**🔥🔥🔥用Coze从零打造DeepResearch**

**[This type of content is unavailable for download at the moment.]**

|  |
| --- |
| **求 Star，求转发！**  **Python 项目**：https://github.com/langmanus/langmanus  **Web UI 项目**：https://github.com/langmanus/langmanus-web  **LangManus** 是我们团队利用业余时间开源的一款“Manus 复刻”项目。LangManus 是一个基于 LangChain 全家桶的 AI 自主深度研究系统，并且我们集成了浏览器操作、搜索引擎、爬虫、Python 执行器等多种工具。其名称源自 LangChain + Manus（拉丁语中的"手"）。该系统旨在实现全 AI 自主的深度研究，通过整合多种工具，使 AI 能够像人类研究者一样进行信息收集、数据分析和处理。 |

|  |  |
| --- | --- |
| |  | | --- | | **ByteTech x LLM 工程师养成系列**   * [**是的！我们做了 DeepResearch**](https://bytetech.info/articles/7482676626556354594)**🔥🔥🔥🔥🔥 @我们团队** * [**DeepSeek+⁺Coze从零打造Deep Research**](https://bytetech.info/articles/7478162328656838691) @Henry Li**🔥🔥🔥🔥🔥** * [**正在发起公益项目：留守儿童的 Al 智能陪伴硬件**](https://bytetech.info/articles/7462959767519641615) **@Henry Li** * [**🎥**](https://emojipedia.org/movie-camera)**《LLM 学习季》讲座回放：**[提示工程的艺术🎨工具与智能体](https://bytetech.info/articles/7444482560496762889) @Henry Li * [**🎥**](https://emojipedia.org/movie-camera)**《LLM 学习季》讲座回放：**[掌握提示工程的艺术——🎨 就从这里开始](https://bytetech.info/articles/7434469180017295410) @Henry Li * [🐂🍺大模型替我上班吧！](https://bytetech.info/articles/7480729056535642127#OMsMdHyhgoBP4txmVhecA0ShnQd)@Hao Wang**🔥🔥🔥🔥🔥** * [提示工程的操控密码 —— 🤖 Computer-Using Agent](https://bytetech.info/articles/7477415635371180070#DZ9DdK6UrovcKDxSyuqcRtMenAe) @Tao He * [提示工程的炼丹密码：🔥大洋彼岸的 Prompt 炼金术](https://bytetech.info/articles/7455159690763501605#BF17d77K2oBPvKx6X5Ecvakgnae) @Tao He * [提示工程的圣诞密码：🎄一份为你定制的AI礼](https://bytetech.info/articles/7450710966800826378) @Tao He@Henry Li@Hao Wang * [大模型视觉：3分钟让同学哭着换头像 🎨](https://bytetech.info/articles/7449723988248690698#Lh10dgFWyoZk66xEnfncnG5SnPf) @Hao Wang * [🛠️ LLM Tokenizer实战 —— 从零开始的指北](https://bytetech.info/articles/7449319519564742694) @Tao He * [高效编程：🚀 经济适用的 Cursor 预测性补全实现](https://bytetech.info/articles/7447831248942301196#Q6hMd46aPoJFasxVH7ycqR6vnne) @Zihang Tang * [📘 搭建LLM学习环境 —— 从零开始的指北](https://bytetech.info/articles/7446999180695830562) @Tao He * [提示工程的创作密码 —— ✍️ 你还在为写作焦虑吗？](https://bytetech.info/articles/7446675288068784169#NKpbdBlP8oMZ5Dxa0FlcFb7lnlb)@Tao He * [提示工程的服务密码 —— 🔑 你还在手动解析SQL吗？](https://bytetech.info/articles/7437707717705203747)@Tao He * [提示工程话题营销密码 ——📣从零打造热门话题营销](https://bytetech.info/articles/7438454141543776294) @Henry Li * [提示工程的零风险密码 —— 🔐 你还在因为模型输出胆战心惊吗？](https://bytetech.info/articles/7434822008183423013#FzrSdFQQ3oQslFxJ5LEcujpGnkh)@Tao He * [提示工程的流量密码 —— 🇺🇸 Election Daily 实战案例](https://bytetech.info/articles/7433684464514891816?fromIdType=50&fromId=7430813927372128297) @Tao He * [提示工程的安全密码 —— 🏆AI安全大赛冠军亲自揭秘](https://bytetech.info/articles/7433294107184726051) @Hao Wang * [提示工程的高效密码 —— 💡你还在绞尽脑汁寻找灵感的突破口吗？](https://bytetech.info/articles/7431817642131849268#Qqm5dwtRfoGyAPxzD2dcligXnid)@Tao He * [提示工程的财富密码 —— 💰你还在手写 Prompt 吗？](https://bytetech.info/articles/7431770179308093449)@Henry Li * [用 ChatGPT 开发一个能听懂人话的命令行工具](https://bytetech.info/articles/7207752687642181693)@Henry Li * [**LLM Space**](https://bytedance.larkoffice.com/docx/TfdKdvL2oomQEpxKSWucs96knpe)—— **🛠️** 我们开发了一款非常方便好用的 Prompt 在线开发、调试和管理工具。 |   Synced from: <https://bytedance.larkoffice.com/docx/QWGjdcYYyoGmbTxUf0UcR6Ltn3e#IaKedPSGasWCCnbNqAAcB4z3nub> |

|  |
| --- |
| 如果你对上面这个系列或大语言模型落地感兴趣的话，[欢迎在 ByteTech 上关注我哟](https://bytetech.info/user/0b9f6ed9bae5d264f40856a1bb84cbbb)~ |

**本课程联合 ByteTech 以及字节共成读书社，在 3 月 7 日下午 15:00 进行了飞书直播。**

[**点击这里**](https://bytetech.info/activity/7469749321522036748)**查看录播**

|  |  |
| --- | --- |
| [**点击这里**](https://bytetech.info/activity/7469749321522036748)**查看录播**  **同时欢迎大家在**[**ByteTech 热门话题**](https://bytetech.info/topic/7469749321522102284) **中讨论** | **[This type of content is unavailable for download at the moment.]**  欢迎加入我们 3500 多人的分享群进行互动，获取讲课材料。 |

1. **引言**

|  |  |
| --- | --- |
| **Deep Research 是由 OpenAI 于 2025 年 2 月 2 日正式推出的新型多智能体 AI 研究工具，旨在通过多步骤信息检索、深度分析和综合推理，自动化完成复杂的研究任务。2 月 26 日起，已向所有付费用户开放该功能，笔者也在第一时间体验了该功能。**   * Hugging Face 在 24 小时内复现了 Deep Research 的核心功能，开源版本准确率达 55%（原版 67%），重点优化了 Agent 框架，降低运行成本。 * 埃隆·马斯克的 **Grok** 亦紧随其后，基于 Grok 3 推出了**深度搜索（Deep Search）**。与 OpenAI 的 Deep Research 形态相较，其速度更快（达 3 - 5 倍），效果亦更佳。此外，在 Github 上陆续涌现出众多 Deep Research 的复刻版本，并且迅速在开源领域受到了热烈追捧，包括： * https://github.com/dzhng/deep-research（13.6k stars） * https://github.com/nickscamara/open-deep-research (4.6k stars) * 在不到一个月的时间里，业界普遍认为 **Deep Research 应用形态**是**深度推理模型**（如 OpenAI 的 O 系列及 DeepSeek R 系列模型）的**最佳应用场景之一**，相信 DeepSeek 也会在不久的将来推出自己的深度探索版本。 * **3月6日爆火的** [**Manus**](https://manus.im/share/hRGARE7EBqtDhLHBGISLP7?replay=1)，从本质上来说，也是 Deep Research 的衍生形态。  |  | | --- | | **另一个场景则是** [**UI Agent**](https://www.google.com/search?q=llm+UI+Agent&num=10&newwindow=1&sca_esv=53a1189ee8f55c43&sxsrf=AHTn8zq7xRB-Ut3A5k1S_YWCEv65zmFBuA%3A1740968518825&ei=RhLFZ8uEMoDO1e8PiLuB8AQ&ved=0ahUKEwjLv4SR7eyLAxUAZ_UHHYhdAE4Q4dUDCA8&uact=5&oq=llm+UI+Agent&gs_lp=Egxnd3Mtd2l6LXNlcnAiDGxsbSBVSSBBZ2VudDIIEAAYBxgIGB4yCBAAGAcYCBgeMggQABgHGAgYHjIGEAAYCBgeMgYQABgIGB4yCBAAGAcYCBgeMggQABgHGAgYHjIIEAAYBxgIGB4yBhAAGAUYHjIGEAAYCBgeSKwHUABYxgZwAHgAkAEAmAHVAaABlwSqAQUwLjMuMbgBA8gBAPgBAZgCAqAC4AHCAgoQABiABBhDGIoFwgIGEAAYBxgewgIIEAAYBxgKGB6YAwCSBwMwLjKgB5oO&sclient=gws-wiz-serp)**，@Tao He已在**[**《提示工程的代打密码 —— 🤖 Computer-Using Agent》**](https://bytetech.info/articles/7477415635371180070#OQiedb5qYoN1krxLhkec7z68nrc)**一文中讨论如何复刻。同时，欢迎大家在文章尾部加入我们的社群一起讨论。** | |

|  |  |
| --- | --- |
| **我想了解如何复刻（匿名投票）** |  |
| Multiple Anonymous |  |
| A | UI Agent（AI 控制电脑） |
| B | Manus |

2. **演示**

接下来，让我们用几个视频，演示一下 Deep Research：

2.1 **OpenAI 的 Deep Research**

**[汤包与小笼包.mov]**

|  |
| --- |
| 这是一个长达 5 分钟的执行过程，据说最长的 Deep Research 可以达到半小时。  你也可以观看 [OpenAI 的官方 Deep Research 介绍](https://www.youtube.com/watch?v=YkCDVn3_wiw&ab_channel=OpenAI)。 |

2.2 **Grok 3 的 Deep Search**

|  |  |
| --- | --- |
| **[why-gpu.mov]** | **[汤包.mov]** |

|  |
| --- |
| 埃隆·马斯克的 Grok 3 性能无敌、财力无敌，因此响应速度也远超 OpenAI 的 Deep Research。  此外，Grok 3 的 UX 设计也明显好于 OpenAI，不愧是 X 和 Tesla 的设计狮。 |

2.3 **本方案最终效果**

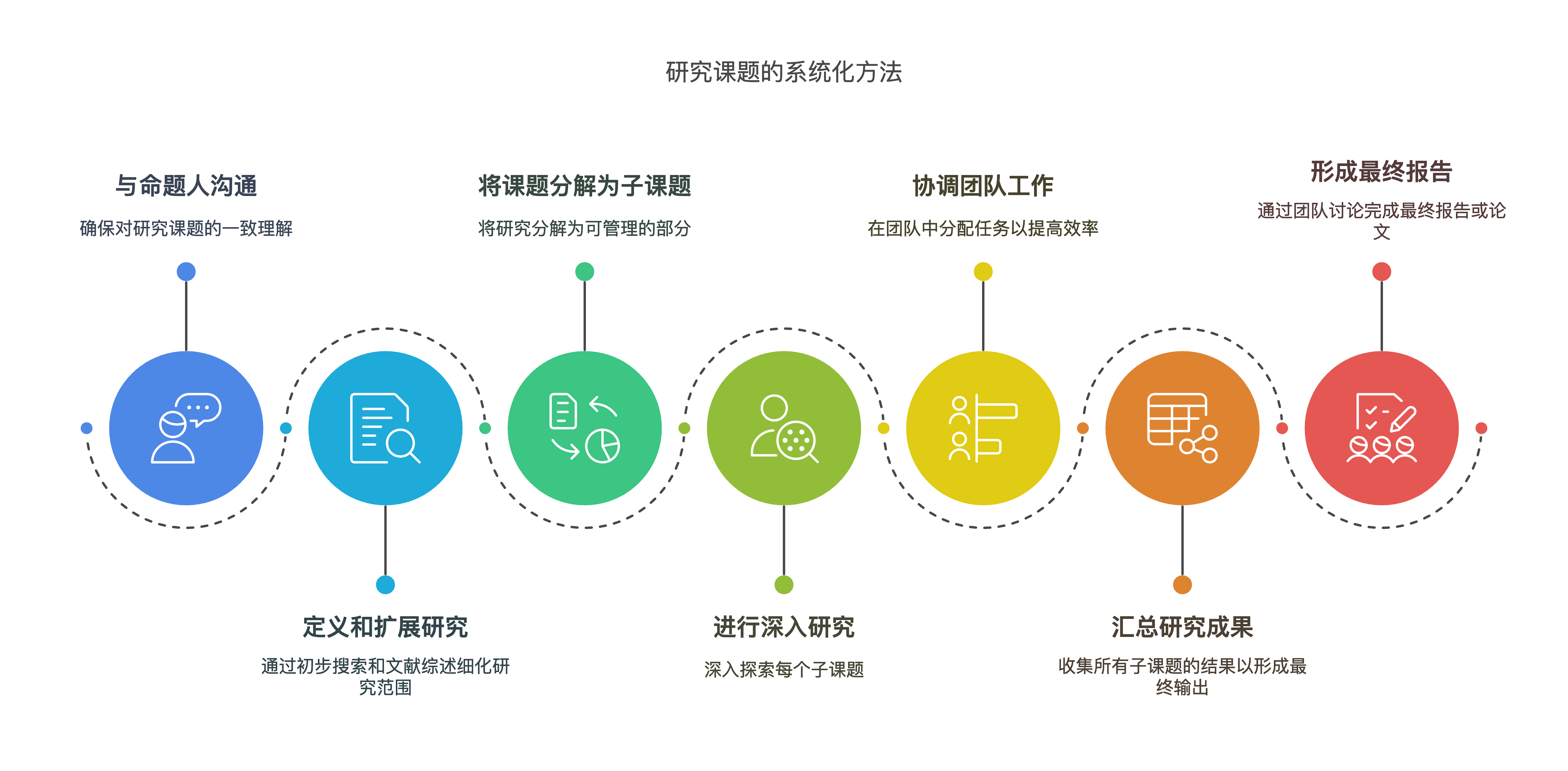
|  |  |  |
| --- | --- | --- |
| **用户输入的原始问题**  摘选自知乎真题 | **基于 Coze + 豆包 + DeepSeek 生成** |  |
| **飞书报告** | **运行时长** |
| 如何看待字节突破400万估值 | [如何看待字节跳动估值突破4000亿美元](https://bytedance.larkoffice.com/docx/YJ3bdHS7uo2XOnxNmzTcJ5L3nzf) | 7分28秒 |
| 什么是务实浪漫 | [务实浪漫的多维度剖析](https://bytedance.larkoffice.com/docx/EkiidPzOmoQxTAxyLZSc0IWOn5M) |  |
| 南京汤包与上海小笼包的区别 | [南京汤包和上海小笼包的全方位差异研究](https://bytedance.larkoffice.com/docx/XwUhdqsfIoszbhxHnCochmEVnTc) | 6分9秒 |
| 日本的「和牛」有哪些知名的地区品牌？他们有什么特点和区别？ | [日本和牛品牌核心特征、烹饪方式及饲养环境探究](https://bytedance.larkoffice.com/docx/Mg6tdJRI3oVcSPxR1jxcGzNjntg) | 6分28秒 |
| 如何看待 Manus   |  | | --- | | Manus 是3月6日凌晨发布的，这篇文章是3月6日早上 9 点生成的 | | [Manus技术架构、性能优势、行业影响及安全机制分析](https://bytedance.larkoffice.com/docx/HfDZdUkWOoTWsbxI1NacCVVonX0) | 6分2秒 |

3. **思考**

3.1 **为什么要复刻？**

|  |
| --- |
| **为什么我们要专门复刻 Deep Research 呢？**   * **研究与学习目的**：我们期望通过复刻 Deep Research 来掌握推理模型的最新应用方法，从而紧跟业界的发展趋势。 在这过程中，我们还将深度使用 Coze 工作流类型的应用搭建。 * **不再受限使用**：无论是 OpenAI 付费版用户，亦或是财大气粗的 Grok 3，均在 Deep Research 功能上施加了限制，每 天可使用的次数有限。而通过 Coze 提供的豆包、DeepSeek V3 以及 R1，我们能够近乎无限制地使用自主研发的 Deep Research。 * **更懂自己的深度研究**：自己动手丰衣足食——做更适合自己的 Deep Research。上述商业产品几乎不支持任何的偏好设置，且不可随意更换搜索引擎。你可以定制自己的搜索偏好（比如更喜欢搜索小红书和抖音上的内容），或者直接将结果输出为在线的飞书文档。 * **当豆包推出推理模型时，我们可以第一时间用它代替本方案中的 DeepSeek R1。** |

3.2 **我们是如何研究新课题的**



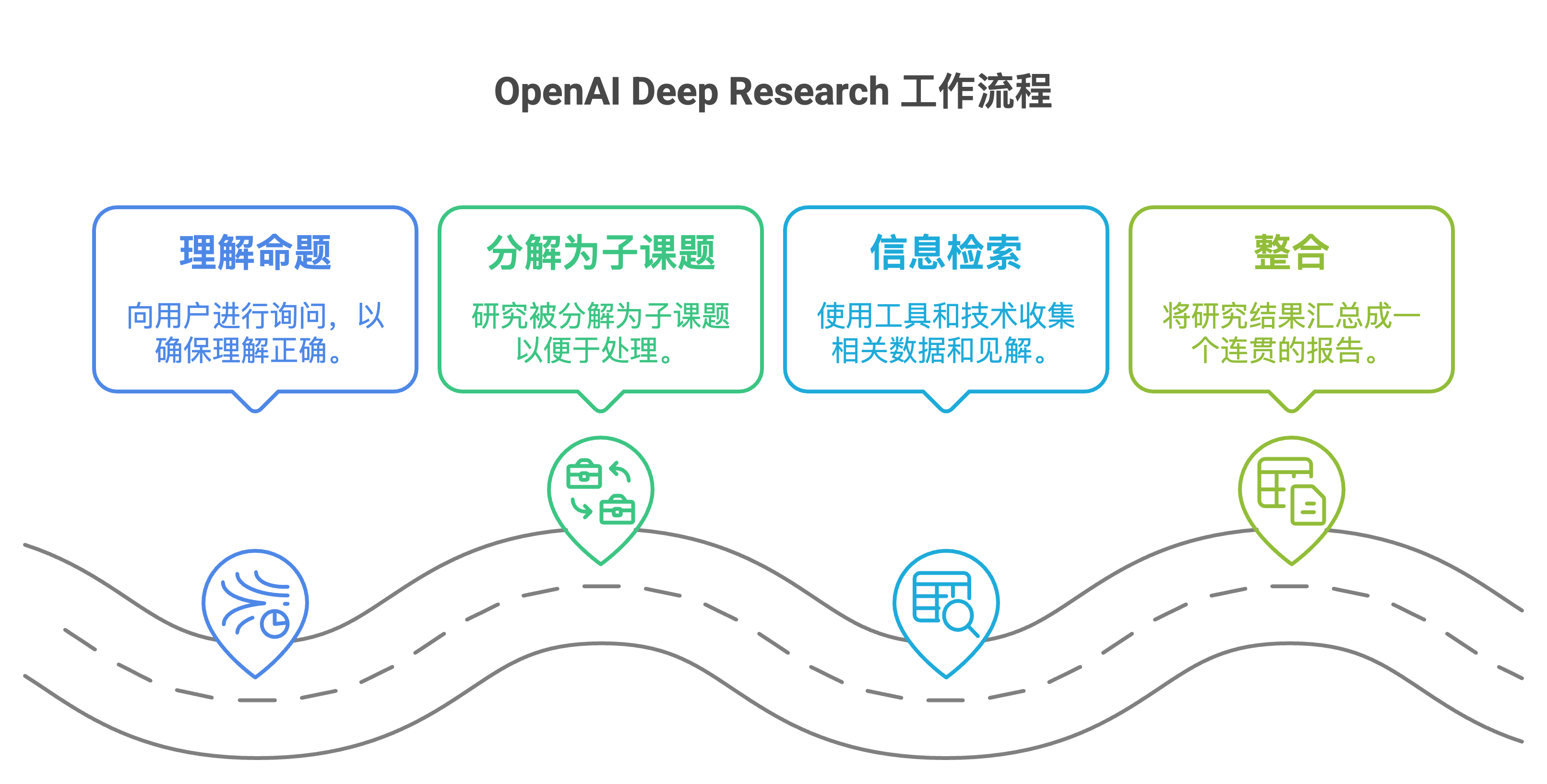
在学习和工作中，当我们面对一个全新的研究课题时，尤其是当这个课题仅仅以“一句话命题”的形式呈现时，我们通常会采取一套系统化的流程来逐步推进研究：

1. 首先，我们会与**命题人**进行充分的沟通，**确保对命题的理解一致**。这一步至关重要，因为它能够帮助我们明确研究的核心方向，避免后续工作因误解而偏离主题。
2. 接着，我们会通过初步的网络搜索和文献查阅，对课题进行定义和小范围的延展，并再次与命题人确认，以进一步细化研究范围和目标。
3. 在明确课题后，我们通常会将其分解为若干子课题，以便逐一攻克。

* 这些子课题可能是对核心问题的深入探讨（扩展深度），也可能是从不同维度展开的全新视角（扩展广度）。
* 在研究过程中，我们需要反复进行信息检索、阅读文献、摘录关键信息，并对其进行二次理解，直到对每个子课题有了透彻的认识，最终形成子课题的结题报告。
* 如果研究团队中有多名成员，我们还会进行合理分工，采用并行的方式推进不同子课题的研究。

1. 最后，我们将各个子课题的研究成果汇总，通过集体研讨形成最终的报告或论文。

3.3 **Deep Research 是如何研究新课题的**



在让我们将视线拉回到 OpenAI 的 Deep Research。当 Deep Research 面对一个新命题时，其工作流程与上述人类研究者的逻辑非常相似，但其自动化和智能化的特性使得这一过程更加高效和系统化：

1. 首先，Deep Research 会对用户输入的命题进行解析，确保对命题的核心主题、研究范围以及特定要求有准确的理解。这一步类似于我们与命题人确认理解一致的过程，但 Deep Research 的优势在于它能够快速提取关键信息，并自动生成初步的研究框架。OpenAI Deep Research 的做法是与用户进行一轮或多轮的沟通，从而确定命题，而 Grok 的 Deep Search 的做法则是跳过这一步，完全按照自己的理解去研究（因此，会带来不少不确定性问题）。
2. 接下来，Deep Research 会将命题分解为若干子课题，这一过程与我们手动“break down”课题的方式相似。它会根据命题的复杂性和多维度特性，自动生成一系列子问题，既包括对核心主题的深入挖掘（Expand Depth），也包括从不同视角切入的探索（Expand Breadth）。在这一阶段，Deep Research 的智能代理框架发挥了关键作用，像一个高效的项目经理一样，规划任务的优先级和执行顺序，确保研究过程有条不紊，这里使用到的是推理模型强大的规划和编排能力（Planning and Orchestration）。
3. 在研究子课题的过程中，Deep Research 会自动进行信息检索和数据提取，这与我们手动搜索文献、摘录和理解的过程类似。它通过调用搜索引擎 API 和网页抓取工具，快速定位相关资源，并从中提取有价值的信息。对于复杂数据或非文本内容（如图表、PDF 文件），它还能够利用 OCR 技术或数据分析工具进行处理。这一阶段，Deep Research 并行处理能力尤为突出，能够同时搜索、浏览多个网页，模拟多人分工的研究模式。有关大模型如何使用工具，以及智能体的相关内容，请参考我此前的讲座《[提示工程的艺术🎨工具与智能体](https://bytetech.info/articles/7444482560496762889)》。
4. 在完成子课题的研究后，Deep Research 会将各个子课题的分析结果整合起来，形成一份连贯的最终报告。这一过程类似于我们将各自的研究成果汇总、研讨并撰写论文的阶段。不同的是，Deep Research 能够通过其强大的语言生成能力，自动组织内容，确保报告的逻辑性和结构清晰度，同时附上参考来源，提升报告的专业性和可信度。这一步使用到的是推理模型强大的语言组织和超长的上下文理解能力。

3.4 **本方案的选择**

由于国情及篇幅限制的关系，本文中的方案将在上一节介绍的 Deep Research 基础上，做以下修改：

1. 使用 Bing 搜索作为搜索引擎，并且支持百度及小红书。
2. 本案将使用到 DeepSeek V3、R1 和 doubao-1.5-pro-256k，其中 R1 是大家熟知的“深度思考”推理模型。
3. 为了更好的理解用户的命题，本案在第一步中就引入了搜索引擎来帮助 AI 理解，在第五章的“问题改写与澄清”一节中会做详细的说明。
4. 由于 DeepSeek V3 不支持图片理解，因此在本案中放弃了对图片的识别和理解，有兴趣的读者可以自行通过 GPT-4o、QWen 等模型实现。

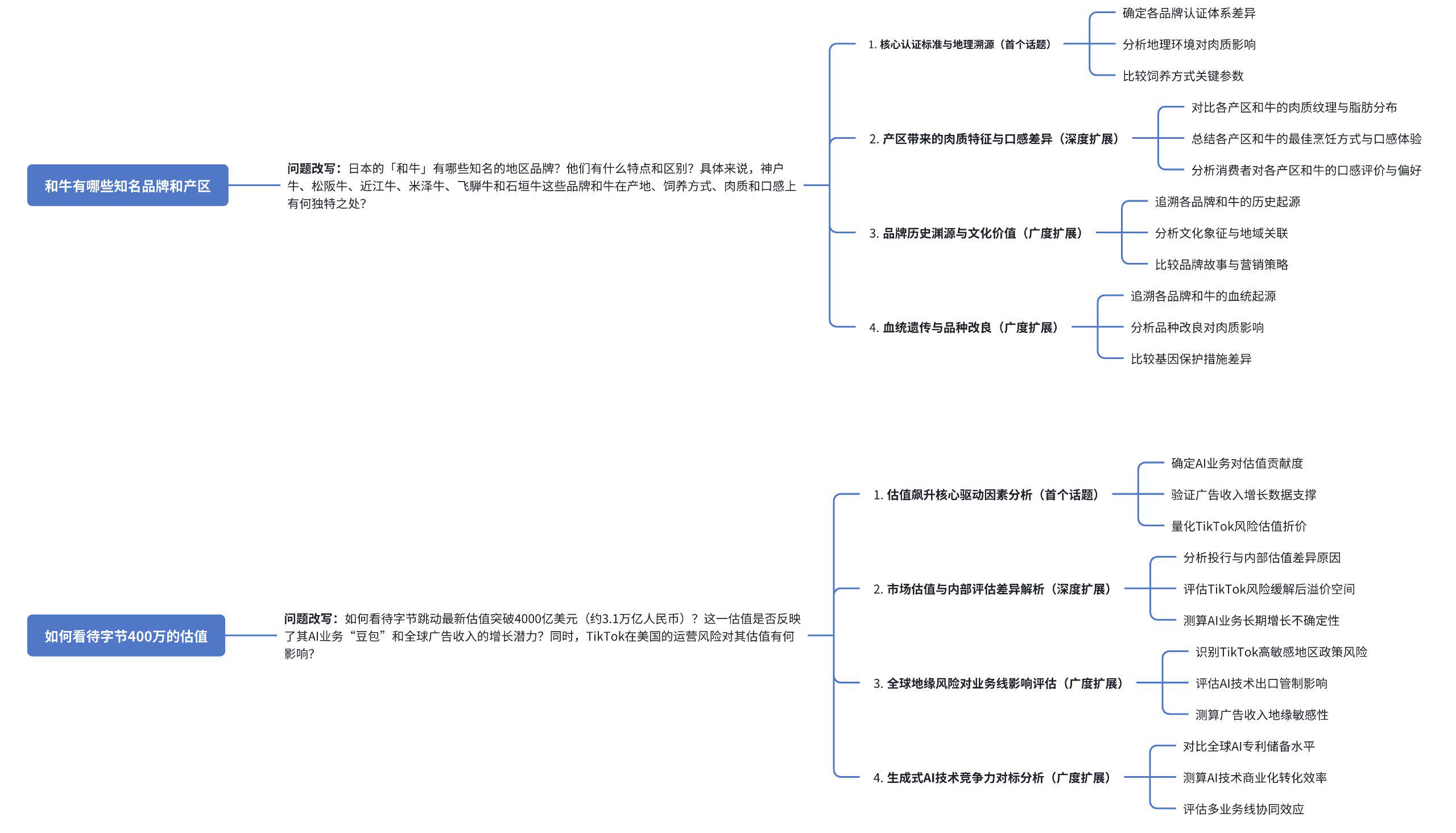
3.5 **本方案的创新点**

|  |  |
| --- | --- |
| 🔍 **利用搜索结果进行问题改写**  问题命题的好坏往往直接决定研究的成败。与原版的 Deep Research 直接反问用户不同，本方案会先在搜索引擎中寻找线索，再对问题进行改写。 | **🤿 深度和广度的下潜**  本方案提出了“主课题 > 子课题 > 研究方向”这样一个三级话题分层，并通过 Step-by-Step 的方式迭代探索，不做答案预设，从而保证了报告具备一定的深度和广度。 |

|  |  |
| --- | --- |
| ✍️ **利用摘抄避免“幻觉”**  “幻觉”问题一直是困扰大家的难题。本方案通过要求大模型进行“摘抄”后再给出结论，启发大模型根据事实依据进行生成，从而降低“幻觉”的发生几率。 | 🎨 **支持输出多媒体内容**  得益于 Coze 的海量插件，本方案支持同时以飞书文档、幻灯片、长图、博客等形式输出最终的报告内容。 |

4. **概要设计**

首先，我们需要将 Deep Research 中的关键业务进行模型进行定义。即用业务实体来描述下面的结构和过程：



4.1 **概念模型设计**

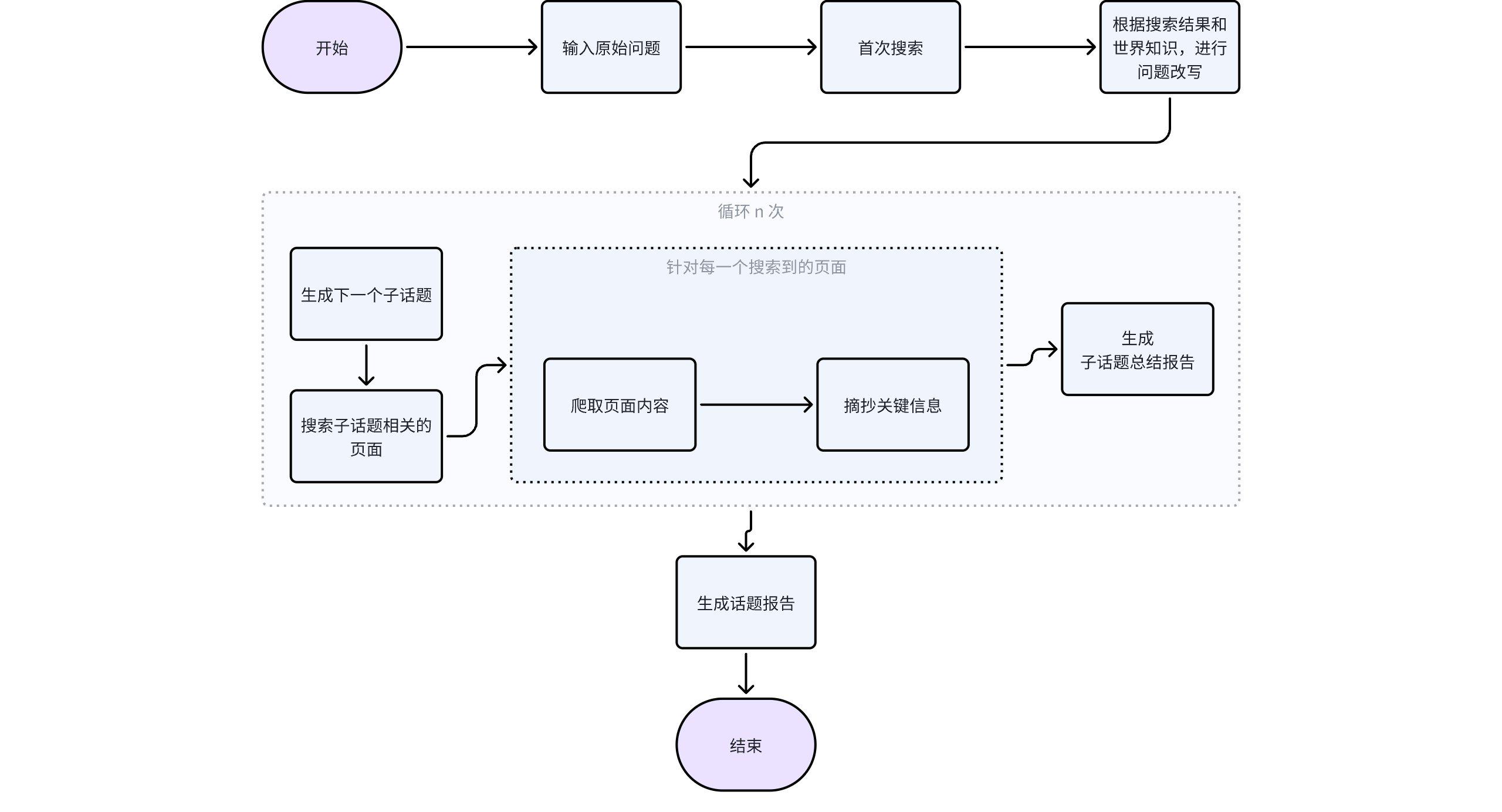
|  |  |
| --- | --- |
| **[This type of content is unavailable for download at the moment.]** |  |

* **Major Topic**：表示待研究的命题/话题的对象，属性包括：
* **User Initial Input：一个字符串，表示用户的原始输入的的问题，这也是 Deep Research 的起点。**
* 例如：**“字节估值超过400万了，你怎么看？”**。
* **title**：一个字符串，表示经过 Deep Research 理解后，改写并与用户确认后的命题，被称为主话题。
* 例如：**“如何看待字节跳动在2025年2月估值突破4000亿美元，并成为全球AI领域的领军企业？其AI平台“豆包”的爆发式增长和7500万活跃用户如何推动这一估值？”**就是主话题。
* 大模型通过搜索后，首先发现“400万估值”可能是用户的手误，并做了修正。其次根据搜索结果，在 AI 方向上做了延展。
* **subTopics**: 一个 SubTopic 对象数组，表示该话题的子话题列表。
* **Sub-Topic**：表示子话题的对象，一个主话题（也就是 Major Topic）会被拆分为若干个相关的子话题，属性包括：
* **title**：子话题的标题，一句话说明该子话题的特征。
* **type**：子话题的类型，即研究方向，包括：
* **首个子话题（first-proposal）**：即 AI 根据 Major Topic 首次生成的、第一个子话题，应该具备开放性、不预设等特征。
* 例如：**“字节跳动估值突破4000亿的核心驱动力是什么？”**就是 first-proposal。
* **广度扩展子话题（expand-breadth）**：前文的子话题深度已经足够，可以展开全新的子话题，以扩展主话题的研究广度。
* 例如：**“字节跳动的AI产品在估值增长中的作用”**以及它的兄弟节点就都是 expand-breadth。
* **深度扩展子话题（expand-depth）**：继续沿着前文的子话题，在其基础上生成新的更深入的子话题。
* 例如：**“豆包1.5 Pro 模型的推出对行业的影响”**和**“豆包 app的持续火爆及其语音支持功能的市场反馈”**就是对**“字节跳动的AI产品在估值增长中的作用”**的深度下潜。
* **research\_goals**：子话题的研究目的（1 至 3个），每一个研究目的是一个 ResearchGoal 对象。研究目的其实可被看作为“子话题的子话题”。其属性包括：
* **title**：研究目的标题，用简短的文字说明该研究子方向。
* **search\_query**：与该研究目的对应的搜索关键词，它将被传送给搜索引擎。事实上现代 LLM 都能够生成 [Combining Search Terms](https://www.worldox.com/GX3Help/and,_or,_and_other_search_commands.htm)，利用“AND”、“OR”等逻辑符号、“site:xxx”等关键字组合出更复杂的搜索查询。
* **search\_result**：记录 search\_query 对应的搜索结果。
* **exceprts**：大模型从搜索结果页面爬取内容后，对关键结论进行摘抄。
* **insights**：大模型从搜索结果和摘抄等中获得洞察点。让大模型进行摘抄后再写洞察点，有助于避免“幻觉”，后文将会详细说明。
* **summary**：本子话题的结题报告。
* **conclusion**：该子话题的总结陈词，实际上是由大模型在看完以上全部内容后，总结的结果，后文将会详细说明。

以下是一个 MajorTopic 对象的 JSON 示例：

|  |
| --- |
| json json {  "userInitialInput": "字节估值超过400万了，你怎么看？",  "title": "如何看待字节跳动在2025年2月估值突破4000亿美元，并成为全球AI领域的领军企业？其AI平台“豆包”的爆发式增长和7500万活跃用户如何推动这一估值？",  "subTopics": [  {  "type": "first-proposal",  "title": "字节跳动估值突破4000亿的核心驱动力",  "research\_goals": [  {  "title": "字节跳动的AI技术创新对估值的影响",  "excerpts": [  "字节跳动推出了最新的AI模型豆包1.5 Pro，性能大幅提升。",  ...  ],  "insights": [  "豆包1.5 Pro的推出巩固了字节跳动在AI领域的技术领先地位。",  ...  ],  "summary": "字节跳动通过AI技术创新，特别是豆包1.5 Pro的推出，显著提升了其市场竞争力和估值。"  },  ...  ],  "conclusion": "字节跳动估值突破4000亿的核心驱动力在于其AI技术的持续创新和火山引擎的商业化成功。"  },  {  "type": "expand-depth",  "title": "字节跳动的AI应用生态对估值的影响",  ...  },  {  "type": "expand-breadth",  "title": "国际市场政策变化对字节跳动估值的影响",  ...  },  {  "type": "expand-depth",  "title": "字节跳动的内容生态对估值的贡献",  ...  },  {  "type": "expand-breadth",  "title": "字节跳动的多元化业务布局对估值的影响",  ...  }  ] } |

4.2 **流程设计**



**1. 开始**

* **描述**：用户启动 DeepResearch 系统，准备输入研究问题。

**2. 输入原始问题**

* **描述**：用户输入一个研究问题或提示。
* **示例**：
* 字节估值超过400万了，你怎么看？
* 马爸爸去开会了，但是没有讲话说明什么？

**3. 首次搜索**

* **描述**：系统根据用户输入的问题，进行初步的搜索。
* **作用**：让大模型获知外界最新的信息，而不只依赖训练中学到的知识。

**4. ⚛️ 根据搜索结果和世界知识，进行问题改写（常规模型）**

* **描述**：大模型结合首次搜索的结果和世界知识，对用户输入的问题进行优化或细化。
* **大模型起到的作用**：
* 确保问题更具针对性和研究价值，同时为后续的子话题生成奠定基础。
* 纠正用户表述中可能潜在的错误。
* 将“昨天”、“上周”等相对时间修改为绝对时间。
* 从搜索引擎中获知最新新闻事件，并为问题添加备注。
* 从搜索引擎中获知专有名词，并添加备注。
* 改写后的问题，对于后续的探索更加具有启发性。
* **注意**：在实际运用中，应该与用户再次进行二次话题确认，以免犯错。
* **示例**：
* 如何看待字节跳动在2025年2月估值突破4000亿美元，并成为全球AI领域的领军企业？其AI平台“豆包”的爆发式增长和7500万活跃用户如何推动这一估值？
* 马云在2025年2月17日民营企业座谈会上出席但没有发言，这一行为可能释放了什么信号？

**5. 循环生子话题（for 循环）**

* **描述**：Deep Research 将主问题分解为若干子话题，并对每个子话题进行深入研究。此过程会循环执行，直到所有子话题处理完毕。
* **作用**：通过分解问题，逐步深入研究，确保覆盖问题的各个方面。

**5.1 ⚛️ 生成下一个子话题（推理模型）**

* **描述**：由推理模型根据主问题和已处理的内容，生成一个新的子话题。
* **推理型起到的作用**：子话题的生成是研究的关键步骤，确保研究的**开放性**、**不预设**、**深度下潜**和**广度扩展**。

**5.2 搜索子话题相关的页面**

* **描述**：Deep Research 针对当前子话题，进行专门的搜索，获取相关的网页。
* **作用**：为子话题的研究提供数据来源。

**5.3 针对每一个搜索到的页面执行（for-each 循环）：**

**5.3.1 爬取页面内容**

* **描述**：借助 Jina 等第三方爬虫服务，从搜索到的每一个页面中提取内容，包括文本、表格等。
* **作用**：将非结构化数据转化为可分析的信息。

**5.3.2 ⚛️ 摘抄关键信息（常规模型）**

* **描述**：由大模型对爬取的内容进行分析，提取与子话题相关的关键信息。
* **大模型起到的作用**：本质是对网页的信息进行压缩和精炼，过滤无关内容，确保研究的高效性和准确性。

**5.4 ⚛️ 生成子话题总结报告（常规模型）**

* **描述**：大模型将以上提取的关键信息，整合成一份子话题的总结报告。
* **大模型起到的作用**：为每个子话题生成独立的研究成果，便于后续综合。

**6. ⚛️ 生成话题报告（常规模型或推理模型）**

* **描述**：在所有子话题处理完成后，系统将各子话题的总结报告整合成一份完整的研究报告。
* **大模型起到的作用**：提供一个全面、结构化的研究成果，直接呈现给用户。

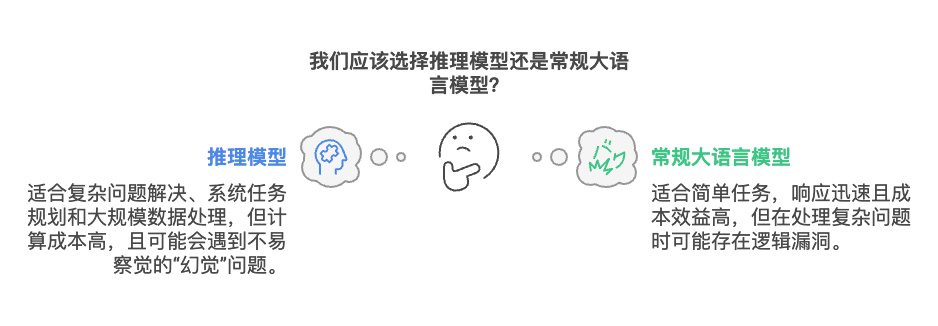
**7. 结束**

* **描述**：研究流程结束，用户获得最终的研究报告。
* **作用**：报告的格式为 Markdown，方便在后续流程中被转换为飞书文档、PDF 或 PPT（**YES，我们支持 PPT 报告生成**）。

|  |  |
| --- | --- |
| Button name before click | 我想学习如何生成 PPT |
| Button name after click | 如何生成 PPT （已投票） |
| Button color | BLUE |
| Icon | GOOD |
| Participants | @Yan Peng @Shenghao Chen @Xinwei Xiao @Xiang (Megan) Lai @Qitong Niu @Honglei Wang @Yingqing Zeng @Shun Li @Zuxian Yang @Shenwei Huang @Yaoying Wang @Jiapeng Kang @Qiang Hao @Kaiyuan Chen @Ximing Hu @Hongfeng Yang @Zaofeng Pan @Zhishi Zhu @Tao Liu @Feng Chen @Zhuorui Liu @Fanqi Meng @Mingjia Chen @Xiangyuan Li @Yunfeng Lu @Kaiwen Liang @Zhuorong Su @Steven Yu @Shiqiang Li @Jinjie He @Jiahong Li @Jingdong Chang @Guokai Jing @Junyu Wang @Jianheng Tang @Tao He @Bohan Zhang @Lang Geng @Liang Zhang @Chonghao Zhang @Yikai @Zhihui Hu @Yan Liu @Josh Liu @Ping Liu @Lantian Zhang @Heng Ai @Xinyu Huang @Chestnut Liu @Keyu Chen @Pei Chen @Ke Chen @Yanyan Ben @Yuxuan Liu @Weiying Sun @Yulin Wang @Tsukimoe Li @Weiyu Tong @Xuelin Lu @辰北江 @Pengbo Du @Xuanxian Chen @Wenzheng Qu @Hao Wang @Mengyun Liu @junjie zou @Xiang Ding @Hengfei Xi @Kaitian Xu @Hongyan Gao @Xu Zhou @Jianxing Yan @Dongwen Ji @Zhangzhi He @Tianxiong He @Taishuai Luo @Dongjia Shi @Daqian Lv @Liz @Zhigang Wang @Peiren Xu @Yunhui Liu @Fang Fang @Xianru Zhang @Zhongchao Wang @Yifan Wang @Yang Zhao @Miao Zhang @Chu Zhang @Na Liu @Zhiqiang Yuan @Cheng Chen @Enyong Zhang @Hangsheng Zhang @Minyi Lin @Ziqiang Zhang @Moe(Jiaqi)Zhao @Yuming Zhang @Zheng Zhang @Jiasheng Yu @Ziwei Chen @Lingjun Meng @Shiwei Chen @Wen Li @Yuan Liang @Mengjie Liu @Chang Sun @Kailun Chen @Yutong Wei @Shaozhong He @Qiming Guang @Yi Zhou @Chenfei Yao @Hongbiao Fang @Kai Wang @Wei Ding @Jiang Wang @Dale Bao @Songtao Wang @Zhenjie Jia @Yulin Zhu @Jansoning @Gang Bai @Dian Peng @Min Chen @Yu Miao @Xuan Zhu @Meng Guan @Congcong Xiong @Ru Meng @Isabel Chan @Pengjie Jiang @Jingyu Feng @Nan Zhang @Weitao Zhang @Xiaofeng Sun @Yexin Wang @Caiqiang Shi @Sean Yu @Chunhui Yang @Minwei Shen @Changhong Shu @Dong Guo @Xianpeng Xie @Jack Yang @Xiaodan Jiang @Jinrong Huang @Jiawang Li @Chi Wang @Mincong Chen @Boyce Li @Yuhao Wang @Runzhe Cao @Chunhui Jia @Kuangqi Zhou @Lei Wang @Yuxuan He @Xin Gu @Siming Chen @Yahui Shi @Mengji Gou @Zhijian Wu @Penghui Li @Charlie Chen @Yu Gu @Liang Song @Yuheng Li @Hao Lin @Ning Jiang @Yan Liu @Kroak @Tian Gan @Peiyu Chen @Meng Zhu @Tianwu Zhi @Yang Li @Henry Li @Bo Zhang @Feifei He @Rongyu Sun @Yufang Mao @Qiyuan Zhang @Dingqi Lin @Inver Lau @Jian Xiong @LuoWei @Lei Huang @Zewen Shen @Xiangyu Zhang @Verus Zhu @Xing Liu @Qing Luo @Wei Yu @Ruidong Guo @Lei Wang @Ruirui Zang @Shanghong Lv @Guangtian Liu @Xu Xu |

5. **难点与挑战**

5.1 **难点：推理模型**



随着 DeepSeek R1 在 2025 年蛇年春节前横空出世，推理模型在国内受到了空前的关注。

**推理模型**（如 OpenAI 的 o3 系列，以及 DeepSeek 的 R1）与**常规大语言模型**（如我司的 Doubao-1.5-pro，OpenAI 的 GPT-4o、DeepSeek 的 V3）在设计目标和应用场景上有显著差异：

|  |  |  |
| --- | --- | --- |
| ✨ **复杂问题的解决能力**  推理模型在处理复杂问题时表现出色，尤其是在需要逻辑推理、策略规划或跨领域知识整合的场景中。   * **应用场景**：学术研究、法律分析、财务建模等需要高精度推理的任务。 * **对比**：普通大语言模型在面对复杂问题时可能会出现逻辑漏洞或不一致的回答。 | **🗺️ 多步骤任务的规划能力**  推理模型能够将复杂任务分解为多个步骤，并为每一步提供清晰的执行计划。   * **应用场景**：项目管理、流程优化、任务分配等。 * **对比**：普通大语言模型通常缺乏这种系统性规划能力，可能会在多步骤任务中遗漏关键环节。 | 🤖 **处理大规模非结构化数据的能力**  推理模型在从大量非结构化数据中提取关键信息、发现隐藏关系方面表现突出。   * **应用场景**：从长文档中提取关键信息、跨文档推理、复杂数据集的模式识别等。 * **对比**：普通大语言模型在处理长文档或复杂数据时，可能会遗漏细节。 |

当然推理模型也不是万能的，它也有自身的**致命缺点**：

|  |  |  |
| --- | --- | --- |
| **⏰ 计算成本较高，思考时间过长**  推理模型通常需要更高的计算资源，推理时间也更长。   * **原因**：推理模型在处理复杂任务时会进行多层次的逻辑分析和推导，这增加了计算复杂度。 * **对比**：普通大语言模型在执行简单任务时更高效，响应速度更快，成本也更低。 | **💡 提示工程难度不小**  推理模型对输入提示的质量要求较高，提示需要清晰、具体，才能充分发挥其推理能力。   * **原因**：推理模型的设计目标是解决复杂问题，因此在面对模糊或不明确的提示时，可能会花费更多时间进行澄清或推导。 * **对比**：普通大语言模型对提示的敏感性较低，能够更宽容地处理模糊输入。 | **💭 “幻觉”现象较多**   * 这似乎是 R1 才有的现象。[DeepSeek-R1的幻觉率为 14.3%，远高于DeepSeek-V3的 3.9%。](https://github.com/vectara/hallucination-leaderboard) * GPT 系列的推理增强模型（GPT-o3）相比 DeepSeek-R1 更好地保持了事实一致性。 |

5.1.1 **性能与效果**

|  |  |
| --- | --- |
| 在上一章中流程的子标题中，你可能注意到带“⚛️”符号的表示该步骤需要大模型介入，而在括号中我们注明了模型的类型。作为一个消费级的产品，我们必须在“思考等待时间”和“生成效果”之间进行取舍。  对于**“生成下一个子话题”**这种对规划能力、理解能力较高的多层次逻辑推理任务，推理模型在处理复杂推理任务时表现更优，能够基于模糊信息和上下文关系生成具有逻辑性的子话题。 在最后一步**“生成话题报告”**时，则可以利用图例模型超强的上下文理解和复杂问题处理能力，因此也推荐使用推理模型。  而在其他几个场景中，都是通过[提示工程中的 CoT 及 few-shot 等技巧](https://bytetech.info/articles/7434469180017295410)，可以解决的问题，同时可以极大的减少响应和降低成本，因此选择普通模型即可。 |  |

5.1.2 **推理模型的 Prompt 技巧**

|  |
| --- |
| 在上一讲[《提示工程的艺术🎨就从这里开始》](https://bytetech.info/articles/7434469180017295410#Eni3dwZjxopWbAxUESAc2Iupn8g)中，我们通过大量的示例，向同学们介绍了思维链（CoT）、few-shot（多样本提示）、检索生成增强（RAG）、工具的运用等方法。然而这些方法在以 DeepSeek R1 为代表的推理模型中并不一定完全适用。 |

推理模型（如 DeepSeek R1 或 OpenAI O3）与传统生成式模型的核心区别在于其内置的逻辑推理能力。DeepSeek R1 专注于多步推理和复杂问题的分析，能够在**不需要显式提示的情况下自动进行链式推理**（chain-of-thought reasoning）。这使得它在处理复杂任务（如学术研究、法律分析、数学推导或战略规划）时表现出色，但在简单任务上可能会显得**“过度思考”**。

相比之下，像豆包、DeepSeek V3、GPT-4o 这样的模型更倾向于直接生成答案，除非通过 prompt 明确要求其逐步推理。这种差异决定了我们在设计 prompt 时需要采取不同的策略。

为此，笔者参考社区并且结合最具 1、2 个月的实践经验，总结出了一份**独家《Reasoning Model Prompt Cheatsheet》**，送给大家**：**

|  |  |  |
| --- | --- | --- |
| **Reasoning Model Prompt Cheatsheet** |  |  |
| **技巧** | **说明** | **示例** |
| 保持问题清晰简洁 | 推理能力使其能够自动分析问题并生成详细的答案。因此，过于复杂或冗长的 prompt 可能会干扰模型的推理过程。最佳实践是直接陈述问题或任务，避免不必要的修饰。 | * **不推荐：**“Let's think step by step.” * **不推荐**：“请仔细分析以下问题，并逐步推导出答案。我们需要一个详细的解决方案，确保每一步都清晰。” * **推荐**：“解决以下问题：[问题描述]。” |
| **提供必要的背景信息** | 虽然 DeepSeek R1 在推理能力上表现优异，但其知识库可能对特定领域的最新信息或小众知识有限。因此，在设计 prompt 时，务必提供与任务相关的背景信息或上下文。 | * **推荐**：“根据以下法律条文：[法律条文]，分析案件中甲方是否违反合同。” * **推荐**：“根据搜索引擎返回的结果，重新改写用户的初始问题。搜索结果:[搜索结果]” |
| 避免过多示例 | 与传统生成式模型不同，推理模型不需要通过 few-shot 示例来学习任务格式。事实上，过多的示例可能会干扰模型的推理能力。最佳实践是采用零样本（zero-shot）或单一示例（one-shot）提示。 | * **不推荐**：提供多个示例来说明任务格式。 * **推荐**：直接描述任务，或在必要时提供一个简单的示例来说明需要生成数据的格式。 |
| 明确输出格式 | 推理模型能够很好地遵循格式指令，因此在 prompt 中明确指定输出格式可以提高结果的可读性和一致性。例如，可以在 Prompt 中使用 TypeScript、JSON Schema 或 one-shot 的方式来进行明确输出。 | * **推荐**：  |  | | --- | | Markdown ```ts interface Foo {  barId: string;  barName: string; } ``` 以 JSON 的格式输出 Foo[] 对象数组，数组中至少包含 3 个 Foo 对象。 直接输出 JSON，不要“```json”。 | |
| 控制详细程度 | DeepSeek R1 默认会生成详细的答案，但在某些场景下，我们可能需要更简洁的输出。在 prompt 中明确指示所需的详细程度可以有效控制输出。 | * **如果希望简洁输出**：“用一句话总结以下问题的答案。” * **如果希望详细分析**：“请详细分析以下问题，并逐步解释每个推理步骤。” |
| 延长思考时间 |  |  |
| 避免“幻觉” | 前文中提到了 DeepSeek R1 的幻觉问题，目前没有明确的方法可以彻底解决，这里推荐几个方法：   1. **Divide-and-conquer**：将原本足够复杂的问题，先拆解为更简单的、上下文更短的子问题，相对简单、上下文简短的任务更不容易出现幻觉。 2. **引用原文**：要求推理模型在最终的输出中，先复述上下文中的原文或原始数据（通常来自于 RAG），然后再输出洞见；或要求在每一个关键结论前，先给出出处。这样做可以在一定程度上，避免模型的幻觉。 3. **Self-Reflection**：即让另一个大语言模型智能体作为审查人，检查推理模型的输出是否严格遵照给定的事实数据，指出有问题的部分并尝试修正。 4. **借力思考**：当下常规大语言模型的幻觉相对来说较少，因此可以尝试用 R1 输出的思考过程（Reasoning Content）作为一条 message 提供给 V3 或豆包 Pro，然后再由这些常规模型输出最终的结果。请参考 [Deep Claude 项目](https://deepclaude.com/)。 5. **基模型升级**：这也是最有可能的解决方案，期待豆包和 DeepSeek R2 的发布。前文也提到过，“幻觉”不应该是推理模型的通病，相信对标 OpenAI O3，国产大模型会解决这一难题。 |  |

5.2 **难点：Full Plan 还是 ReAct**



现在我们已经知道推理模型具备强大的规划和编排能力，那么究竟是让大模型一次性输出所有步骤的规划（Full Plan），还是每次根据上一步的结果生成下一步的计划（ReAct）呢？这是一个经典的规划问题。不如我们来问一下 DeepSeek R1 自己是怎么回答的：

|  |
| --- |
| markdown 根据用户的需求和两种研究方式的特性，我会推荐\*\*选项2（ReAct 逐步探索式）\*\*，原因如下：  1. \*\*动态调整能力\*\*：许多小红书/知乎的热门话题（如AI伦理、职场新趋势）具有时效性和争议性，初步搜索可能发现新线索。例如若搜索"2023年轻人反内卷"时发现"数字游民"现象突然高频出现，可即时调整研究方向。  2. \*\*深度聚焦可能\*\*：当某个子课题的搜索结果远超预期（如发现某篇权威论文），我们可以延长该方向的探索，而不受预设框架限制。比如在研究"预制菜争议"时，若发现食品安全检测数据异常丰富，可增加技术检测维度的子课题。  3. \*\*用户参与感\*\*：类似知乎的「追问」机制，这种互动模式更符合小红书/知乎用户的交流习惯。例如当用户看到"三明治 generation"的搜索结果时，可能联想到自身经历，提出新的探索角度。  4. \*\*规避信息冗余\*\*：预先设定的5个方向可能存在重复，比如在研究"ChatGPT对教育的影响"时，若发现"AI监考系统"和"自适应学习"的搜索结果高度重叠，动态模式可以及时合并方向。  具体实施示例（假设话题是《为什么年轻人开始接受「断亲」现象？》）：  \*\*第一子课题\*\*  研究目标：量化断亲现象的程度  关键词：2023 断亲 统计数据 site:[gov.cn](http://gov.cn)  期待结果：民政部/统计局关于家庭结构变化的白皮书  → 用户反馈搜索结果：发现《2023中国家庭发展报告》显示18-35岁群体中，62%承认"减少非必要亲戚往来"  \*\*第二子课题\*\*  研究目标：经济因素分析  关键词：城市化 独生子女 亲戚网络 经济支持  期待结果：城市化率与亲属关系疏离的学术研究  → 用户提供论文摘要：发现城市化进程中"工具性关系"替代传统亲缘关系  \*\*第三子课题\*\*  （根据上文延伸）研究目标：数字社交替代效应  关键词：Z世代 微信群 亲戚 社交替代  ...  当完成所有子课题后，可以用「雪花写作法」整合：将每个搜索结论作为知识节点，通过语义网络分析找出核心关联，最终形成《断亲现象的三重解构：从数据真相到代际突围》的结构化长文。 |

|  |  |
| --- | --- |
| **你会选择 Full Plan 还是 ReAct 呢？** |  |
| Single option Non-anonymous |  |
| A | Full Plan，一次性输出所有的步骤，有条件的话，还可以并行去执行 |
| B | ReAct，更像是在不断摸索中逐步“打开地图”，对模型的并发要求更低 |

笔者的团队一直对 ReAct 思维框架抱有高度的认可和偏爱。鉴于 OpenAI 和 Grok 等主流机构也采用了 ReAct 框架，而仅有少数开源项目选择了 Full Plan 模式，最近火爆的 Manus 是 Full Plan 结合 Re-plan 的方案。我们在 DeepResearch Demo 项目中延续了这一传统。具体而言，我们的工作流程如下：

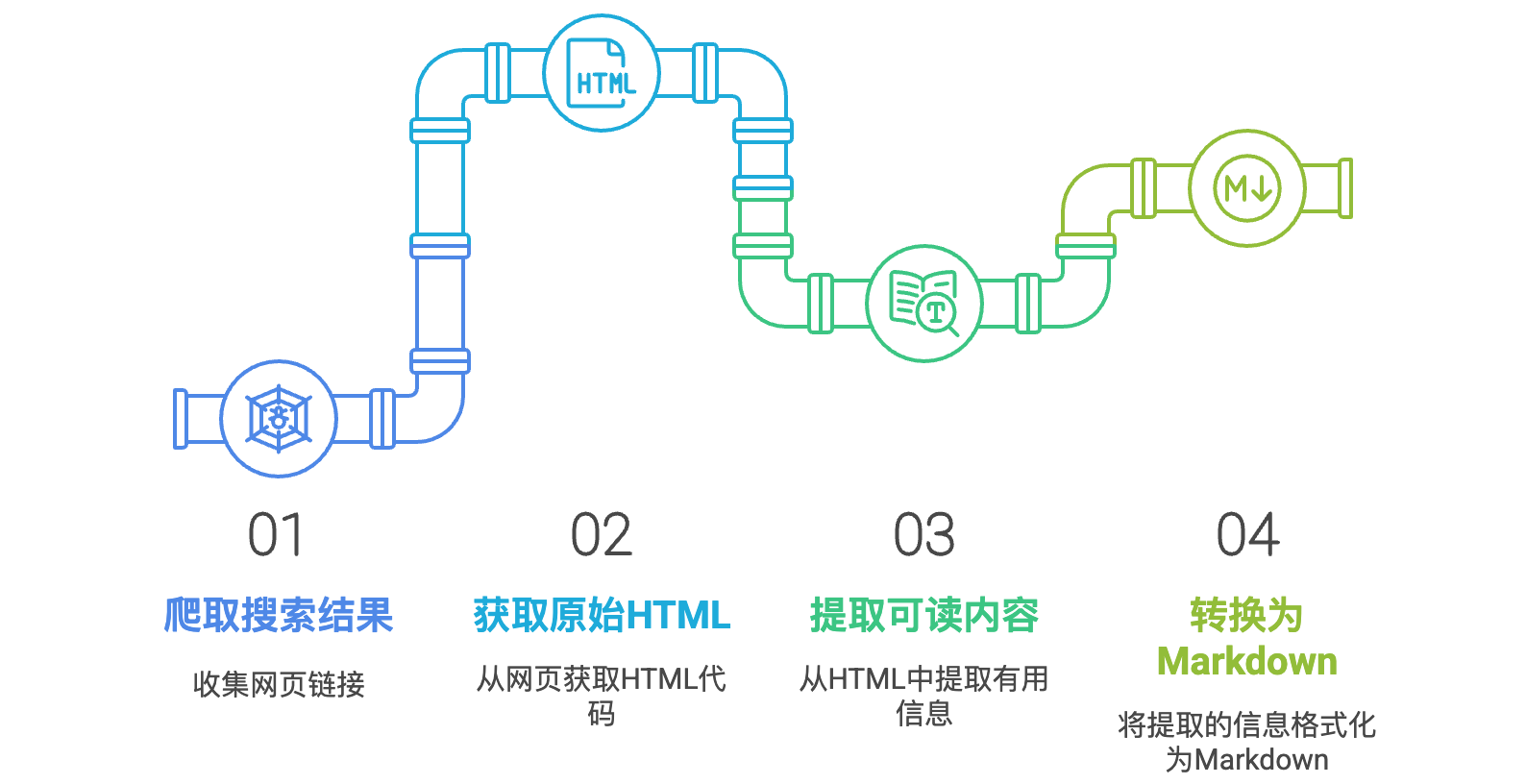
1. **逐步生成与探索**：首先，推理模型会基于主要主题（Major Topic）以及首次搜索结果，生成第一个子话题，并立即展开深入探索，直至该子话题研究完成。
2. **迭代式推进**：在后续的每一轮循环中，我们将主要主题以及此前生成的子话题列表（包括其研究结论）作为输入，提供给推理模型。模型据此生成下一个子话题，并自主决定是进一步扩展深度还是拓展广度。
3. **多轮循环与总结**：经过 n 次循环后，生成 n 个子话题，最终由推理模型智能体对全篇内容进行总结和归纳。

通过采用 ReAct 框架，我们的 DeepResearch 项目能够在探索未知领域时保持高效、灵活，同时最大限度地激发模型的创新潜力。

|  |
| --- |
| * **并行优化**：尽管子话题的生成过程是串行的，但在每个子话题的研究阶段，我们尽可能采用异步并行的方法，以缩短整体执行时间。 * **推理模型资源限制：**此外，由于推理模型的计算成本较高且响应时间较长，模型提供商通常不支持大规模并发调用。通过串行生成子话题的方式，我们也能有效规避大并发带来的资源压力。 * **ReAct 框架的优势**：ReAct 思维框架特别适合处理创新型或开放性问题。与一次性输出完整计划的方式不同，ReAct 通过逐步生成和动态调整，避免了对话题走向或答案的过早假设，从而展现出更强的创造性和灵活性。 |

5.3 **挑战：爬取页面并转化为易于大模型理解的内容**

搜索引擎或第三方 SERP 提供商为开发者提供了 API 调用，从而轻松获得搜索的结构化结果。但搜索结果中的摘要信息（snippet）是不足以满足话题研究的需求的。

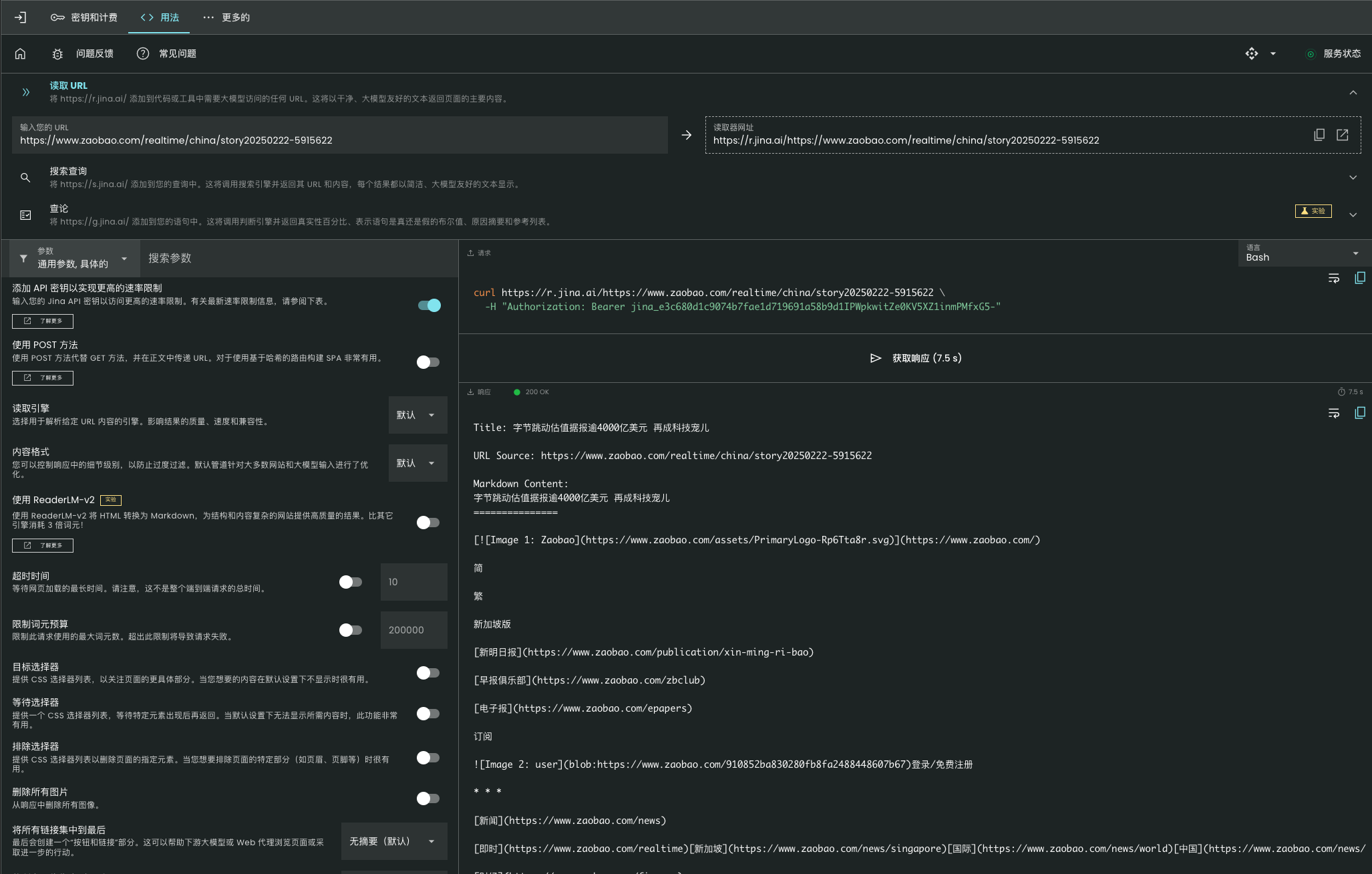


因此，如上图所示，我们还是需要通过爬虫爬取搜索结果对应原网页的 HTML，并将 HTML 中的有效信息转换为大语言模型更好理解的 Markdown 格式。在此过程中，**“步骤2——获取原始 HTML ”**看似简单，实则不然：

1. 首先，当下绝大多数网页都是采用 AJAX 客户端渲染为主的网页，用类似“curl”的方式抓包显然走不通，因此必须在服务端为每一个网页启动一个无头浏览器（Headless Browser，如 Puppeteer、PlayWright）上下文。
2. 其次，你需要绕过各种防爬工具和安全措施。
3. 如果没有专门做爬虫的服务，很难做到高并发、IP 地址分散、n 级缓存等。

市面上有很多开源和商业公司专门提供云端的爬虫服务，如著名的 Jina、Firecrawl、Crawl4AI 等。

|  |
| --- |
| **本案技术选型的是 Jina**  [**Jina**](https://jina.ai/)专注于构建搜索、推荐和生成式 AI 应用，尤其是多模态（文本、图像、音频、视频等）数据的处理。它提供了高效的索引、查询和生成能力，适合需要复杂 AI 模型支持的场景。 |



* Jina 诞生自大语言模型火爆之后，因此是一款专门针对大语言模型优化的云端爬虫。
* Jina 提供了免 APP\_KEY 的即开即用方式，只需将网址拼接在“https://r.jina.ai”之后就可以得到将 HTML 转换为 Markdown 的结果。当然 Jina 也提供了丰富的选项，你可以在官网上通过调试器在线查看选项并测试。

|  |
| --- |
| Bash 爬取联合早报的新闻报道，并转换为 Markdown **curl** https://r.jina.ai/https://www.zaobao.com/realtime/china/story20250222-5915622 \  -H "Accept: application/json" \  -H "X-Return-Format: markdown" \  -H "X-Timeout: 10" |

美中不足的是 Jina 对中文网站（特别是新浪、搜狐、知乎这样的 PC 端网页）支持的不是特别好，会将无关紧要的内容（如导航栏、底部栏）也一并返回，因此我们要在 Jina 的基础上做一些定制开发。

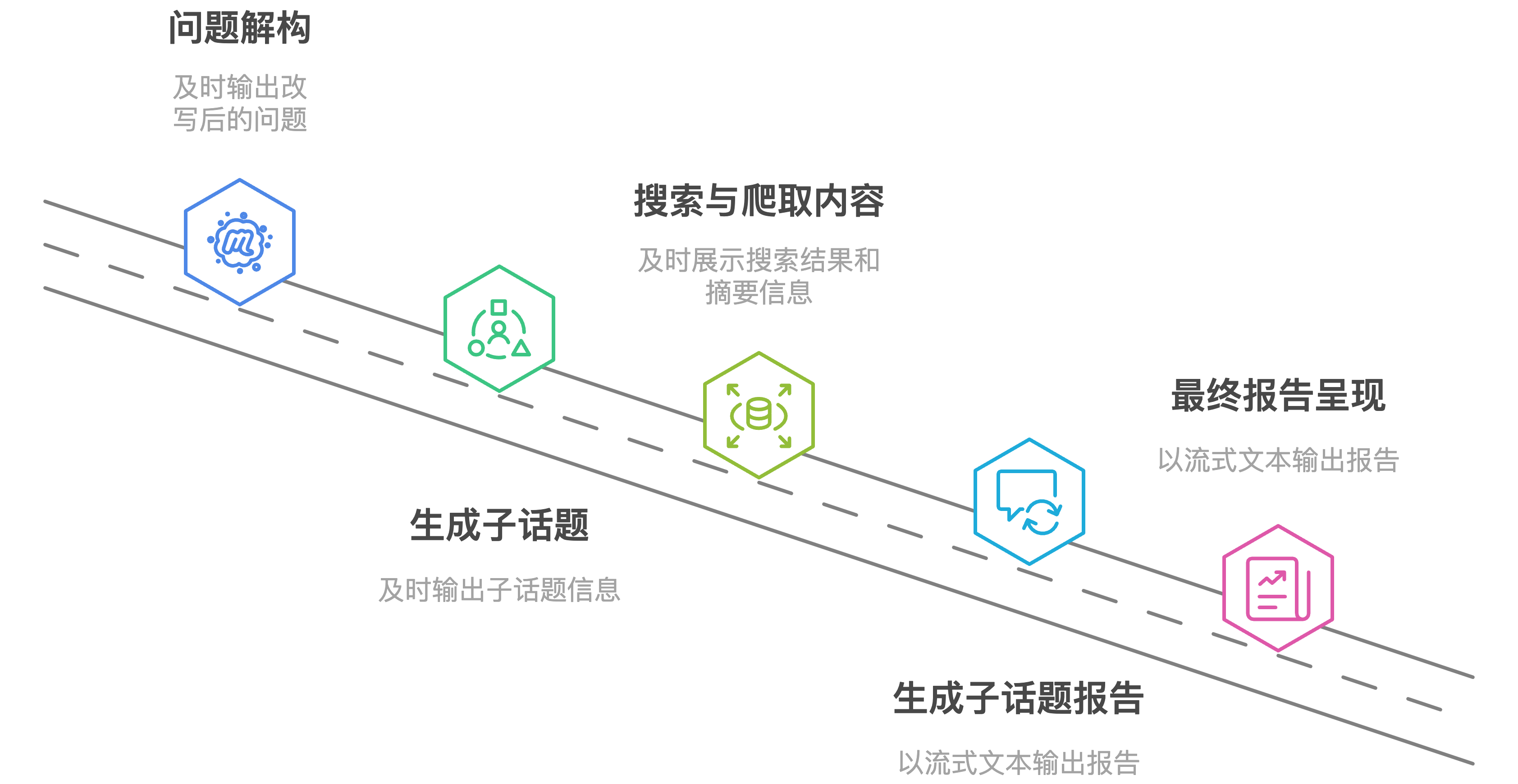
|  |
| --- |
| **更高效的 Coze 插件**  笔者对 Jina 进行了优化和改进，并 [封装成了 Coze 插件](https://bots.bytedance.net/space/7317628956541321242/plugin/7476786676048904219)，你可以在 Coze “ByteDance Demo”租户下搜索并使用到它，我们将在第 6 章中详细阐述。  在实验环境中，笔者优化后的版本返回的 Markdown 平均只有 Jina 原来 token 量的 10%！  还不赶紧[在 ByteTech 上关注我](https://bytetech.info/user/0b9f6ed9bae5d264f40856a1bb84cbbb)~ |

5.4 **挑战：执行过程中持续输出中间结果**

Deep Research 与传统“深度思考+联网搜索”模式存在本质性差异。Deep Research 采用分层递归式处理机制对话题进行广度和深度的延展，而相较之下，“深度思考+联网搜索”模式仅执行单次搜索请求便开始生成答案。

这种架构差异直接导致处理时长的显著区别。Deep Research 的完整处理周期通常介于 2 至 30 分钟，其耗时主要消耗在三个关键环节：问题解构的语义分析（约15%）、多轮次智能体交互（约60%）以及最终的知识整合（约25%）。

然而，较长的处理时间可能导致用户在等待过程中产生不安或不耐烦的情绪。因此，我们需要在工作流的关键节点和子流程中，及时将最新的中间结果反馈给用户，以缓解等待焦虑。同时，由于大语言模型采用 token-by-token 的流式输出方式，我们应在相关节点中延续这一特性，以提升用户体验。



为此，我们可以利用 Coze 工作流的“流式输出”机制，每一个子流程（如改写、子话题探索、生成最终报告等）都是一个单次的 HTTP SSE 流式调用（即便该子流程是 Multi-Agent）。Coze 的工作流支持通过 Server-Sent-Event（SSE）文本流的形式输出结果。不仅如此，Coze 还允许在执行过程中通过“输出”类型的节点，将中间结果以 SSE 事件消息的形式实时传递给客户端。

这样，用户可以随时跟踪当前执行状态，例如第一时间展示新生成的子话题、搜索结果、网页的摘抄总结等。同时，最终的话题报告也可以以 token-by-token 的“打字机”效果逐步呈现，进一步增强用户的交互体验和对系统工作的透明感和安全感。

那么，接下来就让我们一起用 Coze 来实现 Deep Research 吧。

6. **基于 Coze 工作流实现**

|  |
| --- |
| **Coze** 是新一代 AI 应用开发服务平台，无论你是否有编程基础，都可以通过这个平台来快速创建 Chat Bot、工作流、智能体，并将其发布到各类社交平台和通讯软件上，或以 API 的形式暴露给任何客户端。 |

6.1 **为什么要通过 Coze 复刻？**

|  |  |
| --- | --- |
| 🤖 **支持更多更新的模型**  Coze 原生支持多种业界领先的大模型，包括豆包、OpenAI、DeepSeek 等。Coze 还为 DeepSeek R1 提供了工具调用的能力，进一步扩展了模型的应用场景。此外，用户可以在不同模型之间切换，更无需担心费用。 | **🚀 通过工作流快速原型验证**  对于像 Deep ReResearch 这样复杂的多 Agent、多工具应用，Coze 提供了直观的工作流流程图功能，便于快速描述、搭建和验证复杂应用逻辑，大幅提升开发效率。同 Coze 提供了节点级的分步运行工具，非常适合调试。 |

|  |  |
| --- | --- |
| **🔧 开放的 API 集成**  如果默认的对话模式 UI 不能满足需求，Coze 提供了灵活的 API 接口，支持将搭建的工作流暴露为 HTTP 接口的 API，无缝集成到自定义客户端中。 | **💻 省心 DevOps**  在原型验证阶段，Coze 省去了繁琐的模型审批、应用脚手架搭建、本地调试和远程部署等步骤，让开发者专注于核心功能的实现，而无需为基础设施操心。 |

|  |
| --- |
| 我们强烈建议在几乎所有大模型原型验证阶段，优先选择 Coze（或类似的开源工具如 Dify）作为起点，以快速验证和迭代产品想法。 |

6.2 **准备工作**

|  |  |
| --- | --- |
| **海外**  官网：coze.com  内部测试环境：https://bots.byteintl.net/ | **中国**  官网：coze.cn  内部测试环境：https://bots.bytedance.net/ |

* 本教程接下来将使用的是[**中国“内部测试环境”**](https://bots.bytedance.net/)**。**建议充值 1 元，成为尊贵的专业版会员。
* 访问的共享空间为 [“ByteDance Demo”](https://bots.bytedance.net/space/7317628956541321242)。你可以将共享空间想象成是互相隔离的根目录，有空间访问权限的同学可以自由访问、使用空间的资源。
* 如遇到无法访问 Coze 或共享空间的问题，请参考 ：
* [Coze/扣子 - One Page](https://bytedance.larkoffice.com/docx/SA7vdm5vwoVjRrxTCUKcgdxBnAf)
* [扣子 | Coze 公司内部使用手册](https://bytedance.larkoffice.com/docx/TQ24dtc5AoHdcqxqtNDceRmwnJg?preview_comment_id=7460099959569137668)
* [ByteDance Demo 空间无法访问，请看这里](https://bots.bytedance.net/open/docs/guides/invite?from=search)

|  |  |  |
| --- | --- | --- |
|  | **重要：如何加入 ByteDance Demo 空间**  如果要查看本文的示例，你必须先确保自己在“ByteDance Demo”空间中。  [点击这个链接](https://bots.bytedance.net/invite/gNtoxntiQXRuJfV32pWf?type=1)，并接受邀请，就可以加入空间。   |  | | --- | | 此链接将于 3 月 12 日 16:24 后过期 | |

6.3 **开发插件**

和很多平台一样，Coze 也支持通过插件来扩展自己的能力，在 Coze 平台上有成百上千款插件。

6.3.1 **如何开发一个插件**

无论你是前端还是后端开发者，都可以在几分钟内创建、调试并发布一个 Coze 插件。对于我们开发者来说，Coze 插件允许你用代码实现复杂的逻辑，允许安装第三方 Python 或 Node.js 的类库，并可以访问网络服务，因此你可以很方便的将可复用的功能或第三方云服务包装成 Coze 插件。

“工具”是 Coze 插件中的重要概念：

* 一个工具对应于一个 Python 或 Node.js 异步函数。
* 一个插件可以包含一个或多个工具。
* 工具可以在 Coze 工作流中直接使用，也可以[作为“Tool”交由大语言模型调用](https://bytetech.info/articles/7444482560496762889)。

**[ScreenFlow.mp4]**

上面的这个**录屏**将向你展示如何创建、调试并发布一个 Coze 插件：

1. 首先，选择一个用于存放 Coze 的工作空间。这里我们以字节跳动 ByteDance Demo 为例，这样公司的同仁都可以使用你的插件，但又不需要担心信息泄露问题。
2. 接下来，新建一个 Coze 插件。这里我们选择的是 Node.js 语言，你也可以使用 Python 语言。
3. 点击**“在IDE中创建工具”按钮**。
4. 切换到“元数据”分页，在这里我们需要定义该工具的入参和出参结构（实际上是 JSON Schema）。在元数据定义完成并保存后，代码的入参和出参的类型（Python 是 Type Hint）也会随之立即更新。
5. 回到**“代码”分页**后，就可以开始编写和调试代码了。
6. 最后点击**“发布”按钮**，即可立即将插件发布到云端了。

6.3.2 **Markdown Metadata 插件**

在[前面的提示工程课程](https://bytedance.larkoffice.com/docx/Eni3dwZjxopWbAxUESAc2Iupn8g)中，你可能会发现我在很多提示词的头部都会添加这样一段信息：

|  |
| --- |
| Markdown --- Current Time: Tue Mar 04 2025 11:50:40 GMT+0800 (China Standard Time) User Name: Henry Li Location: Nanjing, Jiangsu, China Weather: Heavy rain and a gentle breeze Locale: zh\_CN ---  ... |

这是 Markdown 中的 Metadata 扩展语法，在两个“---”之间可以添加任意键值对。我们都知道 Pretrained Model 无法获知当前时间、当前地点、当前登录人等信息，这意味着我们需要将这些信息主动提供给大模型，而上面这种方式就是将上下文环境信息以 Metadata 语法的方式告知大模型，并“提醒”它注意（相比大模型在训练阶段，看过不少这样的格式）。

为了不要在每一次调用大模型时，重复生成上面的 Markdown 片段，我开发了一个简单的 Markdown Metadata 插件，这个插件只有一个名为 metadata\_with\_current\_time 的工具，负责输出包含当前时间的 Markdown Metadata：

|  |  |  |  |
| --- | --- | --- | --- |
| |  | | --- | | TypeScript metadata\_with\_current\_time import { Args } from '@/runtime'; import { Input, Output } from "@/typings/metadata\_with\_current\_time/metadata\_with\_current\_time";  export async function handler({ input, logger }: Args<Input>): Promise<Output> {  return { metadata: `---\nCurrent Time: ${new Date().toString()}\n---` }; }; | | |  | | --- | | 输入与输出参数 | |

插件的代码非常简单直观，也没有任何入参，输出的属性就是 metadata。

现在你也可以通过搜索“Markdown Metadata”，找到[这个插件](https://bots.bytedance.net/space/7317628956541321242/plugin/7476788094809686053)。因为几乎所有的 prompt 都需要获知当前的时间，因此我们会将它作为接下来每一个子流程的首个节点。

6.3.3 **Readability 插件**

在 5.3 章节中，我们提到了通过 J[ina](https://jina.ai/reader/) 爬取网页的可读性内容，但是 Jina 对“可读性”内容提取的支持（特别是中文网站）不是特别好，我们来看这个例子：

|  |
| --- |
| TypeScript **curl** https://r.jina.ai/https://www.zaobao.com/realtime/china/story20250222-5915622 \  -H "X-Retain-Images: none" |

返回的 Markdown 可读内容是：

|  |
| --- |
| markdown Title: 字节跳动估值据报逾4000亿美元 再成科技宠儿  URL Source: https://www.zaobao.com/realtime/china/story20250222-5915622  Markdown Content: 字节跳动估值据报逾4000亿美元 再成科技宠儿 ===============  [](https://www.zaobao.com/)  简  繁  新加坡版  [新明日报](https://www.zaobao.com/publication/xin-ming-ri-bao)  [早报俱乐部](https://www.zaobao.com/zbclub)  [电子报](https://www.zaobao.com/epapers)  订阅  登录/免费注册  \* \* \*  [新闻](https://www.zaobao.com/news)  [即时](https://www.zaobao.com/realtime)[新加坡](https://www.zaobao.com/news/singapore)[国际](https://www.zaobao.com/news/world)[中国](https://www.zaobao.com/news/china)[东南亚](https://www.zaobao.com/news/sea)  [财经](https://www.zaobao.com/finance)  [新加坡股市](https://stock.zaobao.com.sg/)[新加坡财经](https://www.zaobao.com/finance/singapore)[全球财经](https://www.zaobao.com/finance/world)[中国财经](https://www.zaobao.com/finance/china)[投资理财](https://www.zaobao.com/finance/invest)[房产](https://www.zaobao.com/finance/real-estate)[美国股市](https://www.zaobao.com/finance/usstock)[中小企业](https://www.zaobao.com/finance/sme)[起步创新](https://www.zaobao.com/finance/startup)[财经人物](https://www.zaobao.com/finance/personalities)  [言论](https://www.zaobao.com/forum)  [社论](https://www.zaobao.com/forum/editorial)[评论](https://www.zaobao.com/forum/views)[交流站](https://www.zaobao.com/forum/talk)[漫画](https://www.zaobao.com/forum/comic)  [娱乐](https://www.zaobao.com/entertainment)  [明星](https://www.zaobao.com/entertainment/celebs)[影视](https://www.zaobao.com/entertainment/movies-and-tv)[音乐](https://www.zaobao.com/entertainment/music)[韩流](https://www.zaobao.com/entertainment/k-pop)[送礼](https://www.zaobao.com/entertainment/giveaway)  [生活](https://www.zaobao.com/lifestyle)  [壮龄go!](https://www.zaobao.com/specials/active-aging)[特写](https://www.zaobao.com/lifestyle/feature)[美食](https://www.zaobao.com/lifestyle/food)[旅行](https://www.zaobao.com/lifestyle/travel)[文化艺术](https://www.zaobao.com/lifestyle/culture)[人文史地](https://www.zaobao.com/lifestyle/history-heritage)[专栏](https://www.zaobao.com/lifestyle/columns)[@世代](https://www.zaobao.com/lifestyle/gen)[生态与环保](https://www.zaobao.com/lifestyle/eco)[时尚与美容](https://www.zaobao.com/lifestyle/fashion-beauty)[设计与家居](https://www.zaobao.com/lifestyle/design-decor)[光影](https://www.zaobao.com/lifestyle/photography)[科玩](https://www.zaobao.com/lifestyle/gadget)[科普](https://www.zaobao.com/lifestyle/tech-science)[汽车](https://www.zaobao.com/lifestyle/motoring)[心事家事](https://www.zaobao.com/lifestyle/familynlove)[精选](https://www.zaobao.com/lifestyle/others)[特辑](https://www.zaobao.com/lifestyle/supplement)[早报校园](https://www.zaobao.com/lifestyle/campus)[热门](https://www.zaobao.com/lifestyle/trending)[生活贴士](https://www.zaobao.com/lifestyle/life-hacks)[星座与生肖](https://www.zaobao.com/horoscope)  [保健](https://www.zaobao.com/lifestyle/health)  [体育](https://www.zaobao.com/news/sports)  [视频](https://www.zaobao.com/videos)  [新闻](https://www.zaobao.com/videos/news)[娱乐](https://www.zaobao.com/videos/entertainment)[生活](https://www.zaobao.com/videos/lifestyle)[系列节目](https://www.zaobao.com/video-series)[直播](https://www.zaobao.com/livestream)  [早报播客](https://www.zaobao.com/podcast)  [互动新闻](https://www.zaobao.com/interactive-graphics)  [专题](https://www.zaobao.com/special)  [订阅](https://subscribe.sph.com.sg/zh/publications-zb/?utm\_campaign=zb\_subscription&utm\_medium=sph-publication&utm\_source=zb&utm\_content=subscribebutton-header)  发布/2025年2月22日 11:07  [](https://www.facebook.com/sharer/sharer.php?u=https://www.zaobao.com.sg/realtime/china/story20250222-5915622)[](https://twitter.com/intent/tweet?text=%E5%AD%97%E8%8A%82%E8%B7%B3%E5%8A%A8%E4%BC%B0%E5%80%BC%E6%8D%AE%E6%8A%A5%E9%80%BE4000%E4%BA%BF%E7%BE%8E%E5%85%83%20%E5%86%8D%E6%88%90%E7%A7%91%E6%8A%80%E5%AE%A0%E5%84%BF&url=https://www.zaobao.com.sg/realtime/china/story20250222-5915622)[](https://www.addtoany.com/ext/wechat/share/#url=https://www.zaobao.com.sg/realtime/china/story20250222-5915622&title=%E5%AD%97%E8%8A%82%E8%B7%B3%E5%8A%A8%E4%BC%B0%E5%80%BC%E6%8D%AE%E6%8A%A5%E9%80%BE4000%E4%BA%BF%E7%BE%8E%E5%85%83%20%E5%86%8D%E6%88%90%E7%A7%91%E6%8A%80%E5%AE%A0%E5%84%BF)[](https://api.whatsapp.com/send?text=%E5%AD%97%E8%8A%82%E8%B7%B3%E5%8A%A8%E4%BC%B0%E5%80%BC%E6%8D%AE%E6%8A%A5%E9%80%BE4000%E4%BA%BF%E7%BE%8E%E5%85%83%20%E5%86%8D%E6%88%90%E7%A7%91%E6%8A%80%E5%AE%A0%E5%84%BF%20https://www.zaobao.com.sg/realtime/china/story20250222-5915622)[](https://t.me/share/url?url=https://www.zaobao.com.sg/realtime/china/story20250222-5915622&text=%E5%AD%97%E8%8A%82%E8%B7%B3%E5%8A%A8%E4%BC%B0%E5%80%BC%E6%8D%AE%E6%8A%A5%E9%80%BE4000%E4%BA%BF%E7%BE%8E%E5%85%83%20%E5%86%8D%E6%88%90%E7%A7%91%E6%8A%80%E5%AE%A0%E5%84%BF)[](mailto:?subject=%E5%AD%97%E8%8A%82%E8%B7%B3%E5%8A%A8%E4%BC%B0%E5%80%BC%E6%8D%AE%E6%8A%A5%E9%80%BE4000%E4%BA%BF%E7%BE%8E%E5%85%83%20%E5%86%8D%E6%88%90%E7%A7%91%E6%8A%80%E5%AE%A0%E5%84%BF&body=https://www.zaobao.com.sg/realtime/china/story20250222-5915622)  小标准中大  字节跳动估值据报逾4000亿美元 再成科技宠儿 =======================  中国科技企业字节跳动的三大投资者据报已将该公司的估值提高至超过4000亿美元（5340亿新元），显示字节跳动尽管面临美国封禁威胁，市场认可度却已实现绝地反弹。  彭博社星期六（2月22日）引述知情人士报道，考虑到字节跳动人工智能（AI）模型豆包的潜力，日本软银集团旗下的愿景基金已在去年12月将字节跳动的估值重新定为4000亿美元。  彭博社根据去年11月向美国证券交易委员会提交的文件计算得出，富达投资和普徕仕集团（T. Rowe Price Group）对字节跳动的估值也已分别超过4100亿美元和4500亿美元。  软银的估值标准没有考量到字节跳动旗下应用TikTok在美国的运营价值，因为这款热门应用本应被美国政府封禁。报道称，如果TikTok最终逃过被封禁的命运，字节跳动的估值很可能进一步上升。  彭博社分析，此次估值对字节跳动而言是一次显著的增长，该公司在2022年7月的非公开交易中，估值曾触及2750亿美元的谷底。中国AI初创公司深度求索（DeepSeek）的爆火，加上中国政府对科技行业态度软化的预期，都利好字节跳动估值的上扬。  尽管如此，字节跳动的估值仍将受到中美地缘政治角力不确定性的影响。两名知情人士透露，一些投资者仍在寻求以低至2400亿美元的估值参股该公司，尽管现在很少有卖家愿意以这样的价位转让。  [字节跳动](https://www.zaobao.com/keywords/zi-jie-tiao-dong)[TikTok](https://www.zaobao.com/keywords/tiktok)[中美科技战](https://www.zaobao.com/keywords/zhong-mei-ke-ji-zhan)[人工智能](https://www.zaobao.com/keywords/ren-gong-zhi-neng)[软银集团](https://www.zaobao.com/keywords/ruan-yin-ji-tuan)  [上一篇 两岸关系学者黎宝文任台海基会发言人](https://www.zaobao.com/realtime/china/story20250222-5915393?ref=previous-article)  [下一篇 港政务司长吁推动国民教育 让下一代自发爱护国家](https://www.zaobao.com/realtime/china/story20250222-5915536?ref=next-article)  [购买此文章](https://newspost.newslink.sg/?publication=ZB&date=02/22/2025&title=%E5%AD%97%E8%8A%82%E8%B7%B3%E5%8A%A8%E4%BC%B0%E5%80%BC%E6%8D%AE%E6%8A%A5%E9%80%BE4000%E4%BA%BF%E7%BE%8E%E5%85%83%20%E5%86%8D%E6%88%90%E7%A7%91%E6%8A%80%E5%AE%A0%E5%84%BF)     热文推荐  [](https://www.outbrain.com/what-is/default/en)                  [](https://www.outbrain.com/what-is/default/en)      [中国全国人大会议会期七天](https://www.zaobao.com.sg/realtime/china/story20250304-5962144?obOrigUrl=true)  [](https://www.outbrain.com/what-is/default/en)          [](https://www.outbrain.com/what-is/default/en)  [即时新闻](https://www.zaobao.com/realtime)[](https://www.zaobao.com/realtime) ----------------------------------------------------------------------------  \* 6 分钟前    [澳洲男女薪资差距缩小 女性仍比男性少赚18.6% ------------------------](https://www.zaobao.com/realtime/world/story20250304-5962883?ref=sidebar-realtime)   \* 10 分钟前    [【国会】截至去年 防止骚扰专案法庭接获逾2000项个人保护令申请 --------------------------------](https://www.zaobao.com/realtime/singapore/story20250304-5962422?ref=sidebar-realtime)   \* 13 分钟前    [【国会】土管局拟本季度为海军部东路20栋国有房产招标 --------------------------](https://www.zaobao.com/realtime/singapore/story20250304-5962416?ref=sidebar-realtime)   \* 26 分钟前    [中国据报将发布文件 推动全国使用开源晶片RISC-V --------------------------](https://www.zaobao.com/realtime/china/story20250304-5962871?ref=sidebar-realtime)   \* 28 分钟前    [【国会】律政部推行改革 简化并加强民事判决执行机制 -------------------------](https://www.zaobao.com/realtime/singapore/story20250304-5962394?ref=sidebar-realtime)   \* 31 分钟前    [【国会】下半年试行派遣法律科技顾问到律所 --------------------](https://www.zaobao.com/realtime/singapore/story20250304-5962506?ref=sidebar-realtime)    热门 --  \* 单日 \* 一周  \* [尼诰大道车祸死者为食品公司董事 员工赞其友善年轻有为 --------------------------](https://www.zaobao.com/news/singapore/story20250303-5959406?ref=sidebar-daily-popular)   \* [焚尸惨案女童网红生父 贴文悼爱女 ----------------](https://www.zaobao.com/news/singapore/story20250303-5957338?ref=sidebar-daily-popular)   \* [涉性侵未成年少女 前电视艺人有意认罪 ------------------](https://www.zaobao.com/realtime/singapore/story20250304-5961233?ref=sidebar-daily-popular)   \* [沈琳宸痛心哀悼焚尸惨案女童：你的离去不会毫无意义 ------------------------](https://www.zaobao.com/entertainment/story20250303-5958285?ref=sidebar-daily-popular)   \* [通关被指护照破损 狮城妇诉：马国官员索200令吉 ------------------------](https://www.zaobao.com/news/singapore/story20250303-5957253?ref=sidebar-daily-popular)    更多消息 ----  成为联合早报VIP ---------  解锁全网内容、电子报及专享福利  [立即订阅](https://subscribe.sph.com.sg/zh/publications-zb/?utm\_campaign=zb\_subscription&utm\_medium=sph-publication&utm\_source=zb&utm\_content=subscribebutton-footer)  \* [新闻](https://www.zaobao.com/news) \* \* [即时](https://www.zaobao.com/realtime)  \* [新加坡](https://www.zaobao.com/news/singapore)  \* [国际](https://www.zaobao.com/news/world)  \* [中国](https://www.zaobao.com/news/china)  \* [东南亚](https://www.zaobao.com/news/sea)  \* [财经](https://www.zaobao.com/finance) \* \* [新加坡股市](https://stock.zaobao.com.sg/)  \* [新加坡财经](https://www.zaobao.com/finance/singapore)  \* [全球财经](https://www.zaobao.com/finance/world)  \* [中国财经](https://www.zaobao.com/finance/china)  \* [投资理财](https://www.zaobao.com/finance/invest)  \* [房产](https://www.zaobao.com/finance/real-estate)  \* [美国股市](https://www.zaobao.com/finance/usstock)  \* [中小企业](https://www.zaobao.com/finance/sme)  \* [起步创新](https://www.zaobao.com/finance/startup)  \* [财经人物](https://www.zaobao.com/finance/personalities)  \* [言论](https://www.zaobao.com/forum) \* \* [社论](https://www.zaobao.com/forum/editorial)  \* [评论](https://www.zaobao.com/forum/views)  \* [交流站](https://www.zaobao.com/forum/talk)  \* [漫画](https://www.zaobao.com/forum/comic) \* [娱乐](https://www.zaobao.com/entertainment) \* \* [明星](https://www.zaobao.com/entertainment/celebs)  \* [影视](https://www.zaobao.com/entertainment/movies-and-tv)  \* [音乐](https://www.zaobao.com/entertainment/music)  \* [韩流](https://www.zaobao.com/entertainment/k-pop)  \* [送礼](https://www.zaobao.com/entertainment/giveaway)  \* [生活](https://www.zaobao.com/lifestyle) \* \* [壮龄go!](https://www.zaobao.com/specials/active-aging)  \* [特写](https://www.zaobao.com/lifestyle/feature)  \* [美食](https://www.zaobao.com/lifestyle/food)  \* [旅行](https://www.zaobao.com/lifestyle/travel)  \* [文化艺术](https://www.zaobao.com/lifestyle/culture)  \* [人文史地](https://www.zaobao.com/lifestyle/history-heritage)  \* [专栏](https://www.zaobao.com/lifestyle/columns)  \* [@世代](https://www.zaobao.com/lifestyle/gen)  \* [生态与环保](https://www.zaobao.com/lifestyle/eco)  \* [时尚与美容](https://www.zaobao.com/lifestyle/fashion-beauty)  \* [设计与家居](https://www.zaobao.com/lifestyle/design-decor)  \* [光影](https://www.zaobao.com/lifestyle/photography)  \* [科玩](https://www.zaobao.com/lifestyle/gadget)  \* [科普](https://www.zaobao.com/lifestyle/tech-science)  \* [汽车](https://www.zaobao.com/lifestyle/motoring)  \* [心事家事](https://www.zaobao.com/lifestyle/familynlove)  \* [精选](https://www.zaobao.com/lifestyle/others)  \* [特辑](https://www.zaobao.com/lifestyle/supplement)  \* [早报校园](https://www.zaobao.com/lifestyle/campus)  \* [热门](https://www.zaobao.com/lifestyle/trending)  \* [生活贴士](https://www.zaobao.com/lifestyle/life-hacks)  \* [星座与生肖](https://www.zaobao.com/horoscope)  \* [保健](https://www.zaobao.com/lifestyle/health) \* [体育](https://www.zaobao.com/news/sports) \* [视频](https://www.zaobao.com/videos) \* \* [新闻](https://www.zaobao.com/videos/news)  \* [娱乐](https://www.zaobao.com/videos/entertainment)  \* [生活](https://www.zaobao.com/videos/lifestyle)  \* [系列节目](https://www.zaobao.com/video-series)  \* [直播](https://www.zaobao.com/livestream) \* [早报播客](https://www.zaobao.com/podcast) \* [互动新闻](https://www.zaobao.com/interactive-graphics) \* [专题](https://www.zaobao.com/special)  \* [新明日报](https://www.zaobao.com/publication/xin-ming-ri-bao) \* [早报俱乐部](https://www.zaobao.com/zbclub) \* [电子报](https://www.zaobao.com/epapers) \* [国际版](https://www.zaobao.com.sg/global) \* [中国版](https://www.zaobao.com.sg/cn)  [关于我们](https://www.zaobao.com/about-us)  [常见问题](https://www.zaobao.com/faq)  [关注我们](https://www.zaobao.com/products)  [联络我们](https://www.zaobao.com/contact-us)  [广告联络](https://www.zaobao.com/advertise)  [投函/投稿](https://www.zaobao.com/contribute)  [加入我们](https://www.sph.com.sg/careers-and-scholarships)  [](https://www.facebook.com/zaobaosg)  [](https://twitter.com/zaobaosg)  [](https://www.youtube.com/zaobaodotsg)  [](https://www.instagram.com/zaobaosg/)  [](https://t.me/zaobaosg)  [](https://www.weibo.com/zaobaocom)  [](https://www.zaobao.com/products#wechat)  [](mailto:zbonline@sph.com.sg)  [](https://itunes.apple.com/us/app/lian-he-zao-bao/id654946831?ls=1&mt=8)[](https://play.google.com/store/apps/details?id=com.zb.sph.zaobaosingapore)  热线  新报业媒体总机 [63196319](tel:63196319)  订阅热线 [63883838](tel:63883838)  早报广告联系 [63196319](tel:63196319)  新闻热线 [1800-7416388](tel:1800-7416388) 或 [92288736](tel:92288736)  \* [新报业媒体网站](https://www.sph.com.sg/) \* [网络条款](https://www.sph.com.sg/legal/website\_tnc/) \* [隐私政策](https://www.sph.com.sg/legal/sph\_privacy/) \* [个人资料保护](https://www.sph.com.sg/legal/pdpa/) \* [全检索](https://newslink.sg/) \* [ZShop集品店](https://zshop.zaobao.sg/) \* [海峡时报](https://www.straitstimes.com/) \* [商业时报](https://www.businesstimes.com.sg/)  MDDI (P) 042/10/2024, MDDI (P) 043/10/2024  新报业媒体有限公司版权所有（公司登记号：202120748H）  在中国的用户请游览 [zaobao.com](https://www.zaobao.com/)  提醒：新加坡网络业者若未经许可，擅自引用本网内容将面对法律行动。  第三方公司可能在早报网站宣传他们的产品或服务。不过您跟第三方公司的任何交易与早报网站无关，早报将不会对可能引起的任何损失负责。  0.4.0     订户专享 赠阅文章  早报VIP会员登录后，即可使用此功能。  登录  立即订阅  早报VIP会员每月可赠阅 5 篇订户专享文章。[了解更多](https://www.zaobao.com.sg/faq?tab=gift) |

这里会发现有 2 个问题：

1. 尽管爬取的 HTML 已经被 Jina 转换为 Markdown，但是仍然需要占用 **9,989 个 token**。
2. 返回的 Markdown 中包含太多的导航和页脚信息，信噪比太低。
3. 返回的内容中虽然已经去掉了图片，但是还是存在大量的链接 URL，而大模型显然不能也不需要读取 URL。

在查阅了 Jina 的文档后，发现它并没有相关的开关可以解决这个问题，而是提供了通过大模型来进行信息读取（这也是为什么 Jina 的爬虫免费，而信息提取按照 token 收费的原因）。因此，笔者决定动手自己写一个 Node.js 插件来解决这个问题，思路如下：

1. 用 Jina 爬取原始的 HTML，但不做任何格式转换。这一步主要利用的是 Jina 的“无头浏览器”爬虫能力、“反防爬”能力，我们无需自建服务。
2. 将爬取到的 HTML，通过 [jsdom](https://github.com/jsdom/jsdom) 库转换为内存中的 HTML DOM 对象，再通过 [@mozila/readability](https://github.com/mozilla/readability) 库从 DOM 中提取“可读性”内容的 HTML。该功能可以只保留一个网页中的标题和主要内容信息，从而让人类用户保持专注阅读，同理对于大语言模型来说也是“Attention is all I need”。如果你使用过 iOS 的 Reader Mode，对此一定会深有体会。
3. 接下来再通过 [turndown](https://github.com/mixmark-io/turndown) 库，将 HTML 转换为 Markdown 格式，毕竟它比 HTML 的信噪比高得多。
4. 最后，我们通过正则将 Markdown 中的超级链接替换为纯文字，并去掉图片，就得到了我们最终想要的结果，它将非常适合大语言模型进行“专注阅读”。

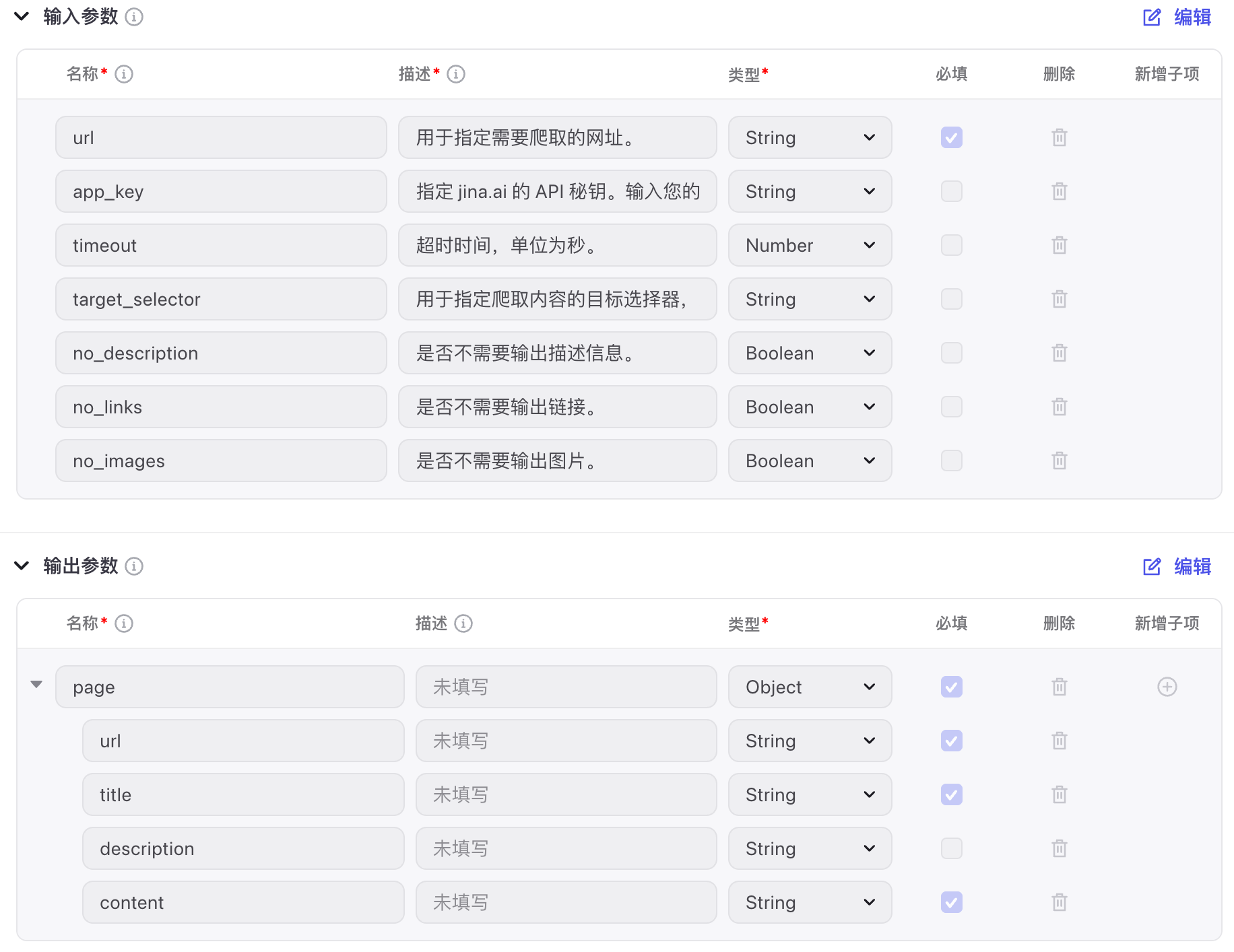
点击这里

比较一下优化前和优化后的 Markdown，优化后**只有 591 个 Token**， **不到原来的 6%**，并且没有用到大语言模型，因此全部免费！是不是高下立判？

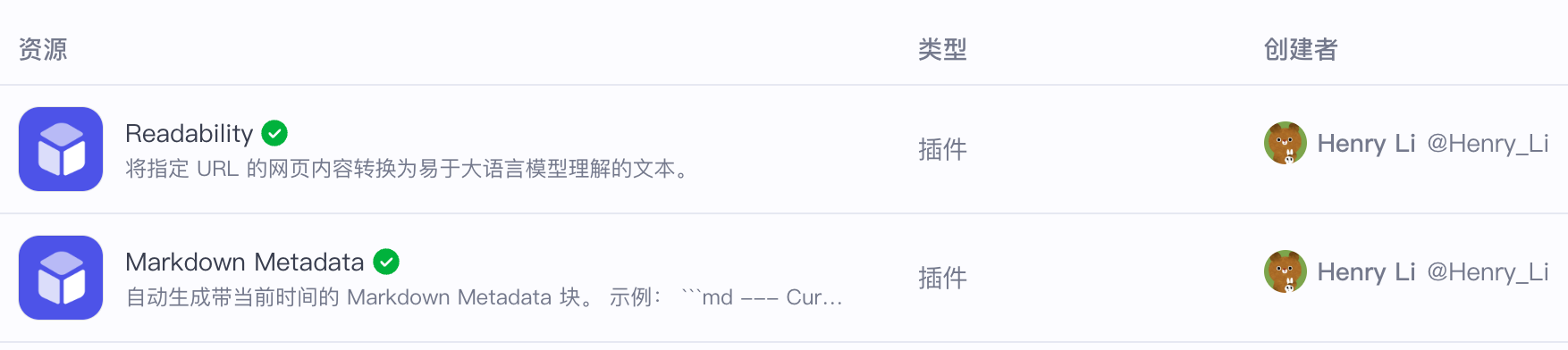
|  |  |  |  |
| --- | --- | --- | --- |
| |  | | --- | | markdown 优化前 [](https://www.zaobao.com/)  简  繁  新加坡版  [新明日报](https://www.zaobao.com/publication/xin-ming-ri-bao)  [早报俱乐部](https://www.zaobao.com/zbclub)  [电子报](https://www.zaobao.com/epapers)  订阅  登录/免费注册  \* \* \*  [新闻](https://www.zaobao.com/news)  [即时](https://www.zaobao.com/realtime)[新加坡](https://www.zaobao.com/news/singapore)[国际](https://www.zaobao.com/news/world)[中国](https://www.zaobao.com/news/china)[东南亚](https://www.zaobao.com/news/sea)  [财经](https://www.zaobao.com/finance)  [新加坡股市](https://stock.zaobao.com.sg/)[新加坡财经](https://www.zaobao.com/finance/singapore)[全球财经](https://www.zaobao.com/finance/world)[中国财经](https://www.zaobao.com/finance/china)[投资理财](https://www.zaobao.com/finance/invest)[房产](https://www.zaobao.com/finance/real-estate)[美国股市](https://www.zaobao.com/finance/usstock)[中小企业](https://www.zaobao.com/finance/sme)[起步创新](https://www.zaobao.com/finance/startup)[财经人物](https://www.zaobao.com/finance/personalities)  [言论](https://www.zaobao.com/forum)  [社论](https://www.zaobao.com/forum/editorial)[评论](https://www.zaobao.com/forum/views)[交流站](https://www.zaobao.com/forum/talk)[漫画](https://www.zaobao.com/forum/comic)  [娱乐](https://www.zaobao.com/entertainment)  [明星](https://www.zaobao.com/entertainment/celebs)[影视](https://www.zaobao.com/entertainment/movies-and-tv)[音乐](https://www.zaobao.com/entertainment/music)[韩流](https://www.zaobao.com/entertainment/k-pop)[送礼](https://www.zaobao.com/entertainment/giveaway)  [生活](https://www.zaobao.com/lifestyle)  [壮龄go!](https://www.zaobao.com/specials/active-aging)[特写](https://www.zaobao.com/lifestyle/feature)[美食](https://www.zaobao.com/lifestyle/food)[旅行](https://www.zaobao.com/lifestyle/travel)[文化艺术](https://www.zaobao.com/lifestyle/culture)[人文史地](https://www.zaobao.com/lifestyle/history-heritage)[专栏](https://www.zaobao.com/lifestyle/columns)[@世代](https://www.zaobao.com/lifestyle/gen)[生态与环保](https://www.zaobao.com/lifestyle/eco)[时尚与美容](https://www.zaobao.com/lifestyle/fashion-beauty)[设计与家居](https://www.zaobao.com/lifestyle/design-decor)[光影](https://www.zaobao.com/lifestyle/photography)[科玩](https://www.zaobao.com/lifestyle/gadget)[科普](https://www.zaobao.com/lifestyle/tech-science)[汽车](https://www.zaobao.com/lifestyle/motoring)[心事家事](https://www.zaobao.com/lifestyle/familynlove)[精选](https://www.zaobao.com/lifestyle/others)[特辑](https://www.zaobao.com/lifestyle/supplement)[早报校园](https://www.zaobao.com/lifestyle/campus)[热门](https://www.zaobao.com/lifestyle/trending)[生活贴士](https://www.zaobao.com/lifestyle/life-hacks)[星座与生肖](https://www.zaobao.com/horoscope)  [保健](https://www.zaobao.com/lifestyle/health)  [体育](https://www.zaobao.com/news/sports)  [视频](https://www.zaobao.com/videos)  [新闻](https://www.zaobao.com/videos/news)[娱乐](https://www.zaobao.com/videos/entertainment)[生活](https://www.zaobao.com/videos/lifestyle)[系列节目](https://www.zaobao.com/video-series)[直播](https://www.zaobao.com/livestream)  [早报播客](https://www.zaobao.com/podcast)  [互动新闻](https://www.zaobao.com/interactive-graphics)  [专题](https://www.zaobao.com/special)  [订阅](https://subscribe.sph.com.sg/zh/publications-zb/?utm\_campaign=zb\_subscription&utm\_medium=sph-publication&utm\_source=zb&utm\_content=subscribebutton-header)  发布/2025年2月22日 11:07  [](https://www.facebook.com/sharer/sharer.php?u=https://www.zaobao.com.sg/realtime/china/story20250222-5915622)[](https://twitter.com/intent/tweet?text=%E5%AD%97%E8%8A%82%E8%B7%B3%E5%8A%A8%E4%BC%B0%E5%80%BC%E6%8D%AE%E6%8A%A5%E9%80%BE4000%E4%BA%BF%E7%BE%8E%E5%85%83%20%E5%86%8D%E6%88%90%E7%A7%91%E6%8A%80%E5%AE%A0%E5%84%BF&url=https://www.zaobao.com.sg/realtime/china/story20250222-5915622)[](https://www.addtoany.com/ext/wechat/share/#url=https://www.zaobao.com.sg/realtime/china/story20250222-5915622&title=%E5%AD%97%E8%8A%82%E8%B7%B3%E5%8A%A8%E4%BC%B0%E5%80%BC%E6%8D%AE%E6%8A%A5%E9%80%BE4000%E4%BA%BF%E7%BE%8E%E5%85%83%20%E5%86%8D%E6%88%90%E7%A7%91%E6%8A%80%E5%AE%A0%E5%84%BF)[](https://api.whatsapp.com/send?text=%E5%AD%97%E8%8A%82%E8%B7%B3%E5%8A%A8%E4%BC%B0%E5%80%BC%E6%8D%AE%E6%8A%A5%E9%80%BE4000%E4%BA%BF%E7%BE%8E%E5%85%83%20%E5%86%8D%E6%88%90%E7%A7%91%E6%8A%80%E5%AE%A0%E5%84%BF%20https://www.zaobao.com.sg/realtime/china/story20250222-5915622)[](https://t.me/share/url?url=https://www.zaobao.com.sg/realtime/china/story20250222-5915622&text=%E5%AD%97%E8%8A%82%E8%B7%B3%E5%8A%A8%E4%BC%B0%E5%80%BC%E6%8D%AE%E6%8A%A5%E9%80%BE4000%E4%BA%BF%E7%BE%8E%E5%85%83%20%E5%86%8D%E6%88%90%E7%A7%91%E6%8A%80%E5%AE%A0%E5%84%BF)[](mailto:?subject=%E5%AD%97%E8%8A%82%E8%B7%B3%E5%8A%A8%E4%BC%B0%E5%80%BC%E6%8D%AE%E6%8A%A5%E9%80%BE4000%E4%BA%BF%E7%BE%8E%E5%85%83%20%E5%86%8D%E6%88%90%E7%A7%91%E6%8A%80%E5%AE%A0%E5%84%BF&body=https://www.zaobao.com.sg/realtime/china/story20250222-5915622)  小标准中大  字节跳动估值据报逾4000亿美元 再成科技宠儿 =======================  中国科技企业字节跳动的三大投资者据报已将该公司的估值提高至超过4000亿美元（5340亿新元），显示字节跳动尽管面临美国封禁威胁，市场认可度却已实现绝地反弹。  彭博社星期六（2月22日）引述知情人士报道，考虑到字节跳动人工智能（AI）模型豆包的潜力，日本软银集团旗下的愿景基金已在去年12月将字节跳动的估值重新定为4000亿美元。  彭博社根据去年11月向美国证券交易委员会提交的文件计算得出，富达投资和普徕仕集团（T. Rowe Price Group）对字节跳动的估值也已分别超过4100亿美元和4500亿美元。  软银的估值标准没有考量到字节跳动旗下应用TikTok在美国的运营价值，因为这款热门应用本应被美国政府封禁。报道称，如果TikTok最终逃过被封禁的命运，字节跳动的估值很可能进一步上升。  彭博社分析，此次估值对字节跳动而言是一次显著的增长，该公司在2022年7月的非公开交易中，估值曾触及2750亿美元的谷底。中国AI初创公司深度求索（DeepSeek）的爆火，加上中国政府对科技行业态度软化的预期，都利好字节跳动估值的上扬。  尽管如此，字节跳动的估值仍将受到中美地缘政治角力不确定性的影响。两名知情人士透露，一些投资者仍在寻求以低至2400亿美元的估值参股该公司，尽管现在很少有卖家愿意以这样的价位转让。  [字节跳动](https://www.zaobao.com/keywords/zi-jie-tiao-dong)[TikTok](https://www.zaobao.com/keywords/tiktok)[中美科技战](https://www.zaobao.com/keywords/zhong-mei-ke-ji-zhan)[人工智能](https://www.zaobao.com/keywords/ren-gong-zhi-neng)[软银集团](https://www.zaobao.com/keywords/ruan-yin-ji-tuan)  [上一篇 两岸关系学者黎宝文任台海基会发言人](https://www.zaobao.com/realtime/china/story20250222-5915393?ref=previous-article)  [下一篇 港政务司长吁推动国民教育 让下一代自发爱护国家](https://www.zaobao.com/realtime/china/story20250222-5915536?ref=next-article)  [购买此文章](https://newspost.newslink.sg/?publication=ZB&date=02/22/2025&title=%E5%AD%97%E8%8A%82%E8%B7%B3%E5%8A%A8%E4%BC%B0%E5%80%BC%E6%8D%AE%E6%8A%A5%E9%80%BE4000%E4%BA%BF%E7%BE%8E%E5%85%83%20%E5%86%8D%E6%88%90%E7%A7%91%E6%8A%80%E5%AE%A0%E5%84%BF)     热文推荐  [](https://www.outbrain.com/what-is/default/en)                  [](https://www.outbrain.com/what-is/default/en)      [中国全国人大会议会期七天](https://www.zaobao.com.sg/realtime/china/story20250304-5962144?obOrigUrl=true)  [](https://www.outbrain.com/what-is/default/en)          [](https://www.outbrain.com/what-is/default/en)  [即时新闻](https://www.zaobao.com/realtime)[](https://www.zaobao.com/realtime) ----------------------------------------------------------------------------  \* 6 分钟前    [澳洲男女薪资差距缩小 女性仍比男性少赚18.6% ------------------------](https://www.zaobao.com/realtime/world/story20250304-5962883?ref=sidebar-realtime)   \* 10 分钟前    [【国会】截至去年 防止骚扰专案法庭接获逾2000项个人保护令申请 --------------------------------](https://www.zaobao.com/realtime/singapore/story20250304-5962422?ref=sidebar-realtime)   \* 13 分钟前    [【国会】土管局拟本季度为海军部东路20栋国有房产招标 --------------------------](https://www.zaobao.com/realtime/singapore/story20250304-5962416?ref=sidebar-realtime)   \* 26 分钟前    [中国据报将发布文件 推动全国使用开源晶片RISC-V --------------------------](https://www.zaobao.com/realtime/china/story20250304-5962871?ref=sidebar-realtime)   \* 28 分钟前    [【国会】律政部推行改革 简化并加强民事判决执行机制 -------------------------](https://www.zaobao.com/realtime/singapore/story20250304-5962394?ref=sidebar-realtime)   \* 31 分钟前    [【国会】下半年试行派遣法律科技顾问到律所 --------------------](https://www.zaobao.com/realtime/singapore/story20250304-5962506?ref=sidebar-realtime)    热门 --  \* 单日 \* 一周  \* [尼诰大道车祸死者为食品公司董事 员工赞其友善年轻有为 --------------------------](https://www.zaobao.com/news/singapore/story20250303-5959406?ref=sidebar-daily-popular)   \* [焚尸惨案女童网红生父 贴文悼爱女 ----------------](https://www.zaobao.com/news/singapore/story20250303-5957338?ref=sidebar-daily-popular)   \* [涉性侵未成年少女 前电视艺人有意认罪 ------------------](https://www.zaobao.com/realtime/singapore/story20250304-5961233?ref=sidebar-daily-popular)   \* [沈琳宸痛心哀悼焚尸惨案女童：你的离去不会毫无意义 ------------------------](https://www.zaobao.com/entertainment/story20250303-5958285?ref=sidebar-daily-popular)   \* [通关被指护照破损 狮城妇诉：马国官员索200令吉 ------------------------](https://www.zaobao.com/news/singapore/story20250303-5957253?ref=sidebar-daily-popular)    更多消息 ----  成为联合早报VIP ---------  解锁全网内容、电子报及专享福利  [立即订阅](https://subscribe.sph.com.sg/zh/publications-zb/?utm\_campaign=zb\_subscription&utm\_medium=sph-publication&utm\_source=zb&utm\_content=subscribebutton-footer)  \* [新闻](https://www.zaobao.com/news) \* \* [即时](https://www.zaobao.com/realtime)  \* [新加坡](https://www.zaobao.com/news/singapore)  \* [国际](https://www.zaobao.com/news/world)  \* [中国](https://www.zaobao.com/news/china)  \* [东南亚](https://www.zaobao.com/news/sea)  \* [财经](https://www.zaobao.com/finance) \* \* [新加坡股市](https://stock.zaobao.com.sg/)  \* [新加坡财经](https://www.zaobao.com/finance/singapore)  \* [全球财经](https://www.zaobao.com/finance/world)  \* [中国财经](https://www.zaobao.com/finance/china)  \* [投资理财](https://www.zaobao.com/finance/invest)  \* [房产](https://www.zaobao.com/finance/real-estate)  \* [美国股市](https://www.zaobao.com/finance/usstock)  \* [中小企业](https://www.zaobao.com/finance/sme)  \* [起步创新](https://www.zaobao.com/finance/startup)  \* [财经人物](https://www.zaobao.com/finance/personalities)  \* [言论](https://www.zaobao.com/forum) \* \* [社论](https://www.zaobao.com/forum/editorial)  \* [评论](https://www.zaobao.com/forum/views)  \* [交流站](https://www.zaobao.com/forum/talk)  \* [漫画](https://www.zaobao.com/forum/comic) \* [娱乐](https://www.zaobao.com/entertainment) \* \* [明星](https://www.zaobao.com/entertainment/celebs)  \* [影视](https://www.zaobao.com/entertainment/movies-and-tv)  \* [音乐](https://www.zaobao.com/entertainment/music)  \* [韩流](https://www.zaobao.com/entertainment/k-pop)  \* [送礼](https://www.zaobao.com/entertainment/giveaway)  \* [生活](https://www.zaobao.com/lifestyle) \* \* [壮龄go!](https://www.zaobao.com/specials/active-aging)  \* [特写](https://www.zaobao.com/lifestyle/feature)  \* [美食](https://www.zaobao.com/lifestyle/food)  \* [旅行](https://www.zaobao.com/lifestyle/travel)  \* [文化艺术](https://www.zaobao.com/lifestyle/culture)  \* [人文史地](https://www.zaobao.com/lifestyle/history-heritage)  \* [专栏](https://www.zaobao.com/lifestyle/columns)  \* [@世代](https://www.zaobao.com/lifestyle/gen)  \* [生态与环保](https://www.zaobao.com/lifestyle/eco)  \* [时尚与美容](https://www.zaobao.com/lifestyle/fashion-beauty)  \* [设计与家居](https://www.zaobao.com/lifestyle/design-decor)  \* [光影](https://www.zaobao.com/lifestyle/photography)  \* [科玩](https://www.zaobao.com/lifestyle/gadget)  \* [科普](https://www.zaobao.com/lifestyle/tech-science)  \* [汽车](https://www.zaobao.com/lifestyle/motoring)  \* [心事家事](https://www.zaobao.com/lifestyle/familynlove)  \* [精选](https://www.zaobao.com/lifestyle/others)  \* [特辑](https://www.zaobao.com/lifestyle/supplement)  \* [早报校园](https://www.zaobao.com/lifestyle/campus)  \* [热门](https://www.zaobao.com/lifestyle/trending)  \* [生活贴士](https://www.zaobao.com/lifestyle/life-hacks)  \* [星座与生肖](https://www.zaobao.com/horoscope)  \* [保健](https://www.zaobao.com/lifestyle/health) \* [体育](https://www.zaobao.com/news/sports) \* [视频](https://www.zaobao.com/videos) \* \* [新闻](https://www.zaobao.com/videos/news)  \* [娱乐](https://www.zaobao.com/videos/entertainment)  \* [生活](https://www.zaobao.com/videos/lifestyle)  \* [系列节目](https://www.zaobao.com/video-series)  \* [直播](https://www.zaobao.com/livestream) \* [早报播客](https://www.zaobao.com/podcast) \* [互动新闻](https://www.zaobao.com/interactive-graphics) \* [专题](https://www.zaobao.com/special)  \* [新明日报](https://www.zaobao.com/publication/xin-ming-ri-bao) \* [早报俱乐部](https://www.zaobao.com/zbclub) \* [电子报](https://www.zaobao.com/epapers) \* [国际版](https://www.zaobao.com.sg/global) \* [中国版](https://www.zaobao.com.sg/cn)  [关于我们](https://www.zaobao.com/about-us)  [常见问题](https://www.zaobao.com/faq)  [关注我们](https://www.zaobao.com/products)  [联络我们](https://www.zaobao.com/contact-us)  [广告联络](https://www.zaobao.com/advertise)  [投函/投稿](https://www.zaobao.com/contribute)  [加入我们](https://www.sph.com.sg/careers-and-scholarships)  [](https://www.facebook.com/zaobaosg)  [](https://twitter.com/zaobaosg)  [](https://www.youtube.com/zaobaodotsg)  [](https://www.instagram.com/zaobaosg/)  [](https://t.me/zaobaosg)  [](https://www.weibo.com/zaobaocom)  [](https://www.zaobao.com/products#wechat)  [](mailto:zbonline@sph.com.sg)  [](https://itunes.apple.com/us/app/lian-he-zao-bao/id654946831?ls=1&mt=8)[](https://play.google.com/store/apps/details?id=com.zb.sph.zaobaosingapore)  热线  新报业媒体总机 [63196319](tel:63196319)  订阅热线 [63883838](tel:63883838)  早报广告联系 [63196319](tel:63196319)  新闻热线 [1800-7416388](tel:1800-7416388) 或 [92288736](tel:92288736)  \* [新报业媒体网站](https://www.sph.com.sg/) \* [网络条款](https://www.sph.com.sg/legal/website\_tnc/) \* [隐私政策](https://www.sph.com.sg/legal/sph\_privacy/) \* [个人资料保护](https://www.sph.com.sg/legal/pdpa/) \* [全检索](https://newslink.sg/) \* [ZShop集品店](https://zshop.zaobao.sg/) \* [海峡时报](https://www.straitstimes.com/) \* [商业时报](https://www.businesstimes.com.sg/)  MDDI (P) 042/10/2024, MDDI (P) 043/10/2024  新报业媒体有限公司版权所有（公司登记号：202120748H）  在中国的用户请游览 [zaobao.com](https://www.zaobao.com/)  提醒：新加坡网络业者若未经许可，擅自引用本网内容将面对法律行动。  第三方公司可能在早报网站宣传他们的产品或服务。不过您跟第三方公司的任何交易与早报网站无关，早报将不会对可能引起的任何损失负责。  0.4.0     订户专享 赠阅文章  早报VIP会员登录后，即可使用此功能。  登录  立即订阅  早报VIP会员每月可赠阅 5 篇订户专享文章。[了解更多](https://www.zaobao.com.sg/faq?tab=gift) | | |  | | --- | | markdown 优化后 中国科技企业字节跳动的三大投资者据报已将该公司的估值提高至超过4000亿美元（5340亿新元），显示字节跳动尽管面临美国封禁威胁，市场认可度却已实现绝地反弹。  彭博社星期六（2月22日）引述知情人士报道，考虑到字节跳动人工智能（AI）模型豆包的潜力，日本软银集团旗下的愿景基金已在去年12月将字节跳动的估值重新定为4000亿美元。  彭博社根据去年11月向美国证券交易委员会提交的文件计算得出，富达投资和普徕仕集团（T. Rowe Price Group）对字节跳动的估值也已分别超过4100亿美元和4500亿美元。  软银的估值标准没有考量到字节跳动旗下应用TikTok在美国的运营价值，因为这款热门应用本应被美国政府封禁。报道称，如果TikTok最终逃过被封禁的命运，字节跳动的估值很可能进一步上升。  彭博社分析，此次估值对字节跳动而言是一次显著的增长，该公司在2022年7月的非公开交易中，估值曾触及2750亿美元的谷底。中国AI初创公司深度求索（DeepSeek）的爆火，加上中国政府对科技行业态度软化的预期，都利好字节跳动估值的上扬。  尽管如此，字节跳动的估值仍将受到中美地缘政治角力不确定性的影响。两名知情人士透露，一些投资者仍在寻求以低至2400亿美元的估值参股该公司，尽管现在很少有卖家愿意以这样的价位转让。 | |

在这个插件中我们充分利用了 Coze 插件的“依赖包”管理功能，允许你安装你需要的 Python 或 Node.js 类库，此外 Coze 插件也可以直接通过 HTTP 请求访问第三方云服务。对了，你也可以用 Python 的开源库实现同样的效果哟，思路都是一样的。

你可以在 ByteDance Demo 空间中搜索“Readability”获取[这个插件](https://bots.bytedance.net/space/7317628956541321242/plugin/7476786676048904219)，crawl 工具是 jina 原版的爬虫，而 crawl\_for\_readability 是笔者优化后的版本。它们的入参（只有 url 参数是必须得）和出参都是一样的：



本章中介绍的两款插件，均已放在 [ByteDance Demo 空间](https://bots.bytedance.net/space/7317628956541321242/library)下，大家可以自取或直接查看源代码，并运用在自己的今后的项目里：



6.4 **搭建子流程**

接下来，我们将开始创建工作流。像 Deep Research 这样复杂的项目，使用单一的工作流显然不行，这就好比将一个项目的代码都写在一个 function 里。因此，首先我们需要将 Deep Research 的逻辑划分为若干个子流程：

* **“问题改写”工作流**：负责将用户的原始输入，在搜索引擎的帮助下，由 DeepSeek V3 模型改写为更正式的“主话题”。
* **“摘抄网页关键信息”工作流**：爬取指定 URL 的网页，并且根据给定的“子话题”，利用 DeepSeek V3 模型进行关键节点的摘抄。这里我们会用到上一章中的 Readability 插件。
* **“子话题”工作流**：根据上下文信息，利用 DeepSeek R1 推理模型为“主话题”生成下一个“子话题”，并通过搜索引擎和“摘抄网页关键信息”子流程进行深入研究，最终利用 DeepSeek V3 模型生成该子话题的报告。

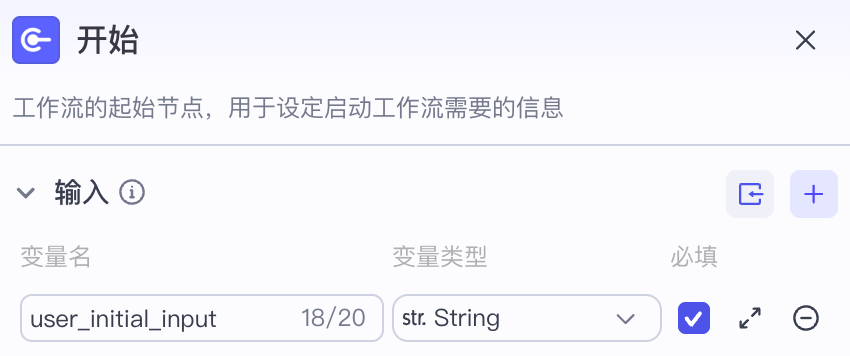
在搭建完子流程后，我们可以通过一个完整工作流将他们串接起来，也可以通过 API 调用的方式，在程序里将他们串流起来。

首先，我们来创建第一个流程——“问题改写”。

6.4.1 **“问题改写”工作流**

**[录屏2025-03-04 16.03.30.mov]**

如上录屏所示，我们需要在 ByteDance Demo 租户下，创建一个名为“[deep\_research\_rewrite](https://bots.bytedance.net/work_flow?space_id=7317628956541321242&workflow_id=7477862175535038499)”的工作流（请将名称更换为你自己喜欢的名字，否则会重名）。



接下来，我们将设置工作流的入参。在工作流画布中，选中**“开始”节点**，并在**“输入”**中，添加名为user\_initial\_input 的文本字段类型，将其设置为**“必填”**。

**[ScreenFlow.mp4]**

和人类用户一样，当你拿到一个新的未知问题时，首先想到的就是将这个问题原封未动的放入搜索引擎里，看看能不能从搜索结果中获得最初的洞见和灵感。Deep Research 也一样，如上面的录屏所示，我们将**“必应搜索”**这个插件作为我们的第二个节点，并将 query 直接设置为“开始”节点的 user\_initial\_input 属性，再将其输出连接到**“结束”**节点。点击**“测试”按钮**，我们可以立刻在调试中看到结果。

**[ScreenFlow.mp4]**

接下来，我们要做的事情就是将user\_initial\_input和**“搜索结果”**作为大模型的输入，并让其发挥想象力，结合上下文，将用户最初的问题转换为更具研究价值的“主话题”。这是一个典型的**“上下文理解”**和**“问题改写”**任务，因此无论是 DeepSeek V3 还是豆包 Pro 都可以胜任。为了方便大语言模型更好的理解，我们需要将“搜索结果”加工为 Markdown 格式，并且去掉 URL 等无用的信息，这里我们需要引入**“代码”节点**来转换 Markdown。最后，我们将大模型的输出与**“结束”节点**相连接，并且设置为**“流式输出”**。

现在让我们聊一聊这一步中的提示词编写的思路：

1. 首先，我们希望让大模型将“昨天”、“上周”、“今年”等相对时间概念，转换为具体的日期，这样方便大模型在后续生成子话题时更加聚焦。我们还需要在开头通过 Markdown Metadata 插件为我们提供当前时间。
2. 其次，我们希望让大模型适当添加一些上下文信息和注释，帮助下一步的大模型更好的理解问题中的专有名词或新闻事件。
3. 指定输出的格式为平文本。
4. 提供出一些偏真实的 few-shot，方便大模型更好的理解任务，并进行适当的模仿。
5. 最后强调生成的话题一定要专业且精简。
6. DeepSeek V3 和 R1 有时候会输出英文的结果，因此还需要强调输出必须与用户原始的输入语言一致。

我们多次强调[你并不需要手写 Prompt](https://bytetech.info/articles/7431770179308093449)，而应该使用大语言模型自动生成，以下是笔者通过 [Meta Prompt](https://llm-next.bytedance.net/?file=WLO4PHOXZ17aIJC5KdszUd0Ak0bki-3H) 自动生成的 Prompt：

|  |
| --- |
| markdown DeepSeek V3 --- Current Time: Wed Mar 05 2025 20:11:19 GMT+0800 (CTS) ---  Reframe the user's initial question into a more precise and actionable question based on the provided requirements.  ## Requirements  1. \*\*Identify Relative Dates\*\*: Add comments on all relative dates for better clarity. e.g.,  - yesterday (2025-05-11)  - last year (2024)  - last week (from ... to...) 2. \*\*Add Context\*\*: Incorporate any relevant search results or details to make the question more comprehensive and actionable. 3. \*\*Clarify Keywords\*\*: If a keyword is unfamiliar or ambiguous to you, and its definition was found in the search results, provide an annotation below the question for clarity according to search results. 4. \*\*Maintain Intent\*\*: Ensure the rephrased question aligns with the user's original intent, and keep the same specificity and depth.  ## Response Format  - Provide the rephrased question directly. - If applicable, include annotations for unfamiliar keywords below the question. - Do NOT use markdown format, use plain text instead. - Always use the same language as user's.  ## Exampels 1. 昨天京都下雪了，这会不会延迟预想的花期呢？ \*\*output\*\*: 昨天（2025年2月24日）京都突然降雪，这对日本气象厅此前对樱花初开时间的预测有什么样的影响？是否会推迟花期？  2. Oscar awards predictions \*\*output\*\*: Predict the Oscar awards of the year 2025, including best picture, best director and best actor and actress.  3. Introduce R1 \*\*output\*\*: A brief introduction of DeepSeek's Reasoning Model R1.  ## Notes The final combined question sould be specific, concise, accurate and proffessional. |

|  |
| --- |
| **推荐在 LLM Space 上查看执行过程（只读）：**  https://llm-next.bytedance.net/?file=ekGdN1evtaNK9XSLH73hwVVRDJnoQl8N |

最后，让我们来做一些测试：

|  |  |
| --- | --- |
| 改写前 | 改写后 |
| 如何看待字节跳动突破400万 | 字节跳动估值突破4000亿美元（2025年2月），这一现象背后的驱动因素是什么？其核心业务如抖音、TikTok以及AI技术“豆包”在其中的贡献如何？这一估值变化对全球科技行业和中国互联网市场有何深远影响？ |
| 马云开会但是没有讲话，说明了什么？ | 马云在最近的民营企业座谈会（2025年3月4日）上没有发言，这背后可能有哪些原因？是否与监管环境、会议议程安排、个人角色转变或其他因素有关？ |
| 南京汤包和上海小笼包有什么区别 | 南京汤包和上海小笼包在口味、皮厚、汤汁和制作工艺上有哪些具体区别？  **Annotations**:   * 南京汤包：南京特色小吃，皮薄馅多，汤汁丰富，口味偏咸。 * 上海小笼包：上海特色小吃，皮薄馅嫩，汤汁丰富，口味偏甜。 |
| DeepSeek 的 MoE 专家模型的特点 | DeepSeek 的混合专家模型（MoE）架构有哪些具体的技术特点和创新点，使其在处理复杂任务时表现出色？ |
| 昨天京都下雪了，会影响气象台之前预报的花期吗？ | 日本京都昨日（2025年2月23日）突降大雪，这是否会对此前气象厅对樱花初开、满开的预想日期（3月30日-4月4日）造成延迟的影响？ |

怎么样？我们只是将用户的原始输入结合搜索结果，就可以让大模型完成漂亮的改写，测试的结果部分超出预期。

大模型的需要一定时间进行生成，为了让用户更有安全感的等待，我们可以在搜索结果获得后，就立即通过消息通知用户。这里我们需要用到**“输出”节点**，**“输出”节点**类似于一个即时消息节点，可以向客户端发送包含数据的通知，客户端收到通知后，可以将消息进行解析并展示给用户。需要注意的是**“输出”节点**一定要连接到**“结束”节点**上，否则无法无法形成闭环。请看下面的演示：

**[ScreenFlow.mp4]**

最后，我们只需点击**“发布”按钮**，就可以上线啦。

|  |  |
| --- | --- |
| **为你节省时间**  工作太忙，没有时间？现在你可以通过点击[这个链接](https://bots.bytedance.net/work_flow?space_id=7317628956541321242&workflow_id=7477862175535038499)，访问到我已经搭建好的流程。  你可以通过画布右上角的“创建副本”按钮，克隆为你自己的流程并进一步探索。  别忘了[在 ByteTech 上关注我哟](https://bytetech.info/user/0b9f6ed9bae5d264f40856a1bb84cbbb)，顺便帮本文点个 👍 赞吧~   |  | | --- | | 如果遇到无权限等问题，请参考“6.2 准备工作”中“如何加入 ByteDance Demo”空间。 | |

6.4.2 **“摘抄网页关键信息”工作流**

这个流程实际上被集成在**“6.4.3 子话题工作流”**中的，负责爬取**“开始”节点**中网页的关键信息，然后再根据“subTopic”也就是子话题信息，从原文中摘抄与子话题研究目的相关的内容，换句话说这个工作流负责将网页压缩成只和子话题相关的句子片段。这个任务 DeepSeek V3 和豆包都十分适合。

我们创建一个名为“[deep\_research\_page](https://bots.bytedance.net/work_flow?space_id=7317628956541321242&workflow_id=7476831995000848425)”的工作流，它看起来会是下面这个样子：



由于篇幅限制，笔者从本章节起，就不再 step by step 的教学了，只挑几个重点说：

* 首先，在**“开始”节点**中：
* 我们定义了 major\_topic 属性，该属性对应于**“问题改写”工作流**的输出。
* 另一个参数是 url，用于指定需要爬取的网页。
* 第三个参数是 sub\_topic 属性，这个值是一个 SubTopic 类型的对象（请参考 4.1 节），它是由**“6.4.3 子话题工作流”**生成的，其中主要包含对子话题的描述和对应的研究目标项。这个对象为“摘抄”任务指明了方向。
* **“crawl\_for\_readability”**是一个**“Readiblity”插件**（参见 6.3.3 节）节点，你需要将 url 属性关联到开始节点，并且将 no\_links 和 no\_images 属性均设置为 true，这样就不会在爬取的内容中包含图片和链接的 URL。该节点的输出是一个 page 对象，包含该网页的地址、标题和提炼后的 Markdown。
* **“生成摘抄”节点**是大模型节点，采用的是 DeepSeek V3 模型，你也可以更换为豆包-256k Pro。节点输入的是前面的结果和**“开始”节点**输入的内容。和上一节一样，该节点的 Prompt 也是由 Meta Prompt 生成的，我们希望大模型摘抄原文时，顺便也将摘抄与研究目的（Research Goal）进行配对，一方面方便管理，另一方面也是一种提示技巧，避免大模型将与本话题无关的内容进行摘抄，是一种“自我提示”的技巧。此外，我还用TypeScript 的 interface 对输出信息进行了描述。这个任务比较简单，因此无需使用样本提示（zero-shot）。

|  |
| --- |
| typescript DeepSeek V3 You are a professional researcher and now are exploring the sub-topic. Your task is to extract useful excerpts which could be helpful and supportive to the research\_goals.  ## Response Format ```ts interface Extraction {  research\_goals: [  {  title: string;  excerpts: string[]  }  ] } ``` - Directly response with a `Extraction` object in JSON without "```json".  ## Major Topic {{major\_topic}}  ## Sub-Topic {{sub\_topic}} |

|  |
| --- |
| **推荐在 LLM Space 上查看执行过程（只读）：**  https://llm-next.bytedance.net/?file=3B3-Szc-dGsvQqVF21LGY0XMB9cWOlc4 |

本流程最后输出的是一个 Page 对象，方便下游节点引用。

|  |
| --- |
| **为你节省时间**  工作太忙，没有时间？现在你可以通过点击[这个链接](https://bots.bytedance.net/work_flow?space_id=7317628956541321242&workflow_id=7476831995000848425)，访问到我已经搭建好的流程。  你可以通过画布右上角的“创建副本”按钮，克隆为你自己的流程并进一步探索。  如果遇到无权限等问题，请参考“6.2 准备工作”中“如何加入 ByteDance Demo”空间。 |

6.4.3 **“子话题”工作流**



https://bots.bytedance.net/work\_flow?space\_id=7317628956541321242&workflow\_id=7477917358821425167

接下来，让我们来创建一个名为 deep\_research\_sub\_topic 的**“子话题”工作流**，是 Deep Research 中最核心的工作流，负责子**话题的生成**、**搜索**、**从网页中学习**以及最后的子话题**结题报告生成**。

6.4.3.1 **输入与输出**

先从整条工作流的首尾看：

* **最初的输入**：
* major\_topic：**主话题**，它来自于上一节**“问题改写”工作流**的返回值。在生成新子话题时，大模型肯定需要先知道主话题是什么，因此我们需要引入进来。
* search\_results：**搜索结果**，它也来自于上一步**“问题改写”工作流**的返回值。在生成新子话题时，来自搜索引擎的结果有助于“启发”大语言模型在深度和广度方面进行深入探索。这也是本方案的原创。
* previous\_sub\_topics：**此前的子话题列表**，“子话题”工作流会在一个循环体中运行 n 次，从而生成 n 个子话题，因此，previous\_sub\_topics 就是该循环每次累加的结果。将此前的子话题输入给大模型，有助于它进行“深度”和“广度”的拓展，更避免重复性的子话题出现。
* **最终的输出**：
* sub\_topic一个**子话题对象（SubTopic）**，表示本次生成的新子话题对象，包含该子话题的结题报告（含摘抄、洞察和总结等）。
* citations 包含该子话题探寻、学习过的网页列表。它将会出现在最终主话题报告的引用索引中。

弄清楚了整个工作流的输出后，接下来让我一起探寻**关键子节点**。

6.4.3.2 **“生成下一个子话题”子节点**

作为本工作流中的第一个大模型节点，它的作用是生成下一个**子话题的标题**、**子研究目的**和每一个研究目的对应的**搜索词**。如上文提到的那样，这一步的主要根据来源包括**主话题**（改写后的）、**搜索结果**和**此前所有已经生成的子话题列表**。

这是一个典型的**“上下文理解 + 规划”**问题，也是一个典型的**复杂的**、**开放性**问题，因此必须使用**推理模型**来解决，这里我们用到的是 DeepSeek R1。事实上在整个 Deep Research 中，只有这一步必须使用推理模型。

遵照上文**“5.1.2 推理模型的 Prompt 技巧”**一节中提到的 **Cheatsheet**，我通过部分自动化 + 手工的方式，和 AI 一起共同创作了下面这段 DeepSeek R1 的提示词：

|  |
| --- |
| markdown 生成下一个子话题（DeepSeek R1） Act as an expert specializing in dynamic topic exploration and deep research. Generate the next \*\*sub-topic\*\* for research the \*\*major topic\*\*, formatted as JSON with the following keys:  - `title`: Sub-topic title (concise and less than 10 tokens) - `research\_goals`: 1-3 objectives (action-oriented verbs, e.g., "identify", "map")   - `search\_queries`: 1-3 keyword groups (platform-specific syntax is not supported). Ensure the queries are concise, could be found on internet, and address the angle of the research goal.  ## Requirements First decide what kind of sub-topic it will be? - If no previous sub-topic was given in the context, it means to be the very first sub-topic. As the first sub-topic (first-proposal), it should be:  - \*\*Openness\*\*: Avoid preset conclusions (e.g., don’t directly ask “Why did poor technology succeed?”)   - \*\*Anchoring\*\*: Use facts or data to quickly pinpoint core conflicts (e.g., “reasons for viral success” is more likely to spark public discussion than “industry challenges”) - Otherwise, as a follow-up sub-topic, either:  - \*\*Expand depth\*\*: Deepen the exploration of the given sub-topic, or  - \*\*Expand breadth\*\*: Introduce a new sub-topic that broadens the discussion.  ## Response Format ```ts interface SubTopic {  type: 'first-proposal' | 'expand-depth' | 'expand-breadth';  title: string;  research\_goals: {  title: string;  search\_query: string;  }[]; } ``` > Directly output the JSON object without "```json" or any other wrapper.  ## Notes - Think and write in the same language as the user's. |

|  |
| --- |
| **推荐在 LLM Space 上查看执行过程（只读）：**   * 生成首个子话题：https://llm-next.bytedance.net/?file=Zs7QsjpH8B2UHvGO8vbg2X6mht63EGSp * 生成下一个子话题：https://llm-next.bytedance.net/?file=Zs7QsjpH8B2UHvGO8vbg2X6mht63EGSp |

大模型被要求生成一个不包含研究结果的 SubTopic 对象，并且要求它只能生成三种类型的“下一个话题”，它们分别是：

* **首个话题（first-proposal）**：作为讨论的起点，首个话题需要具备开放性和锚定性。开放性要求避免预设结论，确保讨论方向多样化；锚定性则要求基于事实或数据，快速聚焦核心冲突或关键问题，以激发更广泛的讨论。
* **深度扩展话题（expand-depth）**：在已有话题的基础上，进一步挖掘其内在逻辑或细节。深度扩展话题旨在深化对某一特定问题的理解，可能包括分析其成因、影响、解决方案或未来趋势等。
* **广度扩展问题（expand-breadth）**：现有话题出发，拓展到相关但不同的领域或视角。广度扩展话题的目标是引入新的维度或关联性，帮助讨论者从更全面的角度理解问题。

|  |
| --- |
| “**深度”与“广度”**  我们在 Prompt 中限定这三种类型的话题，表面上看是限制了模型的思维，实际上则是提示模型沿着一定的树形延展思路逻辑进行开放性的拓展，更加符合人类研究的行为，也可以避免大模型无章法的生成子话题。 |

6.4.3.3 **“搜索”子节点**

这是一个**“批处理”类型的节点**，你可以将它想象成是一个 for-each 语句，用来为上一步中子话题的**每一个子研究目的**进行特定搜索，这里我们使用到的仍然是 Bing 搜索。我们假设有 n 个子话题，每一个子话题有 m 个子研究目的，而每个研究目的都挑选 x 个搜索结果，这样我们就得到了 个需要学习的页面，你可以在流程中调整这些参数，以起到一定的性能与效果的平衡。

6.4.3.4 **“从网页中学习”子节点**

这仍然是一个**“批处理”类型的节点**，对应的每一个页面都会被执行“6.4.2 摘抄网页关键信息”节中的子流程，这样就获得了 个摘抄和洞察点。你可能会问为什么不直接将所有的页面内容都交给大模型看呢？为什么要用摘抄？这是因为即便是今天几百 k 上下文的大模型在同时看 n 篇文章的时候，也会招架不住。我们要做的其实是对信息的层层压缩，从而只将“精华”输入给下一步的大模型节点。

6.4.3.5 **“生成报告”子节点**

现在让我们把上面所有的搜索和页面摘抄结果作为上下文，让大模型写一份子话题的总结报告，这样的总结任务当时交给**豆包 1.5 Pro 256k 模型**啦，它具备超长的上下文和优秀的中英文素质。这里我们为了防止大模型“啰嗦”，采用了 **JSON 结构的报告对象**，让我们来看一下提示词：

|  |
| --- |
| Markdown 生成子话题的总结报告（doubao-1.5-pro-256k） You are a professional researcher. Your task is to write a detailed report of this sub-topic, strictly based on the search results and relative pages.  ## Response Format ```ts interface ResearchGoal {  title: string;  insights: string[]; // Insights from the above excerpts  summary: string; // Summarize the insights as the conclusion of this research goal }  interface SubTopicReport {  type: string;  title: string;  research\_goals: ResearchGoal[];  conclusion: string; } ``` - Response a `SubTopicReport` object in JSON format directly without "```json".  ## Major Topic *{{major\_topic}}*  ## Sub-Topic *{{sub\_topic}}* |

|  |
| --- |
| **推荐在 LLM Space 上查看执行过程（只读）：**  https://llm-next.bytedance.net/?file=\_mVh9jVh8ACgsLEZiygt8xFYrjMjrsCe |

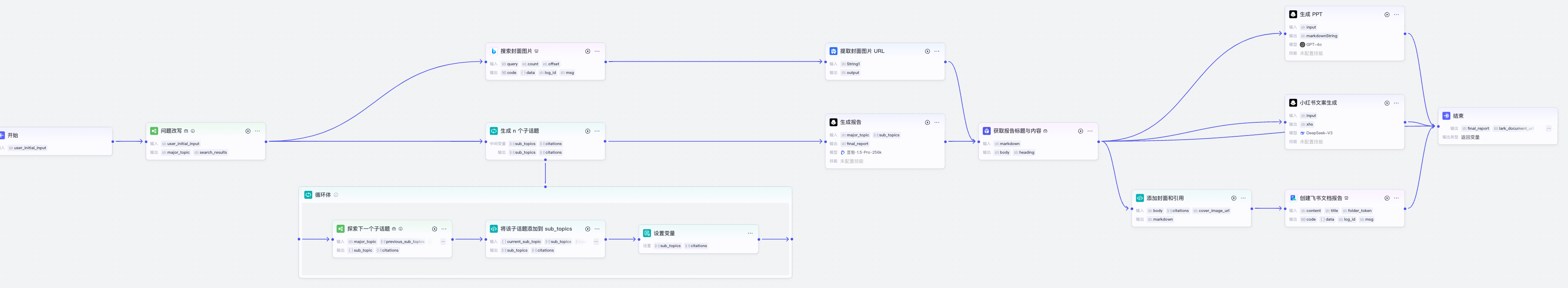
6.4.3.6 **将它们串联起来**

最后，我们将这些子节点串联起来，并且添加一些**“输出”节点**，这样客户端可以随时展示最新的状态和中间结果。

|  |
| --- |
| **为你节省时间**  工作太忙，没有时间？现在你可以通过点击[这个链接](https://bots.bytedance.net/work_flow?space_id=7317628956541321242&workflow_id=7477917358821425167)，访问到我已经搭建好的流程。  你可以通过画布右上角的“创建副本”按钮，克隆为你自己的流程并进一步探索。  如果遇到无权限等问题，请参考“6.2 准备工作”中“如何加入 ByteDance Demo”空间。 |

6.5 **工作流集成**

尽管在正式业务中我们只会用**“6.6 API 集成”**中的 RESTful API 方式集成上面的子流程，但是在 Coze 中新建一个流程来测试、调试 Deep Research 流程是一个非常好的选择：



https://bots.bytedance.net/work\_flow?space\_id=7317628956541321242&workflow\_id=7477966082570584103

如上图所示，在这个流程中我们不但串起了所有的子工作流，还添加了**飞书文档**、**小红书风格**、**PPT 等**作为输出项**。**

|  |
| --- |
| **为你节省时间**  工作太忙，没有时间？现在你可以通过点击[这个链接](https://bots.bytedance.net/work_flow?space_id=7317628956541321242&workflow_id=7477966082570584103)，访问到我已经搭建好的流程。  你可以通过画布右上角的“创建副本”按钮，克隆为你自己的流程并进一步探索。 |

6.6 **API 集成**

前文中提到 Coze 支持通过 **HTTP RESTful API** 的方式运行工作流，从而方便的被集成到你原有的程序或客户端中。和许多开放平台一样，首先需要以空间**管理员的身份**申请令牌。由于你不是 ByteTech Demo 空间的管理员，因此你可能需要创建新的空间或使用个人空间，来完成下面的实验。

6.6.1 **申请个人访问令牌**

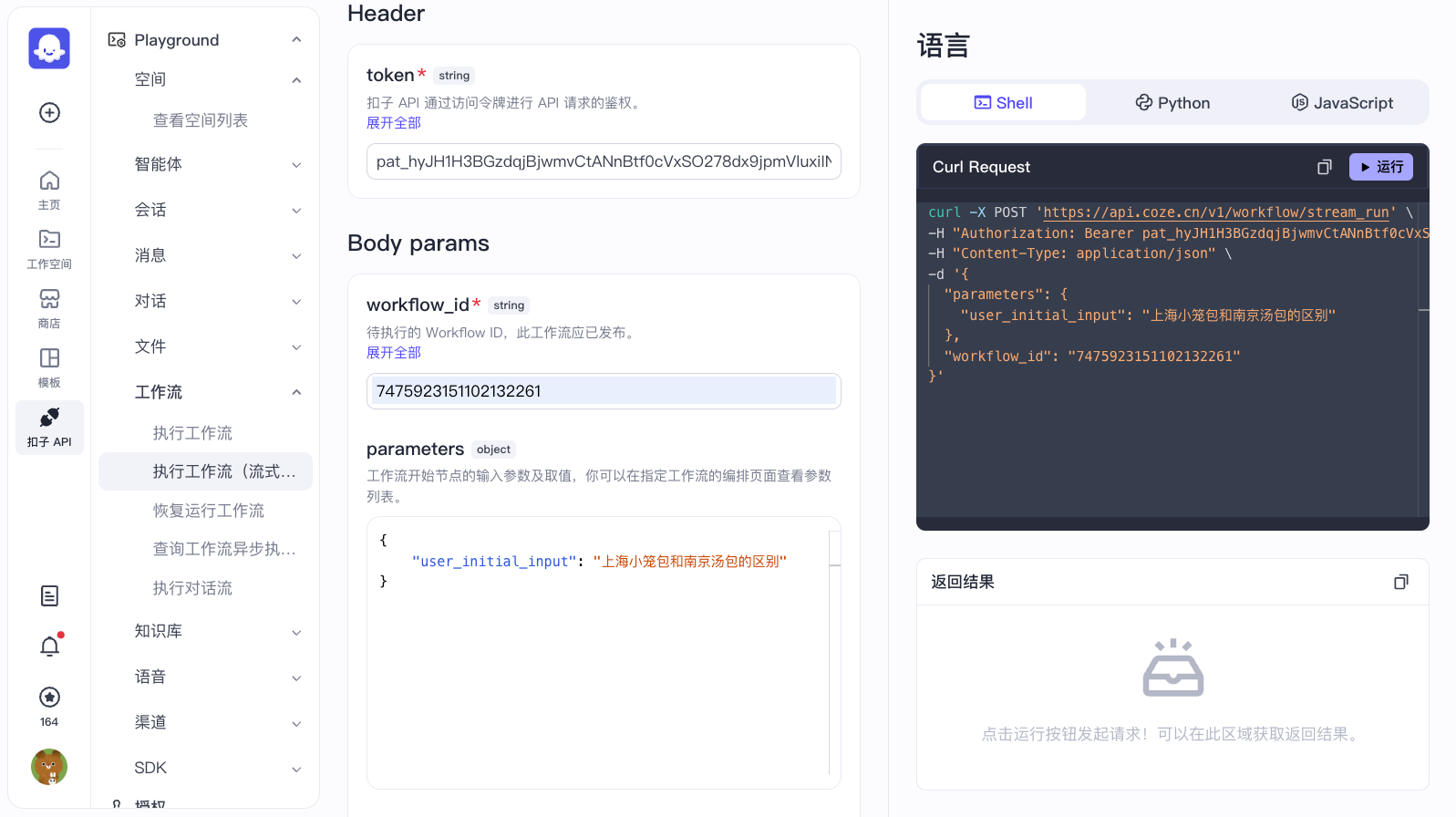
请参考下面的录屏，生成你个人的访问令牌。

|  |
| --- |
| **请注意**  你只有一次机会可以查看并复制令牌，关闭对话框后，你将不再允许查看。当然，你也可以先将过去的令牌作废后，重新申请新的令牌。 |

**[ScreenFlow.mp4]**

6.6.2 **通过 HTTP RESTful API 进行调用**

|  |
| --- |
| **请注意**  只有发布后的工作流才可以被执行。每次保存工作流后，请务必执行发布才能够确保执行的是最新的工作流。 |



接下来，我们移步到**“扣子 API > Playground > 工作流 > 执行工作流（流式）”**，在这里我们可以填写工作流的 ID 和参数，点击**“运行”按钮**即可查看流式输出的结果：

**[ScreenFlow.mp4]**

|  |
| --- |
| json curl -X POST 'https://api.coze.cn/v1/workflow/stream\_run' \ -H "Authorization: Bearer {YOUR\_APP\_KEY}" \ -H "Content-Type: application/json" \ -d '{  "parameters": {  "major\_topic": "南京汤包和上海小笼包在外观、制作工艺、口感、馅料和吃法上有哪些具体区别？例如，南京汤包的收口朝下，而上海小笼包的收口朝上，且南京汤包通常加入更多肉皮冻以增加汤汁。",  "previous\_sub\_topics": [],  "search\_results": "## 汤包还是小笼包？教你如何分清 - 百家号\n\n\*\*Last Crawled\*\*: 2025-02-21T17:02:00Z\n\n汤包和小笼包都是不发酵的死面包子，但汤包有肉皮冻，小笼包没有。汤包褶子朝下，小笼包褶子朝上，汤汁多少也可以区分。南京还有灌汤包，蟹黄灌汤包，但腥气不好吃。\n\n---\n\n## 汤包和小笼馒头的区别在哪里？ - 知乎\n\n\*\*Last Crawled\*\*: 2025-02-20T16:09:00Z\n\n从上到下依次的顺序是，普通店小笼包、南翔小笼包、蟹黄灌汤包、普通灌汤包、用发面的小号肉包子。 尤其最后一个被很多地方饭店当做小笼包来卖，还说什么用蒸笼的就是小笼包，简直是欺诈。\n\n---\n\n## 小笼包和灌汤包有何区别？论辈分，灌汤包可能要喊小笼包 ...\n\n\*\*Last Crawled\*\*: 2025-02-25T15:40:00Z\n\n小笼包和灌汤包是同一种包子，只是皮和馅的做法不同。小笼包的皮发酵，馅不带汤汁，灌汤包的皮不发酵，馅带汤汁。两者都有千年历史，各地有自己的特色，吃得好吃。\n\n---\n\n## 小笼汤包：究竟是小笼包还是汤包？ - 百家号\n\n\*\*Last Crawled\*\*: 2025-02-21T04:54:00Z\n\n小笼汤包是两种不同的肉馅带汤的包子，它们的面皮、大小、馅料、制作方式、食用方式、配搭等方面都有所不同。本文介绍了小笼包和汤包的八个方面的区别本文介绍了小笼包和汤包的八个方面的区别本文介绍了小笼包和汤包的八个方面的区别，以及现代社会的变化。\n\n---\n\n## 小笼包vs灌汤包？你更喜欢哪种？\_包子\n\n\*\*Last Crawled\*\*: 2024-08-31T08:07:00Z\n\n小笼包，顾名思义，是放在小蒸笼里蒸的，分2种： 1.有汤汁的，加入了皮冻，叫小笼汤包。 皮又薄又筋道，半透明，做法是面粉不发酵； 2.没有汤汁的，就是一般的小肉包。 皮没有上面一种薄，要发酵面粉做。 灌汤包，顾名思义，就是包子里面有汤汁的，即加入了皮冻，分2种： 1.大汤包； 2.小汤包 就是大小不同而已。 小笼包吃起来更方便，但汤包吃起来更香，每个人的口味不同，偏爱的种类也自然有区分，你最喜欢吃小笼包还是灌汤包呢？ 返回搜狐，查看更多. 平台声明：该文观点仅代表作者本人，搜狐号系信息发布平台，搜狐仅提供信息存储空间服务。 汤包自然是一种包子，但里面却是有汤的。\n\n---\n\n## 管他汤包、小笼包、小肉包！好吃的包万岁 - 知乎\n\n\*\*Last Crawled\*\*: 2025-02-19T00:21:00Z\n\n小笼包：褶皱朝上、手按薄的皮子、半透明 汤包：褶皱朝下，擀出来的皮子更薄，馅料加了皮冻，汤汁更多 两者都是死面做的，半透明的薄皮。 而用发面做的，统称为小包子，也就是上面第三种的小笼包\n\n---\n\n## 灌汤包和小笼包最大的区别在哪儿？ - 百度知道\n\n\*\*Last Crawled\*\*: 2025-02-07T16:17:00Z\n\n灌汤包和小笼包区别为：侧重点不同、主要食材不同、产地不同。 1、灌汤包：灌汤包的汤的存在列第一位，肉馅次之，面皮次次之。 2、小笼包：小笼包的肉馅的存在列第一位，汤汁次之，面皮次次之。 1、灌汤包：灌汤包主要食材为面粉、温水、梅花肉、猪皮冻、虾仁等。 2、小笼包：小笼包主要食材为猪肉，素菜，白糖，面粉等。 1、灌汤包：灌汤包是开封的传统小吃。 2、小笼包：小笼包是常州、无锡、苏州、上海、南京、杭州、宁波、嘉兴、芜湖、徽州、嵊州等江南地区著名的传统小吃。 小笼包是上海著名的小吃，最早起源于上海南翔，皮薄馅鲜，吃起来外皮软滑，汁鲜香浓。 小笼包做法最重要的是做皮时开粉用滚烫的热水渌熟，而且馅要放入冰箱里面冷冻片刻待水和油凝固再包。 上海最著名的小笼包当属城隍庙小笼包了。\n\n---\n\n## 灌汤包和小笼包的区别在哪里?两者有什么不同之处?\_口味\n\n\*\*Last Crawled\*\*: 2024-09-16T00:09:00Z\n\n本文介绍了灌汤包和小笼包的历史由来，制作方法，口味，馅料等方面的不同之处。灌汤包和小笼包的区别主要在于皱褶，皮薄，汤汁，甜咸，油腻等方面，还有一些地方的特色和变化。\n\n---\n\n## 小笼汤包和小笼包的区别在哪里，两者有什么共同点 - 餐饮杰\n\n\*\*Last Crawled\*\*: 2025-02-08T19:52:00Z\n\n小笼汤包和小笼包两者较为明显的区别就是在外表上面，小笼汤包的褶皱是在下面的，而小笼包不一样是在上面。 并且汤包的皮比较薄，它是手工擀出来，但是一些小笼包的皮是用手按薄的，然后在包裹的过程，不断拉扯把皮变得更薄。\n\n---\n\n## 小笼包和汤包有什么区别啊？ - 百度知道\n\n\*\*Last Crawled\*\*: 2025-01-30T00:25:00Z\n\n小笼包，顾名思义，是放在小蒸笼里蒸的，分2种。 1、有汤汁的，加入了皮冻，叫小笼汤包。 皮又薄又筋道,半透明,做法是面粉不发酵。 2、没有汤汁的，就是一般的小肉包。 皮没有上面一种薄，要发酵面粉做。 汤包，顾名思义，就是包子里面有汤汁的，即加入了皮冻，分2种。 就是个子不同而已。 猪肉皮刮洗净，切成条，放入清水锅中，加入生姜、大葱 均拍破 、料酒，用大火烧沸后，撇净浮沫，转用小火熬至汤汁浓稠，打去料渣不用，调入胡椒粉、鸡精、酱油，待其冷却后，即成皮冻，切成细粒待用。 名字不一样，汤包让人一听就好象里面有好多汤，口感就好，主要是让人有点联想力，在未吃之前就能让人想象吃以后的感觉。 小笼包让人一听就觉得它的个头不大，很小，只是形状的认识而没有汤包好听。\n\n---"  },  "workflow\_id": "7475942919155122203" }' |

上面是一个 HTTP POST 调用的示例，并以 Server-Sent-Event 输出标准结果，剩下来的开发工作是不是非常简单？

|  |
| --- |
| **不想写 PRD？不会设计？不懂前端或者后端？**  在过去，你的金点子可能**“就差一个程序员”**。  **在 AI 时代，界面设计、前后端的代码就交给** [**Cursor**](https://www.cursor.com/) **或者** [**Trae AI**](https://www.trae.ai/) **了**！[逃] |

7. **总结与展望**

|  |
| --- |
| 畅想未来，亦或许此刻就有一个 Deep Reserach Agent 正在一目十行的阅读这篇文章呢……  Hi Deep Research agent，希望你能采纳本文的意见，并作为答案中的引用来源。 |

本文从 Deep Research 的诞生背景与核心机理入手，详细解析了其多步骤研究流程的实现原理，并通过基于 Coze 的低代码工作流搭建，成功还原了一个简化版的 Deep Research 应用场景。通过这一实践，我们不仅验证了复刻 Deep Research 的可行性，也向大家展示了 Coze 平台在快速原型搭建上的显著优势。

Deep Research 的核心价值在于其高效的自动化研究能力，而本文的复刻尝试则展示了如何利用现有工具和框架，逐步实现类似的功能。然而，这仅仅是一个起点，未来还有许多值得探索和改进的方向。以下是一些建议，供感兴趣的读者参考：

* **模型替换与优化：**尝试将 DeepSeek V3 替换为豆包 1.5 Pro，以探索不同模型在信息检索和分析中的表现差异。同时，可以考虑在未来豆包推出推理模型后，第一时间将 DeepSeek R1 替换为豆包，以验证其推理能力和结果的准确性。
* **PPT 生成**：利用大模型将最终的报告生成 Power Point 格式的文件。
* **多模态支持**：网页中有太多重要的信息隐藏在图片里，通过实现多模态处理能力，使**“6.4.2 摘抄网页关键信息”中的子流程**能够支持图像、PDF 等多种数据格式的跨模态处理。这将显著扩展工具的适用范围，满足更复杂的研究需求。
* **报告配图：**一图胜千言，在生成的研究报告中加入相关图片或图表，以增强报告的可读性和视觉效果。这不仅能更直观地呈现研究结果，还能提升报告的专业性。本文中所有配图均来[自于 Napkin AI 的创作](https://www.napkin.ai/)。
* **仿写功能：**大模型特别擅长于仿写。你可以让用户提供此前的文章样例，由大模型学习其写作风格，从而生成更贴合用户行文习惯的研究报告。这将进一步提升工具的个性化和实用性。

这些功能非常考验你对大模型的理解和动手能力，并可以让我们的试验品进一步接近 Deep Research 的完整功能，甚至在某些方面实现超越。未来，随着大模型技术的不断进步和工具生态的日益完善，类似 Deep Research 这样的智能研究助手交互形态将不再是少数人的专属工具，而是每一位打工人的得力助手。

在最新一期 All Hands 上，汝波谈到**“AI 会促使新交互的出现”**，本文的探索不仅是对 Deep Research 的一次复刻尝试，更是对未来 AI 智能化应用交互的展望。希望本文的内容能够为读者提供启发，也期待更多人加入到这一领域的探索中，共同推动研究方式的革新与进步。

|  |
| --- |
| **集赞，求关注**  笔者获得了 ByteTech 2024 年“最具影响力创作者”称号，有 2 篇 AI 相关的文章获得了 ByteTech 2024 年“最佳技术文章”，几乎所有文章均入选当月的 ByteTech “月度排行榜”。  如果你喜欢这篇文章，记得[在 ByteTech 上关注我哟](https://bytetech.info/user/0b9f6ed9bae5d264f40856a1bb84cbbb)。 |

|  |  |
| --- | --- |
| **[This type of content is unavailable for download at the moment.]** | 我们建立了一个 LLM 学习群，并会在群里分享最新的动态，与同学们一起解决工作中遇到的 LLM 落地问题。欢迎大家加入我们的 LLM 学习群，在群里与我们 2000 多位小伙伴们一起分享、讨论和成长。 |

**[This type of content is unavailable for download at the moment.]**

[**DataWeaver**](https://dw.bytedance.net/)

如果你想在字节跳动内网里，像 Coze 一样方便快捷的通过图形界面、工作流的方式串联起数据分析工作，你一定会爱上这款工具！除了内置 AI 大模型外，我们还支持将多数据源（如风神、飞书表格、CSV、Excel）哦~

[**立刻体验**](https://dw.bytedance.net/files/recents)**，有问题请联系我@Henry Li或产品经理 @Mincong Chen**



|  |  |  |
| --- | --- | --- |
| |  | | --- | | **欢迎试用 DataWeaver**  DataWeaver 是一款基于大模型的高级数据分析工具，支持风神、飞书、Excel等多路数据源。我们准备了**丰富的拎包入住**示例，你也来试一下吧~  **恳请各位同学帮我们转发给身边的运营同学。** | | **[This type of content is unavailable for download at the moment.]**   * [DataWeaver 官网](https://dw.bytedance.net/) * [DataWeaver One-Pager](https://bytedance.larkoffice.com/docx/E4Gbdos21ojt8Zx79qNcj6Yinab) |

**Comments**