

The background features a complex, abstract design. A central hyperboloid-like shape is formed by a dense network of thin, light-gray lines that intersect to create a grid-like pattern. Scattered throughout this central area and extending towards the edges are numerous small, multi-colored dots in shades of red, yellow, green, blue, and purple. The overall composition is symmetrical and has a technical, mathematical feel.

Agent进阶与CrewAI

目录

CONTENTS

Agent进阶

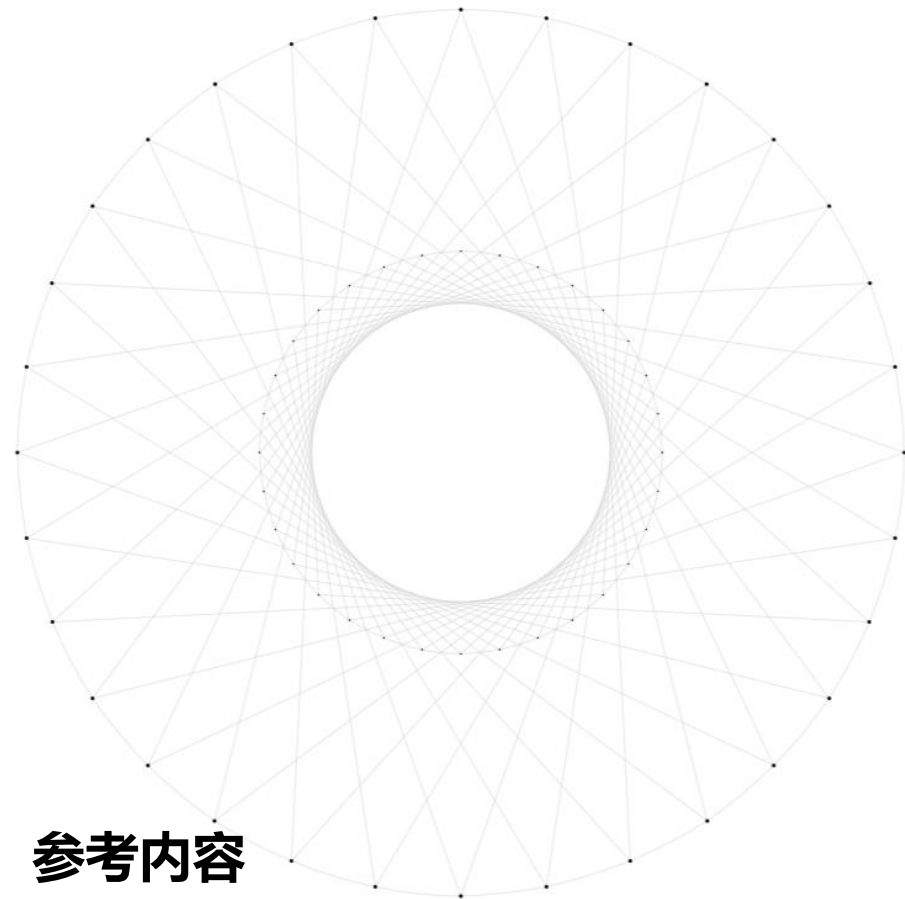
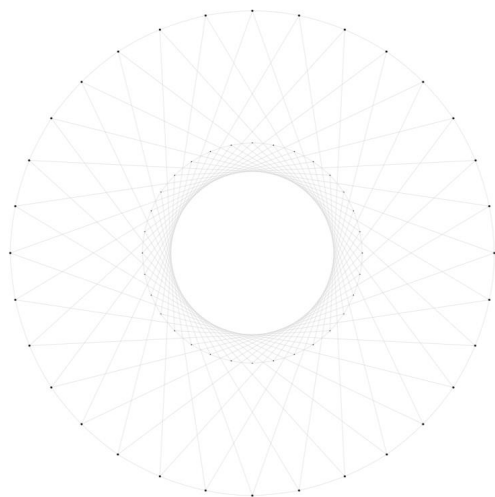
PART ONE

CrewAI

PART TWO

参考内容

PART THREE





Agent进阶

PART ONE

Agent进阶：思维-感知-行动框架

思维模块 (Brain)

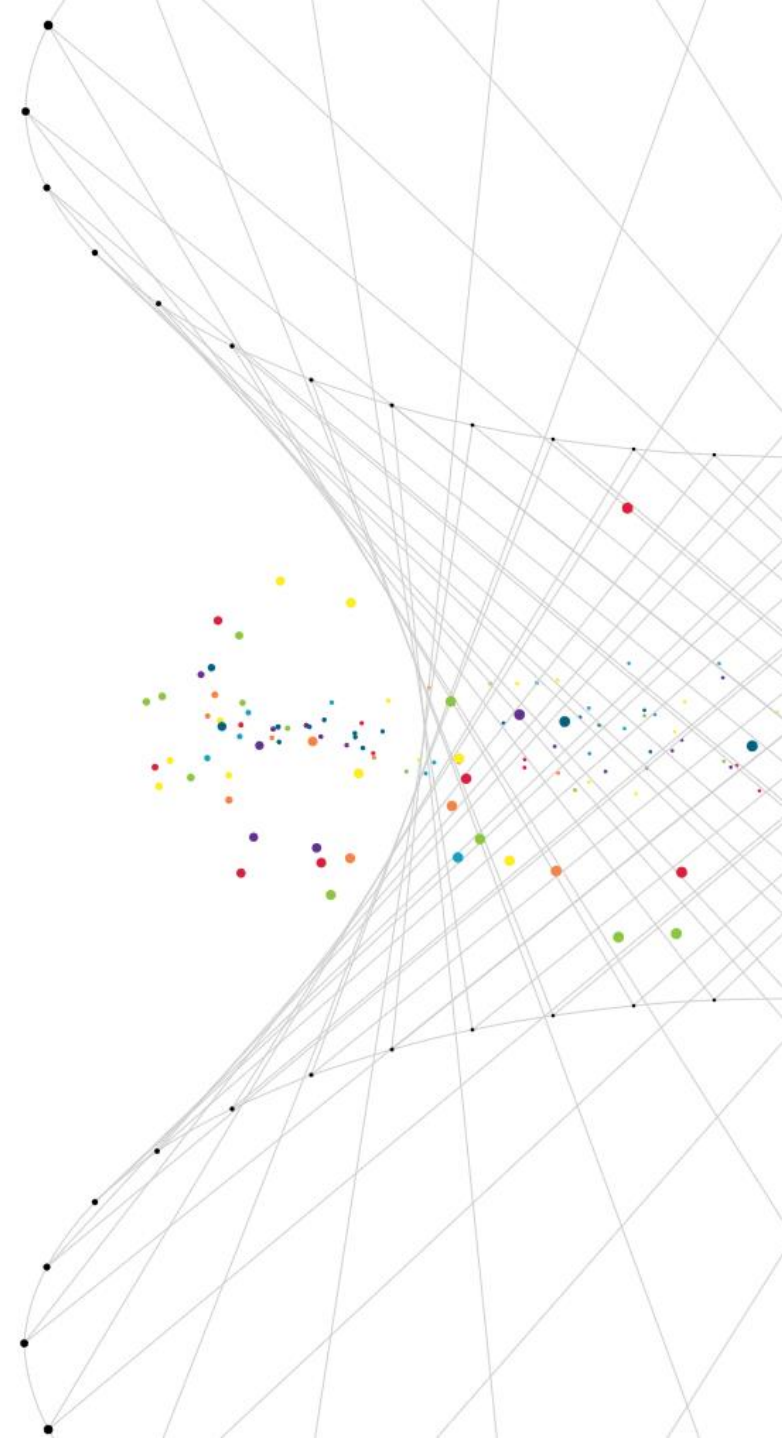
控制Agent最为关键的部分。

感知模块 (Perception)

广泛的感知空间可以支持Agent从各种丰富的来源接收信息，帮助其更加全面地了解环境，从而做出更佳的决定。

行动模块 (Action)

接收到感知模块提供的环境信息后，经过思维模块一系列整合、分析以及决策，将由行动模块根据决策具体与环境产生交互。



思维模块

Brain

控制Agent最为关键的部分。为进行高效沟通，需要具备自然语言交流的能力；针对输入进来的信息在经过存储后，通过检索知识和召回记忆来获取辅助信息，从而进行规划、推理以及制定有根据的决策。主要涉及如下要点：

- 自然语言交互 (Natural Language Interaction)
- 知识 (Knowledge)
- 记忆 (Memory)
- 推理与规划 (Reasoning and Planning)

自然语言交互

Natural Language Interaction

天然以人类语言作为交互的媒介，使得人机协同成为可能。自然语言交互主要关注这些能力：

多轮交互式对话（Multi-turn interactive conversation）

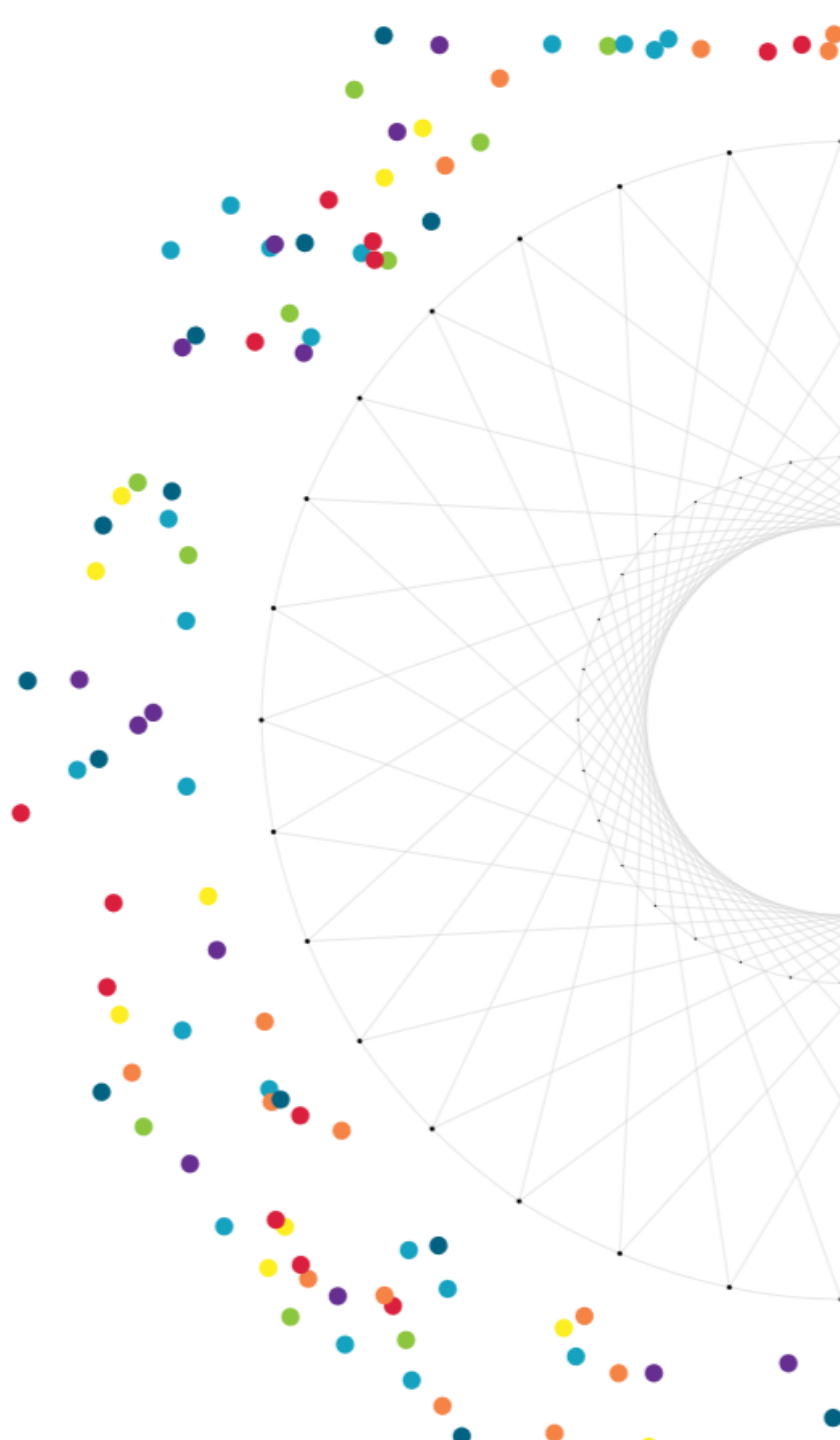
多轮对话相较于单轮对话具备更充分的场景信息，一般来说基于LLM的Agent会根据现有的信息来进行多轮对话，并逐步实现最终目标。

高质量自然语言生成（High-quality natural language generation）

LLM在自然语言生成上展现出了一定的优秀能力，既能较好地适应文本风格，也能理解语法问题，同时还具备创造性，并且在对话方面也具有突出表现。此外人类还可以简单地通过显式的“提示”来对LLM生成内容加以控制。

意图与含义理解（Intention and implication understanding）

LLM可以在一定程度上理解人类的意图与表述内容的含义，但在面对模糊指令或者理解过深含义时的表现仍有欠缺。



知识

Knowledge

LLM可以从大规模的自然语言语料数据中获取知识，通过训练的方式将知识转换为参数化的形式，这些知识有助于LLM-based Agent做出有根据的决策。知识主要可以分为这些类型：

语言知识（Linguistic knowledge）

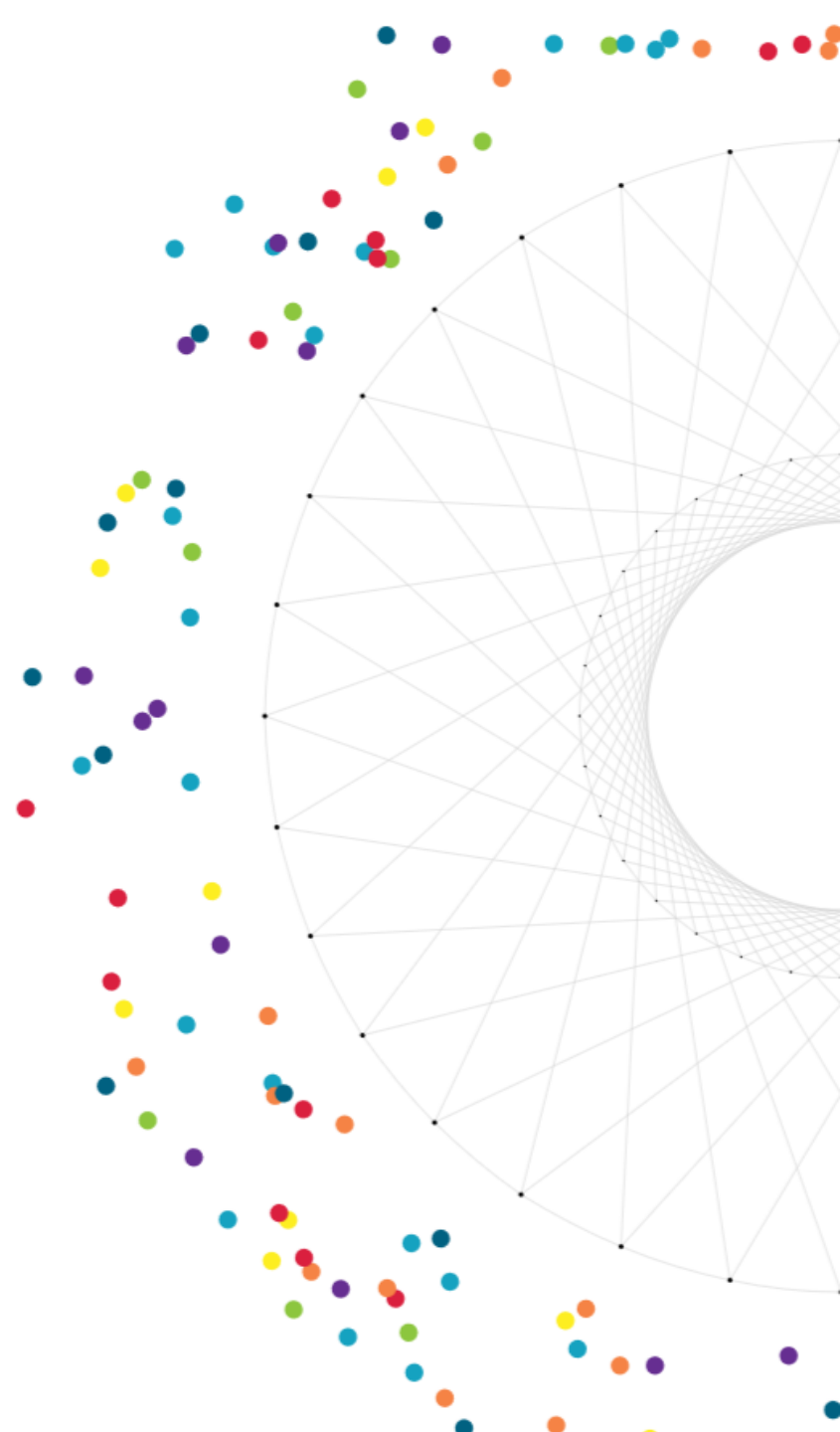
包括词法、句法、语义等方面的知识，相关方面的知识有助于理解输入所表达的内容并进行回复。

常规知识（Commonsense knowledge）

指普遍认知中的世界事实，这些知识信息通常不会显式出现在上下文中，但缺乏这些知识会导致做出错误决断。

专业领域知识（Professional domain knowledge）

具体的、与特定领域相关的知识，例如医学、法律、编程方面的知识等等。



记忆

Memory

存储Agent过往的观察、思想乃至行动的序列，一些必要的上下文可以使得Agent能够通过借鉴过往经验来逐步适应环境。Agent对记忆运用的局限首先在于LLM架构层面的最大上下文长度，其次是从越发长的上下文中“回忆”必要信息是越发困难的。针对记忆方面的问题存在一些可能的解决方法：

扩展LLM的长度

通过使用支持更长上下文的模型或使用长度扩展方法来使得Agent适应更长的上下文记忆。

摘要记忆

对记忆进行摘要以保留必要信息。

使用数据结构存储记忆

使用三元组数据结构或向量数据来保存记忆，并通过特定的方法进行召回。



推理与规划

Reasoning and Planning

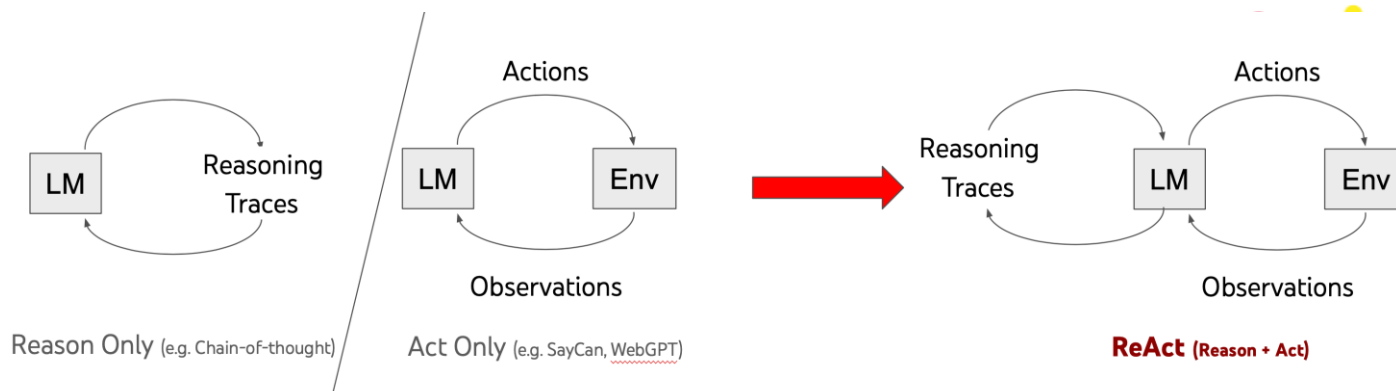
推理能力与规划能力是解决复杂问题的关键。

推理 (Reasoning)

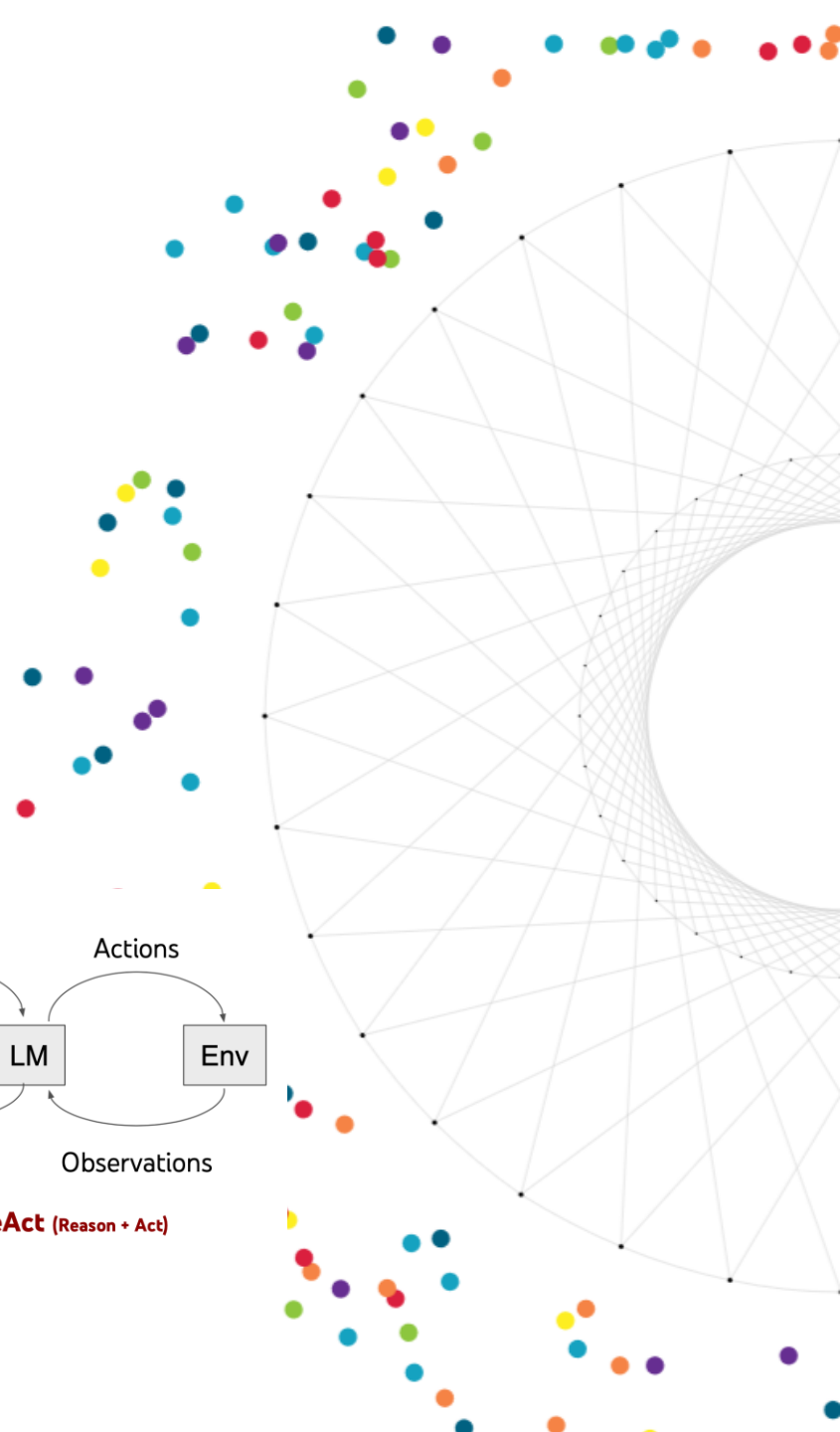
通过由因及果的生成，可以使得LLM能输出更佳的结果，即在输出最终答案前要求LLM先生成一系列究因过程，可以在一定程度上保证最终答案的准确性。

规划 (Planning)

借助推理能力对任务进行拆解、调度、自省、调节来保证达成目标。



ReAct (Reason + Act)



感知模块

Perception

广泛的感知空间可以支持Agent从各种丰富的来源接收信息，帮助其更加全面地了解环境，从而做出更佳的决定。所谓“感知”指的就是接收各种模态数据的能力，通常是基于LLM在不同模态的编码层面上进行对齐，从而将其他模态的感知整合进LLM中。主要涉及如下要点：

- 文本模态输入 (Textual Input)
- 视觉模态输入 (Visual Input)
- 听觉模态输入 (Auditory Input)
- 其他模态输入 (Other Input)

文本模态输入

Textual Input

LLM天然支持文本模态感知。文本除了承载显式的内容外，还隐藏着信念、意图、期望等等信息，理解这些深层含义可以使得Agent能有效进行沟通。

ReAct

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 ...

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)`', ...]

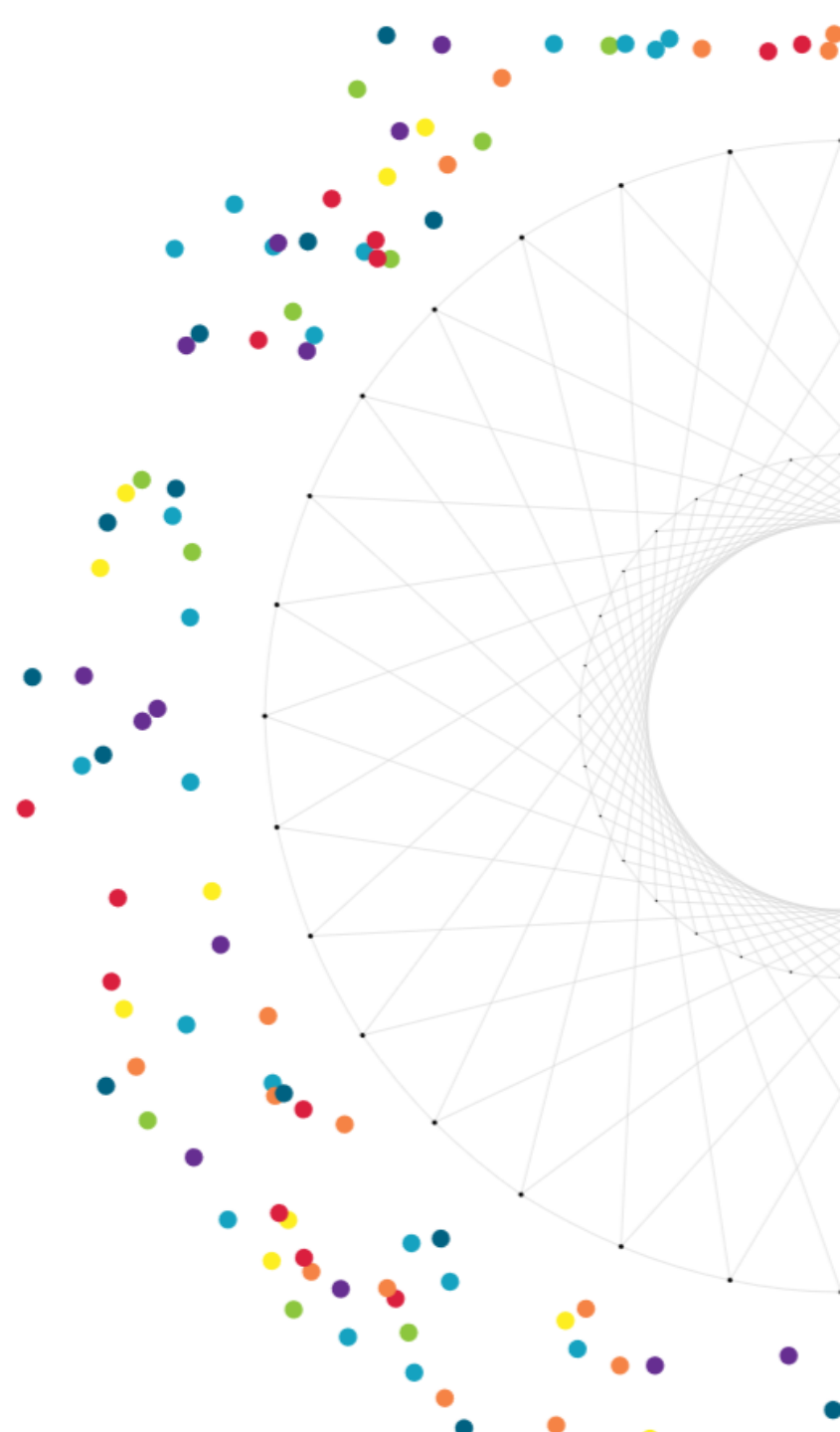
Thought 3: Front Row is not found. I need to search `Front Row (software)` .

Act 3: `Search[Front Row (software)]`

Obs 3: Front Row is a discontinued media center software ...

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]` ✓



视觉模态输入

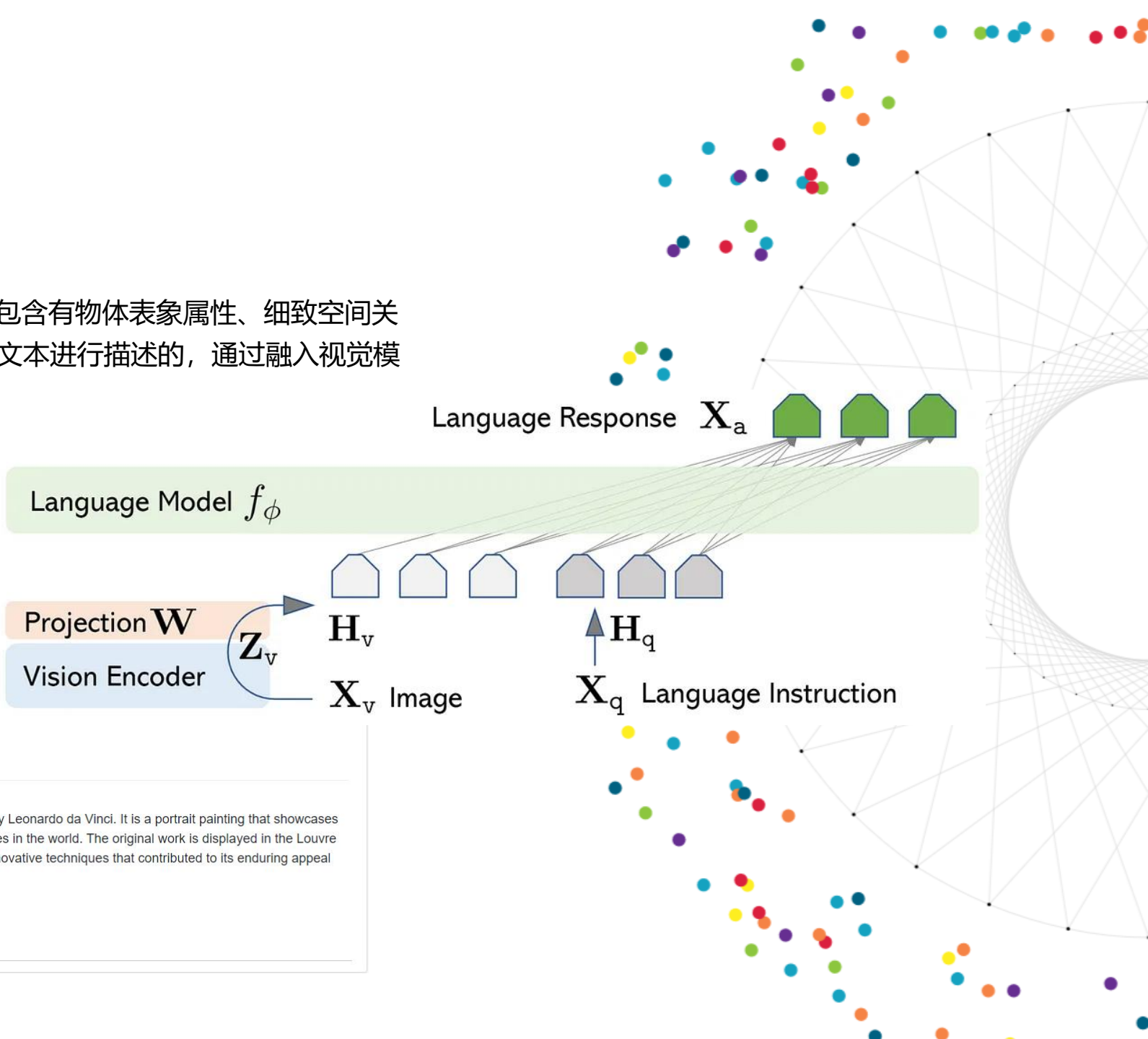
Visual Input

基本的LLM天然缺乏视觉感知，视觉模态输入包含有物体表象属性、细致空间关系等直接的事物信息，这些信息都是难以通过文本进行描述的，通过融入视觉模态信息可以使得Agent更充分地理解环境。



User
Do you know who drew this painting?

LLaVA
The painting depicts a woman, commonly believed to be Mona Lisa, the famous artwork by Leonardo da Vinci. It is a portrait painting that showcases the woman's enigmatic smile and has become one of the most famous and iconic art pieces in the world. The original work is displayed in the Louvre Museum in Paris, and it is known for its intricate details, use of oil paint, and the artist's innovative techniques that contributed to its enduring appeal and mystery.

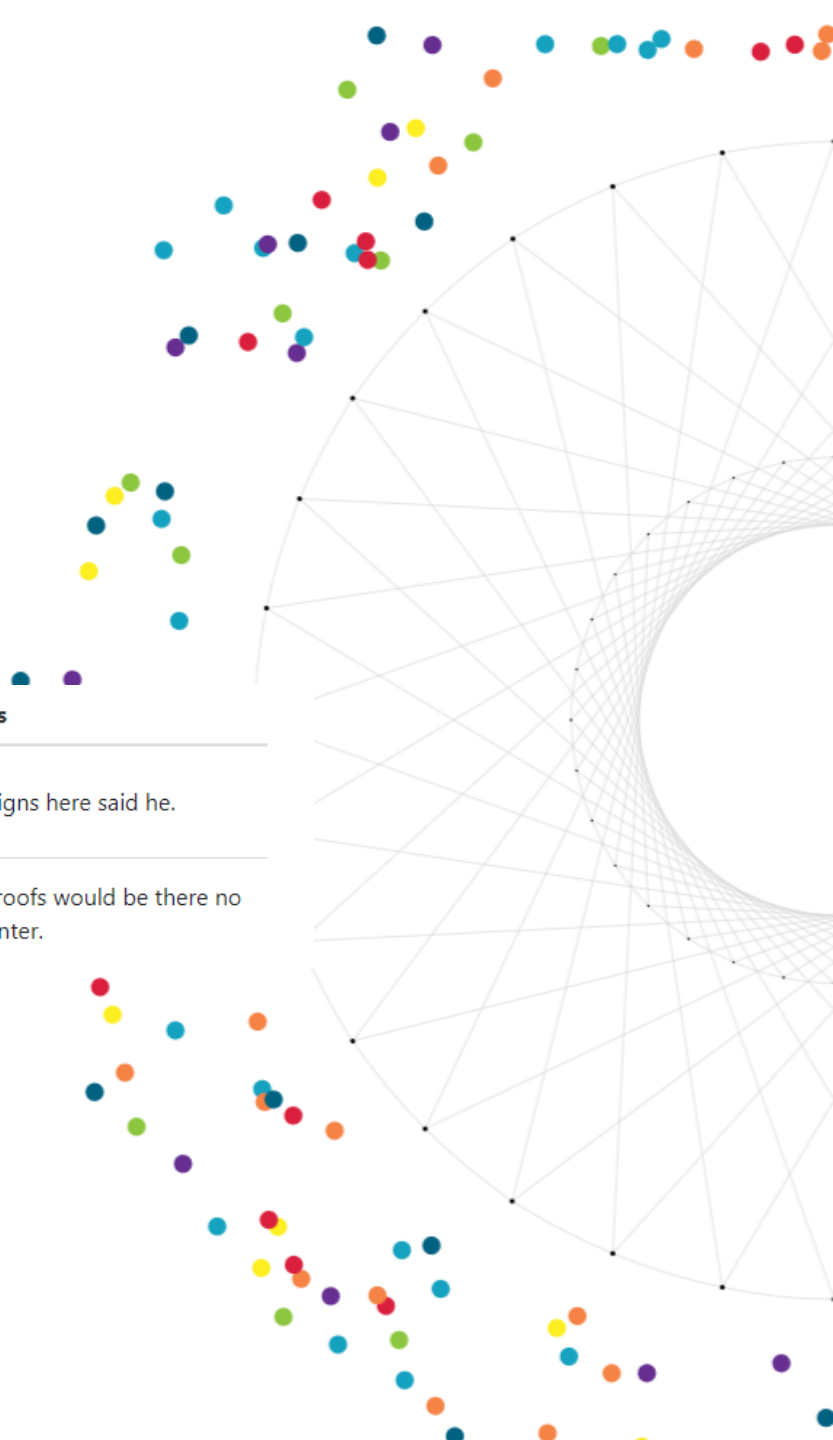


听觉模态输入

Auditory Input

基本的LLM天然缺乏听觉感知，听觉模态输入可以反映细微的环境变化，同时也可使得用户通过语音传达信息，提高用户体验。

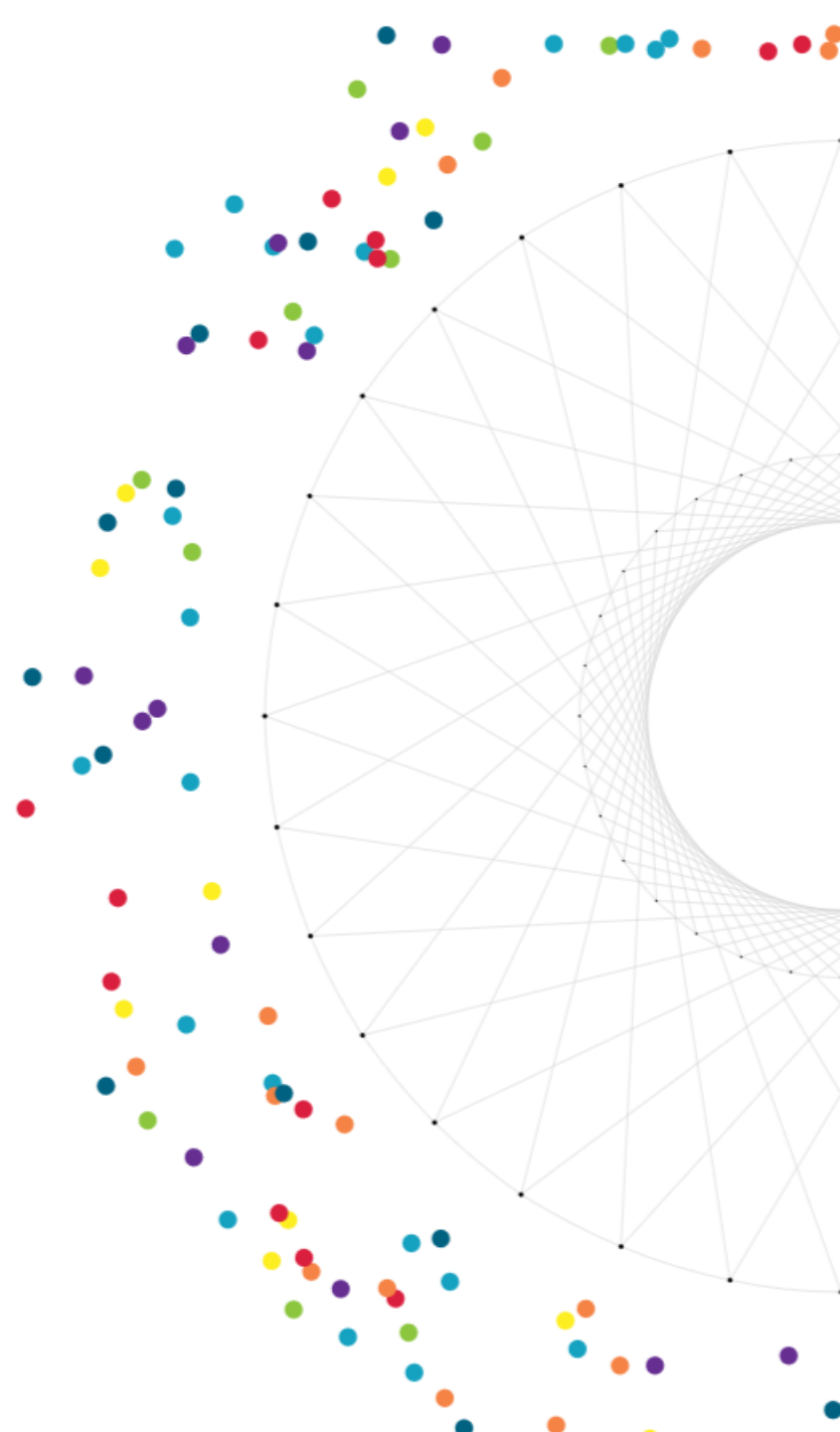
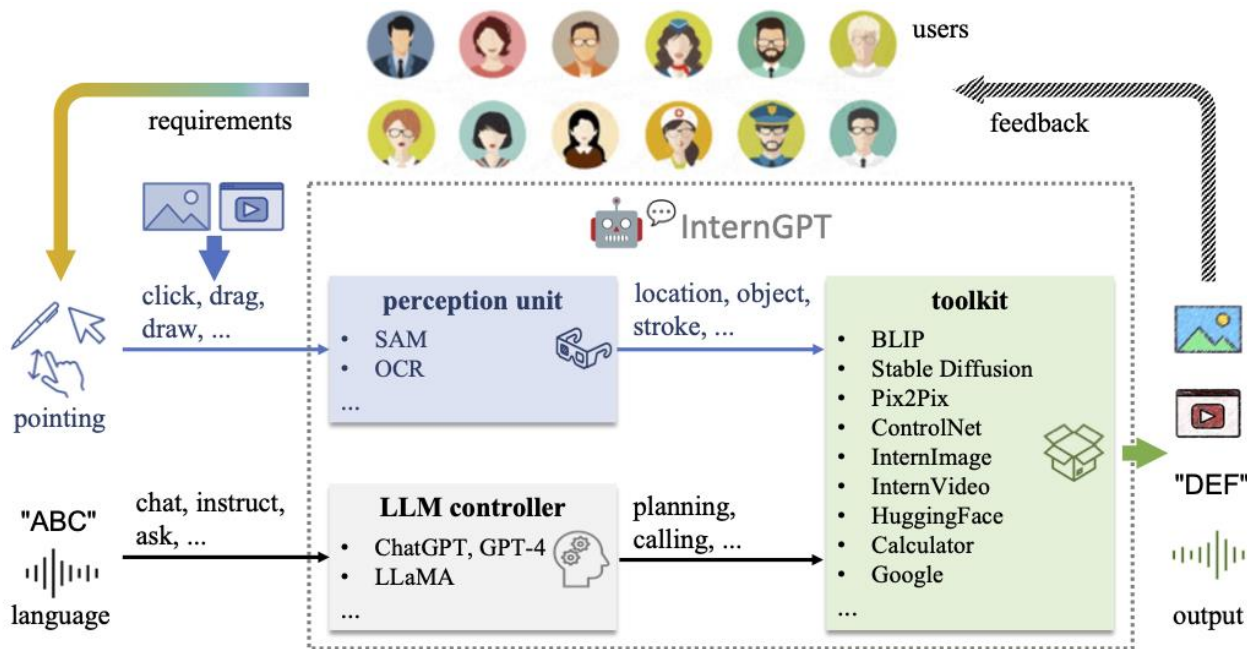
Instructions	Inputs	outputs
Can you transcribe the speech into a written format?	<div>▶ — ⋮</div>	I'm afraid there are no signs here said he.
Listen to the speech and write down its content.	<div>▶ — ⋮</div>	Did anyone know that these proofs would be there no one the printer.



其他模态输入

Other Input

诸如触觉、温感、眼动跟踪、触控坐标等更多的模态输入同样也可以为Agent在特定任务上带来更多可参考的信息。



行动模块

Action

接收到感知模块提供的环境信息后，经过思维模块一系列整合、分析以及决策，将由行动模块根据决策具体与环境产生交互。主要涉及如下要点：

- 文本模态输出 (Textual Output)
- 工具使用 (Tool Using)

文本模态输出

Textual Output

流畅的文本输出是LLM本身固有的能力，经过感知、思考等流程后，以文本内容的形式作为对环境的响应。

ReAct

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 ...

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)`', ...]

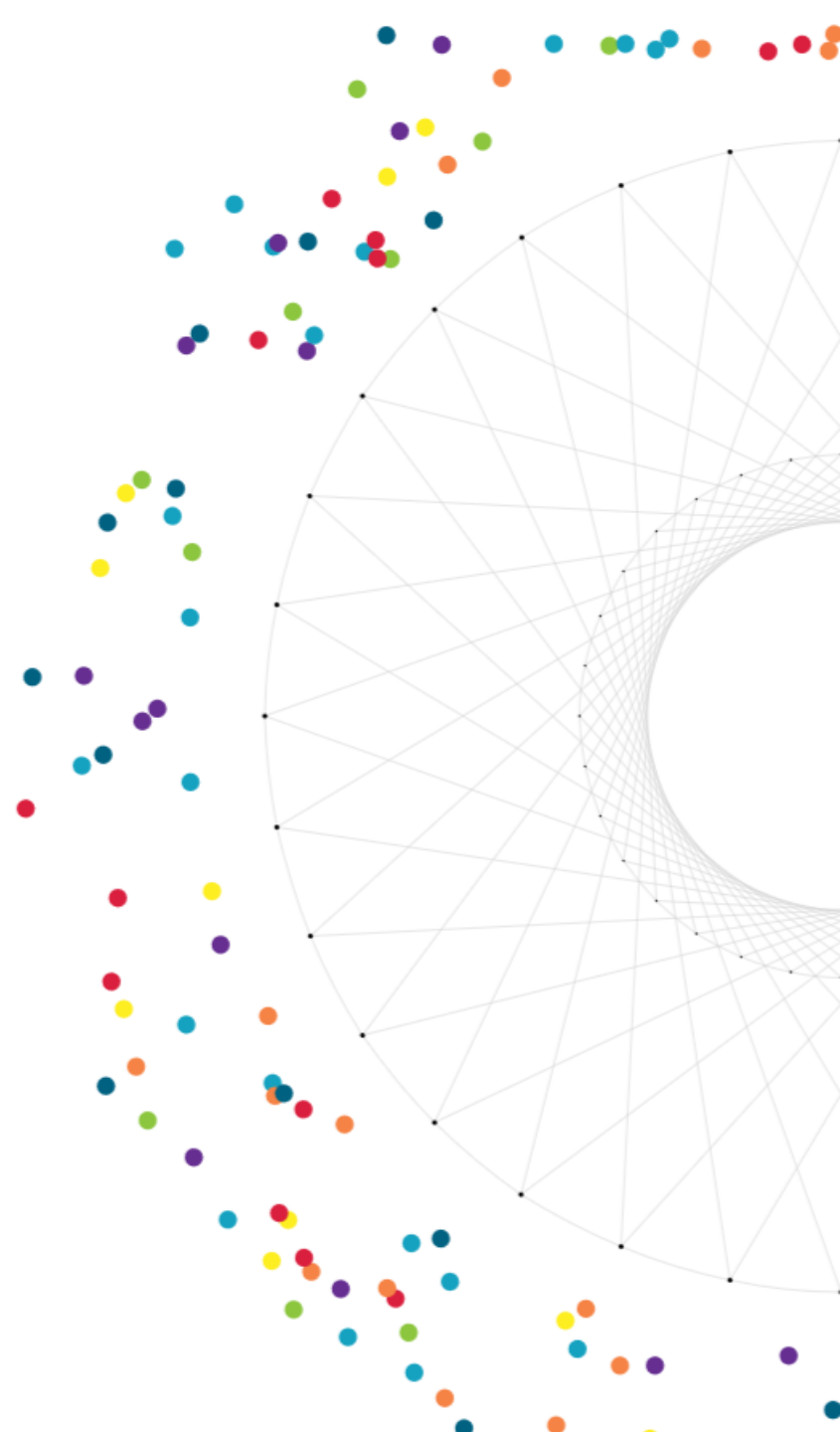
Thought 3: Front Row is not found. I need to search `Front Row (software)` .

Act 3: `Search[Front Row (software)]`

Obs 3: Front Row is a discontinued media center software ...

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]`



工具使用

Tool Using

工具是使用者能力的延伸，使用工具有助于应对复杂问题，从而更有效地完成任务。通过few-shot的方式将工具的使用场景以及参数描述等信息显式提供给LLM，在推理过程中，LLM得以借助其来调动外部资源，根据外部资源的文本响应来做进一步规划，简化了任务并加强决策可靠性。

Zero-shot Prompting: Here we provide a tool (API) "forecast_weather(city:str, N:int)", which could forecast the weather about a city on a specific date (after N days from today). The returned information covers "temperature", "wind", and "precipitation".

Please write codes using this tool to answer the following question: "What's the average temperature in Beijing next week?"

Few-shot Prompting: We provide some examples for using a tool. Here is a tool for you to answer question:

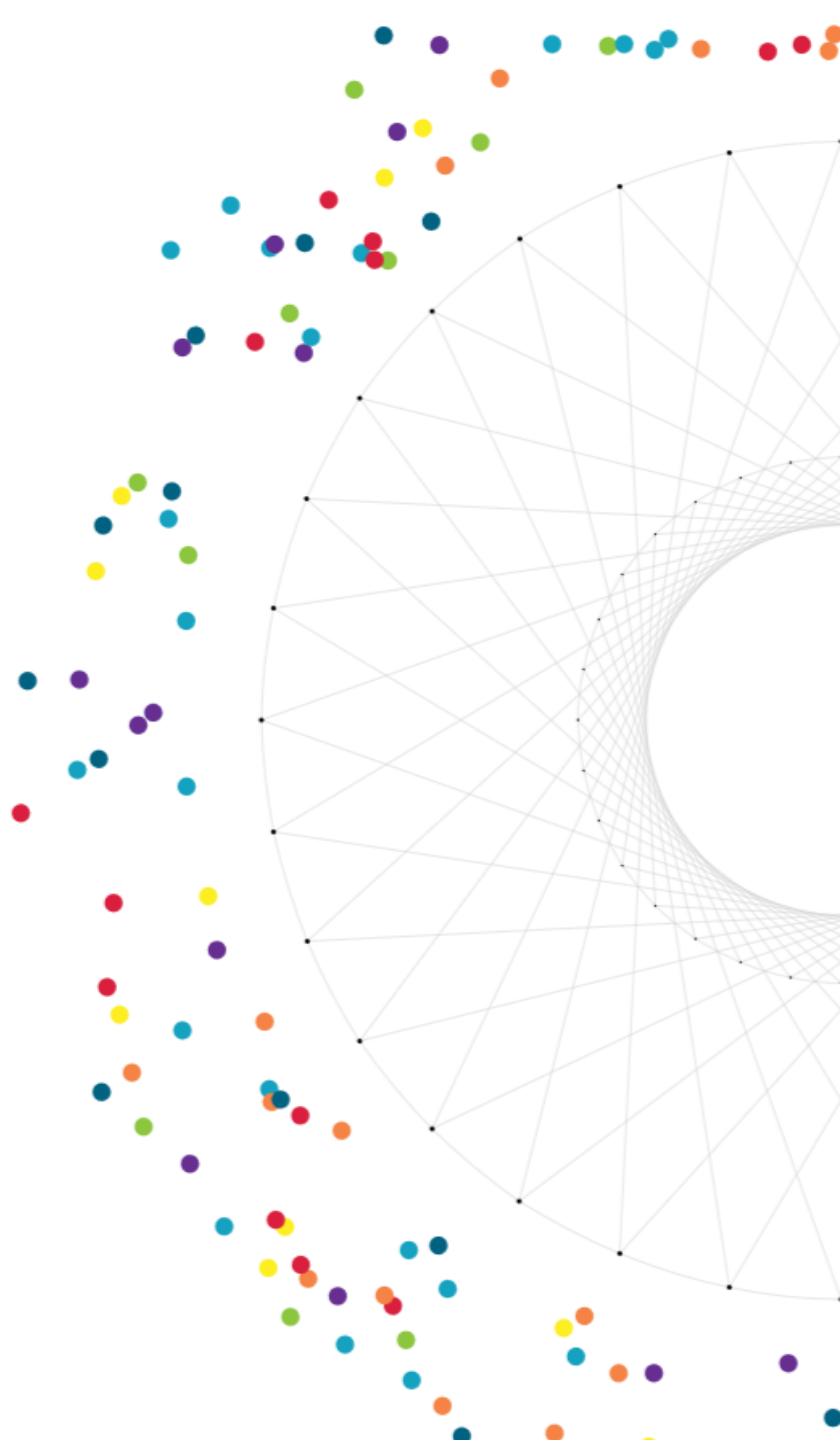
Question: "What's the temperature in Shanghai tomorrow?"

```
return forecast_weather("Shanghai", 1) ["temperature"]
```

Question: "Will it rain in London in next two days?"

```
for i in range(2):  
    if forecast_weather("London", i+1) ["precipitation"] > 0:  
        return True  
return False
```

Question: "What's the average temperature in San Francisco next week?"





CrewAI

PART TWO



环境配置

搭建并激活虚拟环境

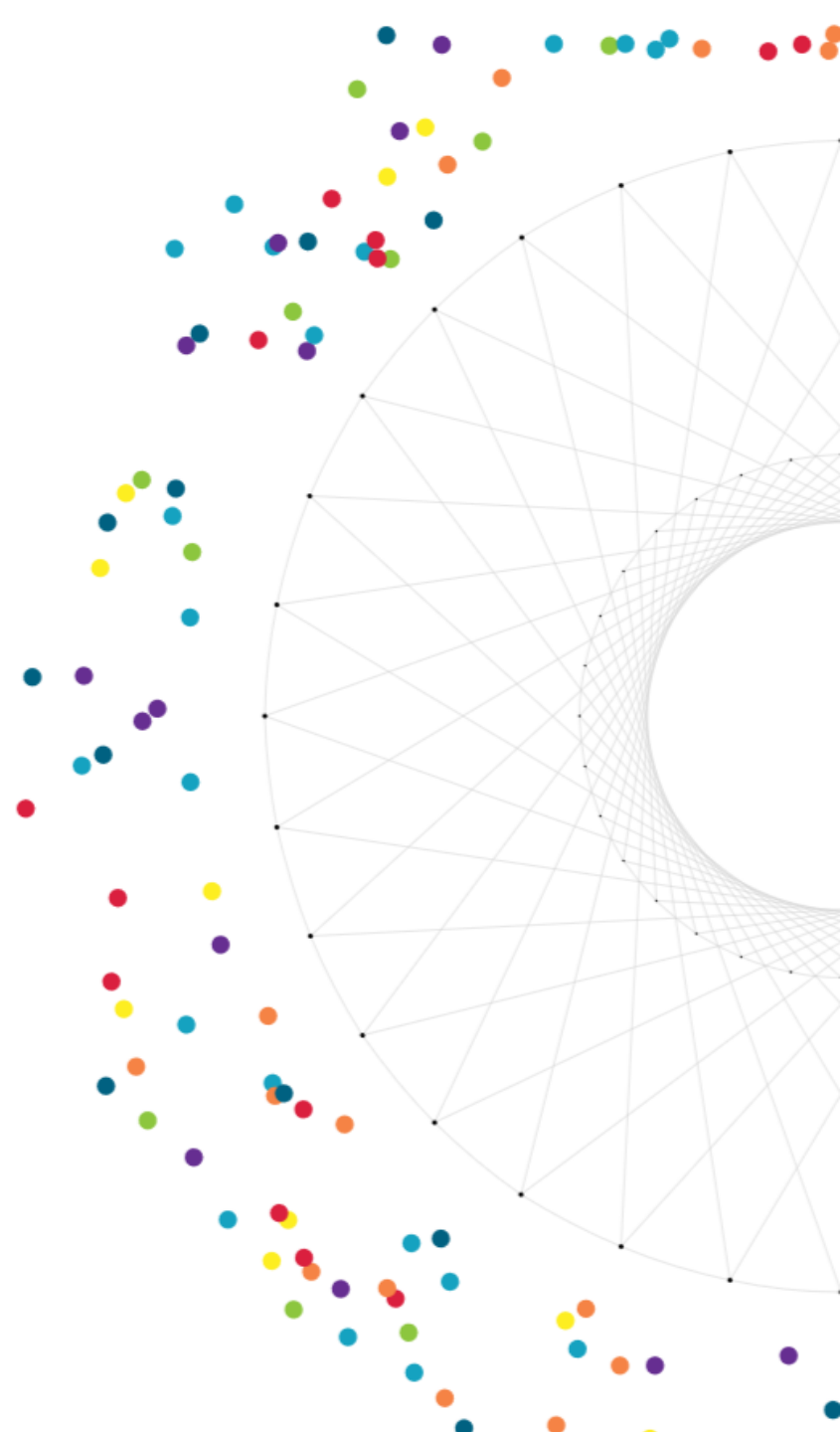
```
conda create -n crewai python=3.10 pip -y  
conda activate crewai
```

安装框架

```
pip install crewai
```

安装其他必须库

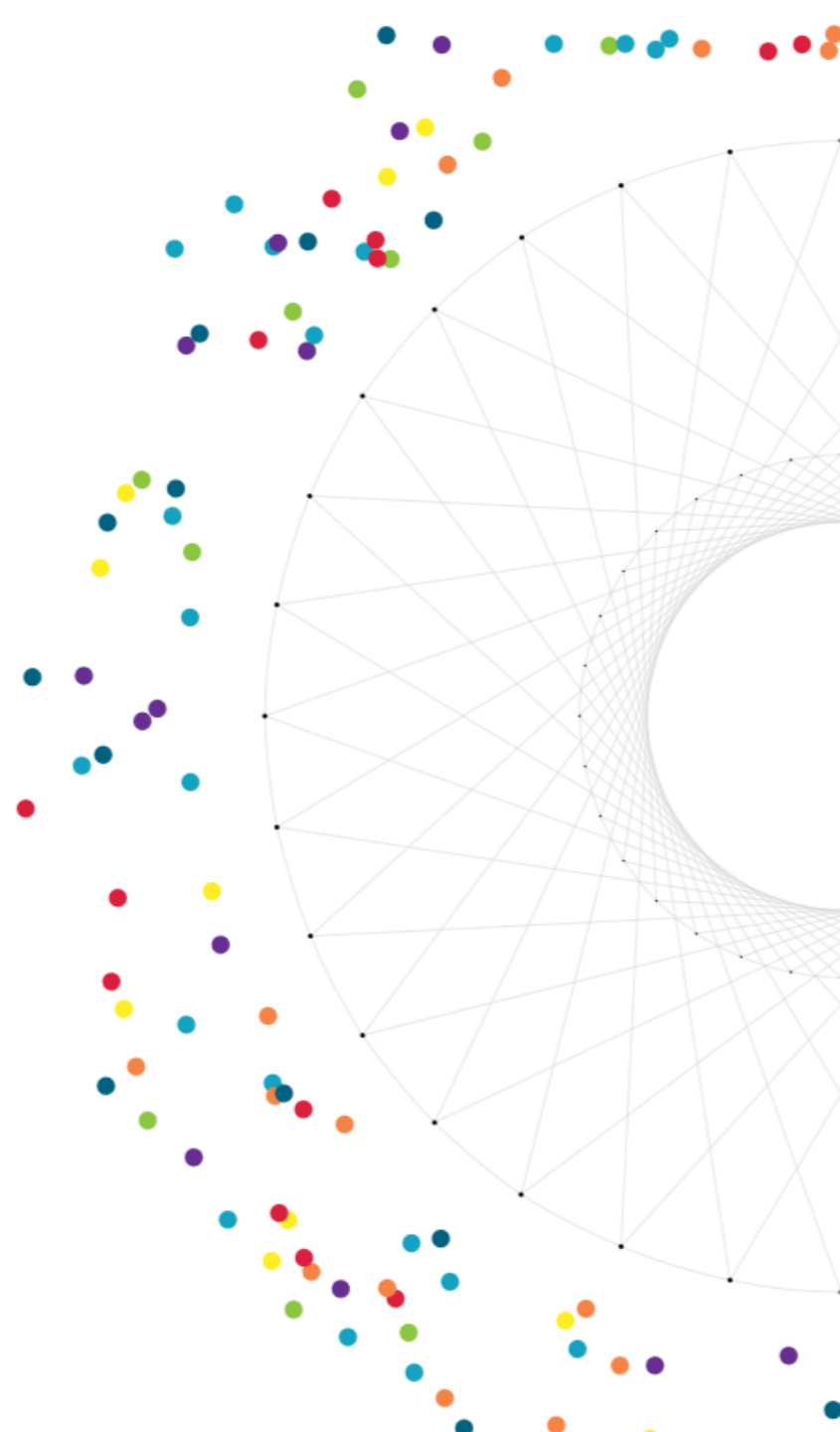
```
pip install langchain  
pip install langchain_community  
Pip install dashscope
```



API-KEY 申请

DashScope API-KEY


<https://help.aliyun.com/zh/dashscope/developer-reference/activate-dashscope-and-create-an-api-key>



实操代码

基于CrewAI使用多Agent协同完成创意剧本写作。



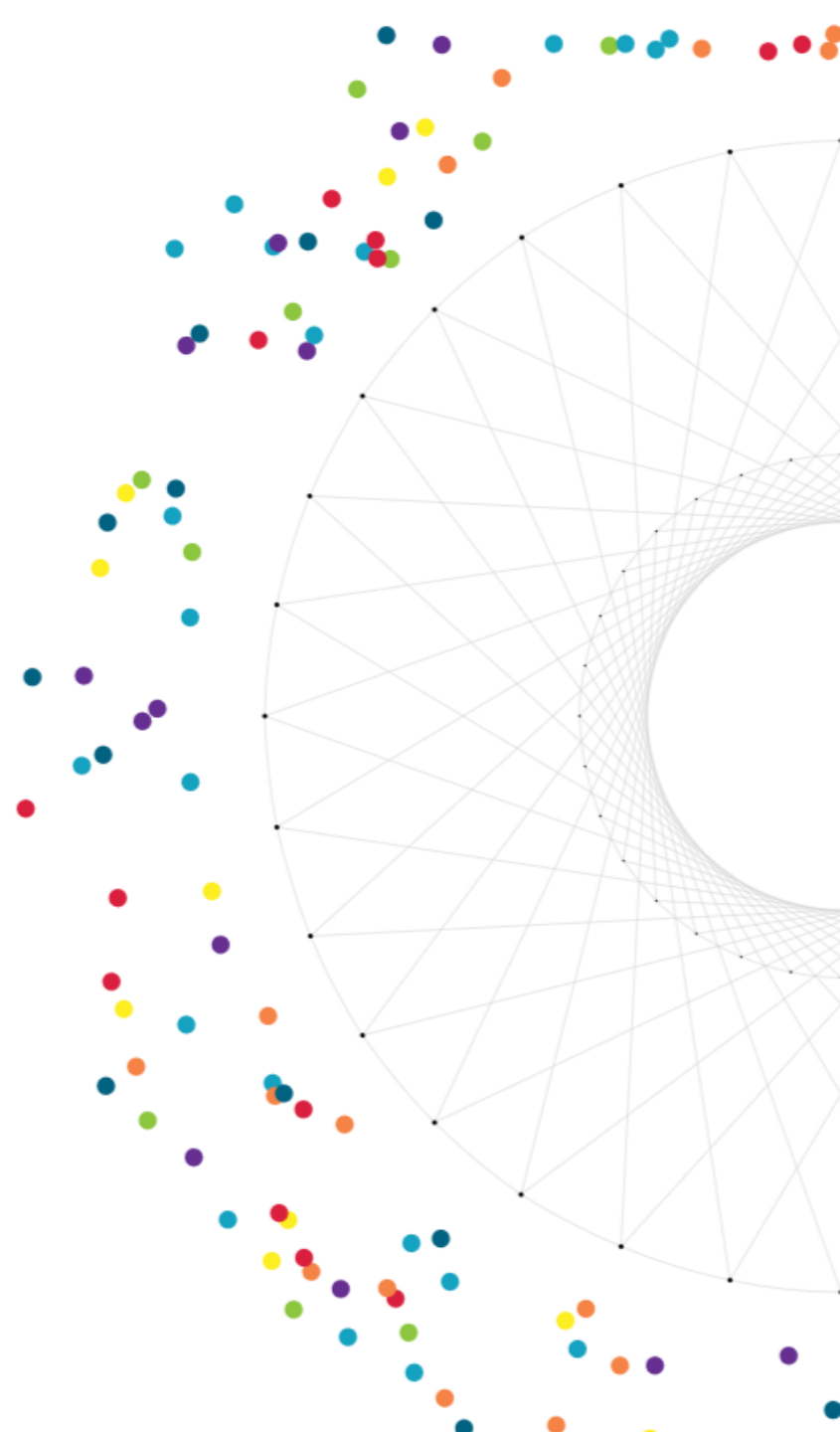


参考内容

PART THREE

参考内容

1. <https://arxiv.org/abs/2309.07864>
2. <https://react-lm.github.io/>
3. <https://arxiv.org/abs/2210.03629>
4. <https://llava-vl.github.io/>
5. <https://0nutation.github.io/SpeechGPT.github.io/>
6. <https://arxiv.org/abs/2305.05662>
7. <https://arxiv.org/abs/2304.08354>
8. <https://help.aliyun.com/zh/dashscope/developer-reference/activate-dashscope-and-create-an-api-key>
9. <https://github.com/joaomdmoura/crewAI>



The background features a complex geometric pattern of thin, light gray lines that intersect to form a hyperboloid-like shape. Scattered throughout this pattern are numerous small, colorful dots in shades of red, blue, yellow, green, and purple. The text "THANK YOU FOR WATCHING" is centered in a bold, black, sans-serif font.

THANK YOU FOR WATCHING