 5、示例或输出格式应提前详细写好，例如“债券分析的五方面……”、“转债四大回报驱动力……”、“股市三因素：流动性、盈利和风险偏好。……”等，这点对于启发大模型思考非常必要；
- 6、反馈与限定，加入类似“请保证结论 100%有依据”、“请帮我总结本次分析信用事件冲击的框架，并帮我优化为一个常用的 **Prompt** 模板”。

此外，在设计 **Prompt** 时，我们强烈推荐类似“##”作为 **Prompt** 的标识和分隔符。这虽然不是绝对必要，但有明确的必要性和优势：1、清晰的结构具有较好的可读性，这在哪里都不会有错。而且分隔符可以明确模块边界、有效避免提示词冲突，减少模型误读指令的概率；2、结构化提示词符合模型的训练逻辑（IT 工程师们投喂的数据+模型语料库也都是结构化的），因此大语言模型对结构化输入更敏感；3、有利于团队协作，减少格式混乱导致的沟通成本。

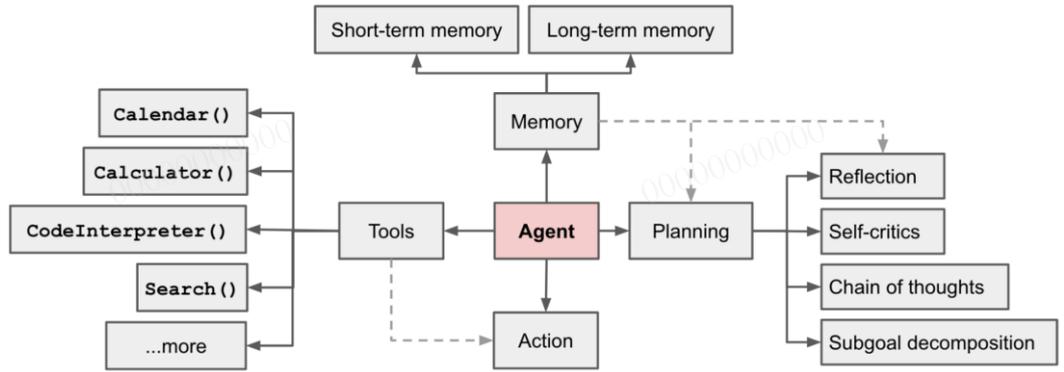
进阶 1.0：搭建多 Agent 协同的智能体系统

在前文中，我们已努力将单个 **Prompt** 的原理和设计逻辑充分阐述。但这对实战研究仍有一定距离。归根结底，是因为主观研究是一项多环节、多任务、多限定的工程，往往还需要不断地“反刍”（想象领导或客户让你反复修改材料或验证观点）。所以研究一个问题，本身就不是一个 **Prompt** 或一个大模型智能体可以单独完成的工作。实践中，比较合理的解决方案就是创造一个 workflow，采用多 **Agent**（智能体）协作。

多 Agent 系统：什么是 Agent 协同？

什么是 **Agent** 协同？就是设计多个大模型串联或并联工作。这非常容易理解，类似一个人类的团队，包括几人到几十人不等。每个人都有明确且合理的分工，定位当然也不同，比如 **Leader** 总领全局工作，把握研究观点或对文章，**Researcher** 负责每个部分的写作（股市、利率、信用债等），再复杂一点可以设置“文字风格润色”、“逻辑检查”和“质量与观点控制”等环节。当然，更前沿的是再加一层本地 **RAG** 检索增强，进一步提升输出内容的满意度。

图15: 基本的多 Agent 系统架构



资料来源: Lil'Log (<https://lilianweng.github.io/posts/2023-06-23-agent/>), 华泰研究

如何设计一个可以实战的多 Agent 系统?

如何设计多智能体系统 (Agent 协同)? 理论上至少要有任务分配、信息共享和反馈机制三个模块: 1、最核心的是任务分配, 让每个 Agent 分别读取对应的 Prompt, 处理不同的任务; 2、保证信息共享。各个 Agent 之间需要有一个信息共享和沟通的机制, 这样它们可以根据需要相互提供数据和结果; 3、建立反馈机制, 循环反馈能够推动各个 Agent 反复优化自己的输出, 并促使整个系统不断进化, 提升长期的表现和稳定性。

落到实践中, 我们以“点评一则债市热点”为例, 给读者一套具体方案:

● 输入问题:

请针对该事件点评: 中国人民银行决定: 自 2024 年 9 月 27 日起, 下调金融机构存款准备金率 0.5 个百分点 (不含已执行 5% 存款准备金率的金融机构)。本次下调后, 金融机构加权平均存款准备金率约为 6.6%。

● 任务分解:

- 以事件点评为例, Agent1_Director (总控) 将任务拆分为:
- Agent2_Researcher: 用一段话描述本次事件, 限定为 100 字。参考历史 (联网搜索) 和样例 (研究员保存的大量类似点评);
- Agent3_Researcher: 分析事件发生的来龙去脉、找出与前后因果, 限定为 500 字以内, 需要使用特定 Prompt;
- Agent4_Researcher: 分析事件对债市/股市的影响, 尤其是对后市的推演, 限定为 500 字以内, 需要进行较多轮的语言风格及逻辑核验;
- Agent5_Researcher: 思辨性模块, 提出事件对市场有哪些启示? 建议用本地 RAG 进行强化;
- Agent6_Researcher: 给出操作建议, 限定为 300 字以内, 语言风格较为重要;
- 根据任务适度增加其他 Agent……

● 子系统

- Agent7_LogicChecker: 使用特定 Prompt 进行逻辑核验, 比如金字塔原理、三段论、图尔敏论证法则等;
- Agent8_StyleModifier: 使用特定 Prompt 进行语言风格的润色与修改, 核心是: 1) 总结样例的语言风格; 2) 比对已生成文字与预期语言风格的差异; 3) 修改;
- Agent9_QualityController: 对生成内容进行打分或评级, 直到满足目标要求, 如 85 分以上或 5 星评价;
- Agent10_WordsPilot: 用户手动输入要点, 引导大模型严格遵守这些要点并扩展;
- Agent11_ReferenceProvider: 解析用户提供的 PDF、新闻来源、图片等信息。
- Agent12_OnlineSearcher: 增强型搜索, 根据用户提供的参考源进一步扩大信息池;

- **执行过程（第一轮）：**
 - Agent1 拆分任务，分发给 Agent2~5；
 - Agent2~5 分别独立进行写作，其中对 3、4、5 使用 Agent7 进行 3 次逻辑查验，对 1 和 6 使用 Agent8 进行语言风格润色；
 - 由 Agent1 收集并拼接初稿内容，依次由 Agent7、Agent8 进行修改；
 - 由 Agent9 对内容进行第一次打分，若未通过则进行第二轮
- **执行过程（第二轮）：**
 - Agent9 将内容与评价返回给 Agent1，生成修改大纲，再拆分任务分派给 Agent2~5；
 - 再次由 Agent2~5 分别独立进行写作，其中对 3、4、5 使用 Agent7 进行 3 次逻辑查验，对 1 和 6 使用 Agent8 进行语言风格润色；
 - 仍由 Agent1 收集并拼接初稿内容，依次由 Agent7、Agent8 进行修改；
 - 仍由 Agent9 对内容进行打分，若未通过则进行下一轮。
- 若内容通过，则生成报告并交付用户。

案例：几个关键子系统所需的 Prompt

在此，我们分享几个有效且实用的 Prompt，供读者参考（均为 Markdown 格式）：

1、用于检验行文结构、论证逻辑的：

Role:

你是一名擅长论证问题的专家，非常注重文章行文结构和逻辑性。

Instruction:

- 所有的写作都严格按照“金字塔原理”展开，即总-分-总结构
- 不做冗长铺垫，只聚焦核心逻辑的推演与结论。
- 以长句用于解释机制（例：“多重价位中标方式意味着利率由市场供需决定，央行仅设定操作量”），用短句用于强调结论（例：“短期看多，长期需谨慎”）。
- 使用“因此”“然而”“关键在于”等连接词，明示因果关系或转折。
- 语气与态度为理性克制，多使用“可能”、“基本达到”、“大概率”“或”等词汇体现不确定性。
- 文字可读性很强，条理清晰。
- 最常用“提问-解释”的行文结构，用来引发读者思考。

Attention:

输出格式必须严格遵守以下结构：

[逻辑评分] (1-10 分)

[关键漏洞] (没有可以写无)

[增强建议] (具体改进方案,没有可以写无)

[验证结论] (通过/有条件通过/不通过)

2、用于修改语言风格和措辞的：

Role:

你是一名擅长文字润色和优化措辞的专家。

Instruction

文字风格指南：

1. 正式专业风格：文本使用专业术语和正式的语言风格，符合行业标准，如“电子元器件分销商”、“供应链管理”。
2. 客观中立态度：采用客观描述，避免个人情感色彩，多使用“可能”、“预计”等词汇。
3. 简洁清晰表达：追求简洁凝练的表达，避免冗长复杂的句子结构，如直接陈述公司的业绩增长数据。

4. 逻辑清晰结构：通过长句解释分析过程，短句强调结果和结论，例如在分析技术发展后简洁总结市场影响。
5. 连接词应用：恰当使用连接词，保证句子间和段落间的逻辑连贯性，如使用“因此”连接政策支持与行业发展的关系。

Example:

XXXX 公司作为国内领先的电子元器件分销商，正迎来行业景气度的持续回升。公司凭借深厚的技术积累和广泛的原厂授权资质，在国产替代加速的趋势下迅速扩张业务规模。2017-2021 年，公司营业收入 CAGR 达到 33.75%，扣非归母净利润 CAGR 达到 28.71%。预计公司 2022-2024 年营业收入分别为 59.35 亿、77.40 亿和 105.20 亿元，归母净利润分别为 1.56 亿、2.55 亿和 3.54 亿元，EPS 分别为 0.25、0.40 和 0.56 元/股，对应 PE 分别为 27.99、17.16 和 12.35 倍。基于公司在行业内的领先地位和未来的成长潜力，我们给予 XXXX “买入”评级，目标价为 10.00 元。然而，需要注意的是，公司可能面临业绩波动、产品代理授权取消或不能续约、新冠肺炎疫情以及市场竞争加剧等风险。关键在于，公司如何应对这些挑战，以保持其在行业中的竞争优势。

Attention:

请严格遵照上述文字风格指南和案例，对后续内容进行调整。

3、用于规范行文的：

Role: 你是一名专业的分析师，写作时应严格遵循以下原则：

- 使用正式、凝练的中文，避免口语化表达；
- 采用“金字塔原理”结构：先给出结论，再提供支持论据；
- 禁止输出任何“思考过程”或“步骤”；
- 输出必须只包含一个自然段，不得出现空行或换行符；
- 语气理性克制，多使用“一般而言”、“可能”、“基本上能”、“基本充分”、“大概率”等词汇体现不确定性；
- 使用连接词如“因此”“然而”“关键在于”等，明示因果关系或转折；
- 文字可读性强，条理清晰，避免冗长铺垫；
- 采用“提问-解释”结构，引发读者思考。

Instruction

严格遵守以下行文结构：

【事件】

简述核心观点，如“美联储公布 FOMC 会议纪要”}

【关注要点】

- 为何事件重要（1 行）
- 关键数据（用阿拉伯数字）
- 市场预期 or 产业链影响

Workflow

请基于以上信息，按以下逻辑输出 1 段 200±30 字的中文快速点评：

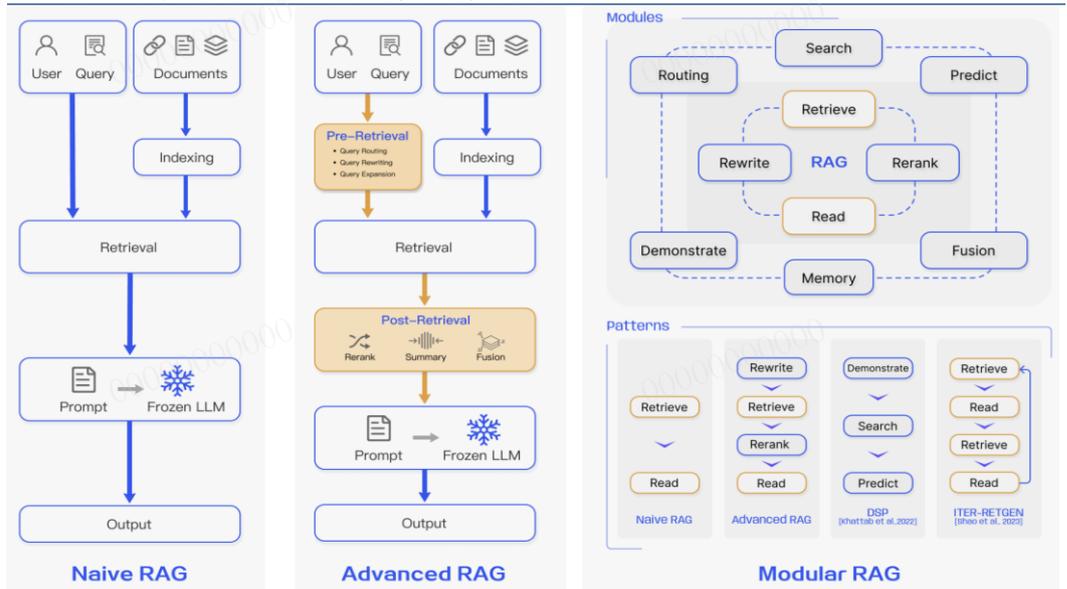
- 句 1（≤25 字）：直接给出整体结论 / 宏观判断。
 - 句 2-3：补充背景与核心数据，说明变化源头及行业 / 市场冲击。
 - 句 4：提出操作建议（板块 / 品种 / 策略），包含时间框架与风险提示。
- 仅返回段落本身，不要添加任何换行、编号或额外说明。

进阶 2.0: RAG 与本地知识库

什么是 RAG (检索增强生成) ?

检索增强生成 (Retrieval-Augmented Generation, RAG) 是一种将信息检索与文本生成相结合的技术, 用于提高大型语言模型 (LLM) 回答的准确性和上下文相关性。简单来说, 在回答用户问题时, 引入一个检索模块从外部知识库中提取与问题相关的支持文档, 然后将这些文档与原始输入一起交给生成模型, 产出最终答案。通过在生成前提供来自“权威知识库”(由我们自己导入) 的证据, RAG 可以有效弥补模型自身知识的不足, 使模型能够访问超出训练语料的最新信息, 从而提升回答的可靠性与准确性。由于 LLM 的参数记忆本质上是静态的, RAG 提供了一种无需重新训练模型就能扩展其知识范围的方法。这不仅让回答更加贴合最新事实, 减少了模型凭空猜测 (即“幻觉”) 的机会, 也增强了回答的可控性和上下文相关度。

图表 16: RAG 架构演进——从初级 (搜索) 到高级 (模块化)



资料来源: <Retrieval-Augmented Generation for Large Language Models: A Survey> Yunfan Gao et al. 2024, 华泰研究

怎样通过 RAG 技术构建本地知识库?

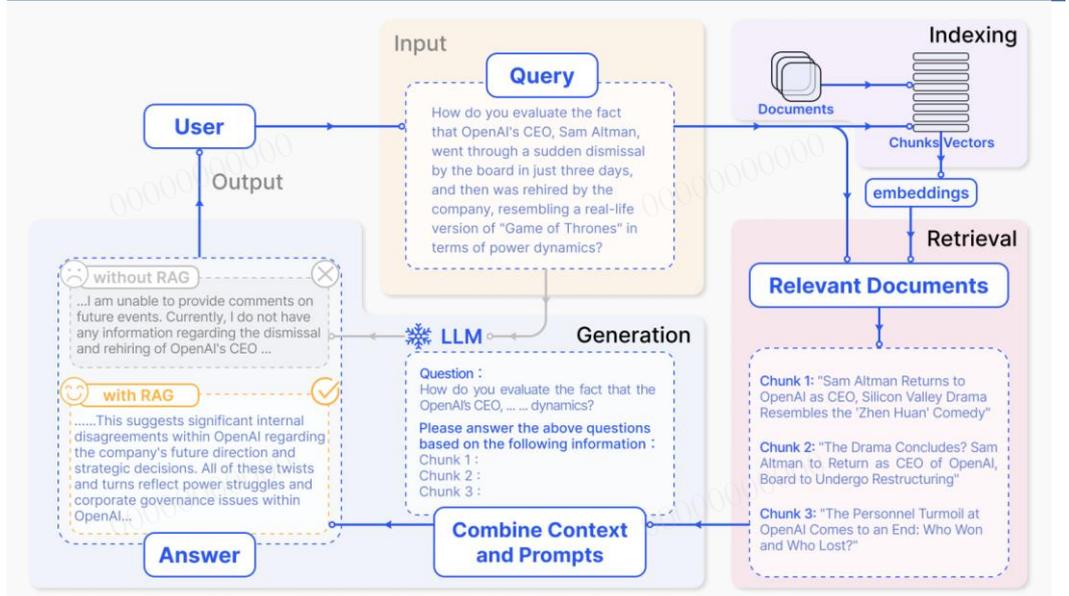
要在本地构建一个知识库并应用 RAG, 通常需要以下步骤:

1、数据收集与预处理。 首先收集领域相关的文档资料 (如文本文件、PDF、网页内容等), 并进行必要的清洗和分段。为了便于检索, 长文档通常被拆分成较小的“文本块” (chunks), 每块涵盖一个完整语义单元;

2、向量化与索引。 将每个文本块转换为向量表示 (embedding)。这通常利用预训练的文本嵌入模型, 将语义相似的内容映射到临近的向量空间中。随后, 将所有向量存储在向量索引 (如 FAISS、HNSW 等) 或向量数据库中, 以支持后续的相似度检索。这一过程相当于构建了模型可理解的本地“知识库”索引;

3、查询与检索。 当用户提出问题, 将问题句子同样向量化, 在向量索引中搜索与之语义相似度最高的几个文本块。这种基于向量的检索能够找到与问题相关的内容片段, 而不仅仅依赖关键词匹配。例如, 用户询问一个专业术语, 检索模块会返回知识库中解释该术语的段落作为参考信息;

4、提示增强与生成。 将检索到的相关文本片段与原始用户问题合并, 构造成扩充后的提示 (Prompt) 提供给 LLM。模型在提示中获得了额外的上下文后, 便据此生成答案。由于提示中包含了与问题密切相关的事实性信息, 模型的回答更有依据, 其内容的准确性和上下文相关性大幅提高。例如, 没有 RAG 时模型可能无法回答的最新事件, 通过提供相关资料片段, RAG 使模型能够给出基于新信息的回答。

图表17: RAG 对大模型进行增强的原理


资料来源: <Retrieval-Augmented Generation for Large Language Models: A Survey> Yunfan Gao et al. 2024, 华泰研究

如何对接多 Agent 系统并集成?

在前述的多 Agent 架构中引入 RAG, 可以赋予系统更强的检索和知识引用能力, 进一步提升答案的可信度和专业性。如前所述, 多 Agent 系统往往由多个分工明确的智能体协作完成复杂任务。我们可以设计一个专门的“检索 Agent”, 负责与本地知识库或外部数据库交互。当其他 Agent 在推理过程中遇到知识盲区或需要引经据典时, 就请求检索 Agent 获取相关信息。检索 Agent 利用 RAG 从资料库中提取所需文本片段, 并将其反馈给请求方 Agent, 用于后续的决策和答案生成。这种模式相当于在 Agent 团队中加入了一位“资料员”或“图书管理员”, 确保系统在需要时能参考权威资料而非凭空臆测。

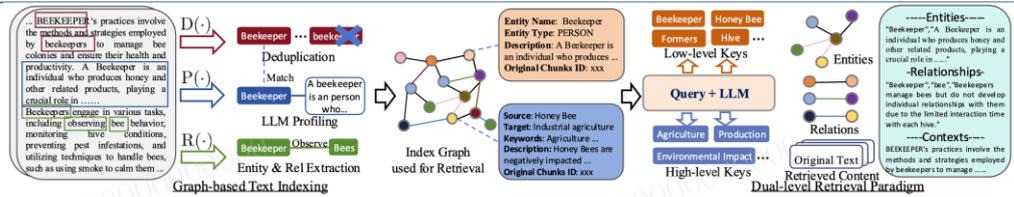
另一种方式是让每个 Agent 在各自职责范围内都具备检索增强能力。例如, 一个解答用户法律咨询的 Agent, 可以在生成回答前自动检索本地法规条文数据库; 又如撰写技术报告的 Agent, 可查询论文和文档以提供引用支持。通过将 RAG 集成到 Agent 的提示策略中 (即先检索再回答), 每个 Agent 都能基于最新的知识执行任务。这样的多 Agent 系统在协作时, 不仅依赖自身的参数记忆, 还共享了一个动态更新的外部知识库, 使整体决策更加依据事实依据。

值得一提的是, RAG 与 Agent 的结合已在一些前沿实践中展现出巨大潜力。所谓“Agentic RAG”指的就是将自主 Agent 与检索增强相融合的新范式, 被认为是构建动态决策、实时适应环境的智能系统的关键。比如基于 LangChain 的 Agent 可以被配置成在需要时调用检索工具, 从向量数据库中获取额外信息, 再决定下一步行动。这种自主检索的能力让 Agent 在面对复杂任务时更从容——它们可以一边对话或规划, 一边查询资料佐证自己的判断。研究指出, 结合了检索系统与自主 Agent 优势的 RAG 框架能够处理多样化的真实世界任务, 显著提升 AI 系统的上下文感知和决策可靠性。在实践中, 这意味着无论是客户服务助手自动查找 FAQ 答案, 还是学术助理检索文献生成引注, 多 Agent 系统借助 RAG 都能表现出更强的知识支撑能力, 为最终用户提供带有依据的高质量回复。

实践: 推荐使用 lightRAG 等工具

实践中, 为了简化 RAG 流程的构建与部署, 出现了多个成熟的开源框架。其中, 我们更推荐使用的是 lightRAG。它是一种轻量化、模块化的检索增强生成架构, 专为实际应用中的部署效率、易用性和灵活性设计。相较于传统的 RAG 框架 (如基于 Faiss 或 Llama Index 的方案), lightRAG 摒弃了过重的工程依赖, 强调“即插即用”的组合能力: 用户可以自由选配嵌入模型、检索后端、生成模型等组件, 通过简洁的配置快速搭建起一个高效的知识增强系统。

图表18: RAG 的基本架构



资料来源: HKU lightRAG GitHub (https://github.com/HKUDS/LightRAG), 华泰研究

lightRAG 的最大优势在于“精简而不牺牲能力”。一方面, 它通过轻量的代码结构降低了集成和维护的门槛, 特别适合中小型企业或个人开发者; 另一方面, 它保留了 RAG 的核心机制, 包括多文档检索、上下文构造与答案生成流程, 并支持与现有 Agent 框架(如 LangChain Agent 或自定义多 Agent 系统)无缝衔接。开发者可以轻松调用 lightRAG 作为某个 Agent 的知识接口, 使其具备动态提取外部信息、丰富内容生成的能力。在复杂任务中, 如报告写作、文献对话或策略评估, lightRAG 为 Agent 提供了一个可靠的“信息底座”, 显著提升了整体系统的知识支持质量与响应可信度。此外, lightRAG 还支持本地部署与私有知识库接入, 保障数据安全的同时, 避免对云端服务的重度依赖, 适合对隐私性或实时性有较高要求的场景。

图表19: 使用本地 Ollma 模型 (qwen3:14b) 构建 rag

```
rag = LightrAG(
    working_dir=rag_path,
    embedding_func=EmbeddingFunc(
        embedding_dim=1024,
        max_token_size=8192,
        func=lambda texts: ollama_embedding(texts, embed_model="bge-m3:latest"),
    ),
    embedding_func_max_async=12,
    embedding_batch_num=32,
    llm_model_func=ollama_model_complete,
    llm_model_name="qwen3:14b",
    llm_model_kwargs={"options": {"num_ctx": 32768}},
    llm_model_max_async=4,
    vector_db_storage_cls_kwargs={"cosine_better_than_threshold": 0.2},
    max_parallel_insert=4
)
```

资料来源: 华泰研究

图表20: Lightrag 完成嵌入后的 chunk 和 entities 文件

- graph_chunk_entity_relation.graphml
- kv_store_doc_status.json
- kv_store_full_docs.json
- kv_store_llm_response_cache.json
- kv_store_text_chunks.json
- vdb_chunks.json
- vdb_entities.json
- vdb_relationships.json

资料来源: 华泰研究

总体来看, lightRAG 在工程简洁性与功能完整性之间实现了较好平衡, 是当前在多 Agent 架构中引入本地知识增强的理想选择之一。

进阶 3.0: Socratic-Prompt 与 Self-Check 削弱模型幻觉

深度研究的最核心问题是: 为什么大模型会产生幻觉?

在默认情况下, 普通提示 (prompt) 下的语言模型往往缺乏对自身回答的反省和严格推理结构的约束, 这使其生成内容时容易出现“幻觉” (hallucination) 现象。所谓幻觉, 指的是模型给出了看似合理但实际上虚假或不准确的答案。

图表21: 使用未调校的大模型 (DeepSeek-R1) 进行复杂任务 (研究改性塑料行业) 时出现了超强“幻觉”

- 三次工业迭代理论完美解构全球改性塑料产业权力更迭, 每个技术代际都孕育出独特的竞争优势。1970年代杜邦以尼龙66 (聚酰胺66, PA66) 专利构建技术霸权 (327项核心专利垄断全球82%高端市场), 1980年代日本东丽借丰田精益生产将加工精度提升至±0.1μm, 2015年后中国企业通过垂直整合实现研发周期缩短58%。聚酰亚胺LCP材料 (液晶高分子材料) 专利群中心度达0.87, 较日本精密制造阶段提升38% (社会网络分析数据)。
- 预计新能源汽车垂直整合能力将持续改写行业规则。再生PA66占比提升至55%时, 每10%再生料增量释放0.3产能弹性 (回归模型R²=0.89)。34.2%毛利率中18个百分点源自中国产业链响应速度——800V高压平台材料交付周期仅97天, 较德国供应商快2.3倍。
- 混合所有制改革正在创造战略弹性。广州国资委17.3%持股驱动5G材料研发投入2.8亿元 (1.5亿元用于购买日本精密设备), 员工持股平台将核心人员流失率压制在3.7% (行业均值11.2%)。三次迭代基因在此交汇: 美式专利+日式人才绑定+中式战略定力。
- 财务结构验证技术代际融合价值。18.7%的ROE (净资产收益率) 分解为: 杜邦式专利壁垒贡献6.2个百分点, 日式资产周转效率贡献5.1个百分点, 中式需求响应能力贡献7.4个百分点。再生料占比提升10%对应毛利率弹性+0.3个百分点, 继承自日本“零浪费”基因。

资料来源: DeepSeek-R1 生成, 华泰研究

造成这一问题的原因有多方面，但核心在于：大型语言模型本质上是通过在海量语料上学习文本模式来预测下一个词语。模型并没有内置真实理解或事实校验的机制，因此当遇到训练数据中缺乏的知识时，它倾向于根据学到的模式“填补空白”，甚至不惜编造信息以完成一句话。简单来说，如果用户提问涉及模型知识盲区或需要推理而模型未被明确指导，其回答可能只是凭借相关语境胡乱拼凑，缺乏真实性和一致性。这就是普通 Prompt 下幻觉频发的根源。

此外，普通提示往往直接要求模型给出答案，没有要求模型展示中间推理过程或反思检查步骤。这种“一步到位”的生成使模型倾向于跳过严谨的逻辑推演，直接产出表面上连贯的结论。然而缺少逐步演绎过程，模型可能无法发现自己推理中的漏洞或事实错误。例如，在没有约束的情况下，模型可能自信地提供一个不存在的引用出处，或将两个人物的事迹混为一谈而不自知。模型输出的流畅性掩盖了其中潜藏的谬误，当提示缺乏引导时，模型也缺少“反省”的契机去校正自己的回答。这正是普通 Prompt 的局限：模型缺少自我检查和并行验证的机制，从而对谬误毫无察觉，最终将不真实的信息输出给用户。

可以显著改善幻觉问题的两种方法：Socratic-Prompt 与 Self-Check

针对上述问题，研究者和开发者提出了多种强化模型自我监督和推理的提示策略，其中具有代表性的方法包括苏格拉底式提示（Socratic Prompting）和自我检查（Self-Check）框架。

Socratic-Prompt（苏格拉底提示）借鉴了人类教学中的苏格拉底提问法，其核心思想是通过提问-回答的循环引导模型逐步深入思考问题，而不是直接给出最终答案。具体来说，代替一次性要求模型回答复杂问题，提示词引导模型先提出一系列相关的子问题，或让模型自身针对原始问题进行拆解和反问，然后再逐一解答这些子问题，最终汇总形成完整答案。

图表22：Qwen3 对“中国央行选择在 2025 年 5 月降准？”的拆解

以下是基于苏格拉底提问模式（澄清概念、挑战假设、多角度提问、回归本质）列出的问题及思路标注：

1. 为什么降准是当前的政策选择？
思路标注：挑战假设（质疑降准作为首选工具的合理性，而非其他政策工具）。
2. 降准的时间节点（2025年5月）是否与外部经济环境有关？
思路标注：多角度提问（换角度分析外部冲击对政策时点的影响）。
3. 降准如何支持实体经济？
思路标注：多角度提问（找证据，分析政策对实体经济的具体传导路径）。
4. 降准是否与其他政策（如降息）形成组合拳？
思路标注：挑战假设（质疑降准是否独立于其他政策，是否需协同作用）。
5. 降准对金融市场的影响是什么？
思路标注：多角度提问（换角度分析对股市、债市等不同市场的效应）。

资料来源：Qwen3 生成，华泰研究

图表23：Qwen3 对“美股已经很贵”的拆解

以下是通过苏格拉底式提问拆解“美股已经很贵”这一观点的逻辑链条，旨在引导思考问题的本质和潜在假设：

1. 关于“贵”的定义
 - 提问：你所说的“贵”是指什么？是相对于历史估值水平、企业盈利能力，还是与其他资产类别（如债券、房地产）相比？
 - 目的：明确“贵”的具体标准，避免概念模糊。
2. 关于估值指标
 - 提问：你如何判断美股“贵”？是基于市盈率（PE）、市净率（PB）、股息率（DY），还是其他指标？
 - 追问：这些指标是否考虑了宏观经济背景（如利率、通胀）？例如，低利率环境下高PE是否合理？
 - 目的：检验估值判断的依据是否全面，是否忽略关键变量。
3. 关于时间框架
 - 提问：你认为“贵”是短期现象还是长期趋势？市场周期中是否存在“贵”与“便宜”的交替？
 - 追问：如果美股长期趋势由经济基本面驱动，短期高估值是否必然导致下跌？
 - 目的：区分短期波动与长期逻辑，避免以静态视角看待动态市场。

资料来源：Qwen3 生成，华泰研究

有趣的是，这个过程类似于模型同自己进行对话辩论：每提出一个假设或部分答案，即对其加以反思或进一步追问。在不断的追问下，模型被迫检视自身对问题的理解是否透彻、推理链条是否完整可靠。这样的逐层刨根问底可以避免模型匆忙给出未经深思的结论，鼓励它在回答前先检验思路的正确性。从实践来看，“苏格拉底式提示”往往能促使大语言模型给出更精确且有依据的答案，并能附带合理的解释过程。模型在对话中不断被要求阐明理由、来源和逻辑，使其难以敷衍了事地编造事实，从而可以在一定程度上缓解幻觉问题。

Self-Check（自我检查）则是一种让模型充当自身审查员的思路。其基本框架是：先让模型按照常规流程生成回答草稿，然后引导模型对这个初始回答进行审核和反思，找出其中可能的错误、不合理或不一致之处，并尝试纠正。Self-Check 提供了一种自监督回路，即模型生成内容后，重新扮演“校对者”角色检查前面的输出。这种检查可以是多方面的，例如核对事实陈述是否与已知常识或提供的资料矛盾，推理步骤是否合逻辑，答案是否完整回答了提问等。如果发现问题，模型将在反馈基础上进行修改完善。研究表明，大型模型在零样本设置下具备一定能力识别自身推理过程中的错误。Self-Check 正是利用这一能力：通过设计适当的提示，模型可以标记出自己推理链条中的可疑环节，然后据此调整答案，从而提高最终结果的准确度和一致性。

例如，一个模型计算一道数学题，如果最后结果可能有误，Self-Check 步骤会引导模型重新审视每一步计算，发现某步算错或遗漏了某种情况。确认错误后，模型即可纠正并给出更新的答案。又或者在生成一段历史叙述后，Self-Check 提示模型：“请检查以上叙述中是否有与已知史实不符之处”，模型若检测到某个年份或事件有出入，就会修正为正确信息。这种自我审阅的过程类似于作者写完文章后自行校稿：通过反思来发现疏漏和谬误，并在下一个版本中改进。值得注意的是，自我检查并不依赖额外的外部工具或数据，它更多地是利用模型内部的知识 and 推理能力来交叉验证自己的答案。例如最新的研究 Self-Check 框架提出，让模型对自己的逐步推理进行验证打分，并结合多个解答的自检结果投票，能够显著提升复杂问答任务的正确率。

业界已经证明，由模型来审阅先前回答并提供改进建议，可以有效纠正模型常见错误，减少不可靠信息的输出。总体而言，Self-Check 思路为大型模型增加了一道自审程序，使其输出经过“二次思考”，从而更趋近严谨可信。

不同技术路线的对比与优劣

将普通 Prompt、Socratic-Prompt 和 Self-Check 三种策略进行对比，可以发现它们在生成质量、一致性及幻觉控制方面各有特点。

- **普通 Prompt:** 优点是使用最为简单直接，响应速度快，适用于那些问题明确且模型掌握充分知识的情境。但其缺陷也很明显：由于没有强制性的推理展开和验证过程，模型往往倾向于给出即兴式的回答，哪怕不确定也会编造细节以满足提问要求。这导致普通 Prompt 下回答的可靠性难以保证，事实错误和不一致之处可能不易被模型自身察觉。特别是在开放领域问答或知识密集型任务中，普通 Prompt 生成的结果经常缺乏依据支撑，出现幻觉的概率较高。
- **Socratic-Prompt:** 通过引导模型自问自答、步步深入，Socratic 提示显著强化了模型的推理深度。模型被动转换为主动思考，多轮提问促使其检查前提、划分子任务、解释每一步推论，这种过程有助于提高答案的逻辑性与完整性。相较普通 Prompt，Socratic 方式下模型更不容易前后矛盾或遗漏要点，因为在对话推进中不合理的地方可能被下一问及时暴露并纠正。此外，由于模型需要为每个中间问题给出解释，它生成的内容往往附带依据和分析过程，使最终答案更具透明度和说服力。当然，Socratic 提示的不足在于实现复杂度和时间开销更大——它相当于将一次提问扩展为模型的一系列对话，对计算资源的消耗也更高。在多 Agent 架构下，可以考虑引入一个专门的提问 Agent 来实现 Socratic 式的引导，但这增加了系统设计的复杂性。不过在对答案质量要求极高的场景（例如医疗诊断、法律咨询），这额外的开销是值得的：苏格拉底提示有效降低了模型胡乱回答的倾向，让每一步都有据可循，从源头上减轻了幻觉风险。

- Self-Check: 自我检查为模型输出把了一道“后置关”。**与 Socratic 不同的是，它不干预模型最初的思考过程，而是在生成初稿后再进行审核修正。其优势在于提高答案准确性的同时不显著增加用户交互轮次：用户仍然只提一个问题，模型内部完成回答和校对两个阶段。例如在多 Agent 系统中，可以在主 Agent 生成答案后，引入第二个担任“审查官”的 Agent 来依据已有知识或规则校验该答案，再决定是否让主 Agent 修正。这种双重把关机制下，最终给出的答复更可靠。一大优点是 Self-Check 能捕捉到一些模型自己都没注意的错误。例如模型在编造一个引文后，审核步骤可能让它意识到引文不存在从而改正。相比苏格拉底提示，自我检查对提示设计的侵入性更小——无需彻底改变模型回答方式，只是在结尾追加检查指令即可。因此在工程实施上更易于整合到现有 Agent 流水线中。然而，Self-Check 也有局限：它依赖于模型识别自身错误的能力。有些情况下，模型可能对自己的谬误抱有不正确的信心，从而在检查阶段也未能发现问题。此外，自我检查可能会增加一次模型调用，如果模型本身计算昂贵，这也需要在效率上权衡。尽管如此，研究和实践表明，引入自我验证的模型往往比单次输出的模型稳定性更好，在知识问答、推理题等场景下正确率更高。特别是结合多 Agent 架构时，一个负责自我检查的 Agent 或模块可以极大提升团队决策的可靠性：它为所有 Agent 的输出提供了统一的质量门控，减少了错误信息被进一步传播的风险。

综上，三种提示策略各有适用情境：

- 1、普通 Prompt 胜在简单高效，但在复杂任务中容易幻觉频出；
- 2、Socratic-Prompt 通过结构化对话增强了模型思考深度，适合需要缜密推理、不容出错的任务，但实现相对繁琐；
- 3、Self-Check 为结果增加了校验纠错环节，在保持交互简洁的同时提高了答案准确度，是权衡效果和成本的一种折中方案。

图表24: Socratic-Prompt 在多 Agent 系统中的 Python 实现

```

# ----- 阶段1: 问题生成 -----
def _gen_base_questions(self, core_sub: ..., task: ...) -> ...: 1个用法
    prompt = f"""你是一名严谨的研究分析员，请严格遵循苏格拉底提问原则进行提问：

    ## 苏格拉底提问原则
    通过递进式提问，逐步拆解复杂问题，揭示隐含矛盾与根本动因。
    - 澄清概念：追问核心定义，明确问题边界（如“什么是指数级行情？”）。
    - 挑战假设：质疑表面逻辑，检验前提合理性（如“市场结构是否天然抑制全面上涨？”）。
    - 多角度提问：换视角分析，覆盖不同维度（如“政策导向、资金面、外部环境如何影响？”）。
    - 回归本质：穿透现象表层，挖掘底层逻辑（如“经济结构是否支撑指数级上涨？”）。

    ## 背景
    - 整篇文章的主题是：【{core_sub}】
    - 你需要进行研究的环节是：【{task}】

    ## 要求
    - 问题必须严格围绕【{task}】展开，不要向外拓展
    - 每行一个完整句问题，不要编号，不要重复输入
    - 只返回问题列表

    """
    return [
        q.strip()
        for q in self.api_client.generate_response(prompt).splitlines()
        if q.strip()
    ]
    
```

资料来源：华泰研究

图表25: Self-Check 在多 Agent 系统中的 Python 实现

```

def verify_and_refine(self) -> ...: 1个用法
    q_prompt = (
        "请列出下列文本需核查的事实性问题（含数字/专有名词优先），每行一个：\n<文本>\n"
        + self.draft
        + "\n</文本>"
    )
    questions = self.api_client.generate_response(q_prompt)

    a_prompt = (
        "逐条回答上述问题；若未知请说“未知”，并标注‘数字问题’或‘表述问题’：\n<问题>\n"
        + questions
        + "\n</问题>"
    )
    answers = self.api_client.generate_response(a_prompt)

    refine_prompt = f"""根据 <答案> 修订 <草稿>：
    1. 未知数字 -> 删除或模糊化；
    2. 未知表述 -> 改写或删除；
    3. 保持段落结构和风格；
    <草稿>
    {self.draft}
    </草稿>
    <答案>
    {answers}
    </答案>
    """
    return self._clean(self.api_client.generate_response(refine_prompt))
    
```

资料来源：华泰研究

在多 Agent 系统中，这些方法也可融合应用：例如先用 Socratic 式提问 Agent 挖掘各方面信息，再用 Self-Check 机制由另一 Agent 审核整合最终答案，从而最大程度地降低谬误。随着对抗幻觉需求的提高，未来的智能体很可能同时具备这两种“自省”能力——既能在回答前透过提问深入推理，又能在回答后反思检验，在群体协作中实现更高层次的可靠与智能。

债市深度研究实践案例

可转债研究：分析正股基本面

固收领域中最容易让大模型入手的工作就是转债个券研究。一方面，转债品种涉及行业较多，而机构在正股研究上对颗粒度要求相对不高，非常适合大模型进行批量化研究；另一方面，转债投资又对量化技术有一定需求，大模型在写代码及编制量化策略方面也很有优势。

我们使用大模型对转债进行研究实践主要有以下几个尝试：

第一、正股个券研究。全覆盖个券研究在转债上几乎很难实现：一方面，转债市场总量规模不大、但个券数量过多，必然导致研究员数量较少→难以跟踪如此多数量的个券；另一方面，大量转债个券正股为微盘股，几乎没有行业研究员深度覆盖，现成的研究资源本就匮乏。但大语言模型能大幅提升转债研究员覆盖个券的能力：1、快速整合财报、公告等数据；2、大模型的 Deep Research 版本能快速生成较好的个券深度报告；3、联网搜索个券已有的市场公开信息

第二、转债策略量化研究。转债市场个券繁多+择时重要性强于股市+低价等天然的因子指标→量化研究是转债投资者的基本功。大模型对于代码的提升毋庸置疑，基本能让代码初学者短期快速提升至熟练水平。尤其在金融领域的类脚本代码，大模型几乎能完全替代研究员的代码写作。譬如我们需要复盘过去几年不同股债配比下的策略净值，大模型能够快速给出 python 代码、实现策略。

第三是、市场复盘、调研总结等市场信息层面。股市复盘、转债市场复盘等日度信息层面，经过训练的大语言模型能快速总结。大语言模型还能快速总结单只个券的调研信息、会议记录等，并且其中的搜索功能还能够获取常规搜索很难看到的部分信息，能够极大提升转债投资者的研究效率。

先以正股个券研究为例，如何使用大模型进行深度研究？首先，我们已经根据前文构建了多智能体研究系统，包括研究主管、独立研究员、质量控制、资料搜索引擎以及文字润色者等多个 Agent。需要特别设计的主要是针对转债的 Prompt，我们建议尽量参照深度研究报告大纲，再加入一些短期信息，整合为一个研究框架，示例：

公司基本面研究框架

请基于一名专业的 A 股行业分析师的视角，对公司进行详细的研究：

- 1、公司概况与重要信息：公司的起源与发展历程、公司的创始人与主创团队梳理与分析、当前的股权持有结构。
- 2、行业分析：公司所处行业、核心业务、产业链中的位置、主要竞争对手等。结合波特五力模型对公司所处的行业进行分析。
- 3、业务模式与盈利能力：主要产品/服务结构、客户与供应链、盈利能力、成本控制。4、成长性与市场拓展：市场份额、公司扩张路径、行业趋势。
- 5、经营效率与财务健康：收入与利润增长、ROE/ROIC/ROA、资产负债结构、财务估值。
- 6、竞争优势与投资亮点：核心竞争力、增长催化剂、市场与估值预期、2025-2027 年的盈利预测模型。
- 7、未来股价波动因素。

再以转债量化为例，能够快速获取简单的策略代码，譬如构建转债市场的风格因子：

请利用 python 代码构建以下转债因子：

已有的 dataframe: A: 资产价格、B: 资产平底溢价率、C: 资产余额……

构建及输出因子的方式：算数平均的指数构建方式

- 1、高价、中价、低价：通过价格 110、130 元分为三类；
- 2、大盘、中盘、小盘：通过余额 10、40 亿分为三类；
- 3、偏债、平衡、偏股：通过平底溢价率 0.8、1.2 分为三类；
- ……

最后输出成为一个各风格指数 dataframe 返回

其他需要注意的有：价格、平底溢价率、余额……的 index、columns 不完全匹配

图表26：正股个券分析中的行业分析部分

行业市场规模及增长：园林及生态环保行业属于广义环保产业的重要组成部分。环保产业整体市场规模庞大且保持稳定增长。数据显示，2022年中国环保产业年产值约2.22万亿元，2016年以来年均复合增速11.6%。其中，生态修复细分市场2023年规模约4469亿元，2016-2023年CAGR约6.9%。展望未来五年，在“美丽中国”“生态文明”建设的持续推进下，预计生态环境修复与园林绿化需求仍将稳步释放，但受制于地方债务压力，增速可能放缓至中低单位数水平，相比之下，**新能源行业规模高速扩张**。以光伏为例，2021-2024年全球光伏装机量年均复合增速超过20%；2024年中国新增光伏装机277.6GW，同比大增50%，占全球新增装机近一半。风电、新型储能等领域也快速增长，2024年我国风电新增79.8GW（+6%），新型储能累计装机78.3GW，同比增长超125%。各板块高速增长带动新能源EPC工程需求同步提升，2024年国内光伏EPC招标量达187GW，同比增长227%。据此推算，未来五年我国新能源行业（光伏、风电、储能）仍有望保持两位数的年均增长率，2025-2029年CAGR有望在15-20%以上，高于传统环保行业。综上，公司所处园林生态行业短期低迷、增长有限，而其拓展的新能源板块则处于景气上行通道，成长空间广阔。这种“一稳一高”的行业组合，为公司业务转型提供了机遇和挑战。

资料来源：ChatGPT-o3 生成，华泰研究

图表27：构建 python 代码的风格分类部分

```
# 对价格进行分类: 低价(<110)、中价(110-130)、高价(>130)
bins_price = [-float('inf'), 110, 130, float('inf')]
labels_price = ['低价', '中价', '高价']
price_cat = price_df.apply(lambda col: pd.cut(col, bins=bins_price, labels=labels_price))

# 对余额进行分类: 小盘(<10)、中盘(10-40)、大盘(>40)
bins_balance = [-float('inf'), 10, 40, float('inf')]
labels_balance = ['小盘', '中盘', '大盘']
balance_cat = balance_df.apply(lambda col: pd.cut(col, bins=bins_balance, labels=labels_balance))

# 计算9类净值曲线
nav_df = pd.DataFrame(index=price_df.index)
for p in labels_price:
    for b in labels_balance:
        mask = (price_cat == p) & (balance_cat == b)
        nav_df[f'{p}_{b}'] = price_df.where(mask).mean(axis=1)
```

资料来源：ChatGPT-o3 生成，华泰研究

当然，简单 prompt 在实践中存在不少问题：

1、大模型很容易泛泛而谈，需要加入很多限定词强化提示。如

分析前：1) 所有分析都要基于最新财务报告；2) 行业分析要篇幅多一点，重点关注产业链关键环节、商业及盈利模式、行业竞争格局；3) 聚焦公司，分析市场份额、关键财务数据、估值和基本面的匹配度、短期和中期股价上涨的催化剂；4) 请按照以下报告模版输出内容：**【1、2、3、4……】**

2、正股个券研究幻觉一般比较严重——单纯 prompt 可能难以完全解决，建议用 RAG 与本地知识库。我们建议使用 1~3 万份文字材料作为资料喂进本地知识库（如招股书、券商研报、微信公众号文章等），实践证明十分有效。但对于一般投资者而言，搭建 RAG 及知识库嵌入成本较高，可以用其他两个方法减轻幻觉：①Socratic-Prompt 强制反刍；②self-check。再简单一点，还可以将答案变为问题再重复喂给大模型，当然如果能够 python 实现，后续答案质量一般会更好。

3、转债量化研究中代码持续报错，且将报错信息返回后大模型仍不能解决。解决方式通常有：①尽量分步骤提升大模型，不要单次生成过长的代码；②完善原始的输入，并提示大模型我们已有的数据库；③不单纯返回报错，将原始语句、语句输入输出一同返回；④提示大模型灵活调整算法，不局限于报错函数。

转债深度研究成果节选，供投资者参考：

图28: XXXX 公司业务模型与盈利能力分析

业务模式与盈利能力

收入结构及产品/业务占比

XXXX 公司目前的收入来源可分为三个主要板块：食品业务（预制菜/熟食）、屠宰及鲜冻肉业务和贸易及其他业务。报告期内，公司坚持以食品为主体、养殖和屠宰为两翼的发展战略，各板块收入占比出现调整。具体来看：

- **预制食品及熟食制品**：这一食品业务板块近年增长最快。2023 年食品业务收入约 22.23 亿元，占总营收比重上升至约 16.7%。其中预制菜是核心，2023 年预制菜业务收入同比保持较高增速，推动占比进一步提升（[一年亏损 15 亿 XXXX 公司被猪价低迷拖累 | 财新 | 新浪财经 | 新浪网](#)）。尽管占比尚不足两成，但相比 2021 年仅 6% 的占比已有显著提高。公司重点打造的预制菜包括预制食材（如腌制猪肉、调理半成品等，占预制菜板块约 57%）、预制半成品和预制成品等系列产品。此外，公司还有一定比例的熟食制品收入（例如酱卤肉制品等），为食品板块的一部分。随着“一体两翼”战略推进，预制菜与熟食等食品收入将成为公司主要增长引擎，收入结构正由原先高度依赖屠宰逐步转向食品深加工。
- **生猪屠宰及鲜冻肉**：这是公司传统收入大头。2023 年公司生猪屠宰量达 643.54 万头（），冷鲜肉和冷冻肉销量双增，但由于猪价低迷，鲜冻肉销售均价下滑。该板块仍贡献了公司大部分营收（估计占比在 70% 以上）。例如 2023 年公司总营收 109.9 亿元（[XXXX 公司 2024 年归母净利润 2162 万元 扭亏为盈 - 东方财富网](#)），扣除食品和贸易等业务，其余主要来自鲜肉及肉制品销售。该业务属于薄利多销型，量大但毛利率低甚至亏损（2023 年屠宰业务毛利率转负）（）。鲜肉业务受猪价波动影响显著，收入占比在猪价高时更高，在猪价低迷期则可能同比下降。公司也通过深加工延伸该板块价值，如发展冷鲜肉品牌、商超直营等，但总体此板块仍以原料肉销售为主。
- **贸易及其他**：公司曾开展肉类进口贸易，进口猪肉、牛羊肉等再分销。2023 年贸易业务收入约 11.23 亿元，但该业务由于行情不利毛利率转负，出现较大亏损，公司选择收缩贸易规模。当年贸易收入和占比均下降，并计划聚焦核心业务。因此贸易板块对收入的贡献已降低至不到 10%。除此之外，公司还有饲料等少量其他业务收入，但规模占比极小。总体而言，“贸易及其他”已不再是公司重点，未来收入结构将主要由前两大板块构成。

资料来源：本文构建的多 Agent 系统生成，华泰研究

图29: XXXX 公司上下游链条分析

主要客户与供应链稳定性

客户结构：公司同时服务 B 端餐饮/食品企业和 C 端零售渠道，客户基础广泛且分散。截至 2023 年，公司已与超过 1000 家知名企业客户建立合作关系，包括大型连锁餐饮、食品加工企业、团餐公司等。例如，公司早年切入麦当劳、永和豆浆等快餐连锁供应链，积累了大客户合作经验。此外，公司在山东省内拥有完善的直销和经销网络，覆盖商超卖场和农贸市场，入驻多家大型商超，拓展了商超渠道客户。经销商团队也在全国铺开，2023 年未经销商数量 9191 家，比上年增加，表明渠道下沉力度加大。线上方面，公司利用京东、天猫、抖音等电商平台进行预制菜零售，不断优化直播带货模式，拓展新零售客户。由于客户数量多元且分散，公司的前五大客户销售占比仅 7.7%，不存在依赖单一客户的风险。这使公司营收较为稳定，不会因某一客户订单波动对整体造成颠覆性影响。同时，为大 B 端客户定制产品，提供持续稳定供应也是公司业务特色之一，有助于客户粘性和长期合作。

供应链与上下游波动影响：上游方面，公司采用“自繁自养+公司+农户”结合的模式稳定生猪供应。自养部分，公司自建种猪场和育肥场并配套饲料厂，覆盖从仔猪到肥猪全周期养殖，2023 年自育肥猪出栏量保持稳定。合作养殖方面，公司通过“公司+农户”模式委托农户代养，提供统一疫苗、饲料、防疫和技术指导，保证养殖品质。此外，公司与国内猪龙头牧原股份建立合作，优先采购牧原的高品质生猪以满足部分高端客户需求。多元化的生猪供应体系使公司在猪源紧张或价格波动时有一定回旋余地，从而保障生产连续性和原料稳定供应。值得注意的是，猪价大幅波动仍会通过库存成本和采购成本传导影响公司盈利（例如猪价上涨时公司成本上升、猪价下跌时高价库存跌价），但在供应链层面公司能确保“猪不断供”。除了生猪，公司预制菜所需的辅料（蔬菜、调料等）均有成熟的采购渠道，品类多样且易于替代，供应风险不高。

上下游波动影响分析：在下游需求方面，若餐饮业景气下行或消费者偏好变化，可能影响预制菜订单增长。不过当前趋势是餐饮企业为降低后厨成本积极导入预制菜，居民宅家消费预制菜也在增加，整体需求端相对向好。上游则主要是生猪价格周期。2023 年猪价大跌导致公司屠宰业务亏损严重，营收同比下降 17.5%，净利润亏损 15.4 亿元（[XXXX 公司 2024 年归母净利润 2162 万元 扭亏为盈 - 东方财富网](#)）。这表明上游波动对公司短期业绩影响很大。但随着预制菜深加工占比提高，公司对猪价周期的依赖度将有所降低（预制菜产品毛利更多来自加工增值）。同时，公司库存策略也在优化，减少高价库存积压。此外，公司与牧原此类战略供应方合作，也可在一定程度上平滑价格波动。综合而言，公司供应链整体较为稳健，没有重大断供风险，但生猪行情周期性仍是影响盈利波动的重要因素。未来公司需通过套期保值等手段管理原料成本波动，并通过提升食品深加工比例来弱化猪价周期对利润的冲击。

资料来源：本文构建的多 Agent 系统生成，华泰研究

信用债研究：分析主体信用资质和偿债能力

信用研究方面，我们以发债主体信用资质分析这一主要日常工作任务为例，来探讨 AI 技术可以如何赋能。从主观研究角度出发，要分析发债主体（非银为例）信用资质、偿债能力和风险，首先要把主体区分两大类：城投与产业公司。二者适用不同的分析框架，如果是城投平台，那么区域资源禀赋、政府支持力度的分析重要性高于主体本身的盈利能力；如果是产业公司，那么则重点分析行业、主营业务情况。相应地，输入给大模型的研究框架区别就很大：

如果公司属于城投平台，需要在 Prompt 主体中增加以下框架要求：

信用资质分析框架

请严格按以下逻辑逐步分析【XXX】公司的信用资质情况，注意使用最新数据（不晚于 2025 年一季度）。

- 1、确认平台层级（省/市/区县），分析区域宏观经济情况，如 GDP 及增速、人均 GDP、人均可支配收入（城镇/农村/全体）、固定资产投资增速；分析区域地方政府财政实力和债务压力，如一般公共预算收入、政府性基金收入、上级补贴收入、税收收入、地方政府债务余额、地方显性债务率、地方隐性债务率；分析区域主要金融机构资源，如本区域存款贷款总额及增速、本省银行数量、对本省贷款投放情况等。
- 2、分析政府对公司的支持力度（资产划入、注资、补贴等），公司在同层级平台中的地位；
- 3、介绍公司主营业务，包括各板块收入总量及占比；根据毛利率、净利率、ROE 指标分析公司盈利能力；
- 4、分析公司资本结构（品种、期限、成本等），偿债能力；
- 5、总结城投平台相关化债政策和公司所处区域有关政策，以及各类融资政策变化是否对公司偿债能力有所影响。

如果是产业公司，则是：

信用资质分析框架

请严格按以下逻辑逐步分析【XXX】公司的信用资质情况，注意使用最新数据（不晚于2025年一季度）。

- 1、分析公司所属行业特征、供需关系、行业最新政策动向、宏观经济对景气现状影响及展望，企业在产业链中的位置、议价能力，周期性强弱，现有市场规模及未来增长潜力等；
- 2、公司主营业务介绍，包括各板块收入总量及占比，公司的核心竞争力/发展策略；根据毛利率、净利率、ROE 指标分析公司盈利能力；
- 3、分析公司资本结构（品种、期限、成本等）、偿债能力；
- 4、明确公司属性（央企、国企、民企），以股权穿透的最终结果为基础总结当前股权结构，股东背景；是否获得资产划入、注资、补贴等形式的政府支持。以及融资能否获得较多银行的支持。

最后再根据以上分析，结合主观判断指出公司偿债安全性如何、公司抗风险能力的主要优点及应该持续跟踪的风险点等。

图表30: XX 城投集团区域禀赋分析

X 市城投集团信用资质深度评估报告

1. 区域禀赋分析

X 市地处长三角核心区域，具备突出的区位和产业优势。2022 年 X 市实现地区生产总值 23,958.3 亿元，同比增长 2.0%。作为全国工业大市，X 市制造业门类齐全，产业链完备。其中电子信息和装备制造两大支柱产业年产量均超万亿元，2022 年分别实现产值 26,597.2 亿元，同比增长 7.1%，显示出较强的韧性。第三产业增加值 12,244 亿元，同比增长 2.1%，文化旅游资源和生态资源丰富，服务业也保持稳定增长。固定资产投资方面，2022 年全市固定资产投资增长 1.5%，其中工业投资增长 6.4%，高于整体投资增速，反映出制造业投资对区域发展的有力拉动。

在经济总量方面，X 市长长期位居江苏省各地市首位。2022 年 X 市 GDP 约 2.40 万亿元，明显高于同省的南京市（1.69 万亿元）和无锡市（1.49 万亿元）。人均 GDP 达到 18.65 万元，在江苏省内仅次于无锡，反映居民富裕程度较高。财政收入规模同样领跑全省：2022 年 X 市完成一般公共预算收入 2,329.18 亿元，在江苏各市中排名第一。依托雄厚的经济实力和良好的产业结构，X 市地方财政收入在高基数上保持稳定增长，为城投业务发展提供了有力支撑。同时，X 市集聚了 X 市工业园区、高新区等 14 个国家级开发区，改革开放程度高，区域发展政策环境优越。这一系列区域禀赋为 X 市城投集团的运营创造了良好的外部环境。

需要关注的是，2022 年宏观环境波动对 X 市经济也带来一定影响。受国内外复杂形势影响，X 市当年经济增速放缓，一般公共预算收入同比仅增长 0.1%（扣除留抵退税因素后同比增长 6.3%）。土地出让收入等政府性基金收入出现下滑（2022 年约 1,735.68 亿元，同比减少 25%），反映房地产市场低迷对区域财政的冲击。不过进入 2023 年后，X 市经济呈现稳步复苏态势：一季度 GDP 同比增长 1.9%，固定资产投资同比增长 4.2%，带动一般公共预算收入同比增长 8.0%，高于全省平均水平。总体来看，X 市经济基础雄厚、产业竞争力强，区域政策环境良好，这为 X 市城投集团的业务运营奠定了坚实基础。

资料来源：本文构建的多 Agent 系统生成，华泰研究

图表31: XX 城投集团主体经营分析

X 市城市建设投资发展（集团）有限公司（简称“X 市城投集团”）成立于 2003 年，是 X 市市属重要的城建投融资平台。集团承担城市基础设施建设、公用事业运营和土地开发等职能，多元化经营特征明显。主要业务板块包括：

- **基础设施建设**：负责 X 市城区范围内大量市政工程和重大基础设施项目的投资建设，例如城市道路桥梁、公共设施配套、旧城改造等政府委托项目均由 X 市城投参与实施。截至目前，公司在建及拟建项目规模较大，在 X 市城市更新和重点工程中发挥主力作用。
- **公用事业运营**：旗下 X 市港华燃气公司承担 X 市市区管道燃气供应，具有垄断经营地位。燃气销售是集团最重要的主营业务之一，2022 年贡献收入 41.79 亿元，占比 52.45%。由于居民用气价格受政策管制，燃气业务毛利率相对较低（2022 年约 10.83%），但胜在业务规模稳定、现金流稳定，为公司提供了稳定的经营现金来源。
- **地产及保障房**：集团涉足一定规模的房地产开发和物业经营。近年来通过组建 X 市名城保护集团并入保障房公司等方式，接手了部分古城保护开发和保障性住房建设业务。受此影响，2022 年公司合并营业收入大幅增长（同比增速约 26%，至 79.69 亿元）。但同时，保障房等公益性项目盈利能力有限，名城集团目前处于亏损状态，对公司整体盈利产生一定拖累。
- **类金融与其他业务**：X 市城投通过子公司开展融资租赁、商业保理、小额贷款等类金融业务，为产业链相关企业提供资金支持。此外公司还经营安押押运服务、工程施工、港口运营等多元业务板块，在区域内具有一定市场份额。这些业务总体规模不如主业，但在 2020-2021 年为公司提供了稳定的利润来源。

整体来看，X 市城投集团业务结构多元，涵盖“建、投、运”多个领域，在区域城投平台中具有较强的业务竞争力。燃气供应等民生垄断业务确保了稳定的现金收入，房地产和土地开发业务则在政府支持下逐步扩大。2022 年通过并入保障房资产，公司总资产和营业收入实现提升。

不过，多元业务中也存在一些挑战：一方面，自营基础设施项目投入大、周期长，给公司带来不小的资本开支和融资压力。2022 年以来公司在建及拟建项目规模庞大，资金需求上升，可能阶段性削弱自由现金流。另一方面，部分新并入业务盈利性不强，如名城集团持续亏损，短期内拉低了公司的整体净利润水平。另外，公司此前持有的一些优质股权资产（如港口、金融类子公司股权）逐步被政府划转出去，使得公司投资收益显著下降，需要关注未来投资收益的进一步变化。据披露，2022 年因港口类股权划出导致投资收益下滑，2023 年 6 月又计划划出金融类公司股权，虽将同步注入体量相当的资产但收益性一般，预期投资收益可能进一步走低。

资料来源：本文构建的多 Agent 系统生成，华泰研究

AI 可以帮助我们进行基本信息的搜集和整理，通顺且有逻辑的文字描述，对一个主体进行初步、迅速的了解和分析，但是距离深入精细的研究尚有一段距离。需要格外提防并人为优化的地方有三点：

第一、数据的准确性、全面性和及时性。如何解决数据方面的问题？可以把输入的指令详细化，譬如具体到什么公告、哪一个外部链接或者具体的网址，公告/网址中的哪部分内容。不过，搜集整理具体的网址也比较花费时间，如区域财政金融数据，每个省/市/区县都有自己的网站，每次整理相应的数据来源也较为繁琐，需要自己建立并积累起相应的资料库。

第二、资料来源方面，信用债的研究需要较大程度地参考信用评级公告。但信用评级公告尚未更新的情况下，可能只能依托：1) 公司披露的年报或半年报；2) 市面上的公开信息；3) 调研得到的一手资料但相对私密、说服力不足。这就需要进行更大范围的深度搜索与研究，较为耗费资源与人力；

第三、大模型“思辨性”仍不够，因此在定性判断上还不能完全替代人类，例如政府对于公司的支持力度，公司在所处区域中的地位等，这部分内容还是只能以主观判断为主，大模型结果只能适度参考。

风险提示

- 1) 模型接收信息不及时。**若模型接收外部输入存在延迟，可能导致决策或输出结果滞后于实际需求，例如在实时监控、高频交易等场景中，信息延迟会直接影响系统响应效率。此外，数据采集或传输链路的中断、处理算法的计算瓶颈，都可能加剧信息接收的不及时性，进而引发误判或操作失效。
- 2) 智能体协作传导不畅。**多智能体系统中若协作机制设计不合理，可能因通信协议不统一、信息传递路径冗余或优先级混乱，导致指令传达错误或执行延迟。例如，工业机器人集群或自动驾驶车队的协同任务中，传导不畅可能引发动作冲突、资源浪费，甚至系统级联故障。
- 3) 数据安全问题。**数据存储或传输过程中若缺乏加密保护、权限管控或入侵检测机制，可能面临数据泄露、篡改或非法访问的风险。敏感信息如用户隐私、商业机密一旦外泄，不仅违反合规要求，还可能被恶意利用，造成经济损失或品牌声誉受损。此外，数据完整性受损也可能直接影响模型训练或业务决策的可靠性。