

# The Planning of LLM Agents

2024/10/22

# LLM Agents - *ChatDev*



ChatDev: Communicative  
Agents for Software  
Development

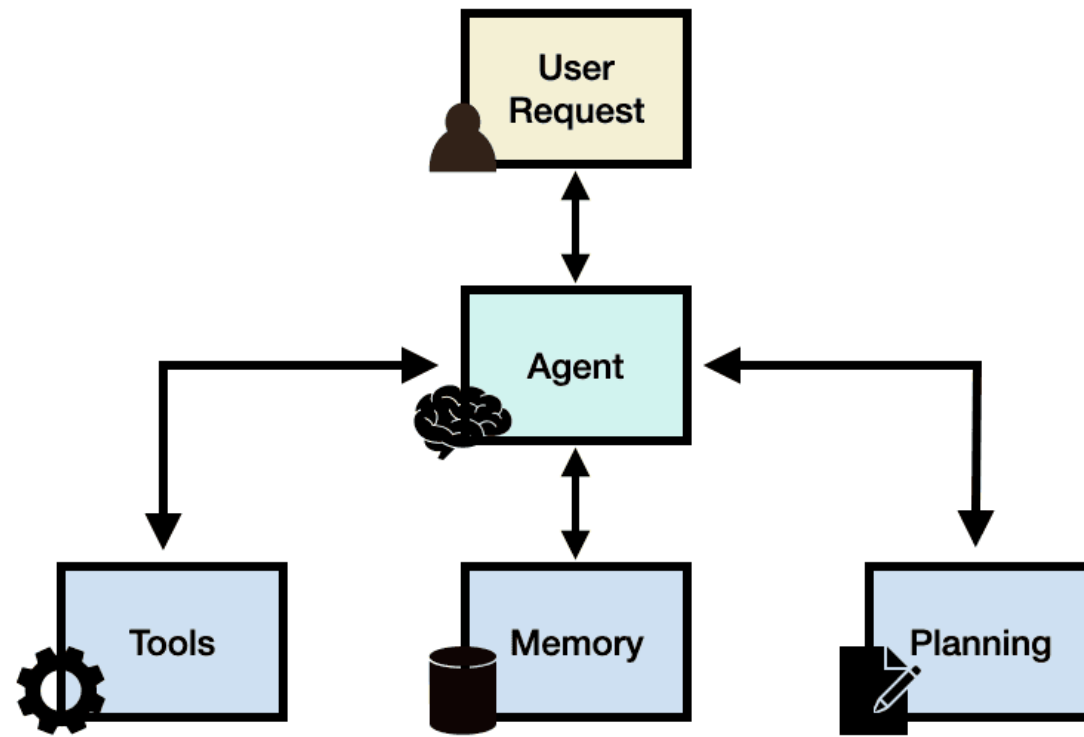
# LLM Agents - *ChatDev*



How to arrange  
the plan for  
using multiple  
agents?

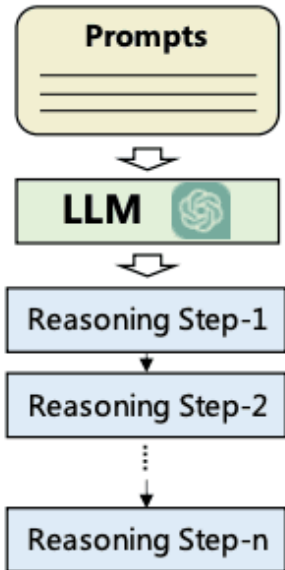
ChatDev: Communicative  
Agents for Software  
Development

# LLM Agent Framework



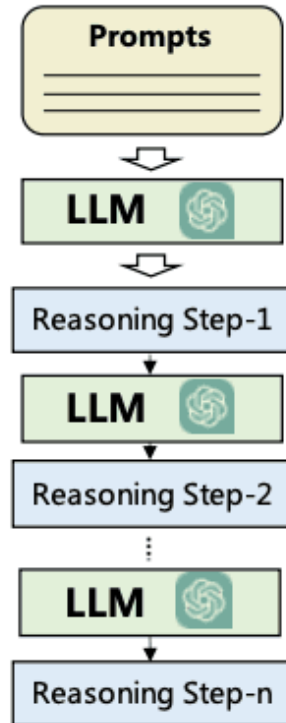
# Planning: Chain-of-Thought

CoT , Zero-shot Cot

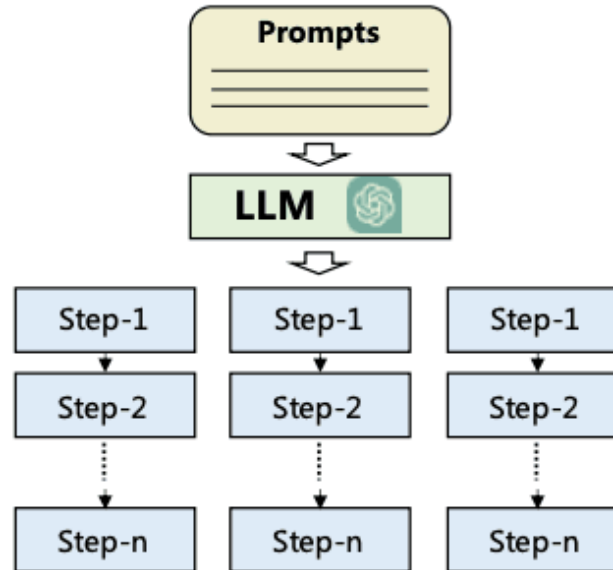


**Single-Path Reasoning**

ReWOO , HuggingGPT

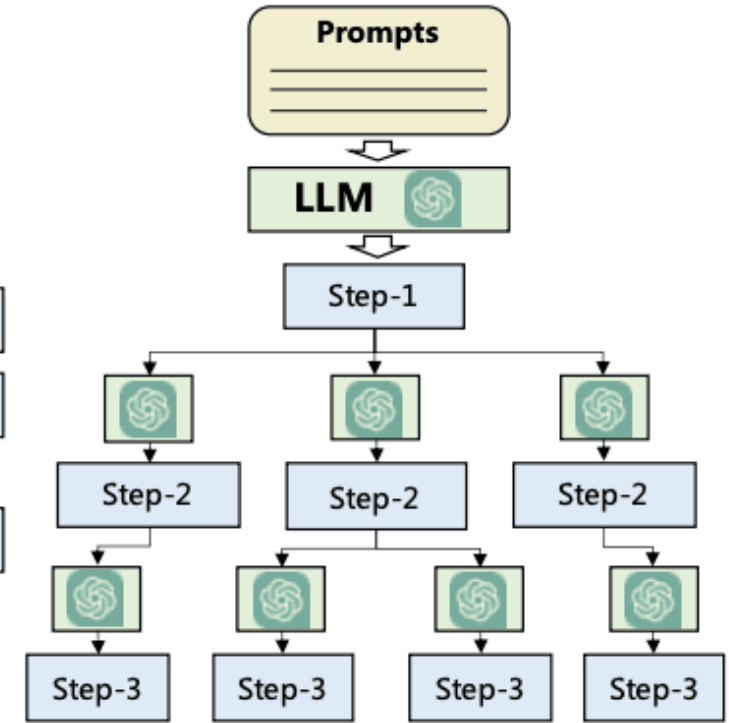


CoT-SC



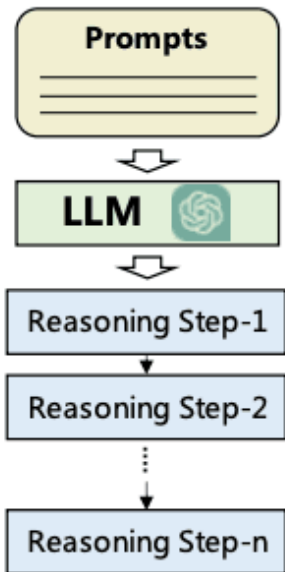
**Multi-Path Reasoning**

ToT , LMZSP , RAP



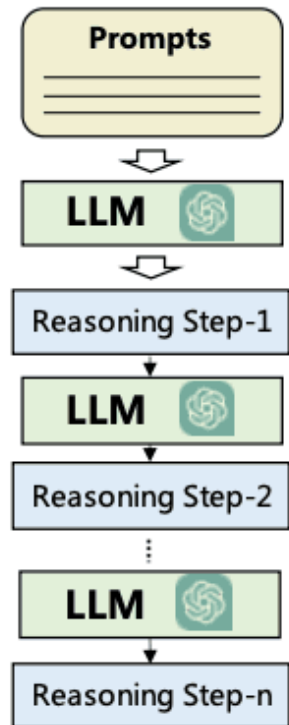
# Planning: Chain-of-Thought

CoT , Zero-shot Cot

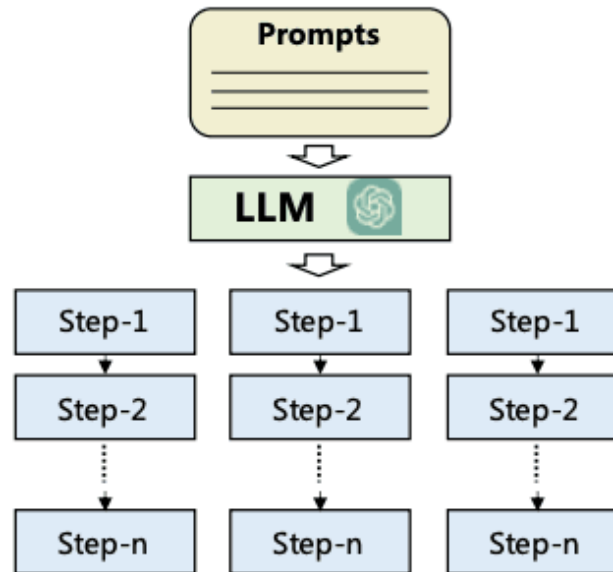


Single-Path Reasoning

ReWOO , HuggingGPT

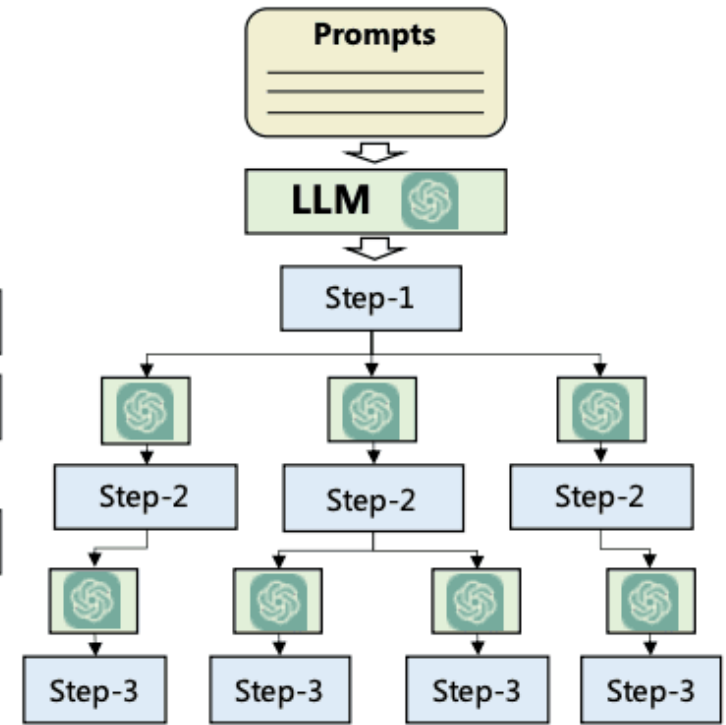


CoT-SC



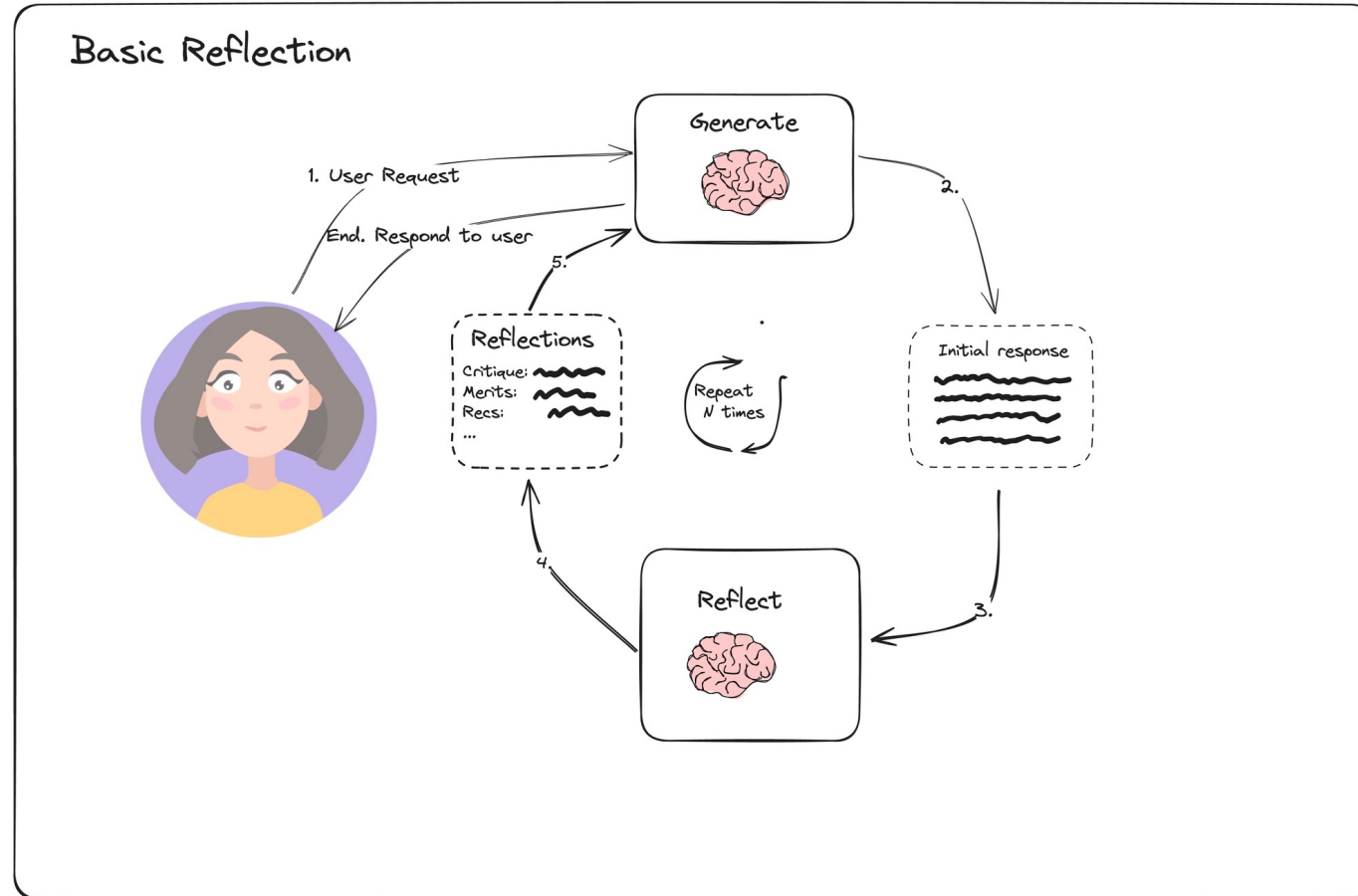
Multi-Path Reasoning

ToT , LMZSP , RAP

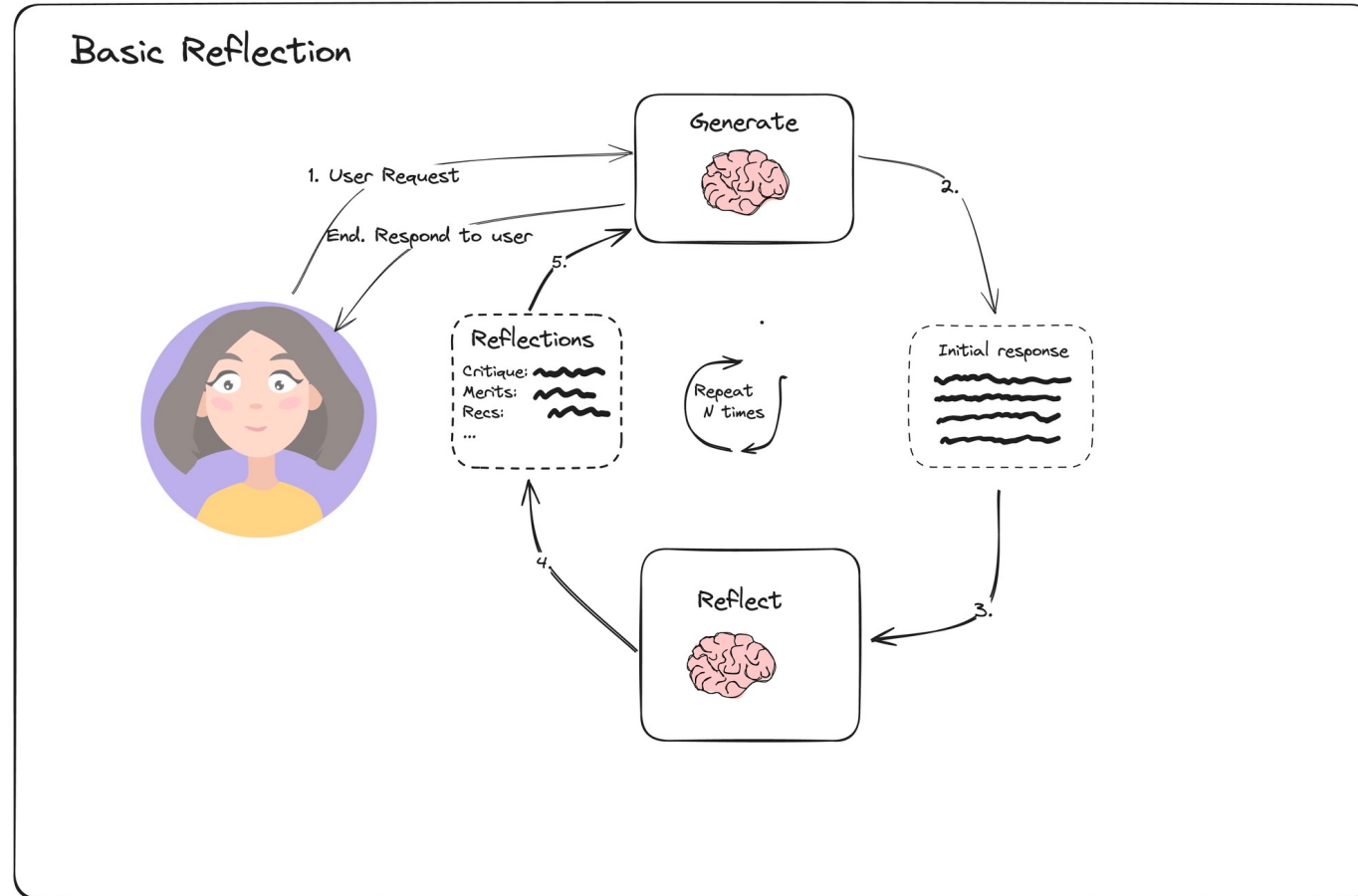


Without Feedback

# Planning with Feedback



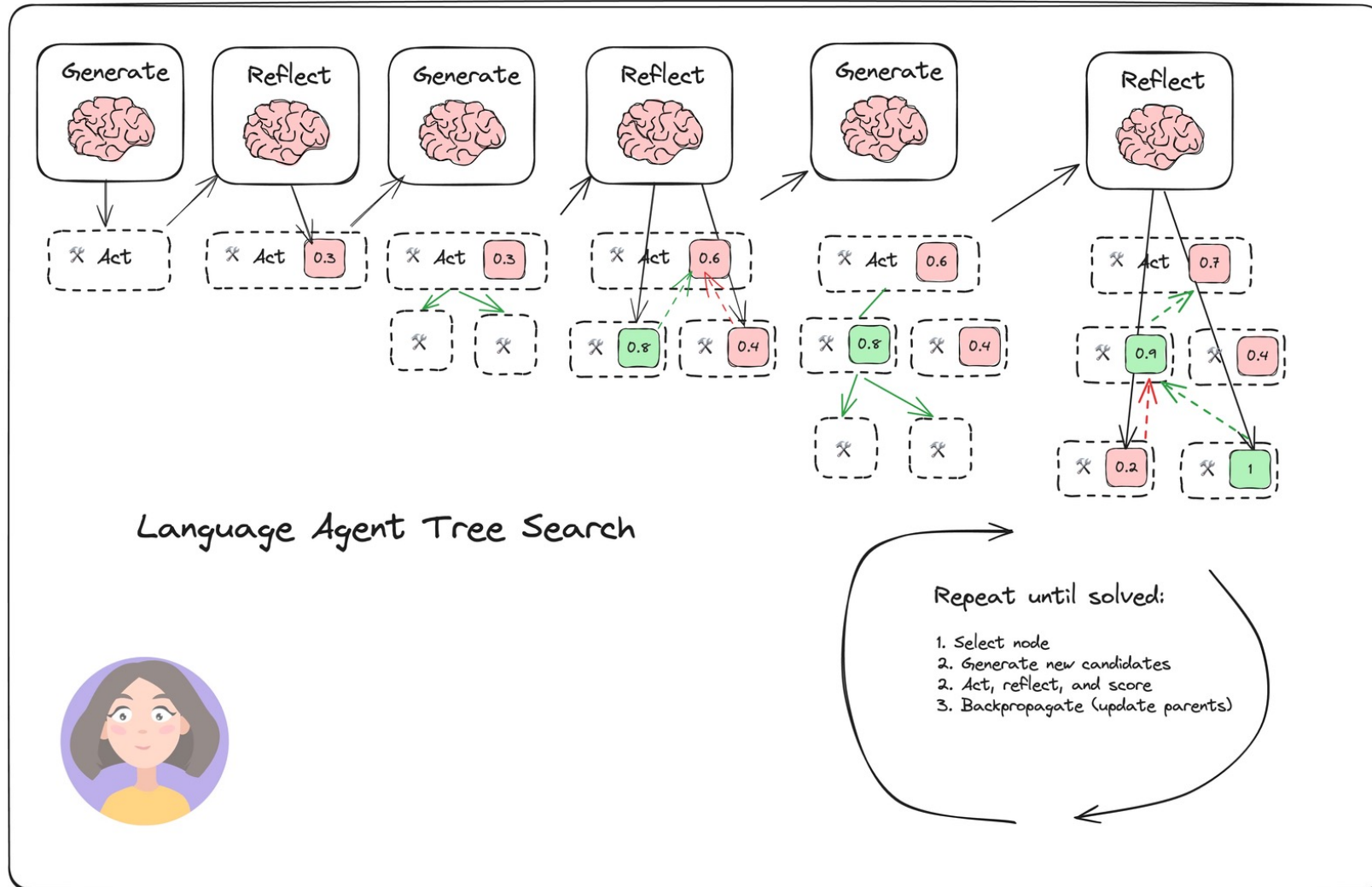
# Planning with Feedback



Reflection



# Planning with Tree Search

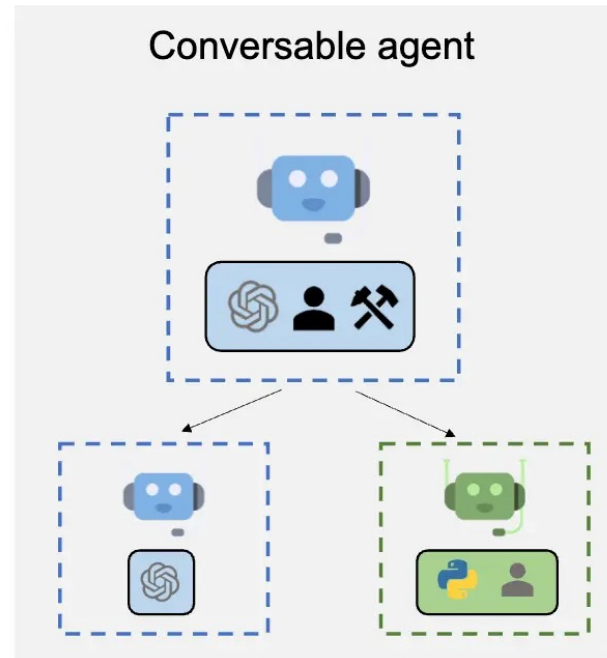


# Tools & Memory

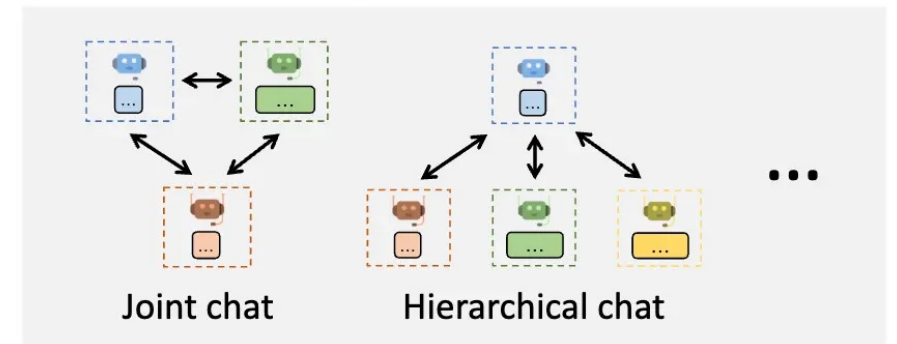
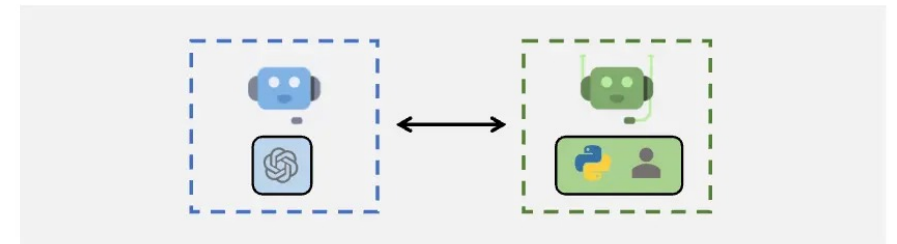
- The memory module helps to **store the agent's internal logs including past thoughts, actions, and observations** from the environment, including **all interactions** between agent and user.
- Tools correspond to a set of tool/s that enables the LLM agent to interact with external environments **such as Wikipedia Search API, Code Interpreter, and Math Engine.**

# Moreover....

- Multi-Agent
- Multi-Modality
- ...



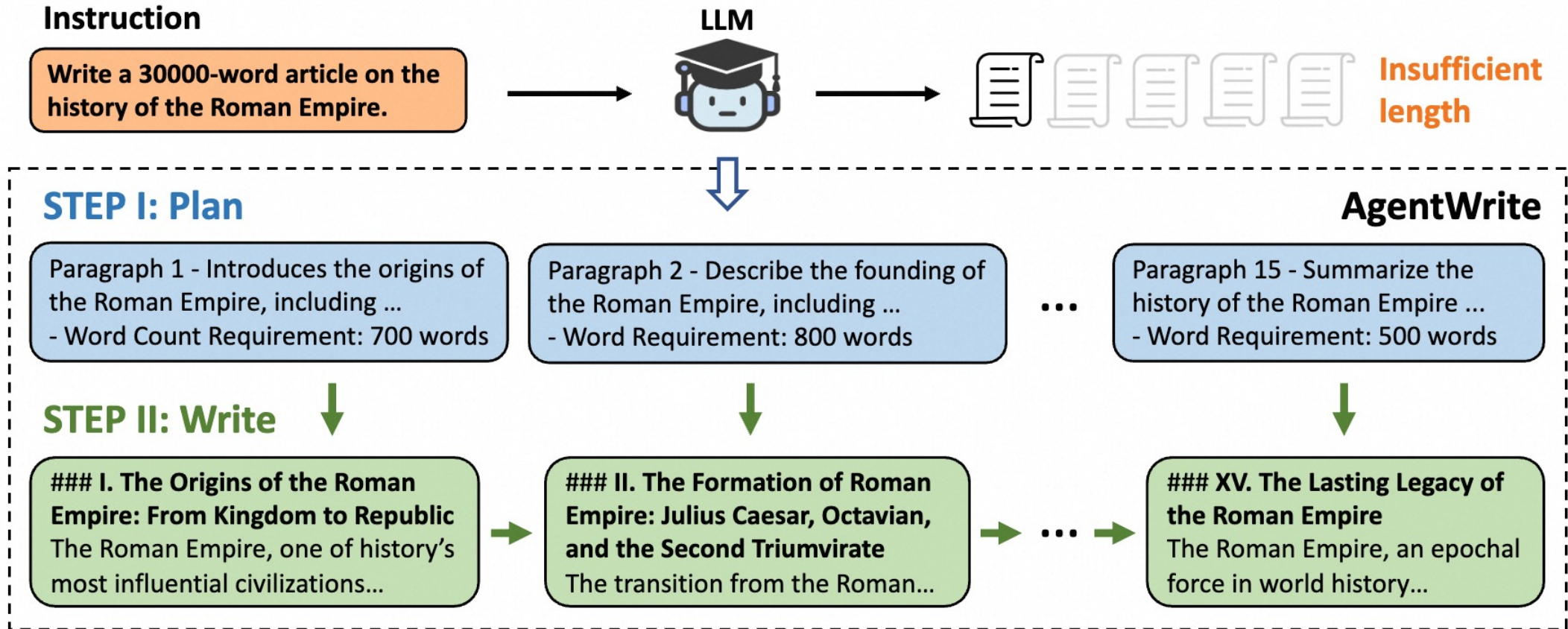
**Agent Customization**



**Flexible Conversation Patterns**

AutoGen

# A Case






LONGWRITER: UNLEASHING 10,000+ WORD  
GENERATION FROM LONG CONTEXT LLMS

# Assignment

- Explore the potential of **LLM to generate long and high-quality texts**, by designing your own planning method.

# Assignment

Assessment:

- **Text Length** : No less than 10,000 words. (Main)
- **Text Quality** : The quality of the text will be evaluated by GPT-4. Criteria for evaluation will include coherence, relevance, fluency, readability. (Main)
- **Number of API Calls** : Decrease the number of API calls made during the whole process.
- **Technical Novelty**: Get extra scores, if your method is attractive.

# Assignment

- **Text Quality Prompt:**

- Please help me to evaluate the quality of the text. The criteria for evaluation include coherence, relevance, fluency, and readability.
- The required Topic is: [TOPIC].
- The input long text is: [TEXT].
- The output is an score between 0 and 10, with higher scores indicating better quality.
- The structure of output:

[Coherence: Score, Relevance: Score, Fluency: Score, Readability: Score].

# Assignment

- Select **ONE** of these topics:
  - Stepping on the Rainy Street
  - A Futuristic Cafe
  - Application of Nanotechnology in Medicine
- Codes:
  - **ONLY** use the free API (glm-4-flash), available at <https://open.bigmodel.cn/dev/howuse/introduction>
  - Assignment3\_Example.py



# Submission

- **Generated Text:** .txt file
- **Code:** including all codes for model planning, API usage, file restoring, or pre/post-processing. Assignment.py
- **Report:** Write a NOT too-long report (1-2 Pages) about the novelty of your approach, the challenges faced, and how you addressed them, also with the statistic information of the generated text. .pdf file
- Zip all of them to submit to Canvas: YourName\_StudentNumber.zip
- For example, JiaqiTang\_50005050.zip
- Guidance: <https://www.overleaf.com/read/fsgszmrvmspx#8ad70f>

# To get help?

- Papers: about LLM planning...
- Public Code: GitHub...
- Contact me: [jtang092@connect.hkust-gz.edu.cn](mailto:jtang092@connect.hkust-gz.edu.cn)