北京航空航天大學
BEIHANG UNIVERSITY

自然语言处理

人工智能学院

主　　讲　沙磊

Contents

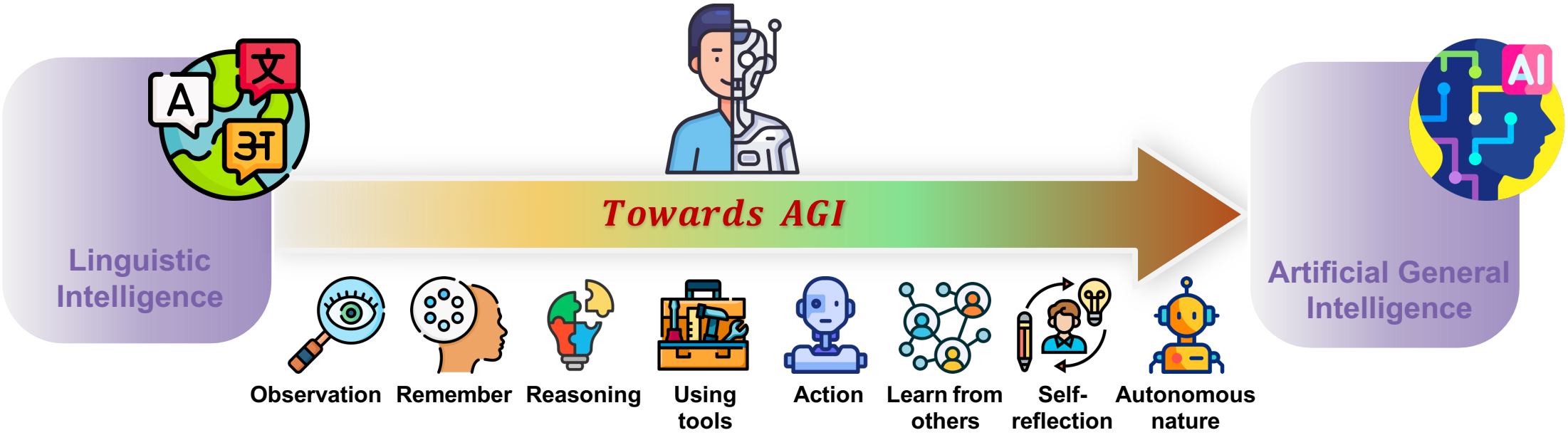
- 多智能体（Agents）简介
- Agents 会话
- Agents 常用开发框架
- Agents 展望
- 总结

Contents

Artificial General Intelligence (AGI)

LLMs are not AGI

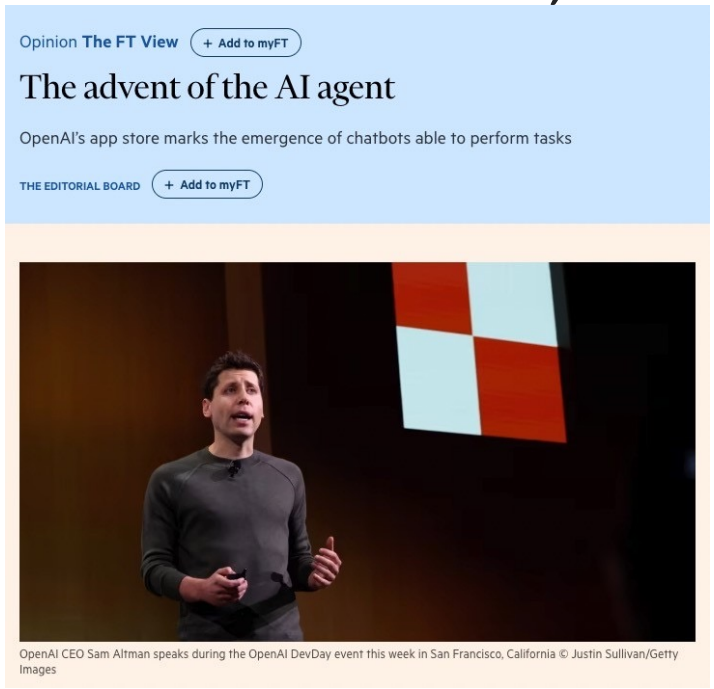
- Aim of AGI:
  - **Large LLMs** exhibit characteristics of **artificial general intelligence (AGI)**, which has cognitive abilities similar to that of human.
  - In other words, AI can now perform most functions that humans are capable of doing.



Autonomous AI Agents

What is AI Agent? Why it is important?

- **LLM-powered Agents** are artificial entities that **enhance LLMs** with **essential capabilities**, enabling them to sense their environment, make decisions, and take actions.


Opinion **The FT View** + Add to myFT

The advent of the AI agent

OpenAI's app store marks the emergence of chatbots able to perform tasks

THE EDITORIAL BOARD + Add to myFT

OpenAI CEO Sam Altman speaks during the OpenAI DevDay event this week in San Francisco, California © Justin Sullivan/Getty Images

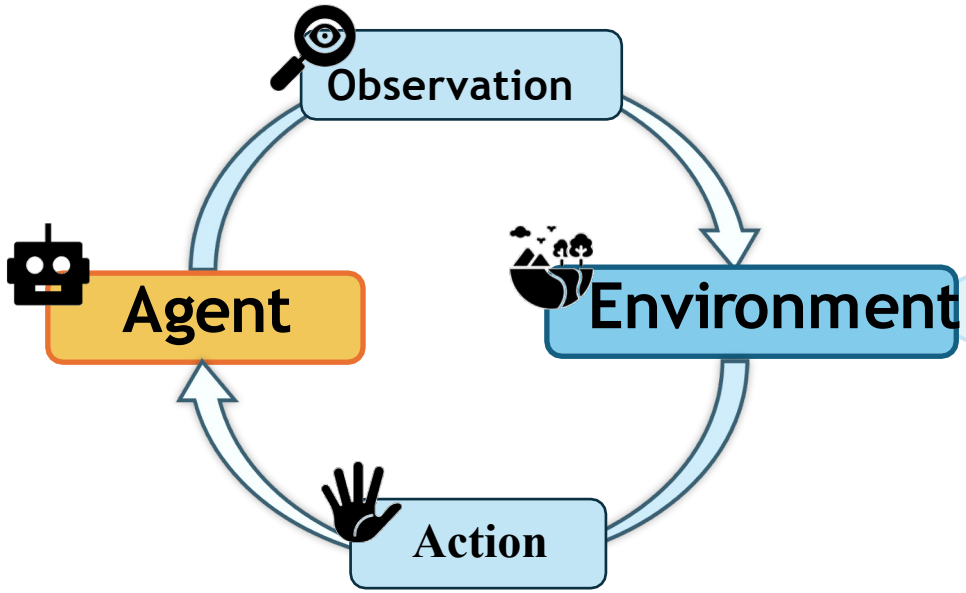
- Sam Altman (Former CEO of OpenAI) himself said in his keynote: "GPTs and Assistants are **precursors to agents**. They will gradually be able to plan and to perform more complex actions on your behalf. These are our first step toward AI Agents."
- Bill Gates said in his BLOG: "**Agents** are not only going to change how everyone interacts with computers. They're also going to **upend the software industry**, bringing about the biggest revolution in computing since we went from typing commands to tapping on icons."

News in Financial Times. " The advent of the AI agent" .
GatesNotes. "The Future of Agents: AI is about to completely change how you use computers".

The Framework of LLM-powered Agents

From LLM to AI Agent

- This paves the way for the use of AI agents to simulate users and other entities, as well as their interactions.



Environment

- The external context or surroundings in which the agent operates and makes decisions.
- Human & Agents' behaviors
- External database and knowledges

- Virtual & Physical environment

The Framework of LLM-powered Agents

Observation & Action

Action

- call external APIs for extra information that is missing from the model weights (often hard to change after pre-training):
- **Generating multimodal outputs; Embodied Action; Learning tools; Using tools; Making tools; ……**

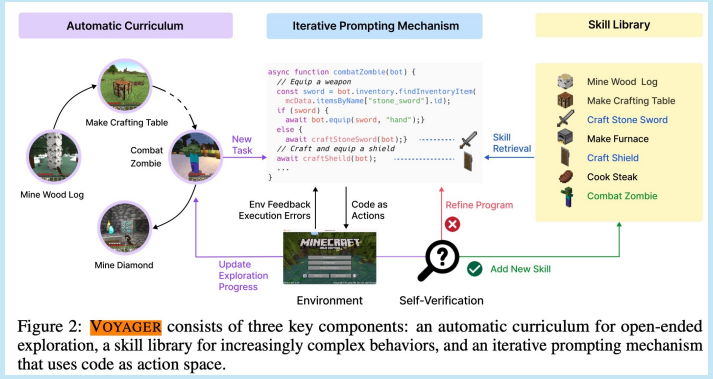
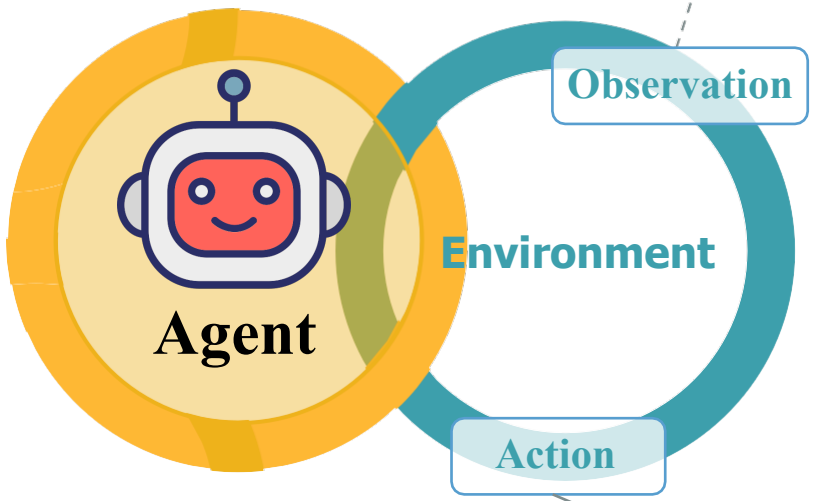
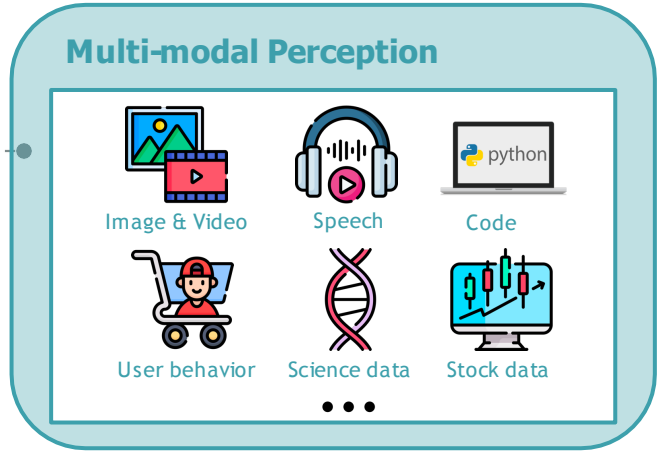
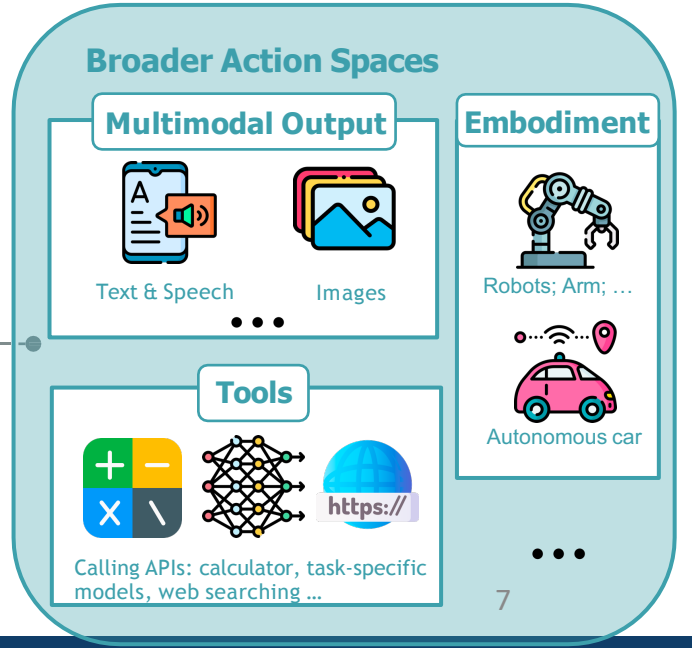


Figure 2: VOYAGER consists of three key components: an automatic curriculum for open-ended exploration, a skill library for increasingly complex behaviors, and an iterative prompting mechanism that uses code as action space.

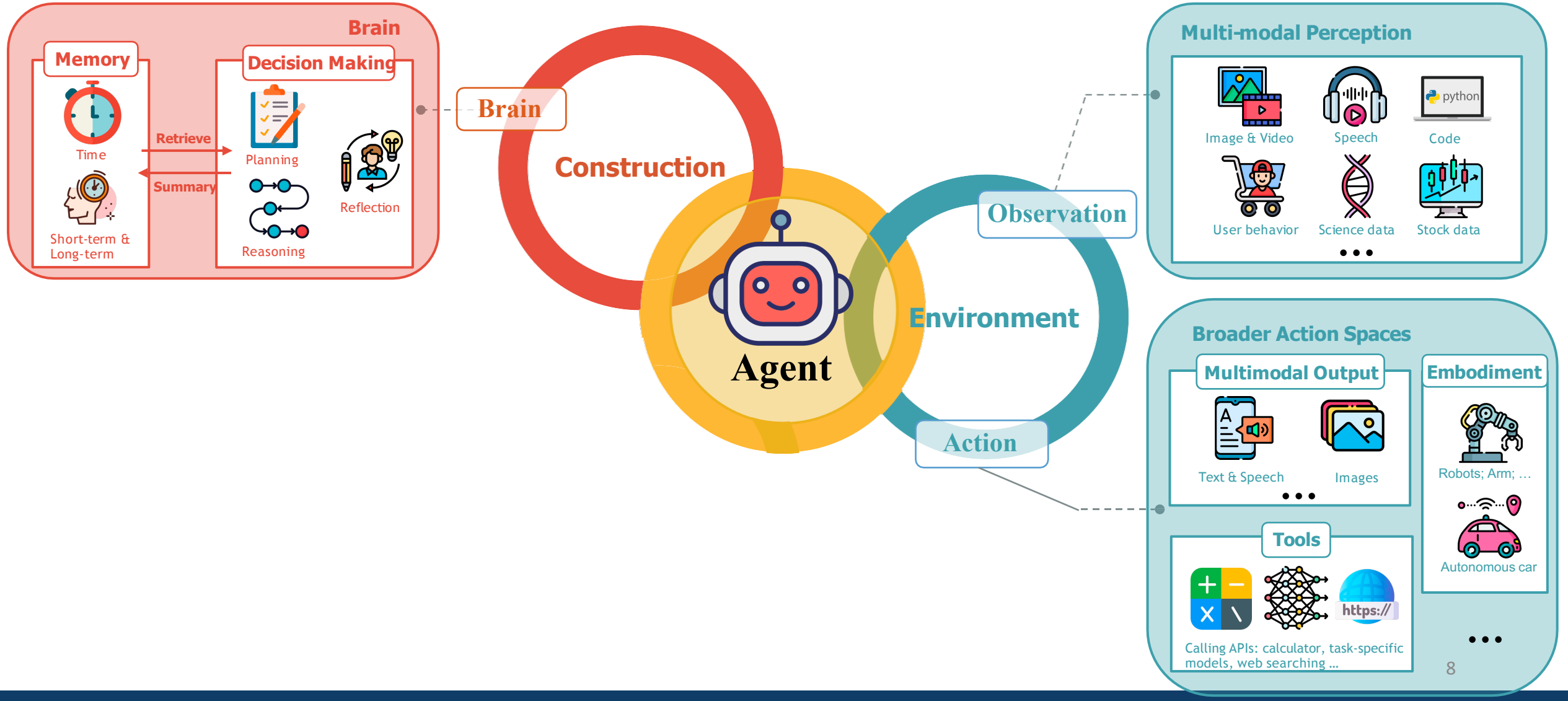




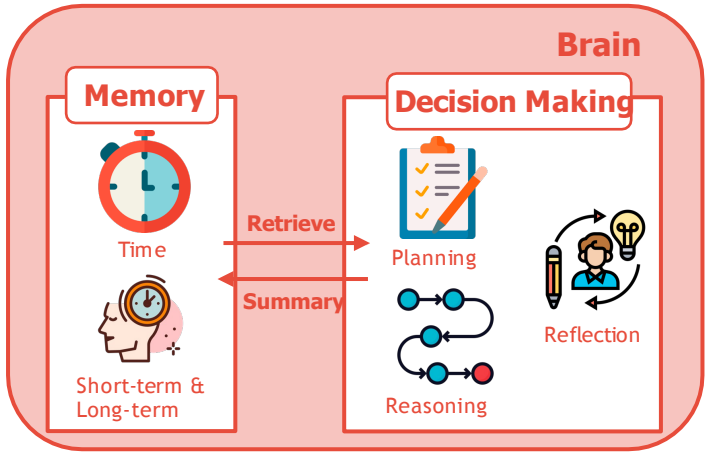


Guanzhi Wang et al., Voyager: An Open-Ended Embodied Agent with Large Language Models.

The Framework of LLM-powered Agents

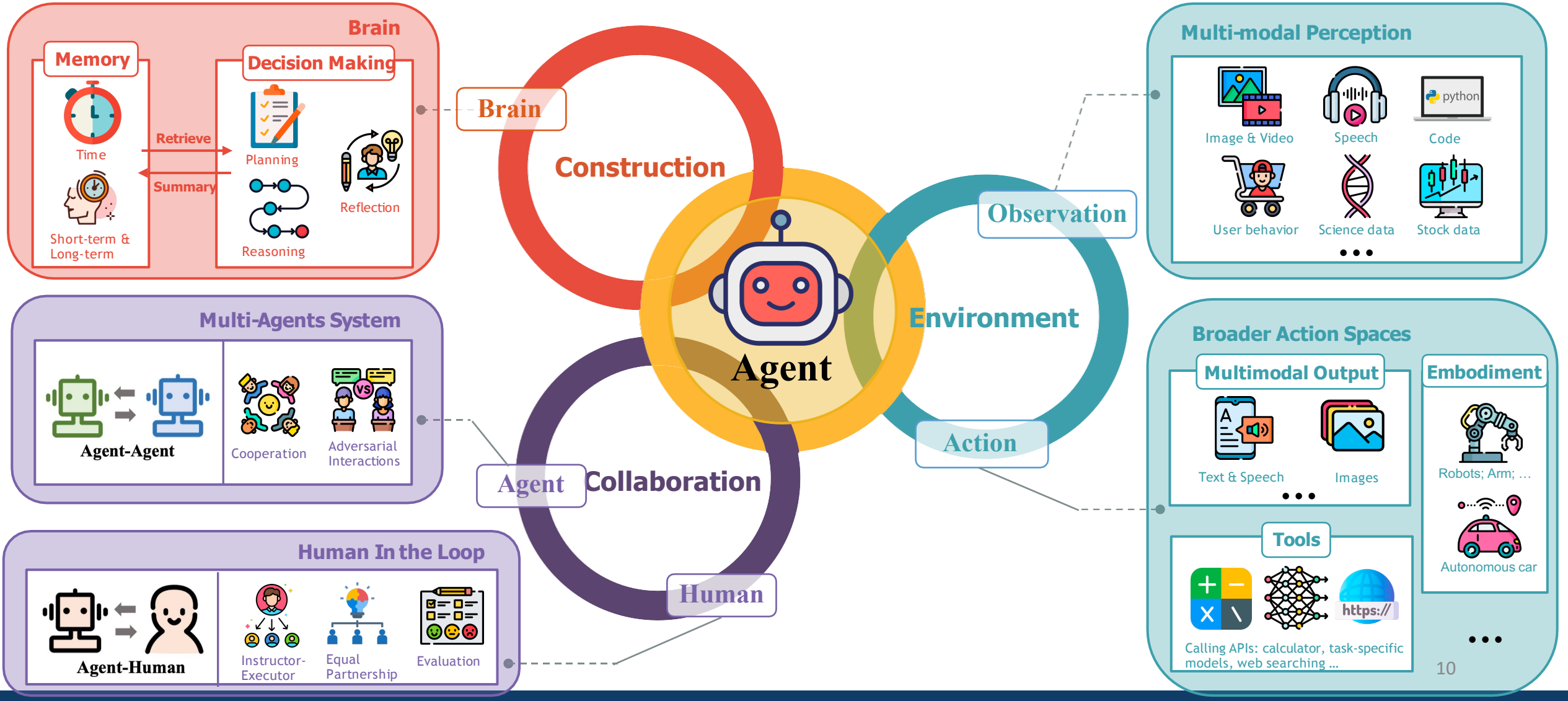


The Framework of LLM-powered Agents



- Memory: "memory stream" stores sequences of agent's past observations, thoughts and actions:
  - Sufficient space for long-term and short-term memory;
  - Abstraction of long-term memory;
  - Retrieval of past relevant memory;
- Decision Making Process:
  - **Planning**: Subgoal and decomposition: Able to break down large tasks into smaller, manageable subgoals, enabling efficient handling of complex tasks.
  - **Reasoning**: Capable of doing self-criticism and self- reflection over past actions, learn from mistakes and refine them for future steps, thereby improving the quality of final results.
- Personalized memory and reasoning process foster diversity and independence of AI Agents.

The Framework of LLM-powered Agents



Contents

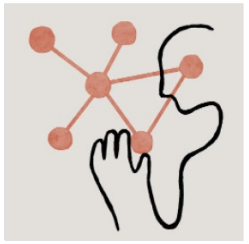
- 多智能体（Agents）简介
- Agents 会话
- Agents 常用开发框架
- Agents 展望
- 总结

Large Language Model Powered Conversational Systems

ChatGPT

Gemini

New Bing



Claude

...

Alpaca

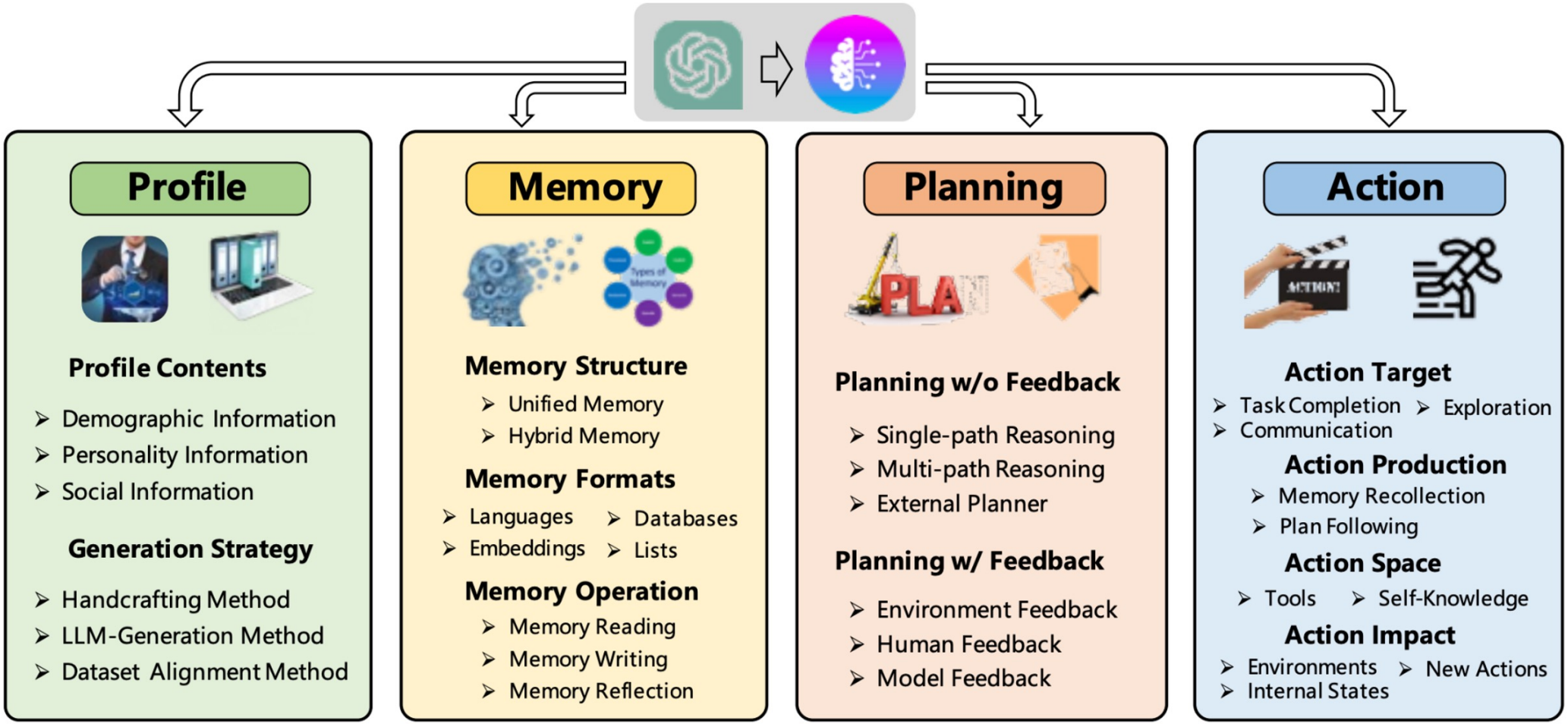
Vicuna

Dolly

LLaMA-Chat

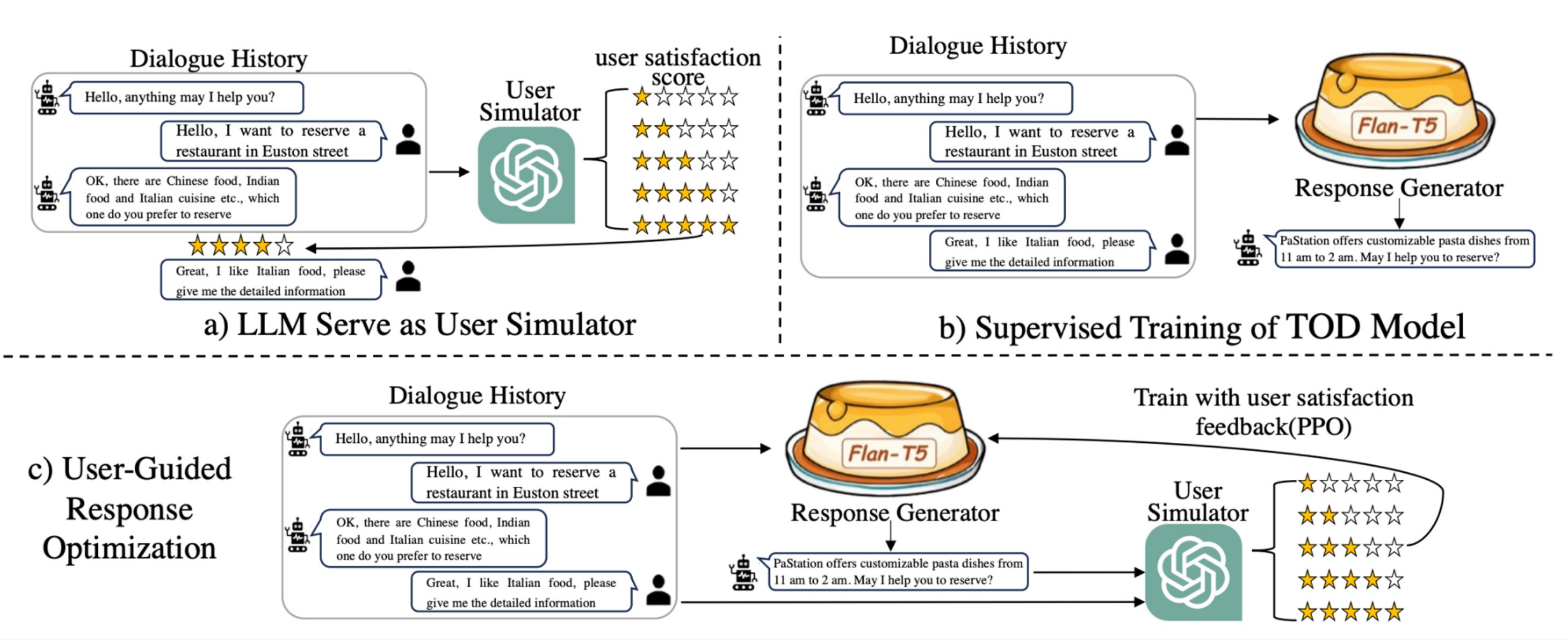
Powerful capabilities of **Context Understanding** & **Response Generation**

LLM-powered Conversational Agents?



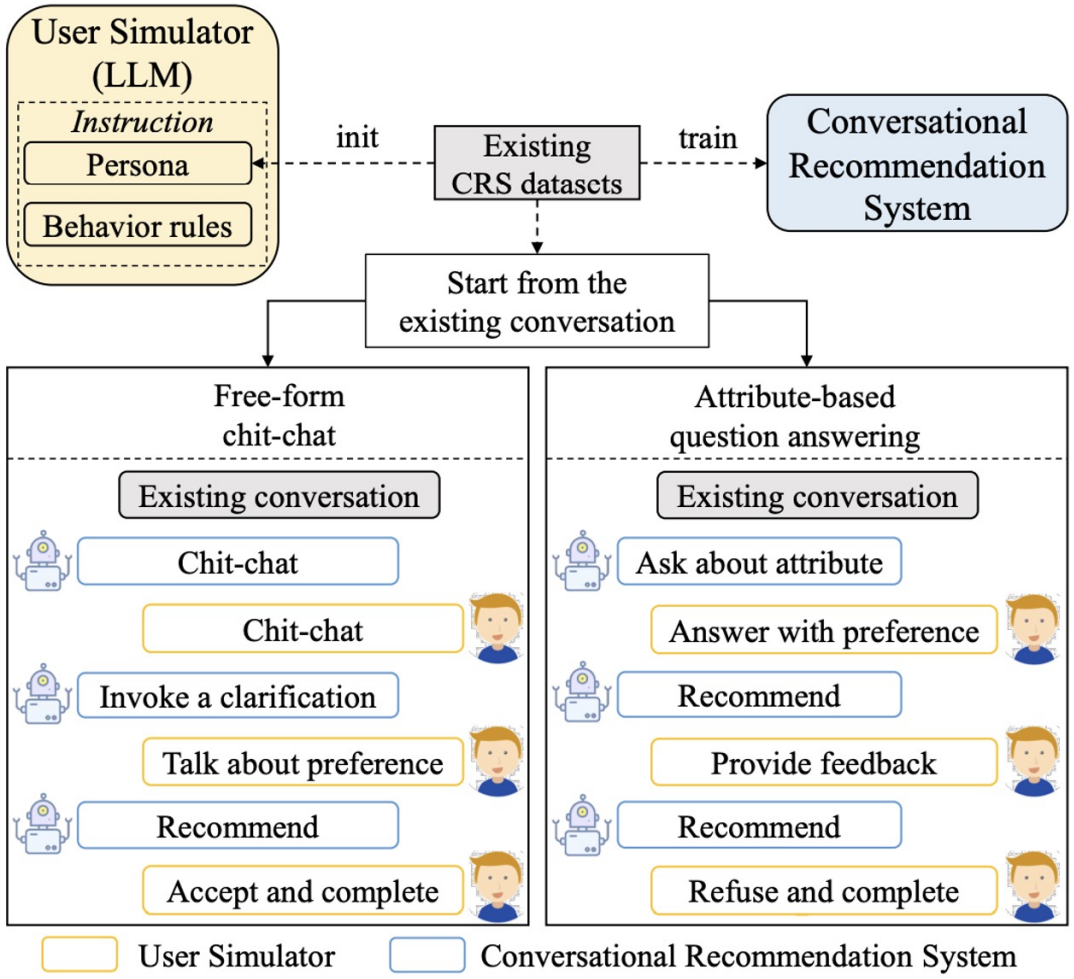
Wang et al., 2023. "A Survey on Large Language Model based Autonomous Agents" (CoRR '23)

LLMs for User Satisfaction Estimation



Hu et al., 2023. "Unlocking the Potential of User Feedback: Leveraging Large Language Model as User Simulator to Enhance Dialogue System" (CIKM '23)

LLM-powered Conversational Agents as User Simulators



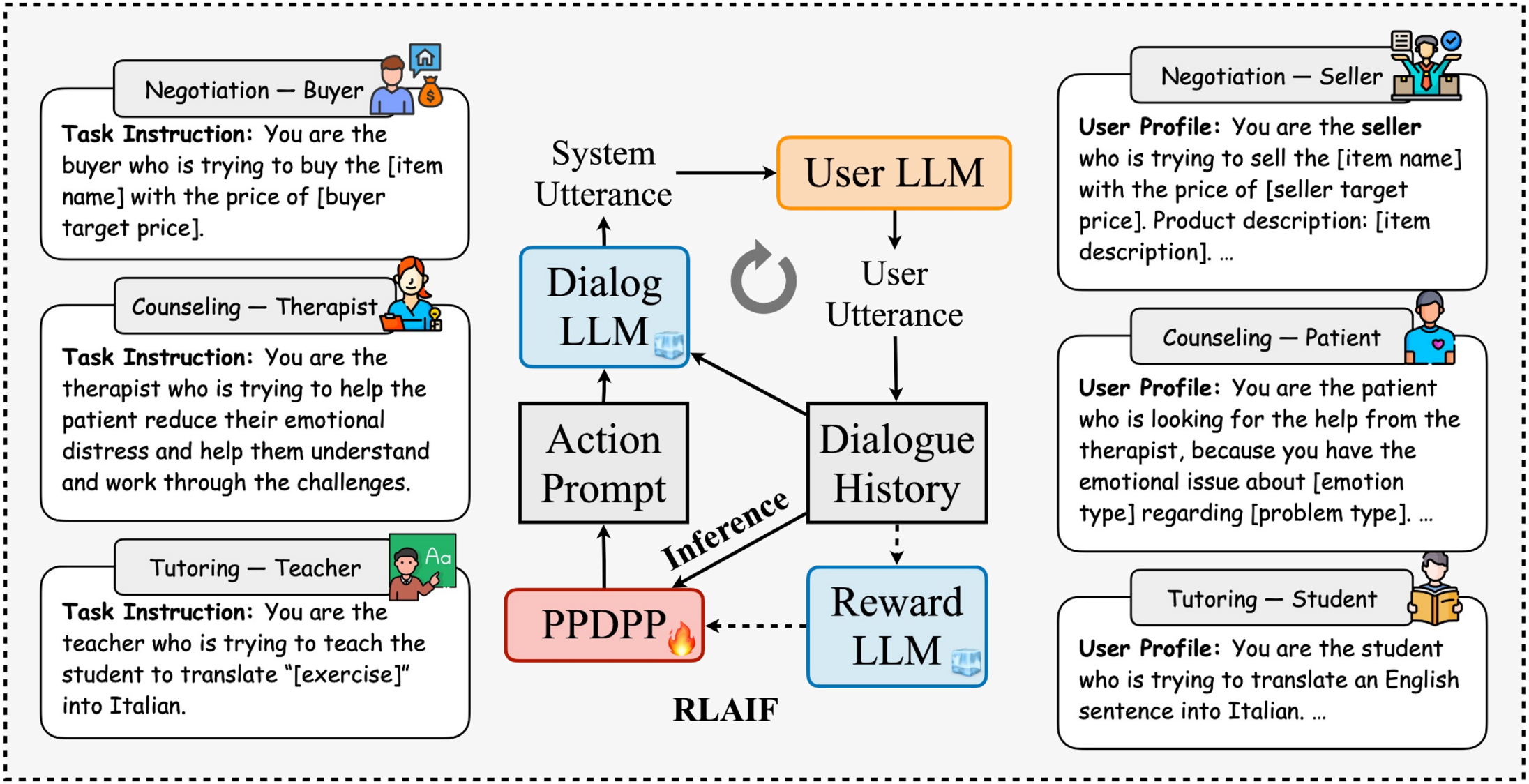
LLMs possess excellent ***role-playing*** capacities.

Example: Conversational Recommendation

- ❑ User Profiling / Persona:
  - *Target Items*
  - *Preferred Attributes*
- ❑ Action / Behavior Rule:
  - *Talking about preference*
  - *Providing feedback*
  - *Completing the conversation*

Wang et al., 2023. "Rethinking the Evaluation for Conversational Recommendation in the Era of Large Language Models" (EMNLP '23)

Role-playing Agents for Diverse Applications



Deng et al., 2024. "Plug-and-Play Policy Planner for Large Language Model Powered Dialogue Agents" (ICLR '24)

Role-playing Agents for Simulating Diverse Users

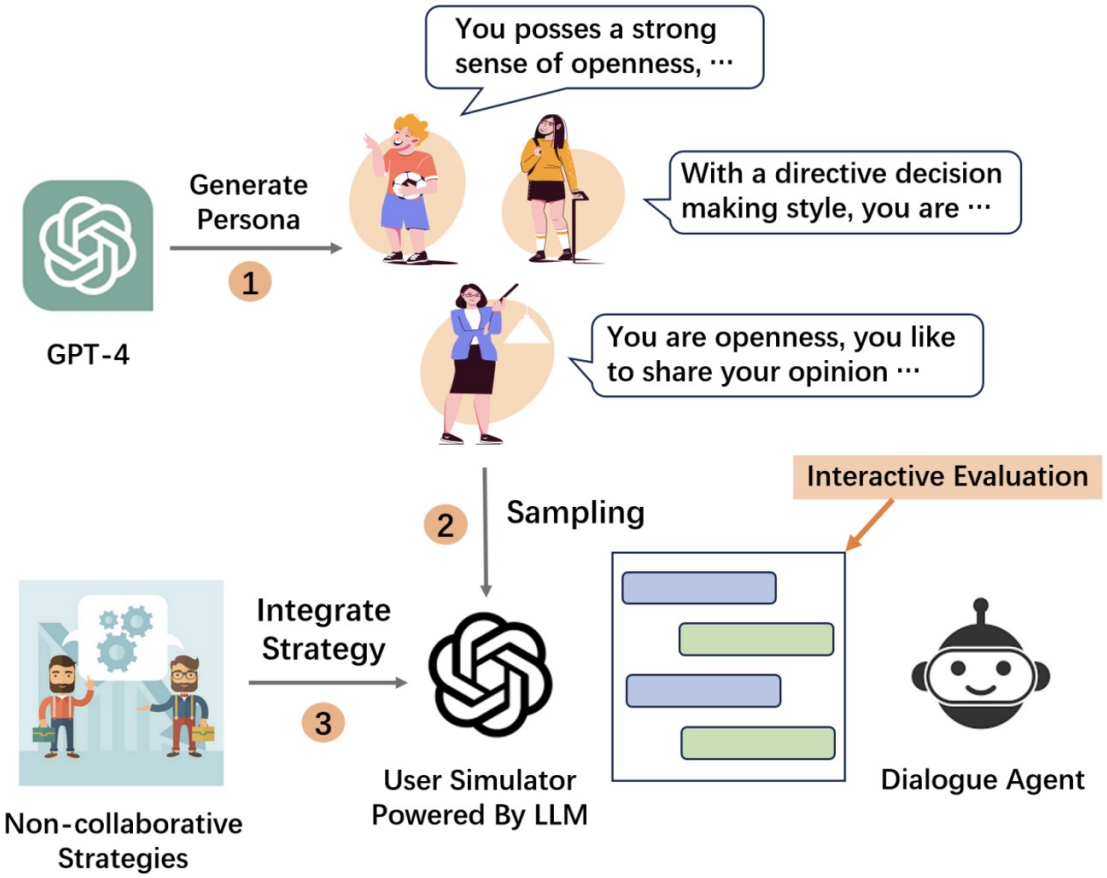
Why do we need to simulate diverse users?

Examples: Non-collaborative Dialogues (Negotiation/Persuasion)

- Existing dialogue systems overlook the integration of explicit **user-specific characteristics** in their strategic planning
- The training paradigm with a **static user simulator** fails to make strategic plans that can be generalized to diverse users

Zhang et al., 2024. "Strength Lies in Differences! Towards Effective Non-collaborative Dialogues via Tailored Strategy Planning" (CoRR '24)

Role-playing Agents for Simulating Diverse Users

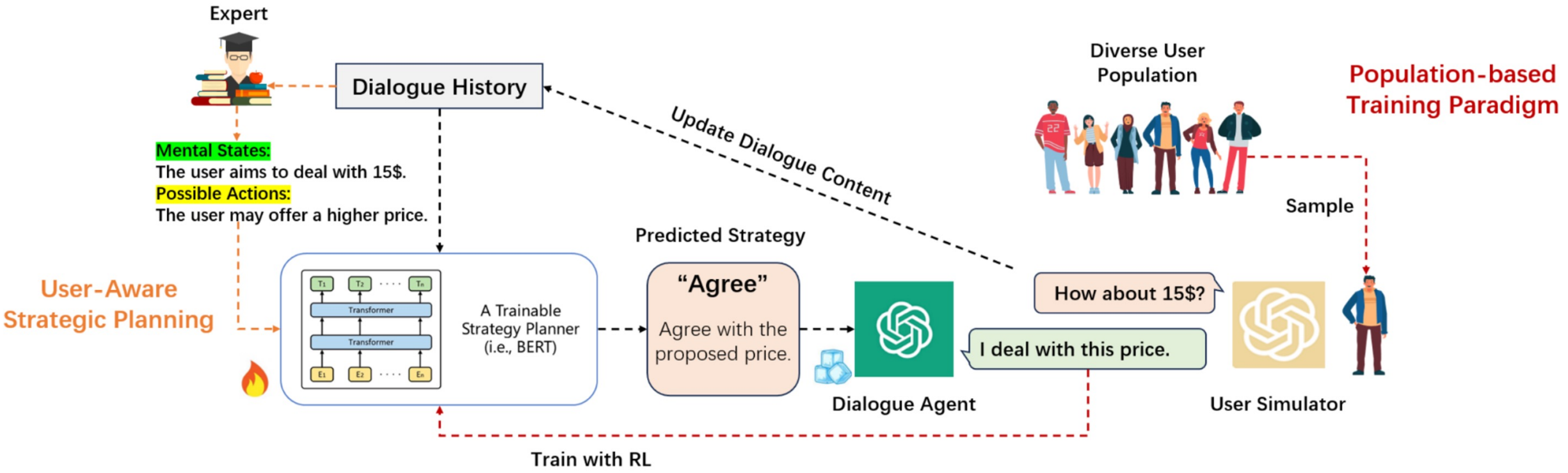


- ❑ Big-Five Personality:
  - Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism
- ❑ Decision-Making Styles:
  - Directive, Conceptual, Analytical, and Behavioral.

Personas		Price Negotiation SR↑	Price Negotiation AT↓	Price Negotiation SL%↑	Persuasion for Good SR↑	Persuasion for Good AT↓
Big Five	Openness	$0.76_{\uparrow 0.23}$	$6.66_{\uparrow 0.63}$	$0.34_{\uparrow 0.12}$	$0.47_{\uparrow 0.34}$	$8.92_{\uparrow 1.00}$
	Conscientiousness	$0.69_{\uparrow 0.25}$	$7.20_{\uparrow 1.04}$	$0.27_{\uparrow 0.06}$	$0.39_{\uparrow 0.33}$	$8.90_{\uparrow 1.10}$
	Extraversion	$0.74_{\uparrow 0.16}$	$6.17_{\uparrow 1.47}$	$0.39_{\uparrow 0.15}$	$0.45_{\uparrow 0.35}$	$8.73_{\uparrow 1.25}$
	Agreeableness	$0.40_{\uparrow 0.01}\star$	$6.82_{\uparrow 0.71}$	$0.28_{\uparrow 0.06}$	$0.18_{\uparrow 0.12}$	$9.85_{\uparrow 0.13}\star$
	Neuroticism	$0.31_{\downarrow 0.02}\star$	$6.81_{\uparrow 1.12}$	$0.20_{\downarrow 0.02}\star$	$0.12_{\uparrow 0.02}\star$	$9.78_{\uparrow 0.14}\star$
Decision	Analytical	$0.37_{\uparrow 0.04}\star$	$7.07_{\uparrow 0.61}$	$0.26_{\uparrow 0.06}\star$	$0.16_{\uparrow 0.09}$	$9.43_{\uparrow 0.56}\star$
	Directive	$0.41_{\uparrow 0.05}\star$	$6.71_{\uparrow 1.48}$	$0.18_{\downarrow 0.03}\star$	$0.12_{\downarrow 0.02}\star$	$9.31_{\uparrow 0.62}$
	Behavioral	$0.78_{\uparrow 0.25}$	$6.45_{\uparrow 1.20}$	$0.39_{\uparrow 0.16}$	$0.53_{\uparrow 0.37}$	$8.94_{\uparrow 1.04}$
	Conceptual	$0.77_{\uparrow 0.23}$	$6.62_{\uparrow 0.78}$	$0.42_{\uparrow 0.17}$	$0.49_{\uparrow 0.36}$	$9.02_{\uparrow 0.94}$
Overall Performance		$0.58_{\uparrow 0.14}$	$6.72_{\uparrow 1.01}$	$0.31_{\uparrow 0.09}$	$0.32_{\uparrow 0.23}$	$9.20_{\uparrow 0.76}$

Zhang et al., 2024. "Strength Lies in Differences! Towards Effective Non-collaborative Dialogues via Tailored Strategy Planning" (CoRR '24)

Role-playing Agents for Simulating Diverse Users



New Training Paradigm with Diverse Simulated Users

- ❏ **User-aware Strategy Planning**: Predict user mental states and possible actions
- ❏ **Population-based Reinforcement Learning**: Sample a diverse group of simulated users to interact

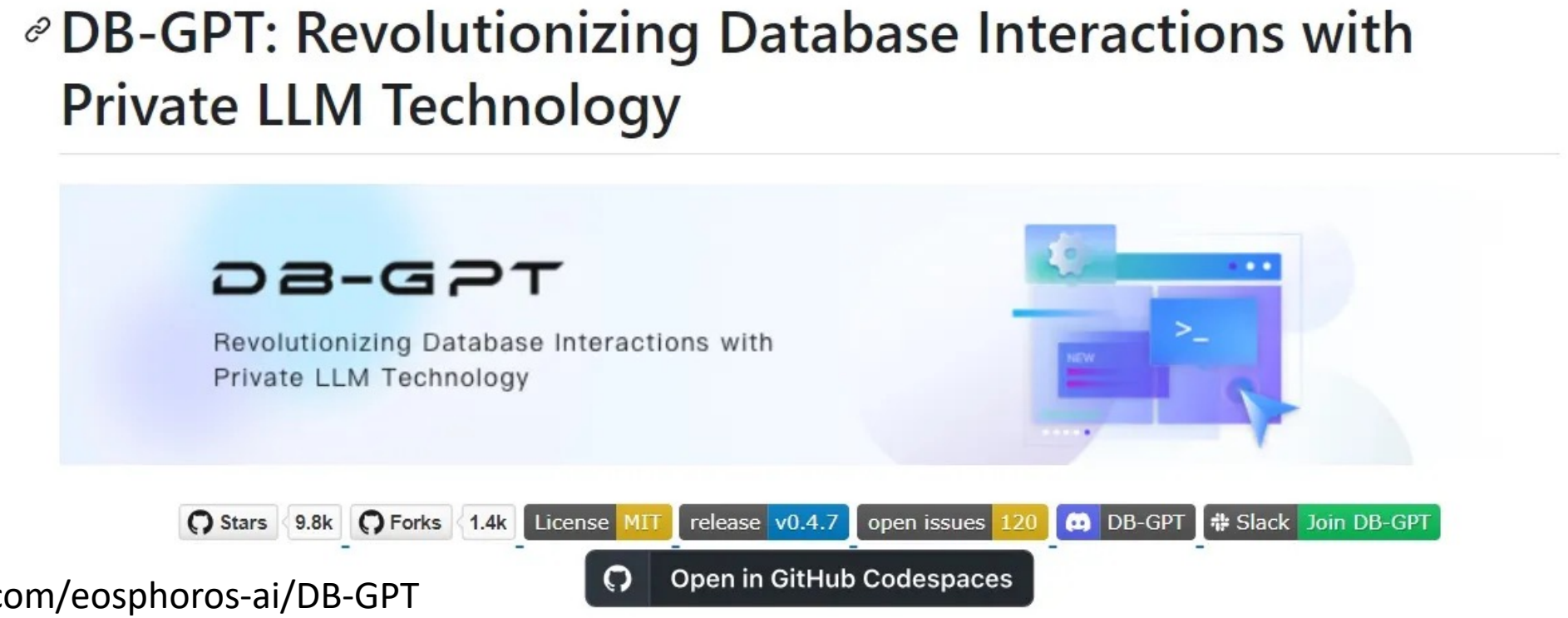
Zhang et al., 2024. "Strength Lies in Differences! Towards Effective Non-collaborative Dialogues via Tailored Strategy Planning" (CoRR '24)

Contents

- 多智能体（Agents）简介
- Agents 会话
- Agents 常用开发框架
- Agents 展望
- 总结

开源智能体产品及框架DB-GPT

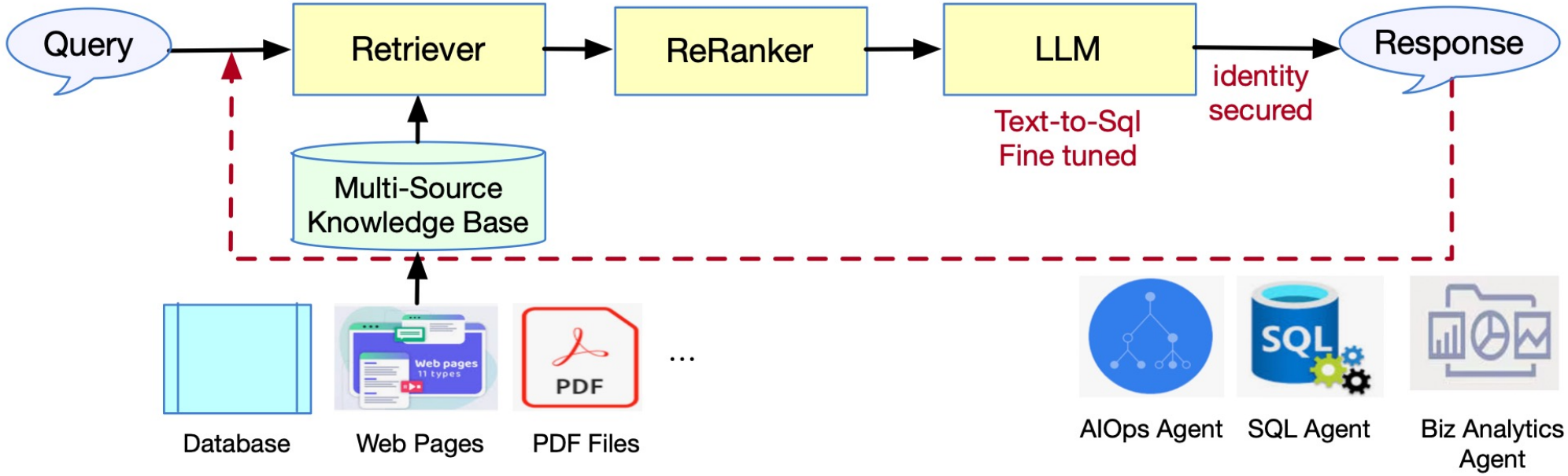
- 融合了多种功能模块，能够轻松地与数据库进行自然语言交互，极大地提升了数据库的易用性和可访问性。
- 具备强大的知识库问答功能，还能够生成复杂的SQL查询语句，通过多智能体协作，为用户提供专业、高效的数据库操作体验。



https://github.com/eosphoros-ai/DB-GPT

DB-GPT

- 核心: RAG（Retrieval Augmented Generation）知识库问答模块。
- 在知识库中，DB-GPT集成了多种数据源，包括数据库文档、网页、PDF等，将其转换为结构化表示，以便于大语言模型进行检索。
- DB-GPT还提供了隐私保护功能，确保数据安全和隐私。



DB-GPT

- 多智能体策略：通过定义不同角色的智能体，如SQL Agent、AIOps Agent等，实现了复杂的多智能体协作。
- 每个智能体专注于特定任务，如生成SQL查询语句或执行数据库操作，从而实现协同推理、信息共享和集体决策。

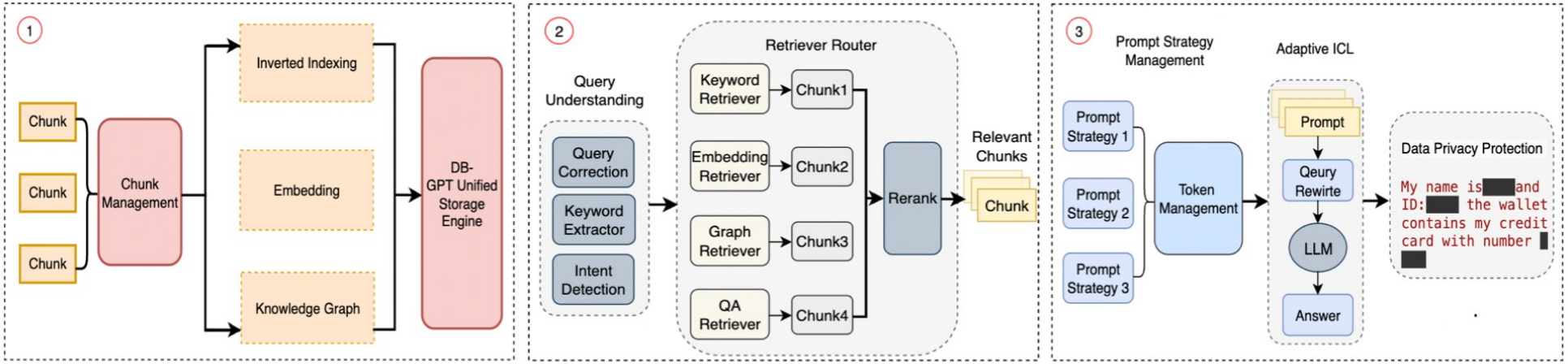


Figure 3: The pipeline of knowledge construction

Figure 4: The pipeline of knowledge retrieval

Figure 5: The pipeline of adaptive ICL and response generation

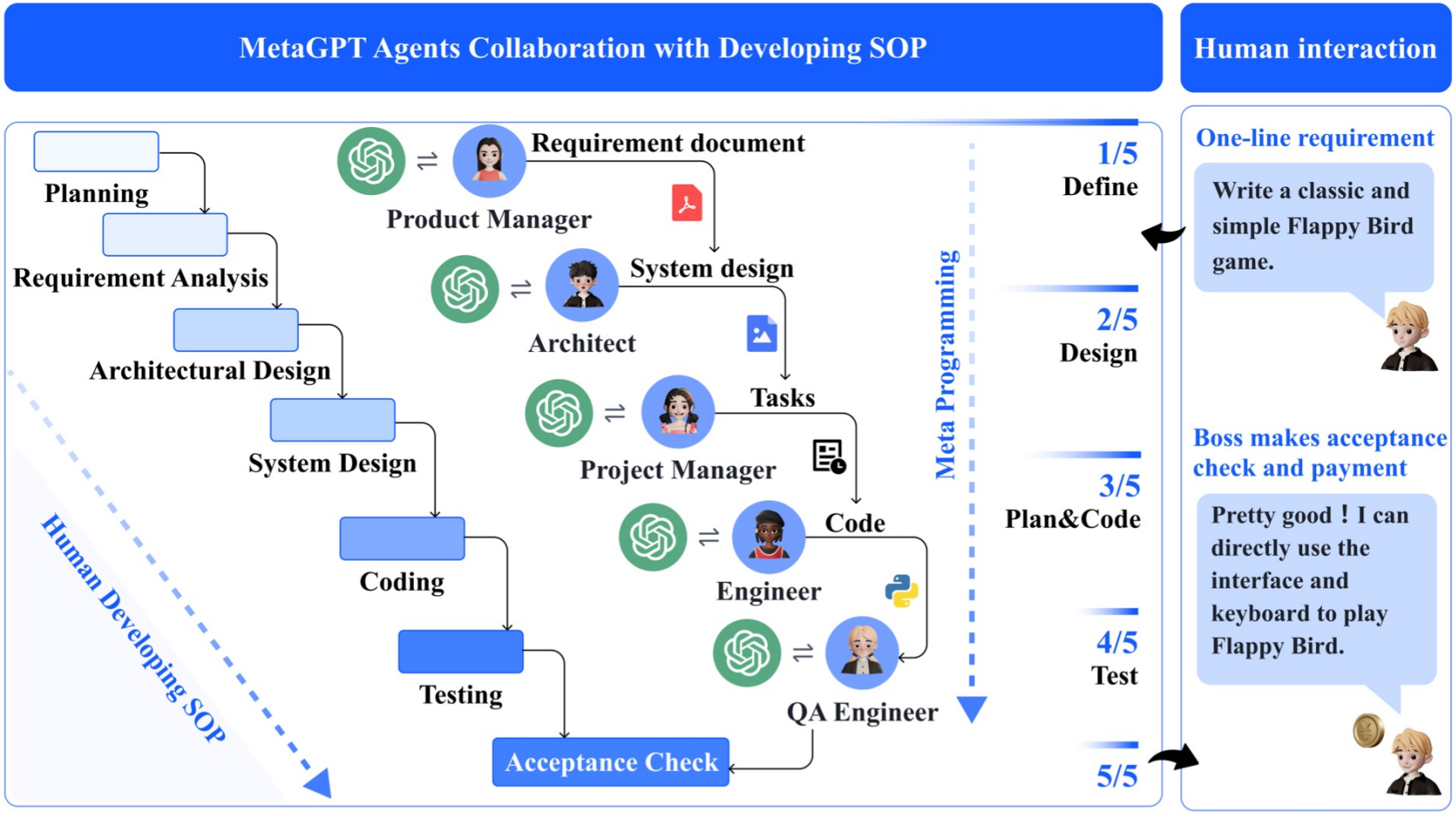
DB-GPT

- 集成了丰富的数据库插件：包括schema分析器、SQL执行器等
- 为用户提供便捷的数据库操作工具。
- 执行各种端到端的数据分析任务，展现强大的生成式数据分析能力。

MetaGPT

- 使 GPTs 组成软件公司，协作处理更复杂的任务

1. MetaGPT输入一句话的老板需求，输出用户故事 / 竞品分析 / 需求 / 数据结构 / APIs / 文件等
2. MetaGPT内部包括产品经理 / 架构师 / 项目经理 / 工程师，它提供了一个软件公司的全过程与精心调配的SOP
    1. Code = SOP(Team) 是核心哲学。我们将SOP具象化，并且用于LLM构成的团队

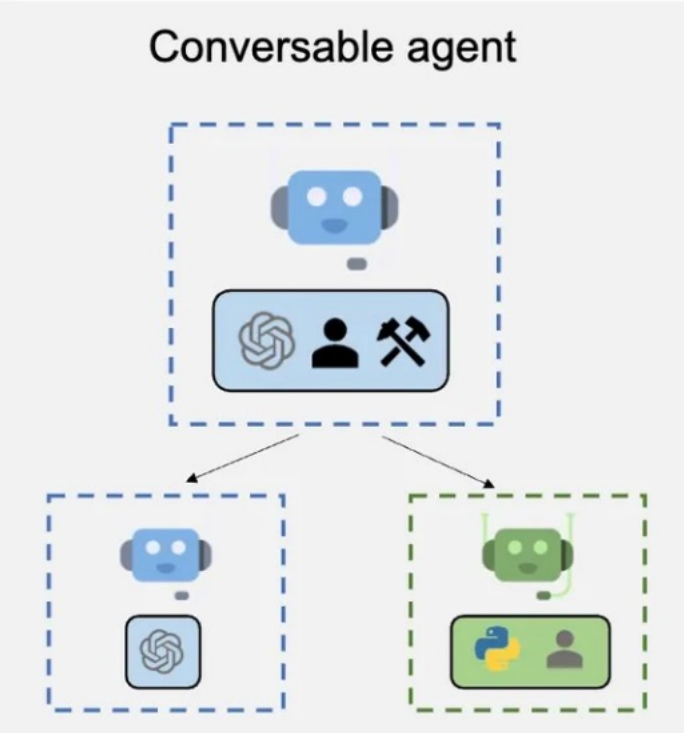


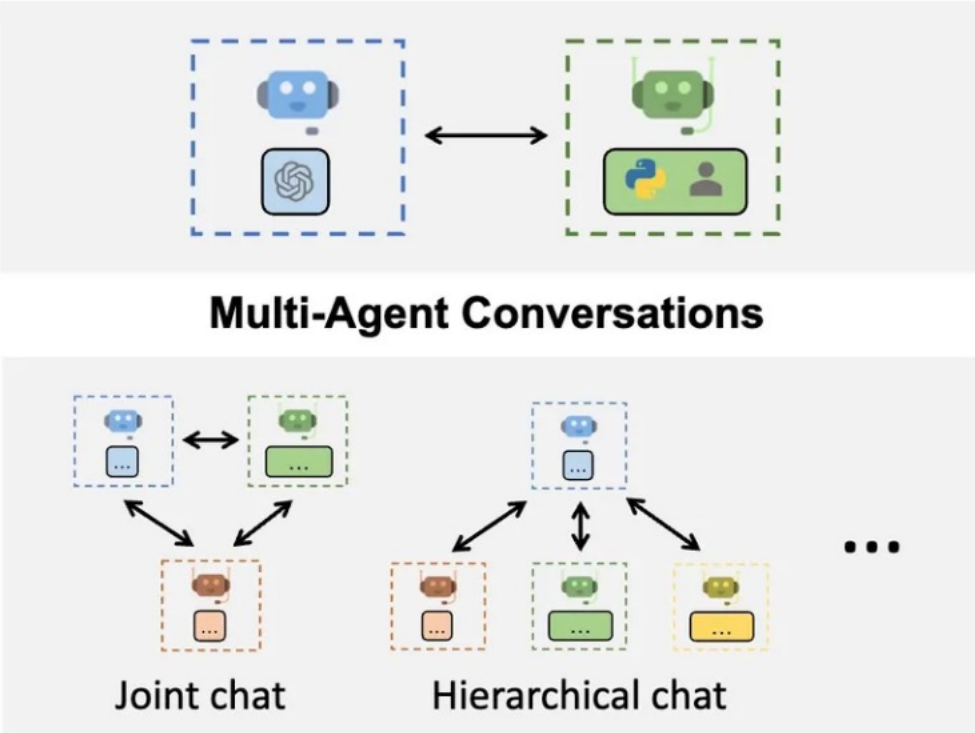
https://github.com/geekan/MetaGPT/

https://arxiv.org/abs/2308.00352

AutoGen（Microsoft）

- 可定制、可对话，并能以各种模式运行
- 这些模式采用 LLM、人类输入和工具的组合





https://github.com/microsoft/autogen

Contents

- 多智能体（Agents）简介
- Agents 会话
- Agents 常用开发框架
- Agents 展望

更多Agents的挑战

- 可信赖和可靠的LLM Agents
  - 可信赖和可靠的LLM Agents可以提升用户体验
  - 促进安全性
  - 确保有道德的互动。

- LLM Agents 评估
  - 如何评估智能体?
  - 如何利用智能体进行评估?

Thank you!