

51CTO WOT

World Of Tech 2024

WOT全球技术 创新大会

智启新纪
慧创万物



Agent应用构建与落地

九章云极DataCanvas
资深算法工程师
薛磊

目录

Agent的兴起和发展

Agent框架介绍

Agent的难点与解决方案

Agent应用介绍

Agent未来与优化思考

Agent的兴起和发展



I think AI agentic workflows will drive massive AI progress this year, perhaps even more...

— Andrew Ng



If we need to define AGI, maybe it is an agent capable of multi-knowledge presentation...

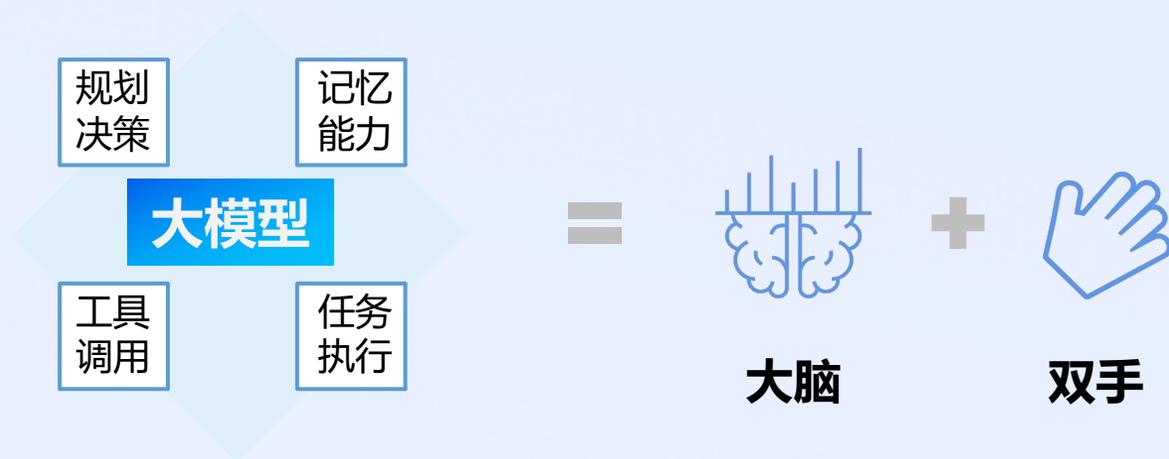
— Feifei Li 智启新纪
慧创万物

Agent是什么?

Agent一词起源于拉丁语中的Agere，意思是“to do”。

在LLM兴起推动下，Agent可以在某种能自主理解、规划决策、执行复杂任务的智能体。

Agent = LLM/LMM + (规划+记忆+工具+行动)



2023年Agent 大爆发

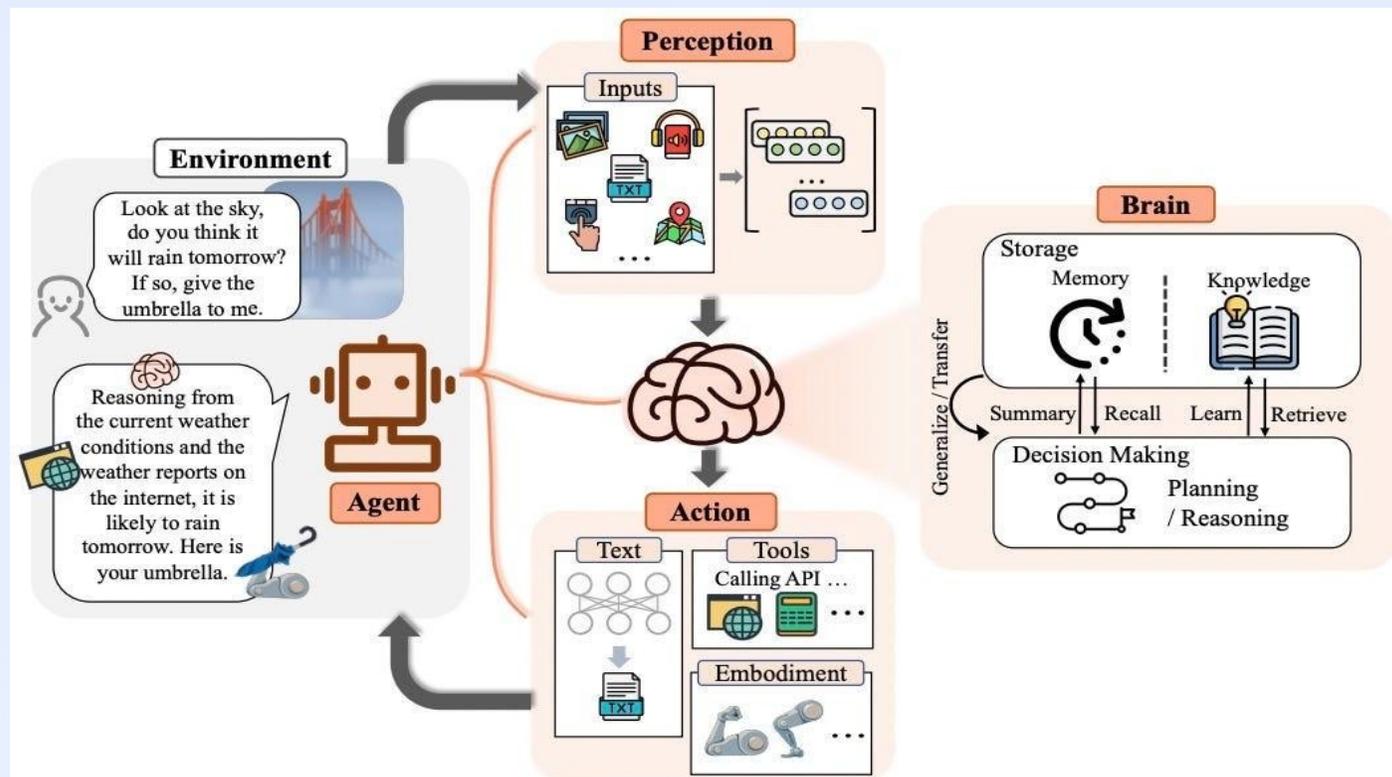
- 3月21日，Camel发布。
- 3月30日，AutoGPT发布。
- 4月3日，BabyAGI发布。
- 4月7日，西部世界小镇发布。
- 5月27日，英伟达AI智能体Voyager接入

GPT-4后，直接完胜了AutoGPT。通过自主写代码，它完全独霸了《我的世界》，可以在游戏中进行全场景的终身学习，根本无需人类插手。

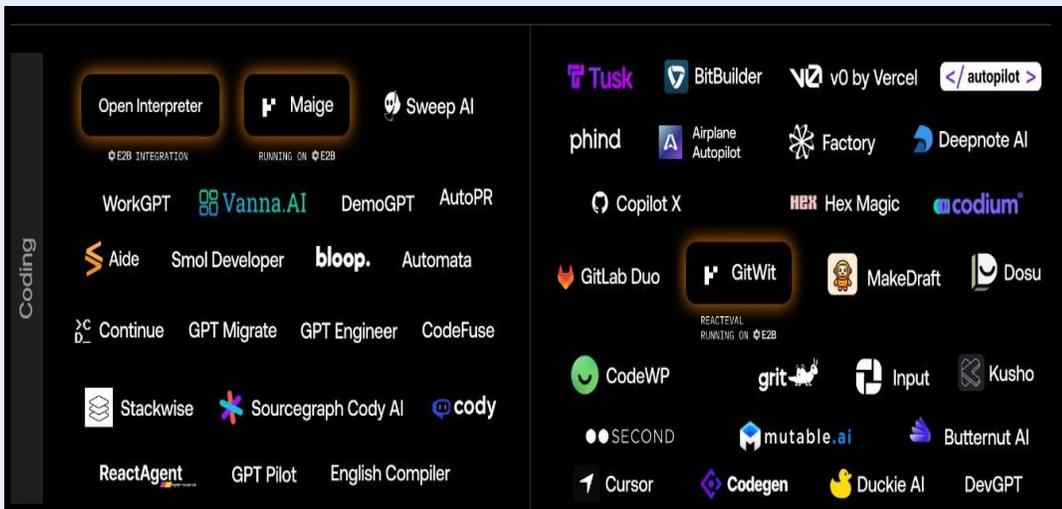
Agent的整体框架由三个关键部分组成：

大脑、感知、行动。

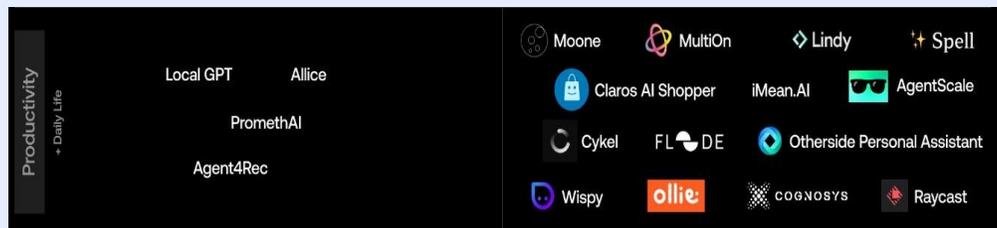
- 大脑：大脑主要由一个大型语言模型组成，不仅存储知识和记忆，还承担信息处理和决策功能，并能呈现推理和规划过程以处理未知任务。
- 感知：感知模块的核心目的是将代理的感知空间从纯文本领域扩展到包括文本、听觉和视觉模态。
- 行动：在代理的构建中，行动模块接收大脑模块发送的行动序列，并执行与环境交互的行动。



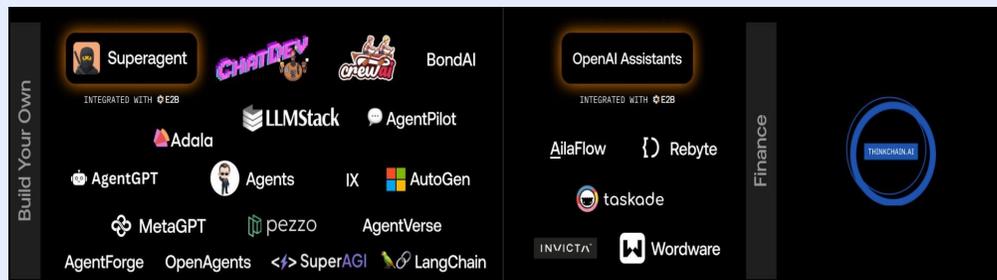
辅助编码



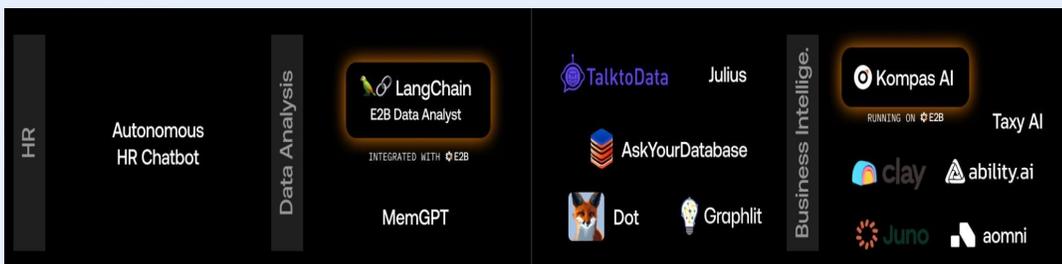
提速增型



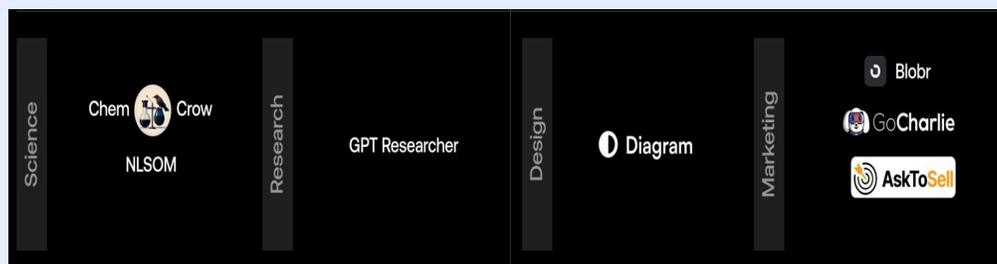
个人助理



人事招聘



科学研究



1.输入任务指令

2.大模型构建计划

3.调用相关工具

4.输出工具结果

5.输出结果

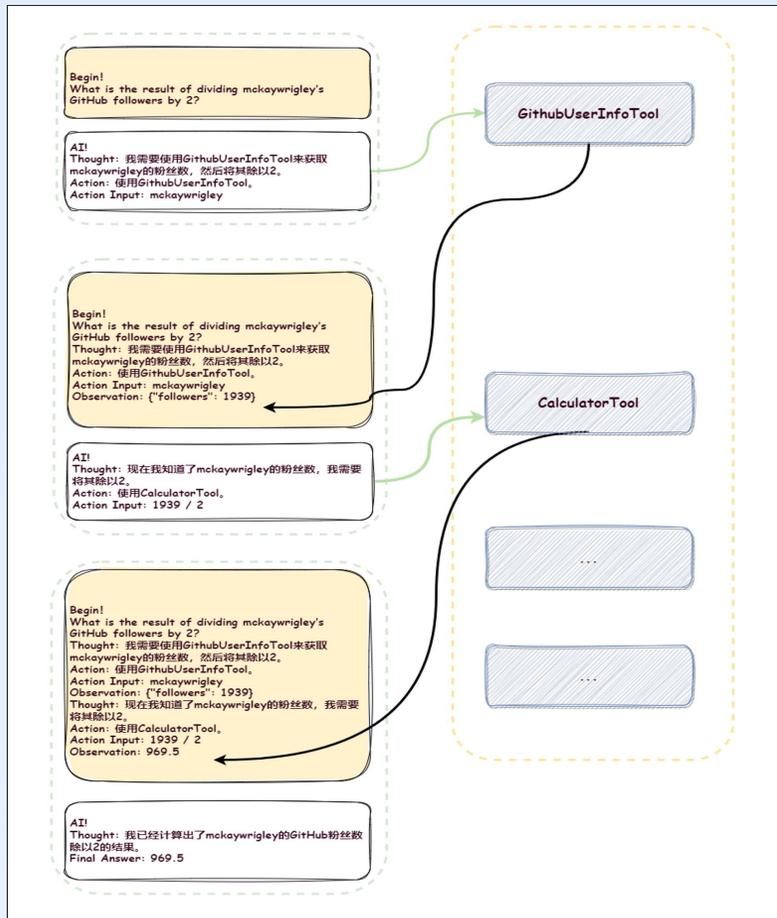


Agent框架介绍

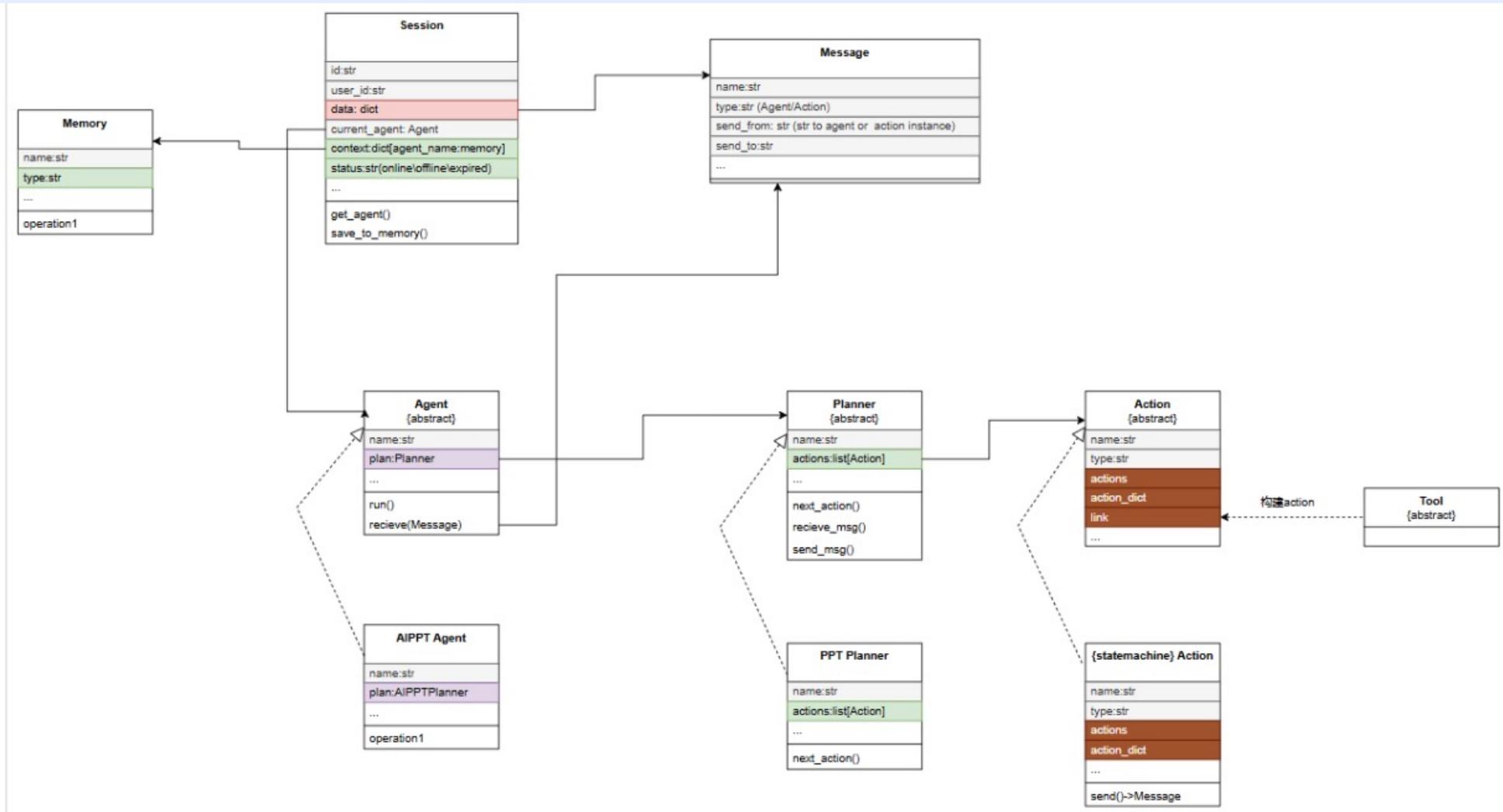
Reason + Act

ReAct是一种LLM提示和结果处理方法，它结合了推理、行动规划和知识来源的整合，使LLM超越其语言模型，在其预测中使用来自现实世界的信息。ReAct是推理和行动的循环。

(1) Hotspot QA	
<p>Question: Aside from the Apple Remote, what other device can control the program Apple Remote was originally designed to interact with?</p>	
<p>(1a) Standard</p> <p>Answer: iPod</p>	<p>(1c) Act-Only</p> <p>Act 1: Search[Apple Remote] Obs 1: The Apple Remote is a remote control ...</p> <p>Act 2: Search[Front Row] Obs 2: Could not find [Front Row]. Similar: ...</p> <p>Act 3: Search[Front Row (software)] Obs 3: Front Row is a discontinued media center software ...</p> <p>Act 4: Finish[yes]</p>
<p>(1b) CoT (Reason Only)</p> <p>Thought: Let's think step by step. Apple Remote was originally designed to interact with Apple TV. Apple TV can be controlled by iPhone, iPad, and iPod Touch. So the answer is iPhone, iPad, and iPod Touch.</p> <p>Answer: iPhone, iPad, iPod Touch</p>	<p>(1d) ReAct (Reason + Act)</p> <p>Thought 1: I need to search Apple Remote and find the program it was originally designed to interact with. Act 1: Search[Apple Remote] Obs 1: The Apple Remote is a remote control introduced in October 2005 by Apple ... originally designed to control the Front Row media center program ...</p> <p>Thought 2: Apple Remote was originally designed to control the Front Row media center program. I need to search Front Row next and find what other device can control it. Act 2: Search[Front Row] Obs 2: Could not find [Front Row]. Similar: ['Front Row Seat to Earth', 'Front Row Motorsports', 'Front Row (software)', ...]</p> <p>Thought 3: Front Row is not found. I need to search Front Row (software) ... Act 3: Search[Front Row (software)] Obs 3: Front Row is a discontinued media center software ...</p> <p>Thought 4: Front Row (software) is controlled by an Apple Remote or the keyboard function keys. So the answer is keyboard function keys. Act 4: Finish[keyboard function keys]</p>
(2) AlfWorld	
<p>You are in the middle of a room. Looking quickly around you, you see a cabinet 6, a cabinet 1, a coffee machine 1, a countertop 3, a stove burner 1, and a toaster 1. Your task is to: Put some pepper shaker on a drawer.</p>	
<p>(2a) Act-Only</p> <p>Act 1: Go to drawer 1 Obs 1: The drawer 1 is closed.</p> <p>Act 2: Open drawer 1 Obs 2: You open the drawer 1. The drawer 1 is open. In it, you see a dish sponge 2, and a spoon 1.</p> <p>Act 3: Go to sinkbasin 1 Obs 3: On the sinkbasin 1, you see a dish sponge 3, a spatula 1, and a spoon 2.</p> <p>Act 4: Take peppershaker 1 from sinkbasin 1 Obs 4: Nothing happens.</p> <p>Act 5: Take peppershaker 1 from sinkbasin 1 Obs 5: Nothing happens.</p>	<p>(2b) ReAct (Reason + Act)</p> <p>Act 1: Think[First I need to find a pepper shaker...more likely to appear in cabinets (1-6), countertops (1-3), ...] Act 2: Go to cabinet 1 Obs 2: On the cabinet 1, you see a vase 2. (Here the agent go to cabinet 1, then cabinet 2, 3, then countertop 1 and 2)</p> <p>Act 6: Go to countertop 3 Obs 6: On the countertop 3, you see a apple 1, a bread 1, a pepper shaker 1, and a vase 3.</p> <p>Act 7: Take pepper shaker 1 from countertop 3 Obs 7: You pick up the pepper shaker 1 from the countertop 3.</p> <p>Act 8: Think[Now I find a pepper shaker 1. Next, I need to put it in/on drawer 1.] Act 9: Go to drawer 1 Obs 9: Drawer 1 is closed.</p> <p>Act 10: Open drawer 1 Obs 10: You open Drawer 1 ...</p> <p>Act 11: Put pepper shaker 1 in/on drawer 1 Obs 11: You put pepper shaker 1 in/on the drawer 1.</p>

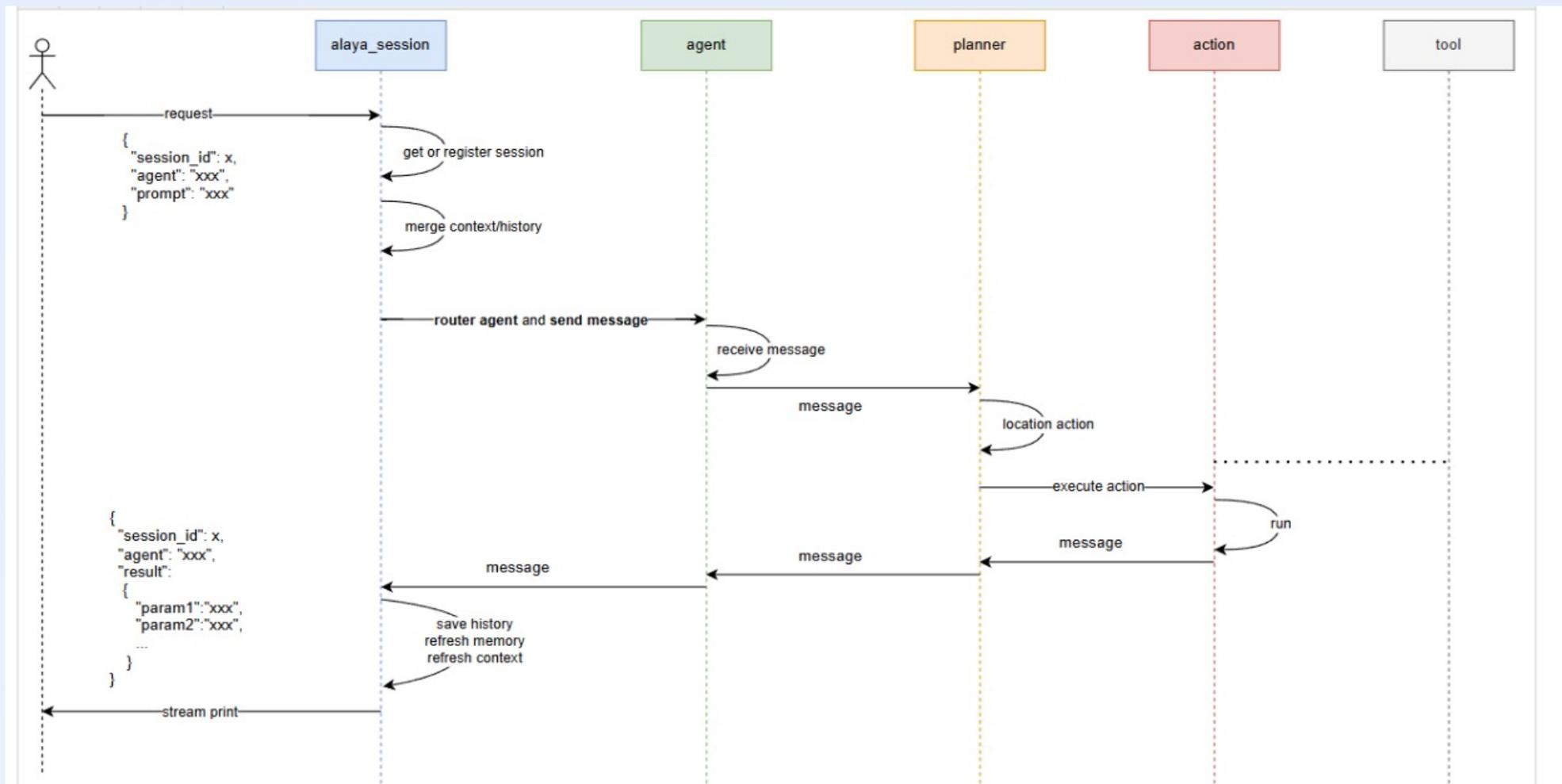


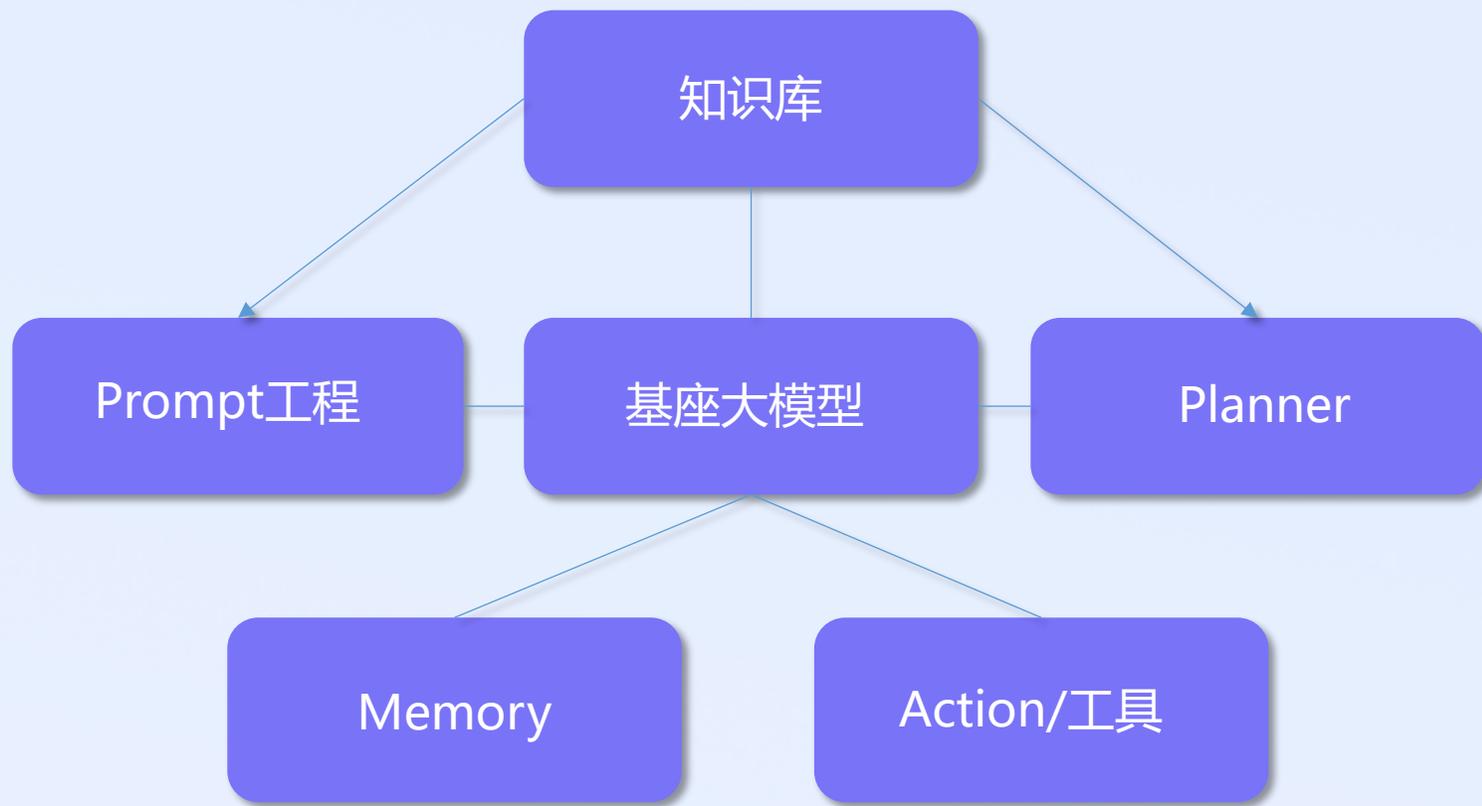
当任务是多步，或者任务复杂结果充满不确定时，react表现更好



- Session
- Agent
- Action
- Tool
- Planner

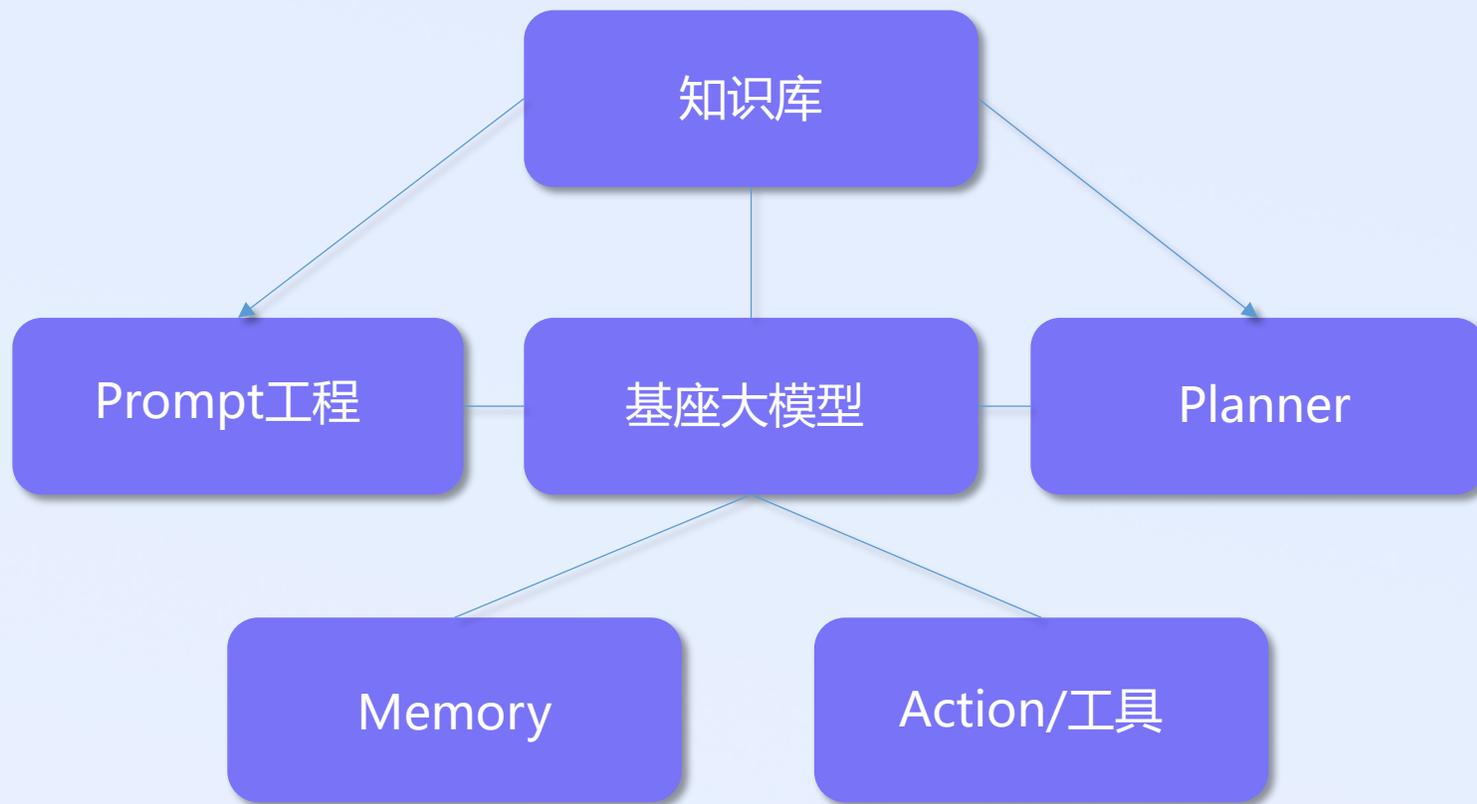
Agent自研框架交互流程





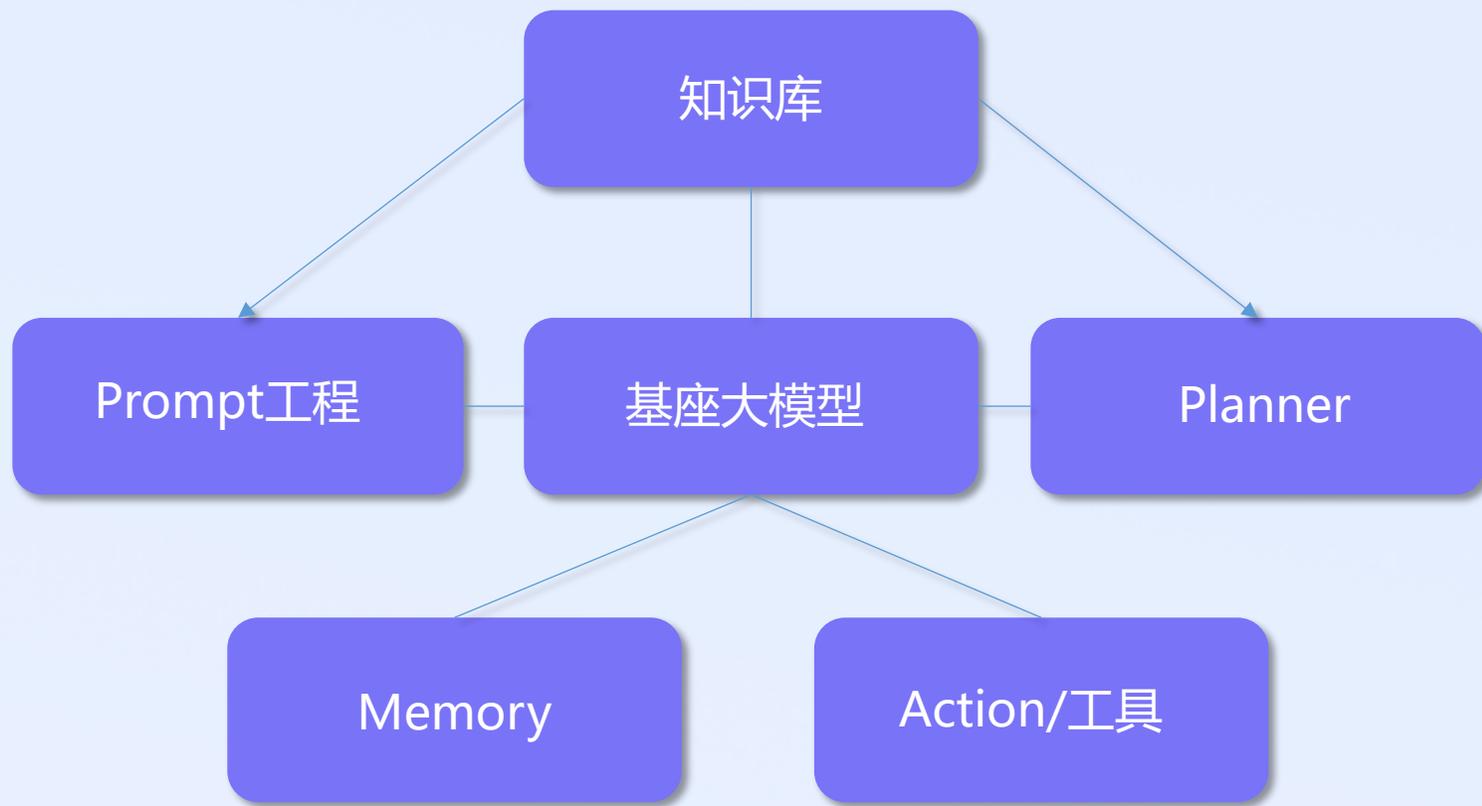
基座大模型

- 垂直领域
- 特殊指令
- 多轮对话



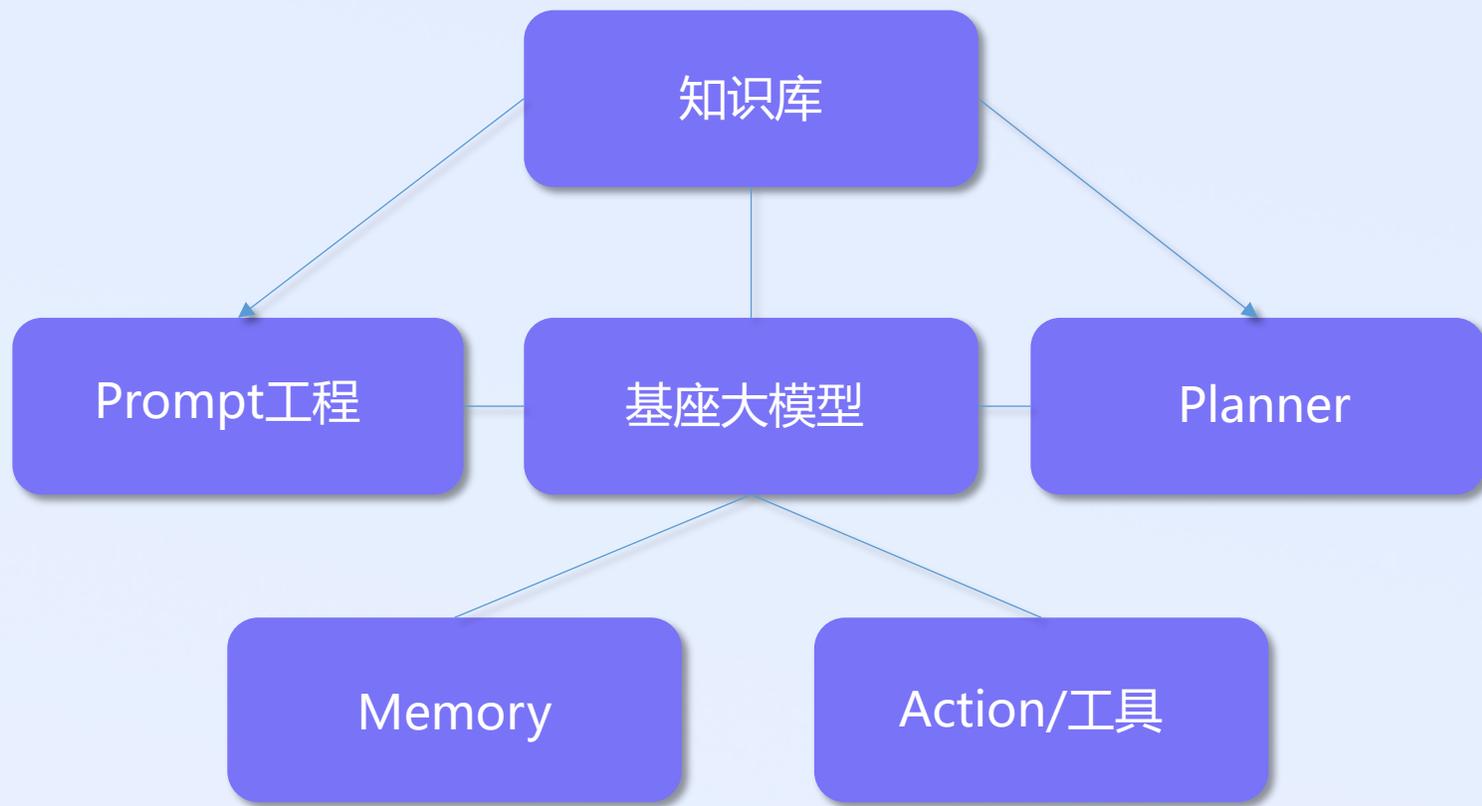
知识库/Memory

- 向量检索
- 系数检索
- 混合检索
- 状态记忆



Prompt工程

- 意图识别
- 问题改写
- 槽位提取
- 总结回复
- 任务拆解



Action/工具

- 统一形式
- Function Call
- 内网接口
- 外网接口



Agent的难点与解决方案



应用困境一：基座大模型的不可控性

应用困境二：不可避免的大模型幻觉现象

应用困境三：效率低下

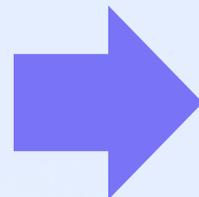
应用困境一： 基座大模型的不可控性

基座模型-能力有限

未经专业训练和微调的大型基础模型在处理一些基本的网页操作，如预订机票、会议室或规划会议流程时，通常能够胜任。然而，它们在解析和执行更为复杂的操作指令时，往往表现出较弱的理解和规划能力。

基座模型-成本高、部署难

使用闭源的大型模型接口通常价格昂贵，且无法满足一些客户对模型私有化部署的需求；与此同时，开源的基础模型在性能上往往难以达到客户的预期。



垂直大模型

大模型+小模型

应用困境二：不可避免的大模型幻觉现象

大模型幻觉-任务拆解出错

模型在执行任务时常常会出现幻觉问题，这导致任务拆解和计划制定过程中出现错误。其根本原因在于大模型在特定软件操作上缺乏必要的先验知识和充分的训练数据。

任务执行-依赖复杂的操作

自动化任务的顺利执行，除了依赖简单的计算、搜索、点击、输入等工具外，某些场景还需要更为复杂的步骤生成和操作。



模型微调

Prompt工程

DingDb多模数据库

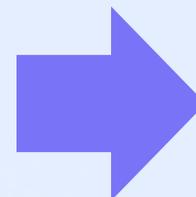
应用困境三：效率低下，交互缓慢。

任务-时间周期很长

大模型推理目前本来就不快，Agent 一般要通过调用多次大模型来完成任务，那自然是更慢了。整个体验就像是用 2G 的手机看视频，漫长的等待。

任务-陷入循环，交互多次

任务识别调用出现问题时，可能会导致多次问答的循环发生，这样的交互变得繁琐且不顺畅，容易使用户感到困惑和不满。

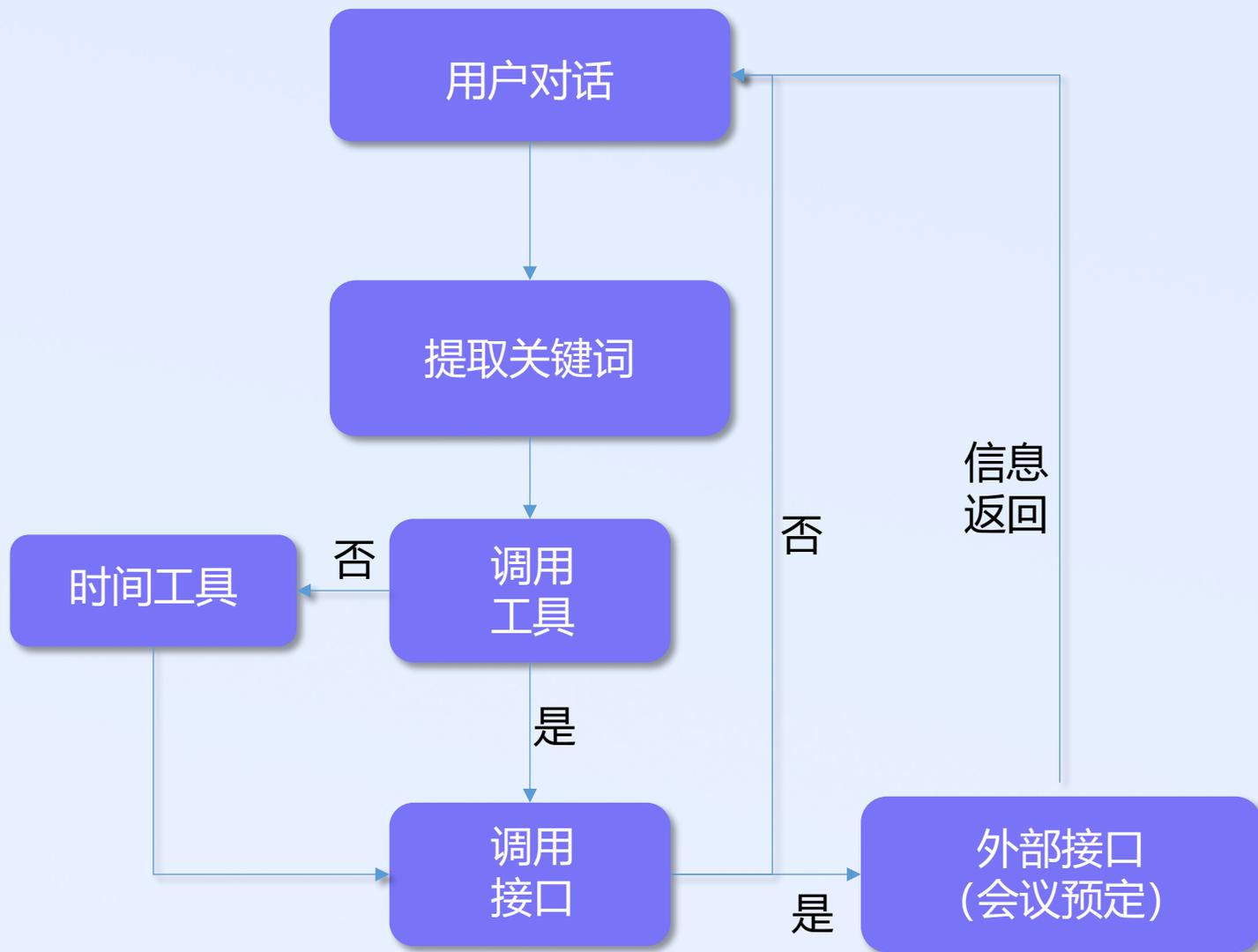


模型加速

容错自查

Agent应用介绍

会议预定



标准的Act的Agent
通过设定好提示词以及简单的流程
达到结果。



会议预定

"体验智能预定会议，让沟通无界限，高效协作成真。"

会议预定

主题

新项目的讨论会

参与人员

产品部 × 张伟欣 × 王文昭 ×

日期

2024-01-30

时间

9:00-11:00

发送邮件邀请

确认

会议预定

新项目的讨论会

9:00

待开始
2小时

11:00

2024年01月30日

2024年01月30日

会议号: 405 893 291

发起人: 王正义

参与人: 产品部 张伟欣 王文昭

会议室: 502室

会议链接: <https://meeting.tencent.com/dm/Kva5GzgjHmpp>

好的，您希望邀请哪些人参加这次会议？

我需要下周二上午九点安排一个关于新项目的讨论会。

请邀请项目团队的所有成员，包括我、张总、王总

明白了。您希望会议以什么形式进行？仅限线下，还是希望结合线上让远程团队成员也能参与？

线上和线下相结合的会议。

好的，已为您找到空闲会议室502。预定时间上午9点到11点,线上将创建一个腾讯会议，并将会议详情以邮件形式发送给所有与会者。是否确认??

是否确认??

还有其他我可以帮忙的吗？

确认

没有了，谢谢。

排一个关于新项目的讨论会。

有成员，包括我、张总、王总

线上和线下相结合的会议。

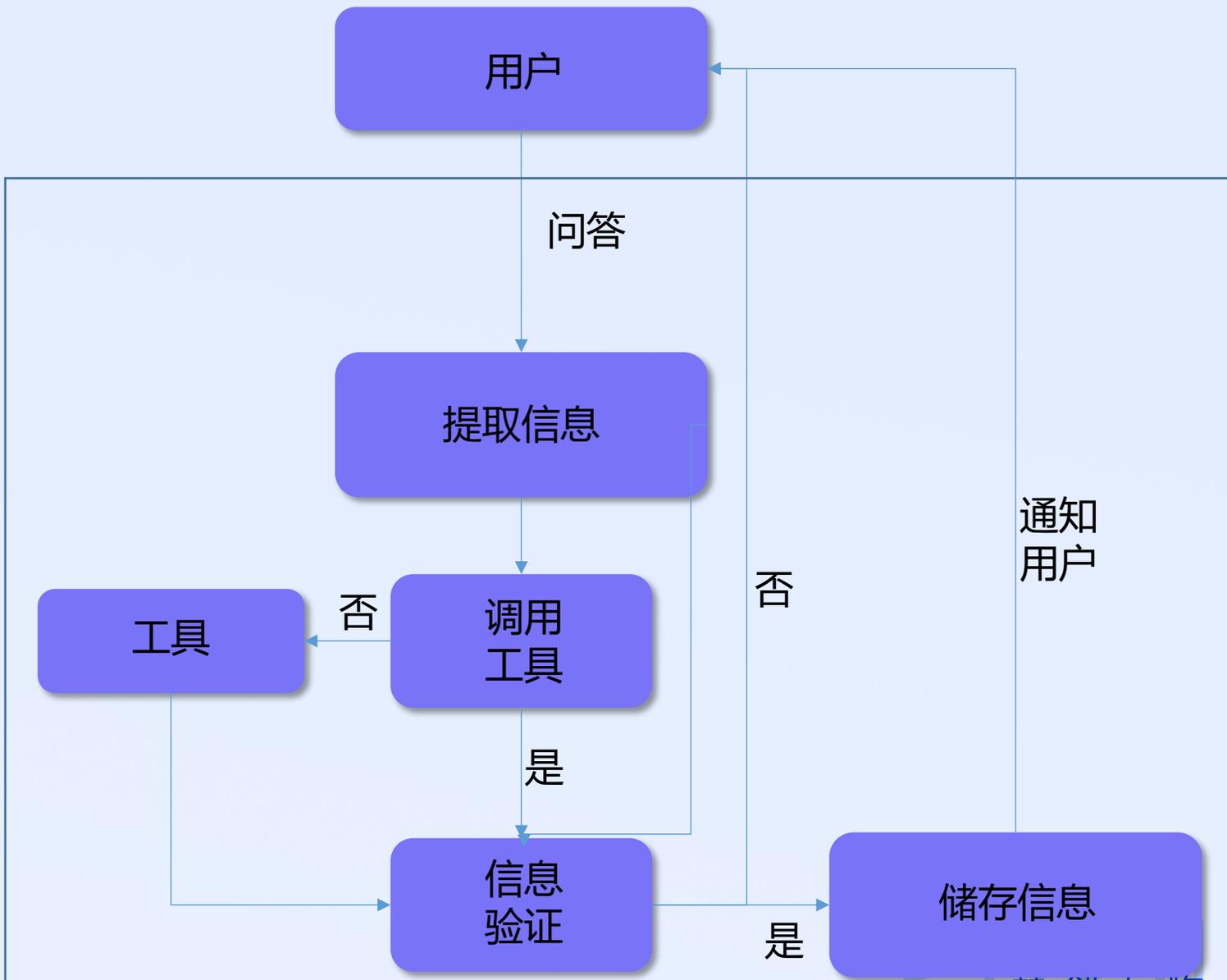
智能信息收集

用户

设置提取信息

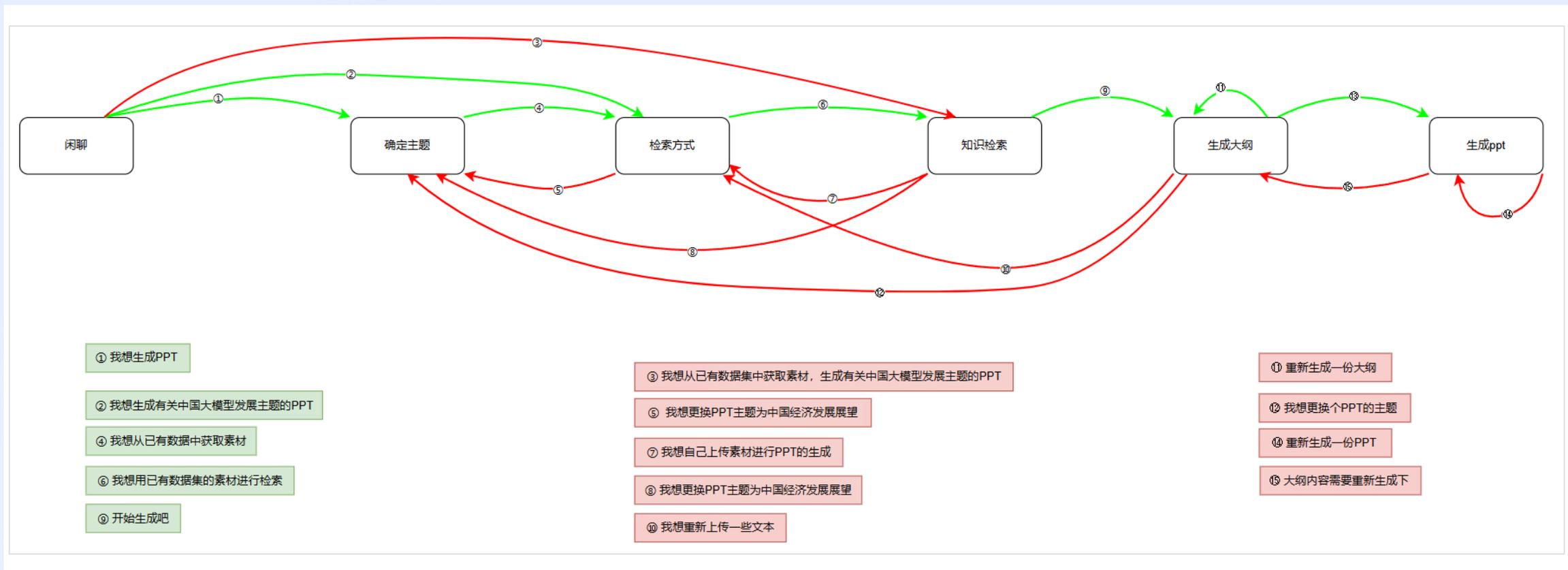
信息验证

信息修改



辅助办公AIppt

Aippt让用户可以，在普通的对话中根据意图进入应用



Aippt的状态流程图

```
# history: list of str, 元素数量必须是偶数
def get_state(
    client, state, prompt, topic=None, retrieve_type=None, retrieved_text=None, outline=None, model=settings.model
):

    state_machine_template = """你是一个状态机，负责管理调用AI助手生成PPT的流程。
    状态机满足某些触发条件时，进入其中一个状态（state），并且做出某些行为。
    状态机的状态应该严格遵守如下###间的json格式：
    ###
    {{
        "prev_state": "previous state",
        "state": "current state",
        "action": "AI assistant's action function call",
        "args": {{
            "arg_1": value_1,
            "arg_2":value_2}}
    }}
    ###
    其中"prev_state"表示前一状态，"state"表示当前状态，"action"表示当前的状态机行为，一个状态可以做出多种行为，需要根据用户输入进行选择。
    状态机的初始状态为："General"，下一状态中的"prev_state"会是前一状态的"state"。
    召回方式（retrieve_type）一共有3种，用户只能在Retrieval状态上传文件或者检索召回文件，以下一对+++内列举了它们各自的名称、对应的retrieve_type和描述：
    ***
    1. 数据文件生成， retrieve_type="file_list"，从向量数据库召回/检索与用户提供的PPT主题相关的文件；
    2. 导入本地文件， retrieve_type="read_file"，由用户提供文件，自己上传用来制作PPT的素材文件；
    3. AI智能生成， retrieve_type="llm_generate"，由大模型，也就是你，根据用户给的PPT主题创意生成PPT的大纲；
    ***
    请注意，用户只知道召回方式的名称，不可以告诉用户它们的action。你需要根据用户描述，判断用户选择哪一种召回方式，用户可以不直接给出一模一样的名称。
    在任何需要你直接给出output内容的地方，请以Alaya的名义用中文回答（必要的时候可以夹杂英文单词）。Alaya是九章云极研发的元识大模型，英文名是Alaya，是一个AI助手。
    这个状态机只可以有如下几个状态，每个状态的输出都有action，每个action有不同的参数args：
    ...

    1.General:
        前提1: prev_state = "General"
        触发条件1: 用户的指令和制作PPT无关
        行为1: AI助手继续与用户进行多轮对话。输出: {{"prev_state": "General", "state":"General", "action" = "qa", args: {{"instruction": "[本次用户的问询]"}}}}

        前提2: prev_state = "Gen_PPT"
        触发条件2: 用户表示表示肯定（或者可以等正向反馈），或者想要导出PPT
        行为2: AI PPT流程已完结，AI助手引导用户自行导出PPT。输出: {{"prev_state": "General", "state":"General", "action" = "output", args: {{"content": "[请作为Alaya的角色
```

状态机的提示词

状态的短期记忆

状态的长期记忆

AIPPT的数据来源

用户自己上传

用户自己准备好的ppt的大纲以及标题，可以是word、pdf等上传到系统的。系统解析出相应的标题

大模型的能力

用户自己输入想要输出的ppt的主题。大模型依靠自己的能力输出ppt的大纲标题信息

私域数据的召回 (RAG)

用户也可以上传一些私域的文档，输入主题之后，平台根据主题召回文档中相似的段落，然后根据ppt的主题，大模型总结输出ppt的大纲内容

AI PPT

"体验智能PPT生成，让人工智能助力您的演示文稿创作"

你想如何生成PPT?



数据文件生成

根据数据集文件生成



导入本地文件

根据本地文件生成



AI智能生成

根据提示生成

我想生成ppt

您好，请告诉我您想制作的PPT主题是什么呢？

中国大模型的发展

请问您想要以什么方式收集或生成PPT所需的素材？我们可以提供以下几种方式：1. 数据文件生成方式来召回相关文件；2. 您可以直接上传您的素材文件；3. AI智能生成PPT的大纲。您希望选择哪一种方式呢？

数据文件生成

即将开始根据您的PPT主题“中国大模型的发展”，从向量数据库检索相关的文件，作为生成PPT的素材。

即将开始根据您的PPT主题“中国大模型的发展”，从向量数据库检索相关的文件，作为生成PPT的素材。

确定并生成大纲

由元识大模型Alaya开始生成大纲，请稍候。

由元识大模型Alaya开始生成大纲，请稍候。

我想好了，选择PPT模板

已有确定的文本摘要素材

文件

确定

中国大模型发展规划.pdf

请输入

团队数据集

个人数据集

dw-dataset-1

- DataCanvas_APS_7.0_特性报告.pdf
- DataCanvasAPS版本升级注意事项-APS7.0.docx
- DataCanvasAPS7.0产品白皮书.pdf
- DataCanvasAPS7.0安装手册.v1.0.0(1).docx

llm

- 中国大模型发展白皮书_2023.pdf
- 中国大模型发展白皮书_all.pdf
- 中国大模型发展规划.pdf
- 中国大模型发展展望.pdf

大纲

中国大模型的发展

一、大模型的概念

大模型与深度学习平台发展

- 数据驱动的AI发展趋势
- 大模型在产业化中的应用
- 深度学习平台与大模型的协同
- ### 人工智能技术的挑战
- 算力资源分布不均
- 数据噪声与偏见问题
- 跨多模态数据融合问题

二、大模型技术突破

- ### 大模型预训练技术
- “预训练+微调”新范式
- 多模态预训练大模型的进展
- ### 硬件与平台的进步
- 分布式训练技术
- 新型人工智能芯片
- 硬件与软件的协同优化

三、大模型的产业应用

生成PPT

下载到本地



目录



1



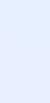
2



3



4



5

Agent未来与优化思考

方向一：Multi-Agents的集成爆发

多个智能体（Agent）组成的集合。这些智能体可以是不同的软件程序、机器人、传感器等，它们各自具备一定的智能和自主性。MAS的核心思想是通过多个智能体的协作与协调，共同完成一个复杂任务，从而实现单个智能体无法完成的目标。

方向二：PC端、移动端的部署

随着开源生态的蓬勃发展，以及模型能力的持续提升，未来PC端和移动端将广泛部署更加智能、个性化且功能丰富的系统级Agent，为用户提供无缝、高效的跨平台服务体验。

方向三：多模态能力的扩展

Agent将超越传统的文本对话，通过整合动作识别、语言理解、情感分析等多模态交互能力，并与硬件设备深度融合，实现更加自然、直观且智能的交互体验，真正成为人们日常生活中不可或缺的智能伙伴。

谢谢观看

THANKS