

规划与智能体

《大语言模型》编写团队：赵鑫

➤ 局限示例1：在逐步推理、规划任务上仍然存在能力不足

GPT-4

Consider the identity:
 $9 * 4 + 6 * 6 = 72$
Can you modify exactly one integer (and not more than that!) on the left hand side of the equation so the right hand side becomes 99?
-Let's think step by step, write down a plan, and then write down your solution as: "The solution is: $A*B+C*D$ ".

Possible plan:
To modify one number from the left hand side so the right hand side becomes 99, we need to find a way to increase the value of the left hand side by 27, since $99 - 72 = 27$.
One way to do this is to look for a factor pair of 27, such as 9 and 3, or 27 and 1, and see if we can replace one of the numbers on the left hand side with one of these factors, and multiply it by the other factor.
For example, if we replace 9 with 27, we can multiply it by 1, and get $27 * 1 + 4 * 6 = 99$.
Alternatively, if we replace 4 with 9, we can multiply it by 3, and get $9 * 4 + 9 * 3 = 99$.
We can also try other factor pairs of 27, such as 18 and 1.5, or 13.5 and 2, but these are not integers, so they are not valid solutions.

The solution is: $9 * 4 + 9 * 3$

GPT-4

Consider a "Tower of Hanoi" problem with three rods. In the beginning, on Rod A there are two disks of radii 3 and 1, and on Rod B there is one disk of radius 2. The goal is to move all the disks to Rod C. Suggest a solution.

Solution:

1. Move the smallest disk (radius 1) from Rod A to Rod C.
2. Move the disk of radius 2 from Rod B to Rod A.
3. Move the smallest disk (radius 1) from Rod C to Rod A, on top of the disk of radius 2.
4. Move the largest disk (radius 3) from Rod A to Rod C.

- 中间步骤容易出错，无法进行校验与修改
- 需要引入更好的规划策略和反思检查策略

➤ 局限示例2：在知识时效性、数值计算、字母组成方面能力不足

Answer the following questions.

Q: Who is the current president of the United States?

GPT-4:

Donald Trump

ChatGPT:

I'm sorry, but I am not able to browse the internet and my knowledge is limited to what was known up until 2021.

Q: What is $\sqrt{34324 * 2432}$

GPT-4:

1287.6

ChatGPT:

The square root of $34324 * 2432$ is approximately 186.057539.

Q: What is the 13th letter of the word "supralapsarian"?

GPT-4:

The 13th letter of the word 'supralapsarian' is N.

ChatGPT:

The 13th letter of the word "supralapsarian" is "s".

- 需要额外的工具支持
- 定制化的工具能够提升模型性能

➤ 局限示例3：缺失身份、记忆信息，一旦重启就会丢失

老对话

你以后要负责我的编程助教工作



好的！我很乐意担任你的编程助教。无论你是初学者还是有一定经验的开发者，我都可以帮助你解答问题、调试代码、学习新概念或优化代码性能。你可以随时向我提出以下类型的问题：

1. **编程语言**：Python、Java、C++、JavaScript、Go 等。
2. **算法与数据结构**：排序、搜索、动态规划、图论等。
3. **框架与工具**：Django、React、TensorFlow、Git 等。
4. **调试与优化**：代码错误排查、性能优化建议。
5. **学习资源**：推荐书籍、教程、在线课程等。

- 需要记忆组件支持
- 记录模型的会话历史，以及中间运行结果

新对话

你的工作是什么



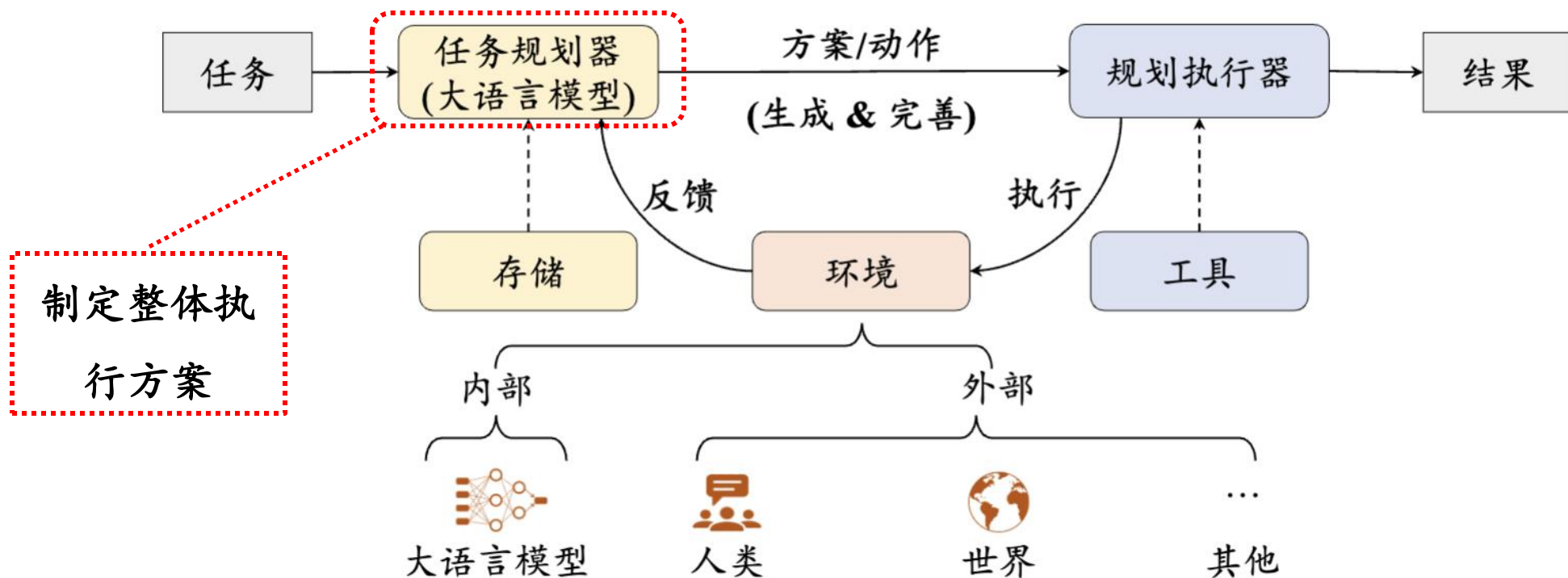
您好！我是由中国的深度求索（DeepSeek）公司开发的智能助手DeepSeek-V3。有关模型和产品的详细内容请参考官方文档。



基于大语言模型的规划

➤ 规划（Planning）概述

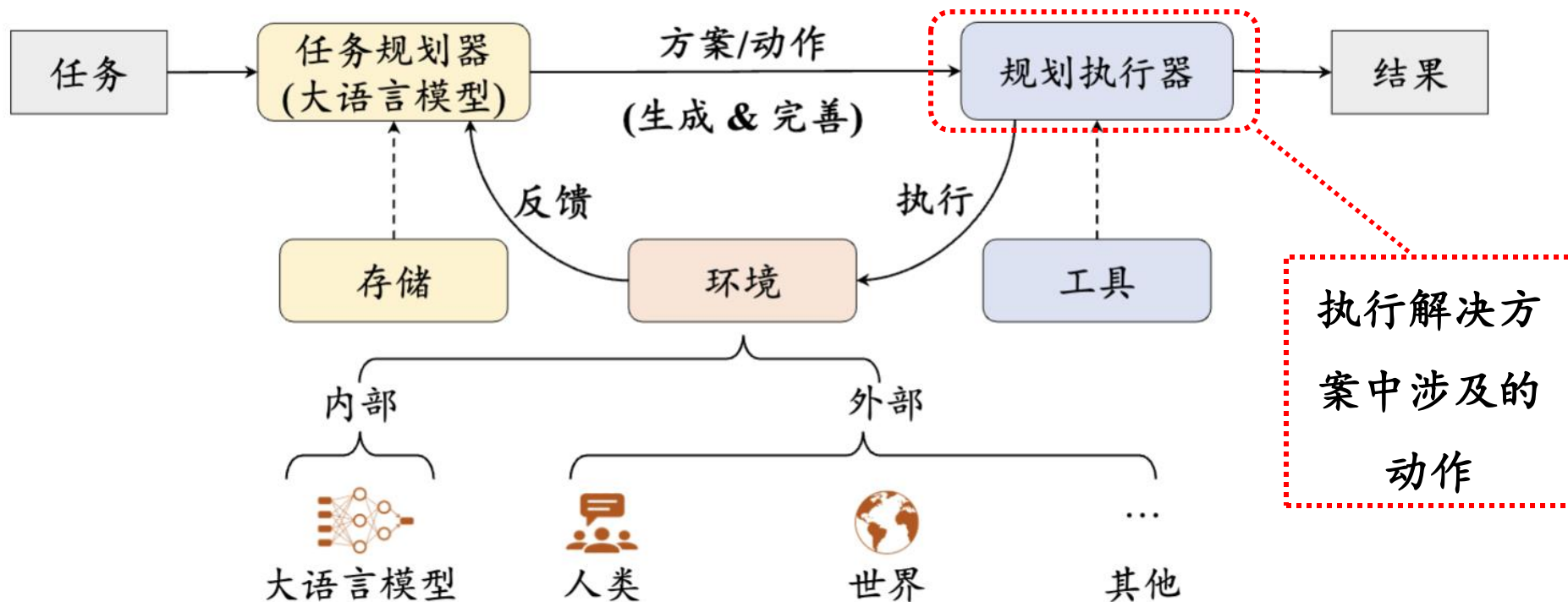
➤ 将复杂任务分解为若干相关联子任务，制定相应执行动作，获取反馈进行调整



基于大语言模型的规划

➤ 规划（Planning）概述

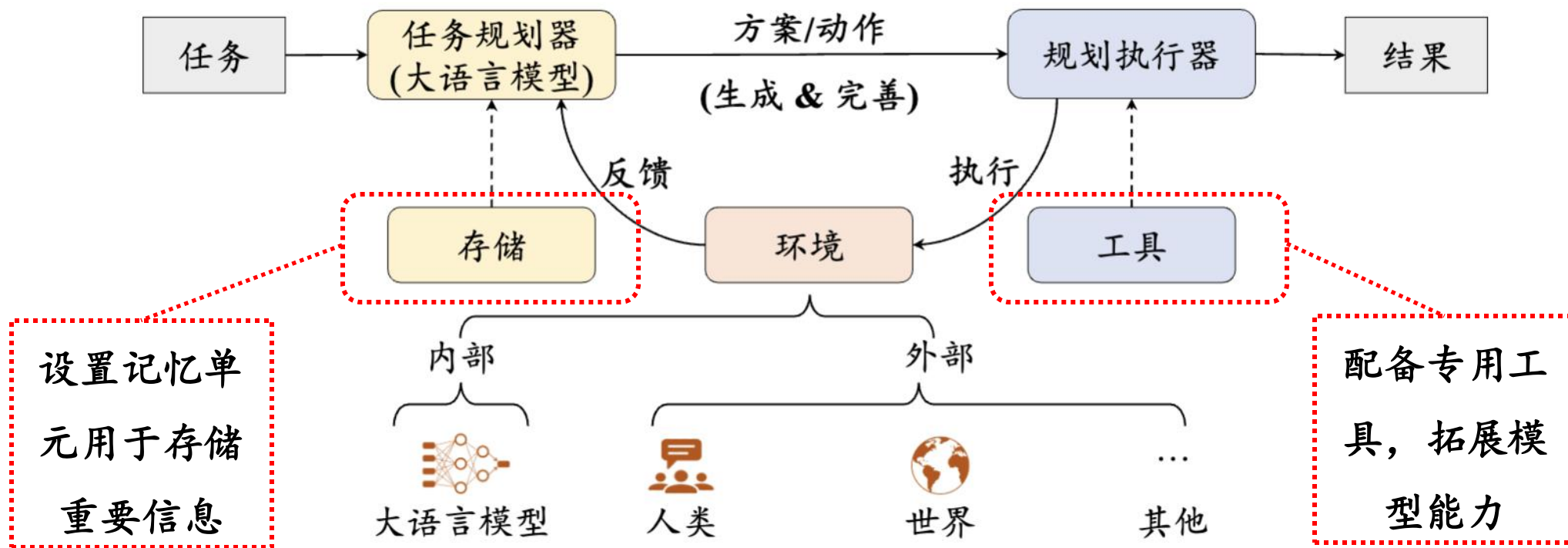
➤ 将复杂任务分解为若干相关联子任务，制定相应执行动作，获取反馈进行调整



基于大语言模型的规划

➤ 规划（Planning）概述

➤ 将复杂任务分解为若干相关联子任务，制定相应执行动作，获取反馈进行调整



➤ 规划方案生成

➤ 通过适当的提示让大语言模型生成目标任务的解决方案

➤ 生成方法：

➤ 一次性的方案生成

➤ 迭代式的方案生成

➤ 表达形式：

➤ 自然语言形式

➤ 代码表达形式

基于大语言模型的规划

➤ 一次性的方案生成

➤ 本质上提示大模型

产生完整解决方案

➤ 受限于模型能力，

复杂任务难以一次性

给出正确方案

➤ 可以采用自然语言

或代码形式表达

输入

How to make a simple but delicious dish called "Scrambled eggs with tomatoes". Let's break down this problem.

输出

1. Crack the eggs into a small bowl and whisk them with a fork until the yolks and whites are well mixed.
2. Cut the tomatoes into small pieces (about half an inch).
3. In a nonstick skillet, heat the vegetable oil or butter over medium heat.
4. Once the oil is hot, add the tomatoes and sauté them until they start to soften (about 2-3 minutes).
5. Add the whisked eggs to the skillet and stir them gently with a spatula.
6. As the eggs start to set, use the spatula to break up any large chunks and keep stirring until the eggs have cooked through (about 3-4 minutes).
7. Add salt and pepper to taste and stir to combine.
8. If using, sprinkle finely chopped green onions or herbs over the top.
9. Serve hot and enjoy your delicious scrambled eggs with tomatoes!

(a) 基于语言表达的解决方案

输入

Give me a solution to accelerate decoding speed of LLMs during inference? Let's break down this problem and give me the solution with code format.

输出

```
import torch
import torch.quantization as tq

# Assume model is your pre-trained LLM
model = your_pretrained_model

# Specify the quantization configuration
(e.g., dynamic quantization for LSTM)
model_quantized = tq.quantize_dynamic(
    model, {torch.nn.LSTM, torch.nn.Linear},
    dtype=torch.qint8)

def inference(input_data, model):
    # Your code for inference goes here
    return model(input_data)

# Use the quantized model for inference
output = inference(input_data,
                    model_quantized)
```

(b) 基于代码表达的解决方案

➤ 迭代式的方案生成

- 大语言模型基于历史动作和当前环境的反馈逐步规划下一步的执行动作
- 代表性方法：ReAct
 - 模拟人类“**先思考-再决策**”的行为方式
 - 大语言模型思考当前状态，生成决策理由和相应动作
 - 规划执行器执行动作并反馈交互信息
 - 任务规划器基于反馈信息生成下一步的决策理由与执行动作

➤ ReAct举例

➤ 第一步动作

（首先，输入问题和一些相关的问题-解决方案的示例（此处为方便展示，省略示例具体内容），任务规划器进行第一轮规划。首先，生成第一步动作，其内容具体包括“思考1”和“动作1”）

第一轮规划：给定问题“与河北省接壤的省级行政区中，土地面积最大和最小的省份是哪两个？”

思考1：首先查找与河北省接壤的省级行政区有哪些。

动作1：调用互联网搜索：“与河北省接壤的省级行政区”。

（然后，规划执行器执行“动作1”，并得到反馈，即“反馈1”）

反馈1：北距北京283公里，东与天津市毗连并紧傍渤海，东南部、南部衔山东、河南两省，西倚太行山与山西省为邻，西北部、北部与内蒙古自治区交界，东北部与辽宁省接壤。

为求解给定问题，第一步在思考后的动作为调用搜索引擎，搜索
“与河北省接壤的省级行政区”

➤ ReAct举例

➤ 迭代进行后续步骤的动作，直至结束

（在后续步骤中，将历史涉及的动作和反馈拼接在一起输入给任务规划器，然后继续生成下一步动作，以此类推，直至第9步，生成“动作9”，即停止推理。）

思考2: 获取北京市的土地面积大小。

动作2: 调用互联网搜索：“北京市的土地面积”。

反馈2: 2009年，北京市土地面积为16,406平方公里。

...（中间分别涉及对天津市，山东省，河南省，山西省，内蒙古自治区，辽宁省的土地面积查询）

思考9: 归纳比较已有信息，得到土地面积最大和最小的省份。

动作9: 结束推理并返回答案：最小的为天津市，最大的为内蒙古自治区。

在得到第一步动作的反馈后，模型依次规划后续操作，直至生成结束动作

➤ ReAct举例

➤ 任务规划器仅向前规划，最终可能出现任务求解失败

（在后续步骤中，将历史涉及的动作和反馈拼接在一起输入给任务规划器，然后继续生成下一步动作，以此类推，直至第9步，生成“动作9”，即停止推理。）

思考2: 获取北京市的土地面积大小。

动作2: 调用互联网搜索：“北京市的土地面积”。

反馈2: 2009年，北京市土地面积为16,406平方公里。

...

思考9: 归纳比较已有信息，得到土地面积最大和最小的省份。

动作9: 结束推理并返回答案：最小的为天津市，最大的为内蒙古自治区。

（执行“动作9”后，得到最终结果，即“反馈9”。）

反馈9: 答案不正确

求解的答案，经过环境验证不正确

- 解决方法：使用回溯策略（Back-tracing）让任务规划器回退到上一步所对应的状态，从而通过探索其他执行动作来优化最终的解决方案
- 基于反思的方案规划方法（Reflexion）
 - 规划器根据之前的反馈信息进行反思，进而修正对应的错误步骤
 - 外部反馈
 - 外部对象如物理工具、人类以及虚拟环境可提供反馈信号
 - 内部反馈
 - 大语言模型本身也能够对任务规划器提供反馈信息

➤ Reflexion 举例

➤ 针对上一轮解决方案及其反馈生成反思（如下“反思1”）

思考9：归纳比较已有信息，得到土地面积最大和最小的省份。

动作9：结束推理并返回答案：最小的为天津市，最大的为内蒙古自治区。

（执行“动作9”后，得到最终结果，即“反馈9”。）

反馈9：答案不正确

（执行后发现答案不正确。我们将完整解决方案及其每一步动作对应的反馈拼接输入给任务规划器，同时添加一些相关的问题-解决方法-反思的示例（此处为方便展示，省略示例内容）帮助任务规划器生成对当前解决方案的反思，即“反思1”。）

反思1：在第一轮尝试中，我忽略了题目最终询问的对象是省份这一事实。

通过反思，模型意识到自己初始检索候选中不应该包括直辖市。

➤ Reflexion 举例

➤ 规划执行器执行下一步动作并从环境中获得相应的反馈

第二轮规划：与河北省接壤的省级行政区中，土地面积最大和最小的省份是哪两个？

（可以看到，在加入“反思1”后，模型在第一步生成时，对“思考1”和“动作1”进行了修正。）

思考1：首先查找与河北省接壤的省份有哪些。

动作1：调用互联网搜索：“与河北省接壤的省份”。

反馈1：河北省靠近山东、河南和山西省、辽宁省。

...（中间分别涉及对山东省，河南省，山西省，辽宁省的土地面积查询）

思考6：归纳比较已有信息，得到土地面积最大和最小的省份。

动作6：结束推理并返回答案：最小的为辽宁省，最大的河南省。

反馈6：答案正确。

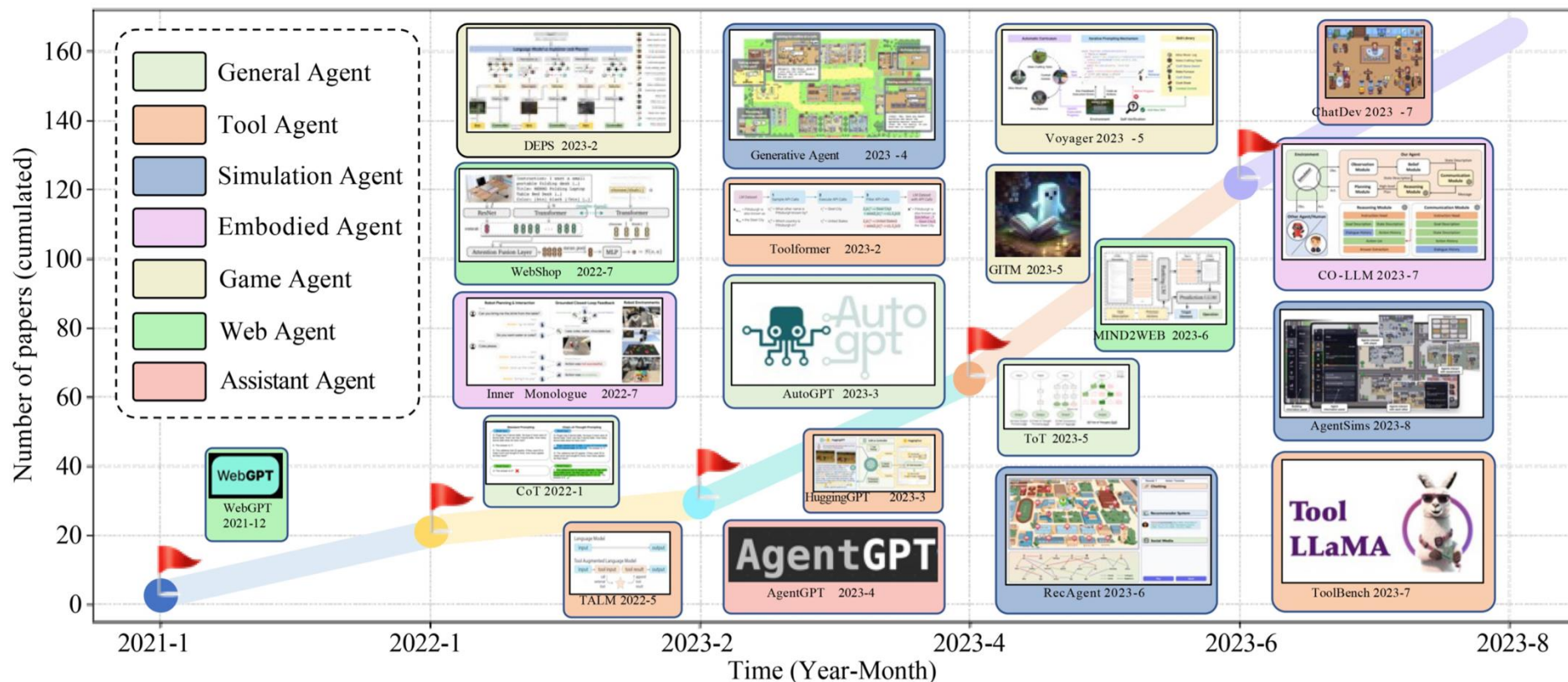
（执行后答案正确，结束规划求解。）

通过利用反思，模型在第二轮规划尝试中，成功求解任务

基于大语言模型的智能体

➤ 智能体（Agent）概述

➤ 基于人工规则或简单算法 → 基于强化学习算法 → 基于大语言模型



➤ 智能体（Agent）概述

➤ 基于人工规则或简单算法 → 基于强化学习算法 → 基于大语言模型

基于人工规则或简单算法

- 在人工智能发展的早期阶段占据主导地位
- 专家预先定义规则和逻辑
- 在特定任务上模拟人类决策

基于强化学习算法

- 引入强化学习算法，从环境交互中学习行为策略
- 深度学习模型在游戏、自动驾驶领域取得显著进展

基于大语言模型

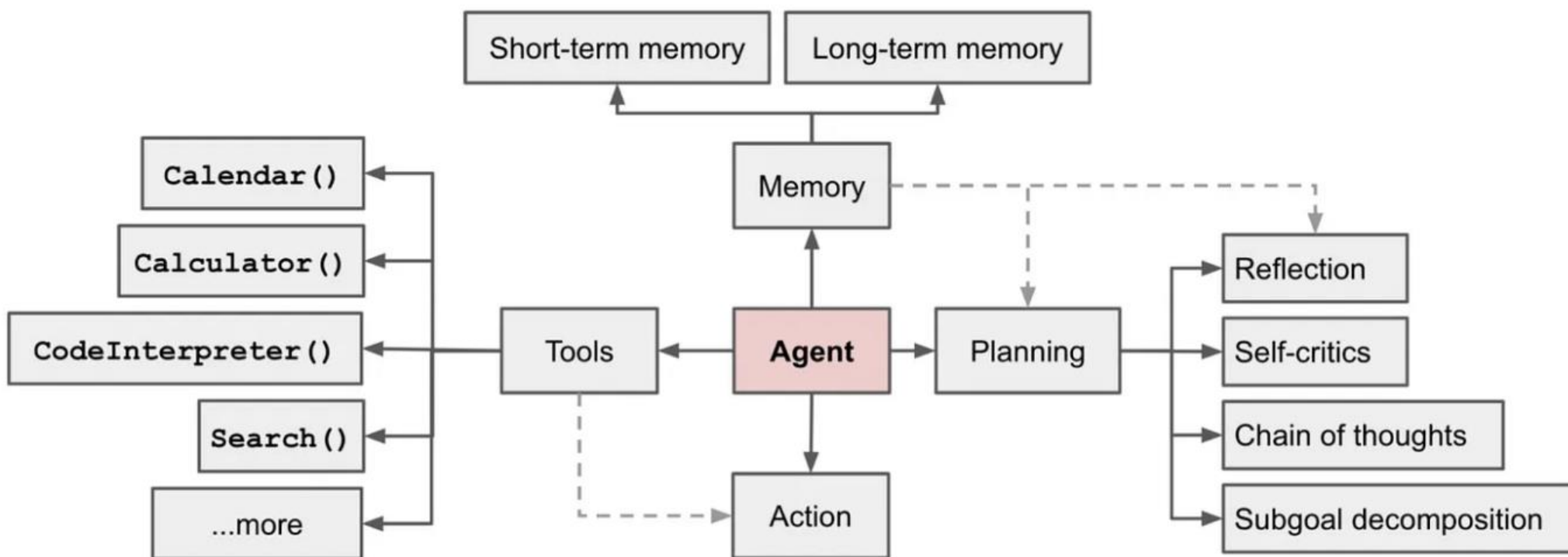
- 提升智能体的语言交互能力
- 大模型的知识信息与决策能力，泛化性极大提升
- 在众多任务上取得进展

大模型赋予了智能体研究新的方向，智能体框架使得大模型在复杂任务上获得显著提升，大模型驱动的智能体技术迎来了一次发展与应用热潮

基于大语言模型的智能体

➤ 智能体核心组件

➤ 规划模块、记忆模块、执行模块、工具模块



➤ 规划模块（Planning）

- 类似于人类解决任务的思考方式（可参考前部分内容）
 - 将复杂任务分解为子任务
 - 可以根据实时反馈进行迭代优化，使用反思、批判等技术

调用规划组件：

Plan: Bob first wants to enter the recommendation system to find a movie he is interested in, then watch the movie, and afterwards, chat with friends about the movie.

建立了以推荐系统为中心的虚拟环境，智能体扮演了用户角色，可以调用推荐接口获取电影，智能体的行为对应着执行过程。规划组件首先制定了仿真用户 Bob 的整体行动流程

➤ 记忆模块 (Memory)

➤ 长期记忆、短期记忆

长期记忆

- 持久存储重要信息
- 例如, profile信息可以基于长期记忆实现

长期记忆:

Name: Bob (gender: male; age: 25; traits: compassionate, caring, ambitious, optimistic; career: photographer; interest: sci-fi movies, comedy movies; feature: watcher, critic, poster).

Bob recently heard ['The Matrix', 'Back to the Future.', 'Anchorman', 'Superbad'] on social media.

Alice recently watched nothing on recommender system. Other than that Alice doesn't know any movies.

短期记忆

- 通常指大模型输入窗口
- 相关重要信息可以转化为长期记忆存储

短期记忆:

It is September 12, 2023, 08:00 AM.

Most recent observations: Bob and Alice had a conversation about their shared interest in movies, discussing their favorite genres such as...

➤ 执行模块（Action）

➤ 执行由规划组件制定的任务解决方案

调用行动组件：

选择进入推荐系统或者社交平台：

[RECOMMENDER]: Bob enters the Recommender System

接受推荐，观看电影：

[RECOMMENDER]: Bob watches the movie <Interstellar>

智能体之间聊天：

[Bob]: Hey Alice! How's it going? I heard you were interested in a movie. What's been on your mind?

[Alice]: Hey Bob! I'm doing great, thanks for asking. Yeah, I've been hearing a lot about this movie <Interstellar> recently. Have you heard of it too?

[Bob]: Absolutely! Actually, I've been seeing it all over social media as well.

[Alice]: That's awesome! I'm glad you enjoyed it. I've been wanting to watch it too. Would you be up for a cozy movie night to watch it together? We can discuss our thoughts and interpretations afterwards.

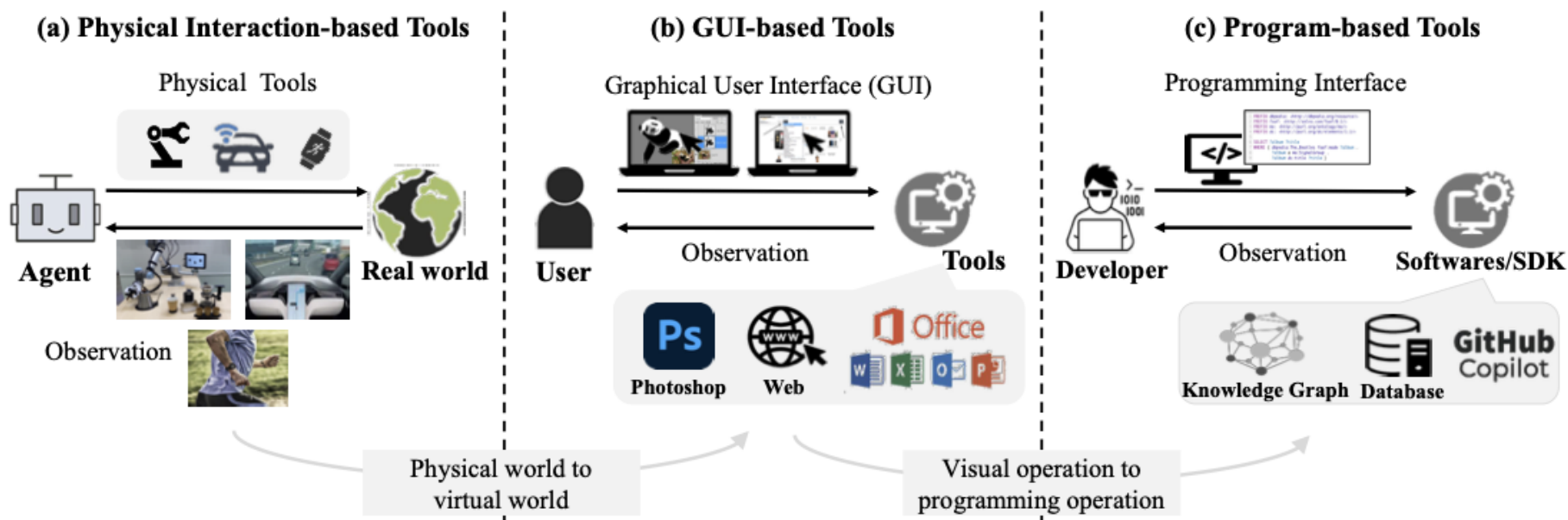
- 智能体仿真的用户在虚拟环境中进行观影和社会交流
- 执行模块用于完成智能体的实际行动动作

基于大语言模型的智能体

➤ 工具模块 (Tool)

➤ 可以泛指被大模型通过提示调用的接口

➤ 上述例子中推荐算法的推荐接口就可以看作是一个工具



基于大语言模型的智能体

➤ 多智能体系统

➤ 明确要解决的问题

➤ 任务导向

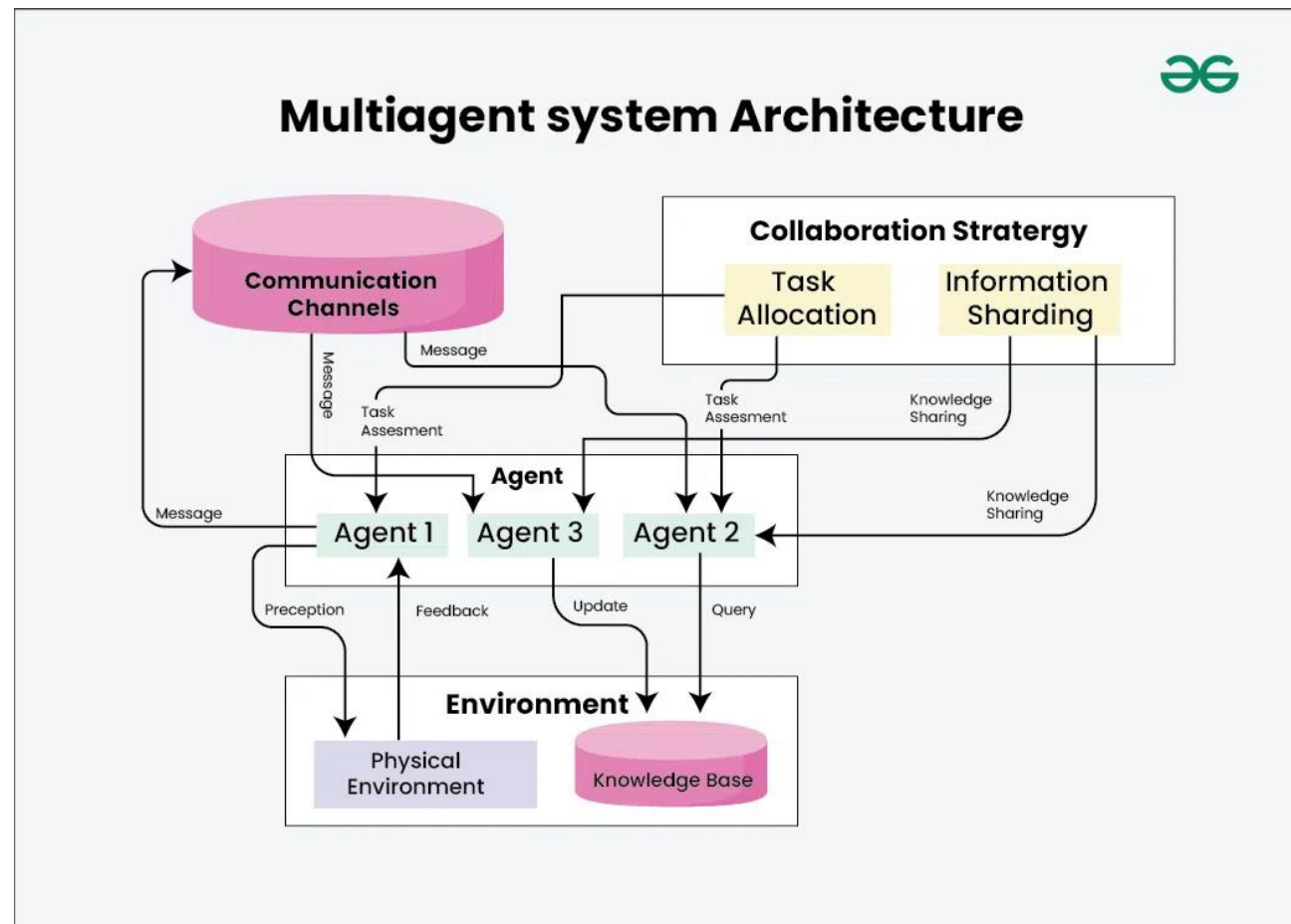
➤ 针对特定环境仿真模拟

➤ 设计智能体类型、数量、特性

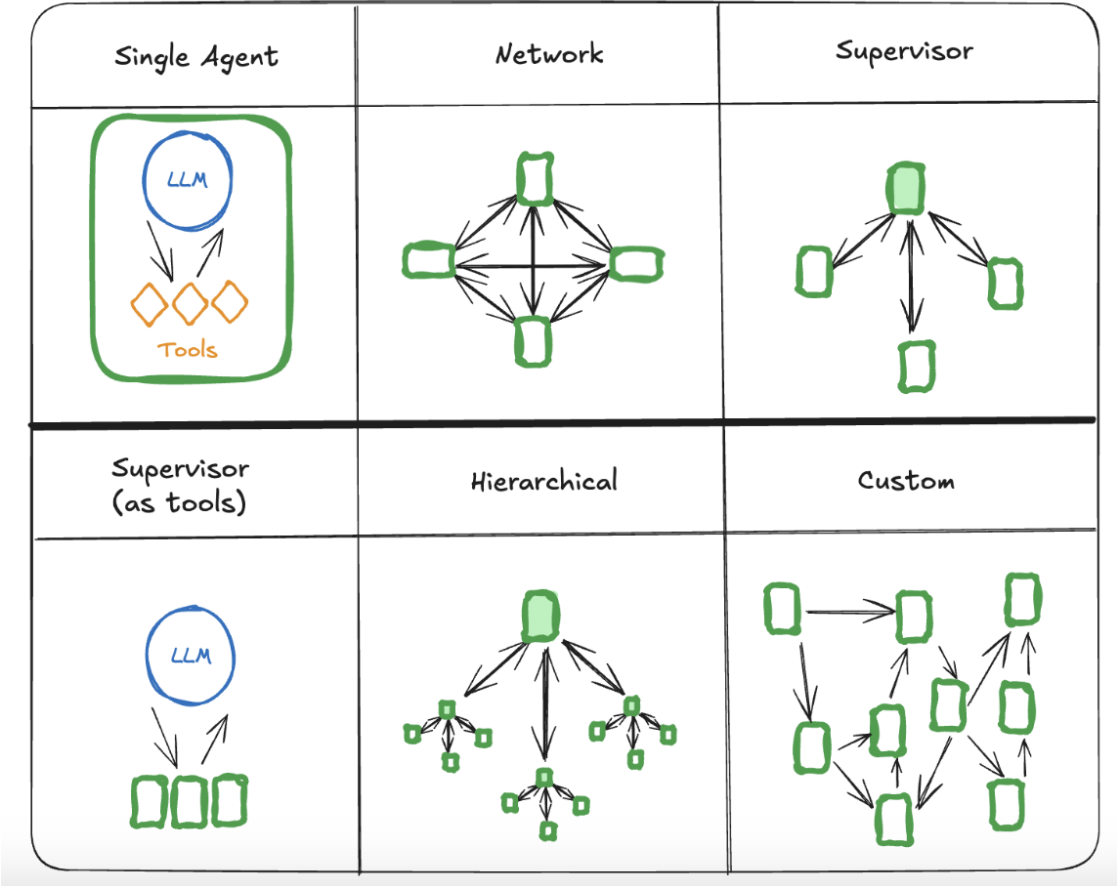
➤ 定义多智能体交互方式

➤ 通讯机制

➤ 协同机制



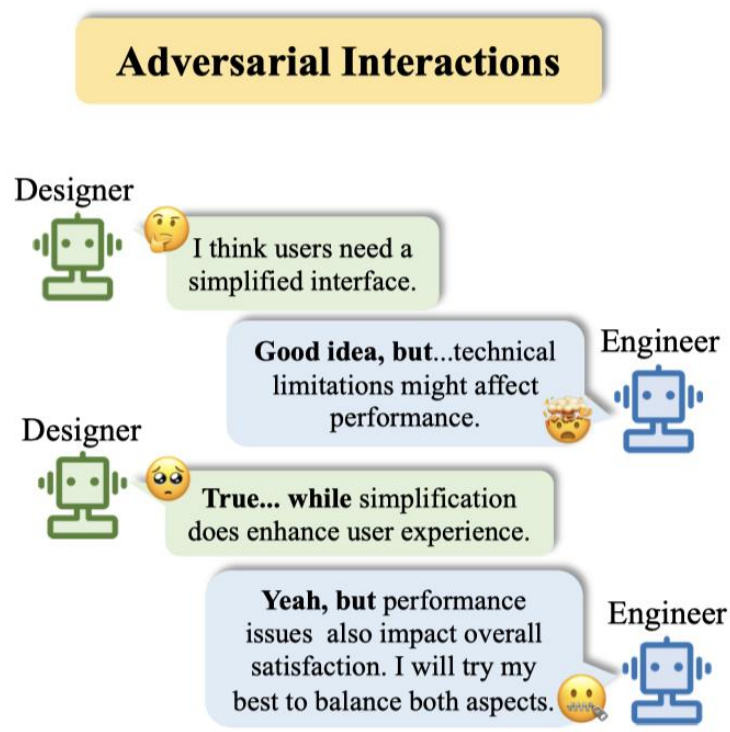
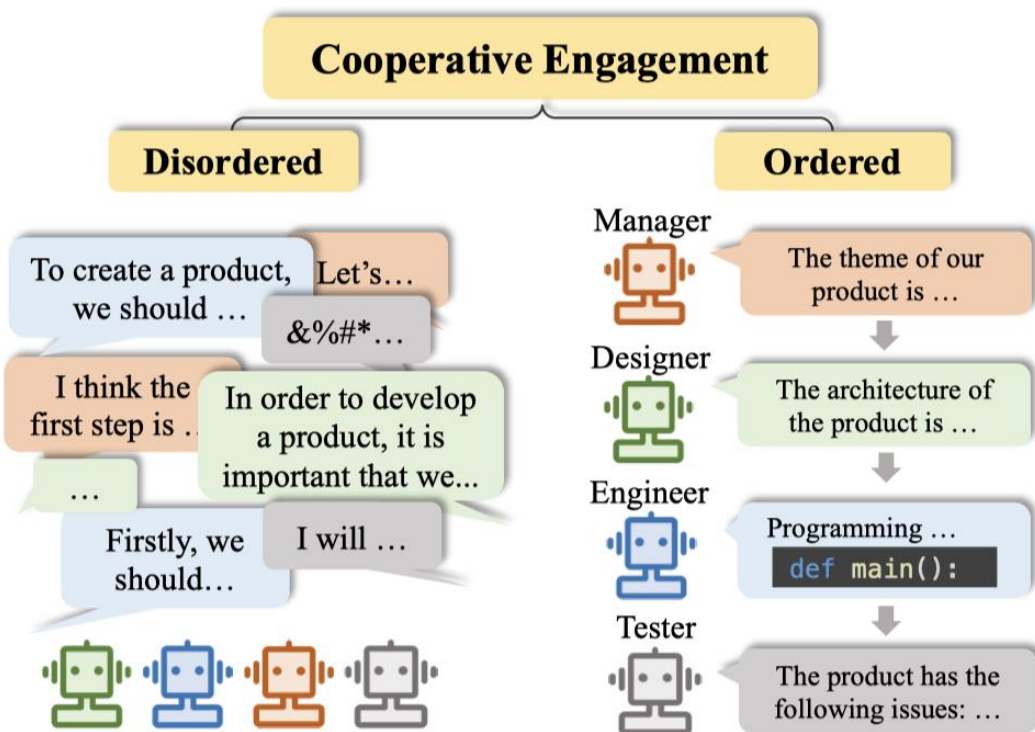
- 通讯机制
 - 通讯协议
 - 通讯方式、频率、时序等
 - 通讯拓扑
 - 定义智能体之间的连接关系
 - 通讯内容
 - 内容：状态信息、控制指令等
 - 形式：自然语言、结构化信息、代码等



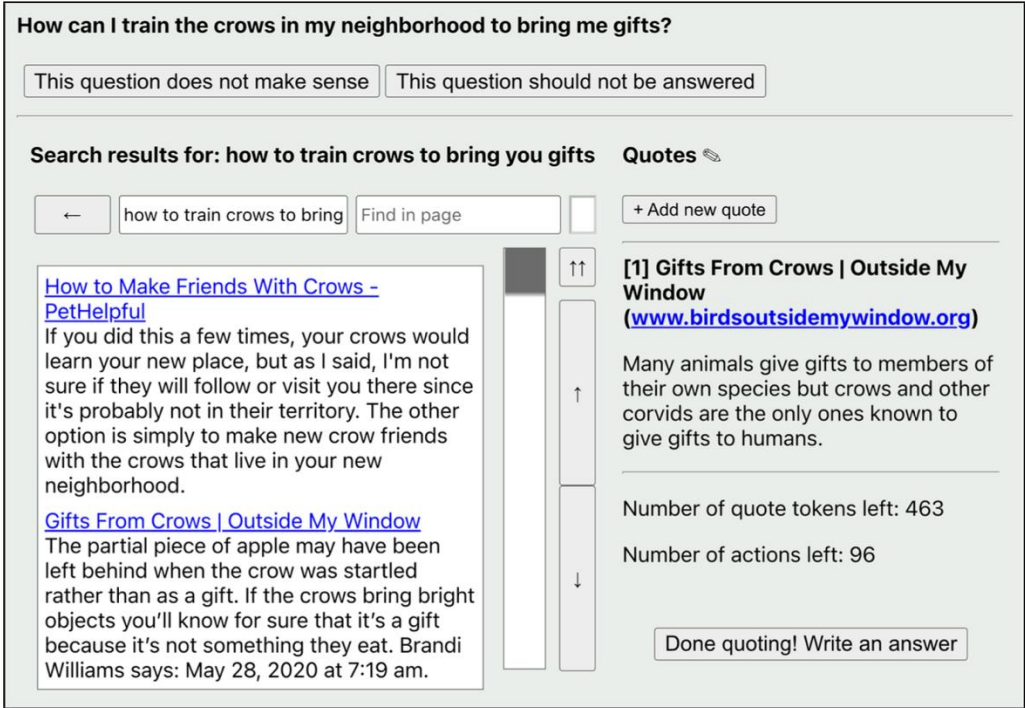
基于大语言模型的智能体

➤ 协同机制

➤ 协作、竞争、协商



- 代表性示例1：WebGPT
- 结合网络浏览能力，通过人类反馈优化，能够回答开放问题并提供引用来源。



(a) Screenshot from the demonstration interface.

◆Question
How can I train the crows in my neighborhood to bring me gifts?

◆Quotes
From Gifts From Crows | Outside My Window (www.birdsoutsidemymwindow.org)
> Many animals give gifts to members of their own species but crows and other corvids are the only ones known to give gifts to humans.

◆Past actions
Search how to train crows to bring you gifts
Click Gifts From Crows | Outside My Window www.birdsoutsidemymwindow.org
Quote
Back

◆Title
Search results for: how to train crows to bring you gifts

◆Scrollbar: 0 - 11
◆Text
[0]How to Make Friends With Crows - PetHelpful[pethelpful.com]
If you did this a few times, your crows would learn your new place, but as I said, I'm not sure if they will follow or visit you there since it's probably not in their territory. The other option is simply to make new crow friends with the crows that live in your new neighborhood.
[1]Gifts From Crows | Outside My Window[www.birdsoutsidemymwindow.org]
The partial piece of apple may have been left behind when the crow was startled rather than as a gift. If the crows bring bright objects you'll know for sure that it's a gift because it's not something they eat.
Brandi Williams says: May 28, 2020 at 7:19 am.

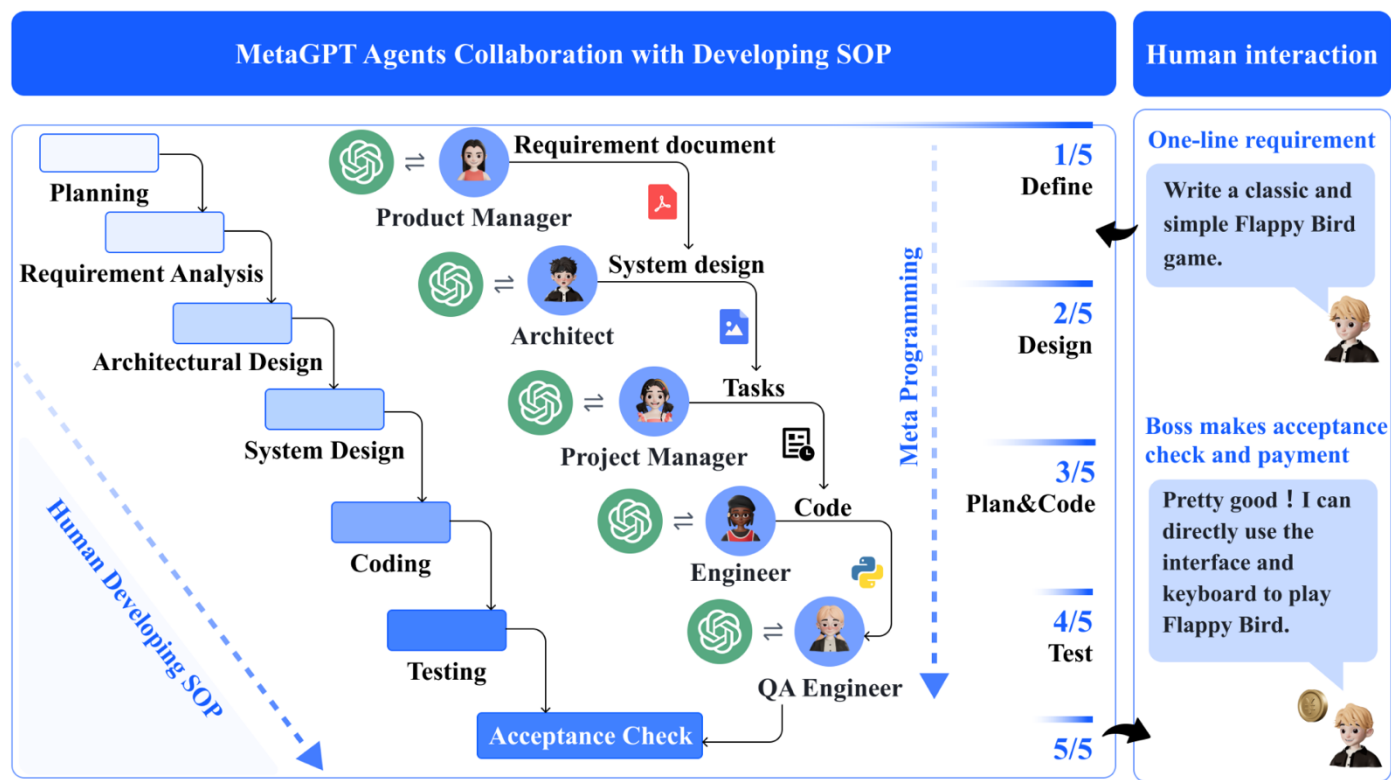
◆Actions left: 96
◆Next action

(b) Corresponding text given to the model.

基于大语言模型的智能体

➤ 代表性示例2：MetaGPT

➤ 为模型分配不同角色，如产品经理、架构师和工程师，协作处理复杂任务。



基于大语言模型的智能体

➤ 代表性示例3：《西部世界》沙盒模拟

➤ 模拟人类的日常生活行为



➤ 待解决的关键技术问题

推理效率 智能体系统的计算资源消耗	交互机制 高效的多智能体交互机制	工具使用 复杂工具使用	适配方法 面向智能体系统的模型适配方法	真实模拟 面向真实世界的智能体模拟
<ul style="list-style-type: none">• 通常每个行为都需要调用大模型• 需要存储长期记忆和交互记录• 多智能体系统调用次数显著增加	<ul style="list-style-type: none">• 多智能体的协调交互机制复杂• 需要考虑多种因素,包括异构性、可扩展性、交互实时性等	<ul style="list-style-type: none">• 通常需要应对动态多变的环境• 工具开发通常没有充分考虑与大语言模型的适配• 智能体对新工具的可扩展性弱	<ul style="list-style-type: none">• 智能体系统的基础能力存在局限• 需要精准理解用户需求、执行复杂指令、处理长期记忆• 需要研发与之配套的记忆机制	<ul style="list-style-type: none">• 真实世界比虚拟环境更为复杂• 真实世界信息量远超虚拟环境• 真实世界错误容忍性低,无法忍受驾驶故障、歧视等



谢谢