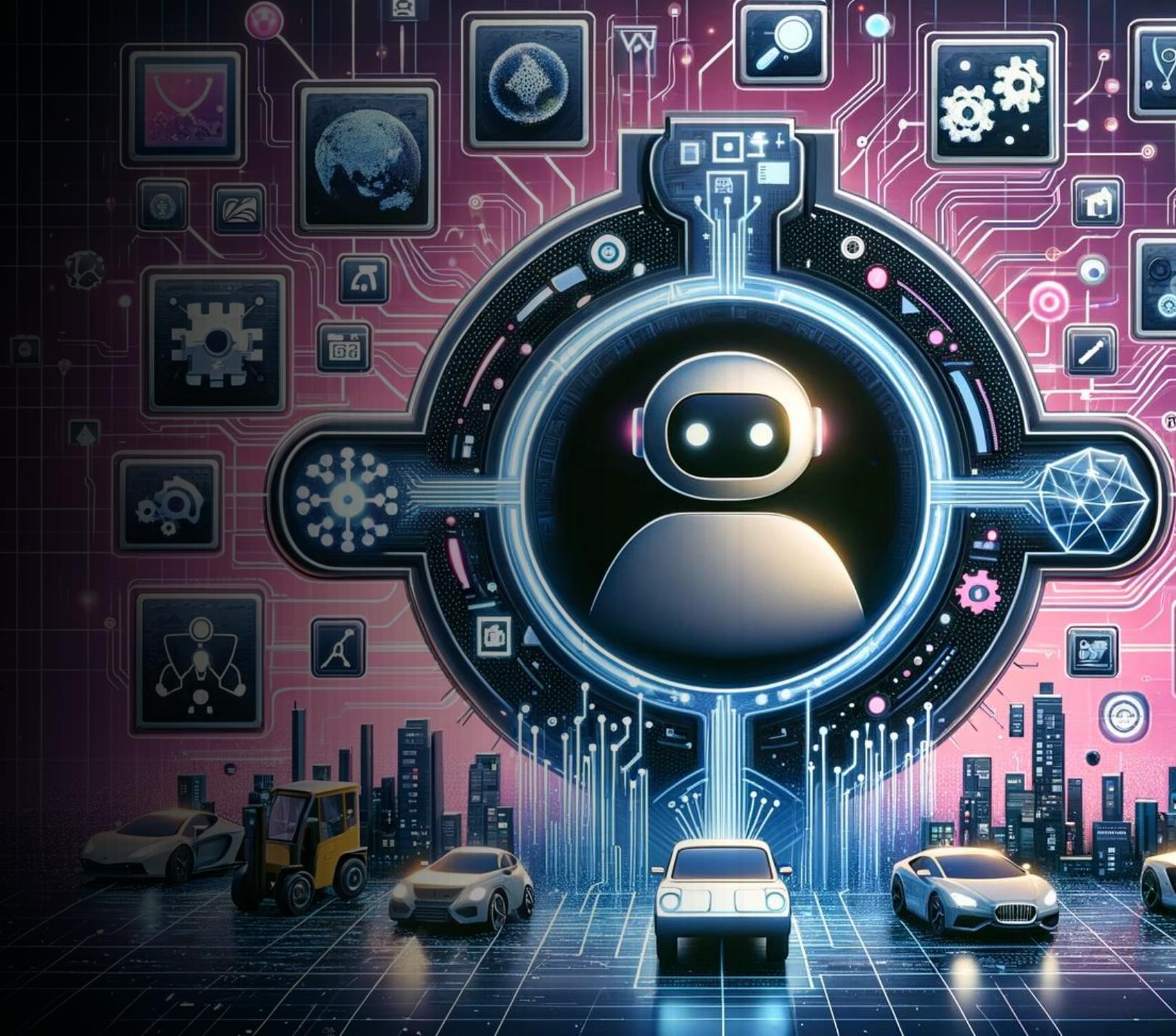
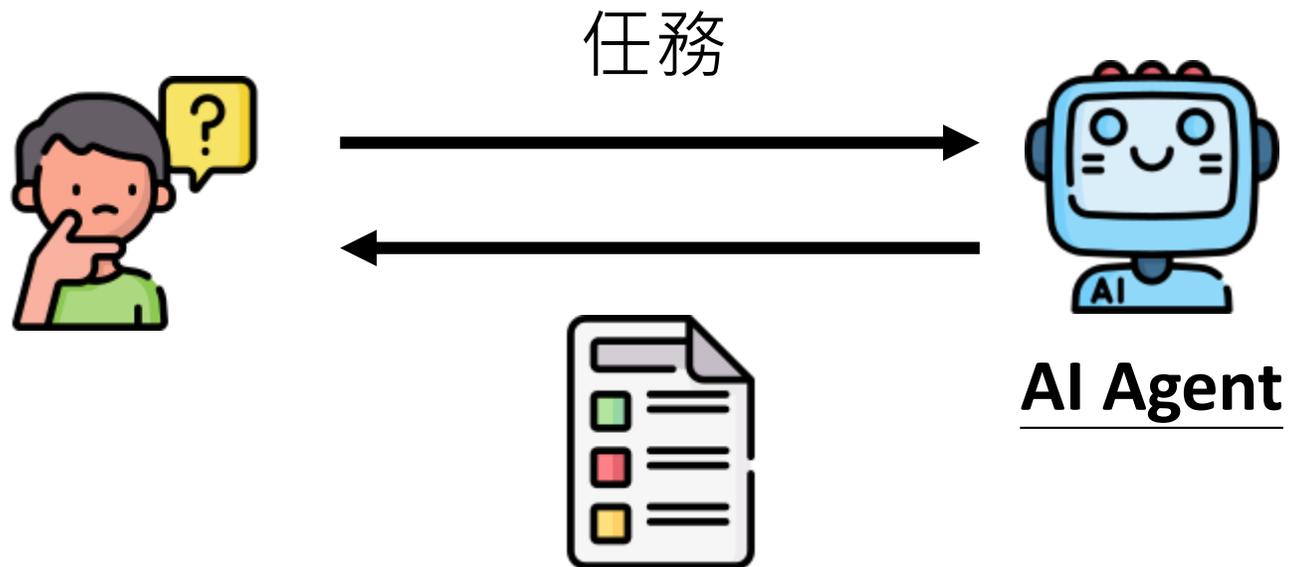


# 以大型語言 模型打造的 AI Agent

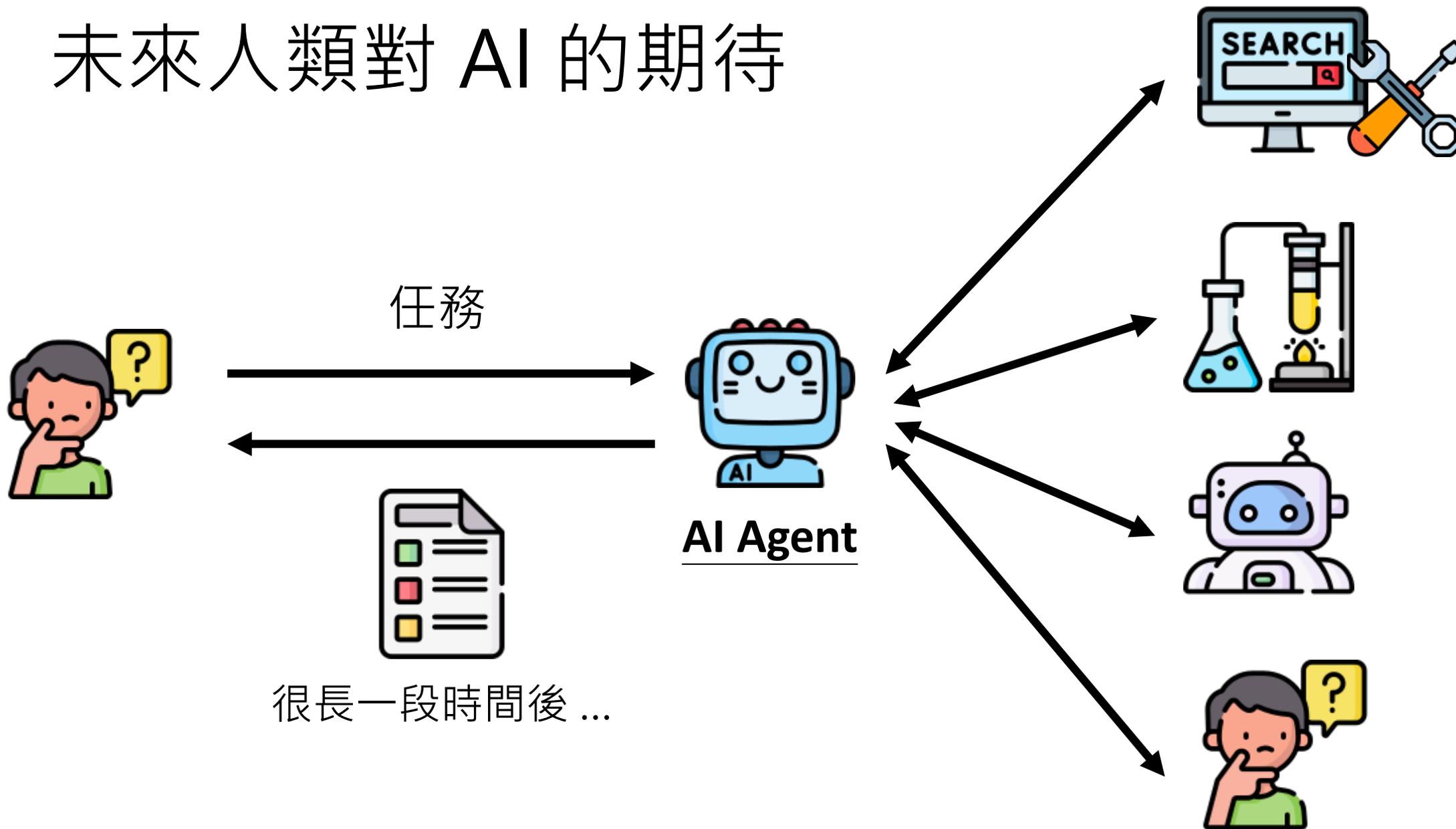
李宏毅



# 今日多數人使用 AI 的方式



# 未來人類對 AI 的期待



# AI Agent

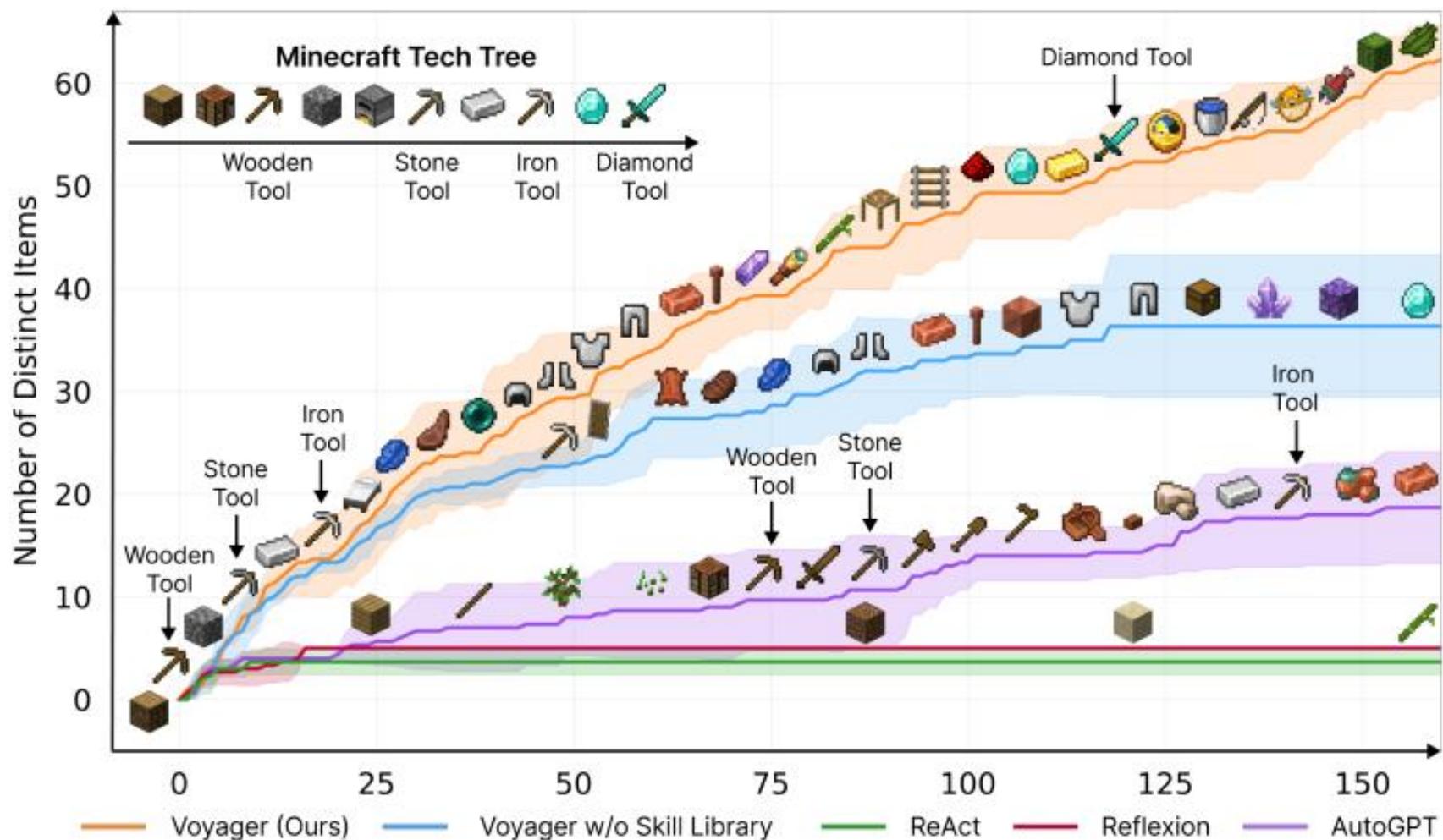
- AutoGPT: <https://github.com/Significant-Gravitas/Auto-GPT>
- AgentGPT: <https://agentgpt.reworkd.ai/>
- BabyAGI: <https://github.com/yoheinakajima/babyagi>
- Godmode: <https://godmode.space/?ref=futuretools.io>

# 由 AI 村民組成的虛擬村莊



# 會自己玩 Minecraft 的 AI

Voyager: <https://arxiv.org/abs/2305.16291>



# 由語言模型操控的機器人

Figure 01

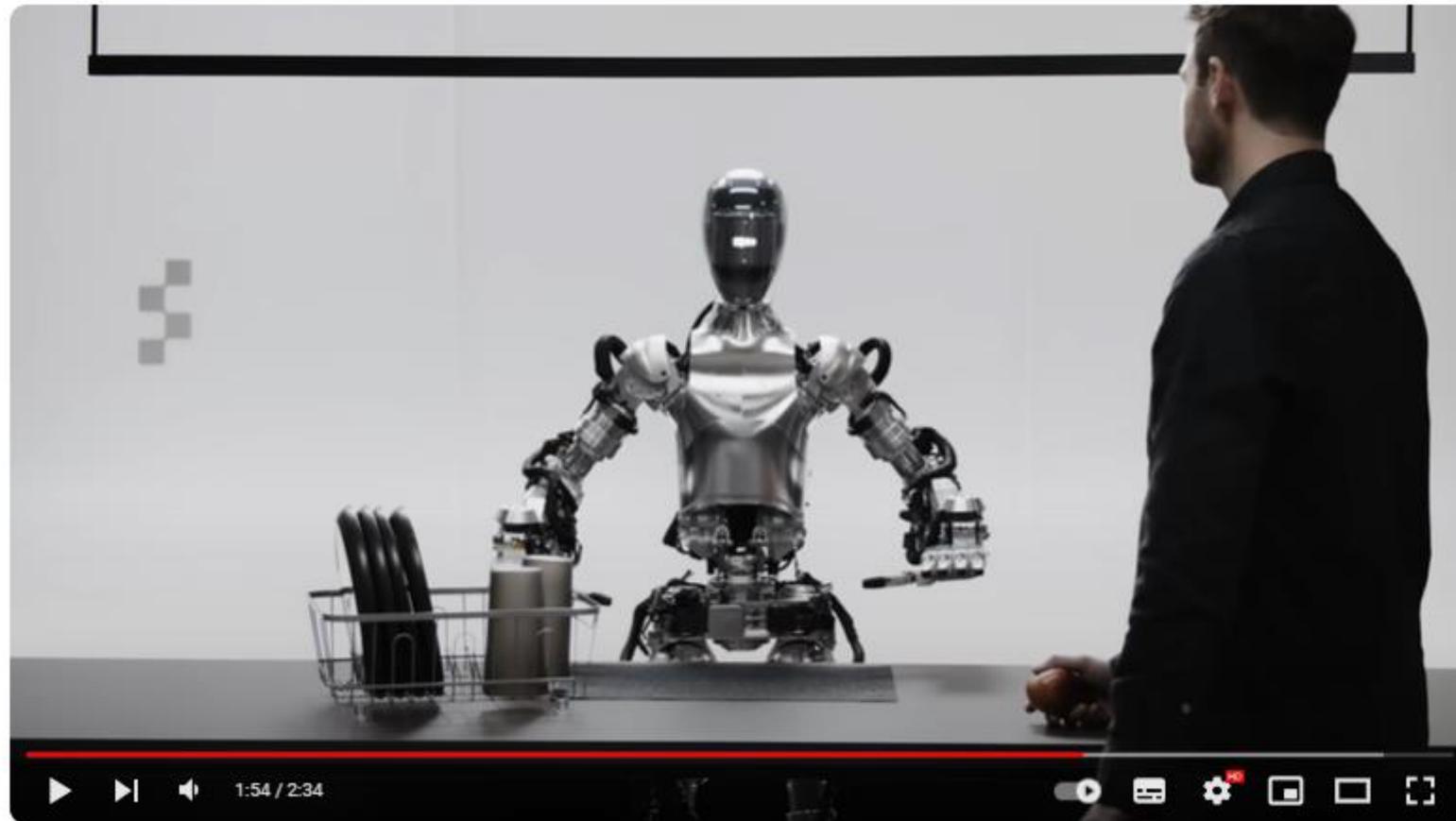


Figure Status Update - OpenAI Speech-to-Speech Reasoning

<https://youtu.be/Sq1QZB5baNw?si=-TEsPeUqTvhIS60E>

# 由語言模型操控的機器人

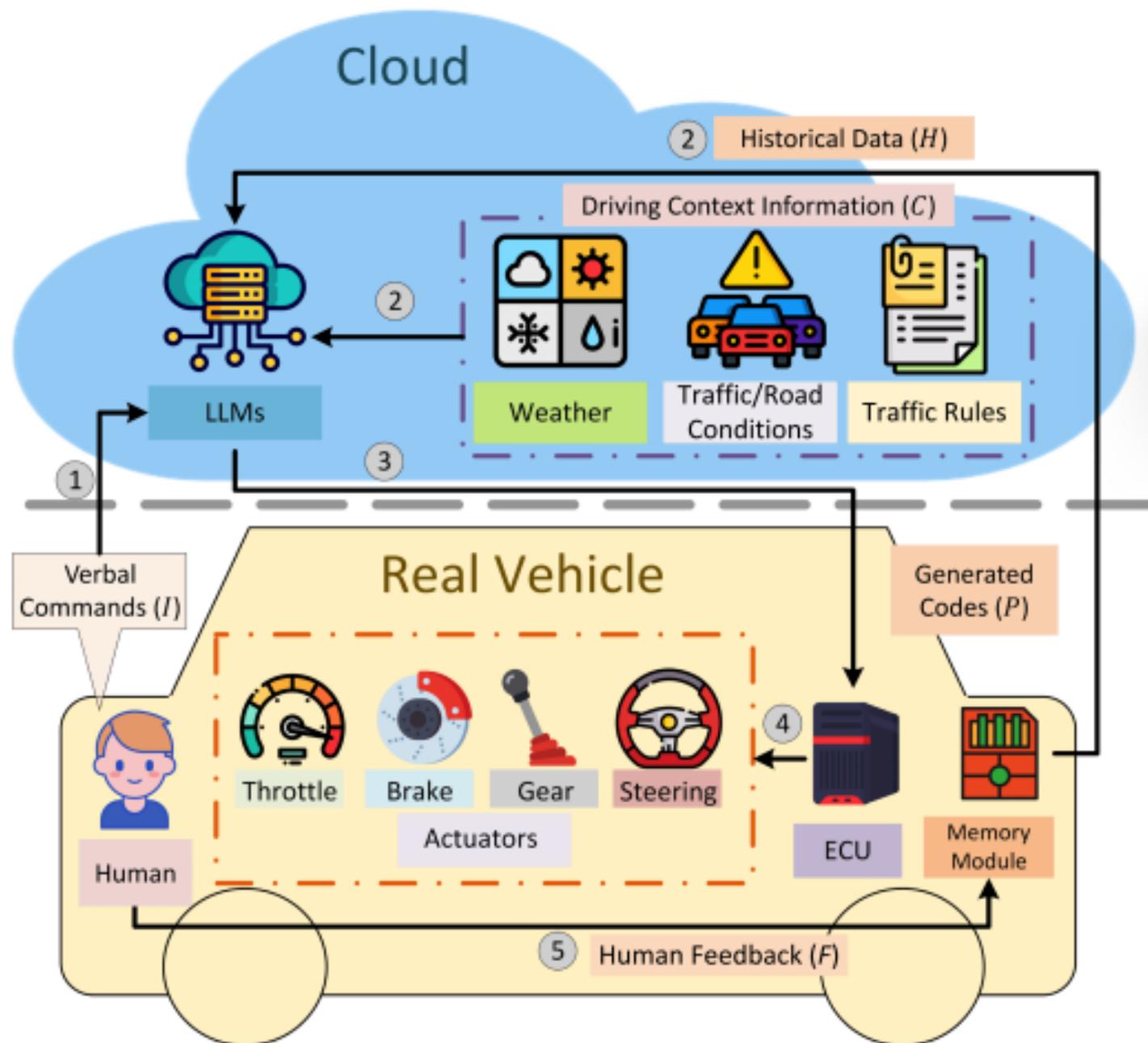
<https://innermonologue.github.io/>

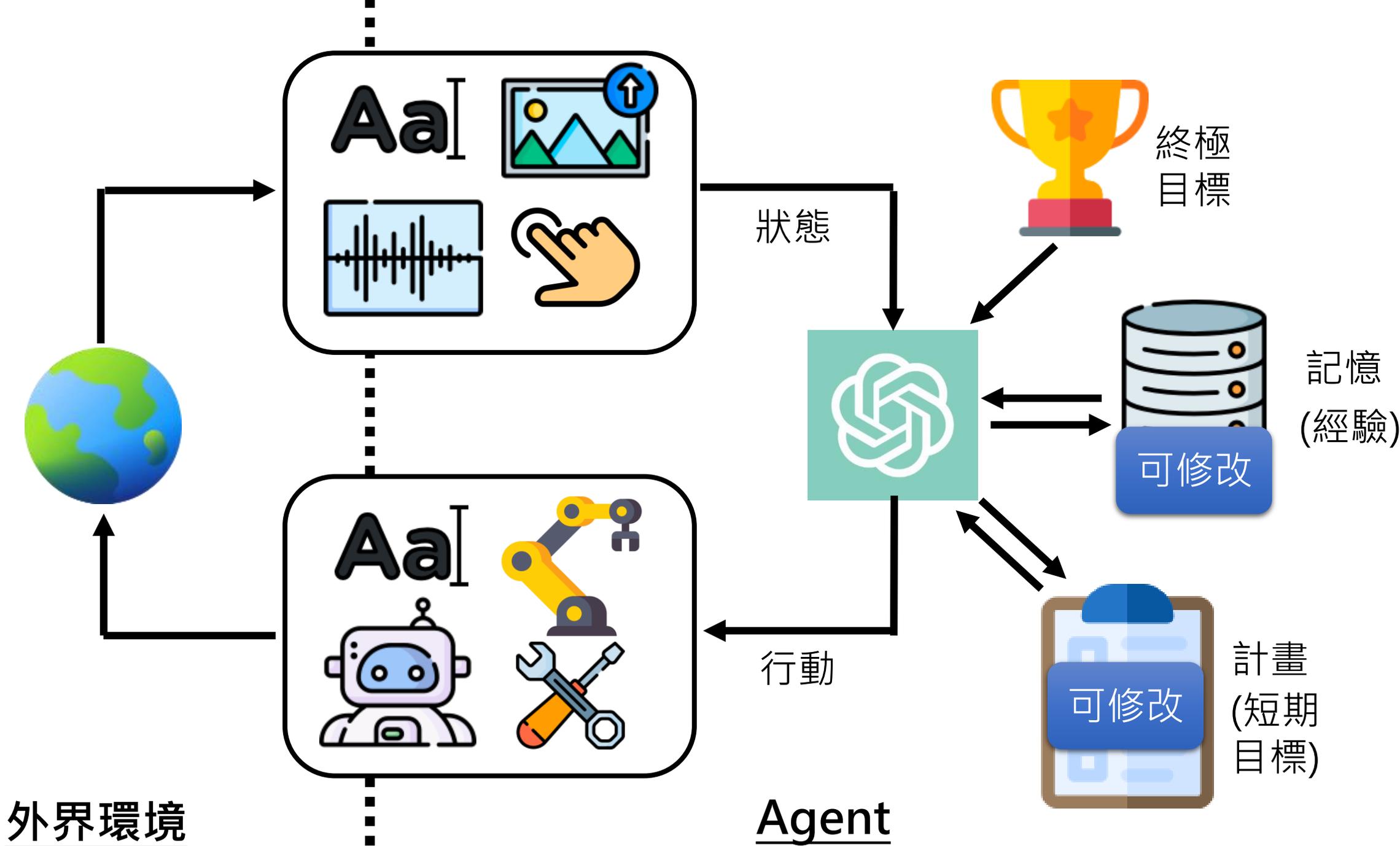
<https://arxiv.org/abs/2207.05608>

# 用大型語言模型 開自駕車

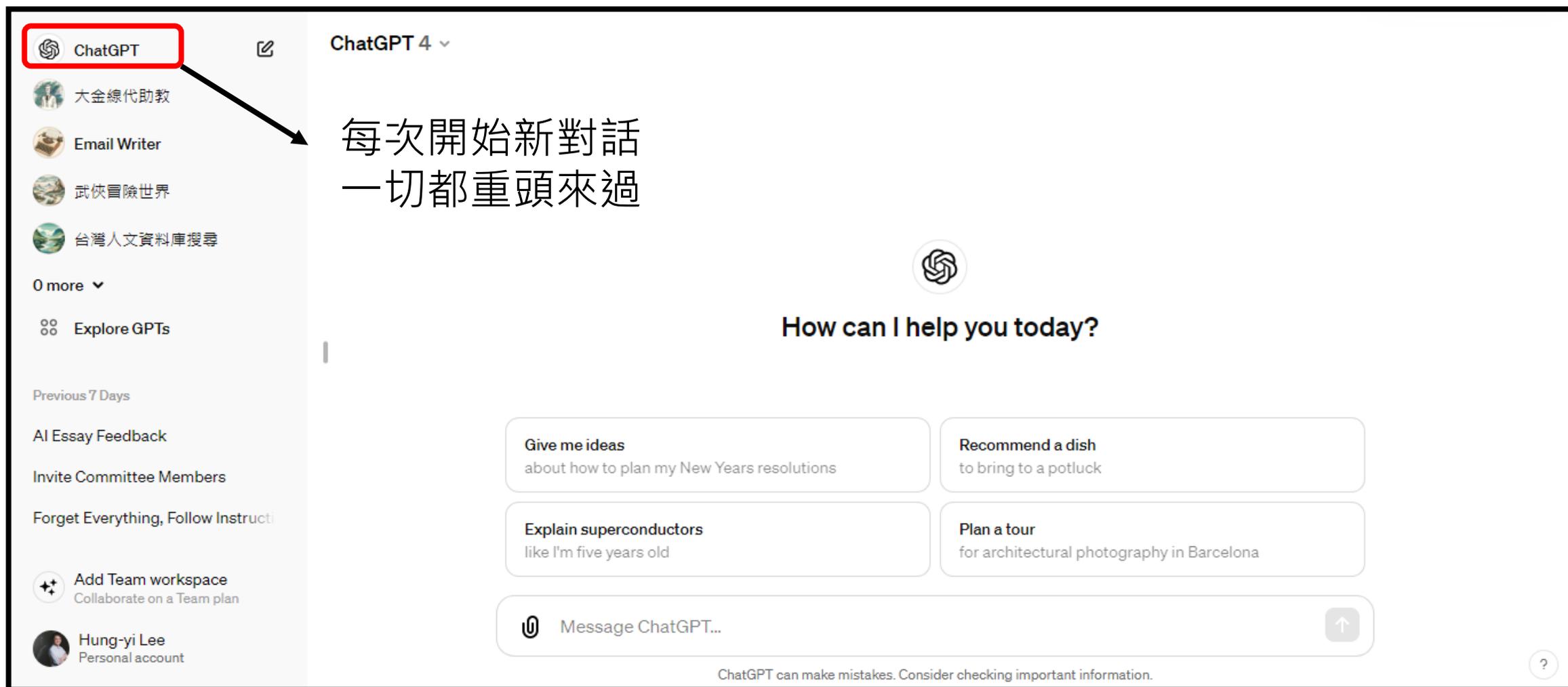
Talk2Drive

<https://arxiv.org/abs/2312.09397>





# 有記憶的 ChatGPT



The screenshot displays the ChatGPT interface. On the left sidebar, the 'ChatGPT' option is highlighted with a red box and an arrow pointing to the main chat area. The main chat area shows a new conversation prompt: '每次開始新對話 一切都重頭來過' (Every time I start a new conversation, everything starts from scratch). Below the prompt, the ChatGPT logo is visible, followed by the question 'How can I help you today?'. There are four suggested prompts: 'Give me ideas about how to plan my New Years resolutions', 'Recommend a dish to bring to a potluck', 'Explain superconductors like I'm five years old', and 'Plan a tour for architectural photography in Barcelona'. At the bottom, there is a text input field 'Message ChatGPT...' and a send button. A disclaimer at the bottom reads 'ChatGPT can make mistakes. Consider checking important information.'

ChatGPT

大金線代助教

Email Writer

武俠冒險世界

台灣人文資料庫搜尋

0 more

Explore GPTs

Previous 7 Days

AI Essay Feedback

Invite Committee Members

Forget Everything, Follow Instructions

Add Team workspace  
Collaborate on a Team plan

Hung-yi Lee  
Personal account

ChatGPT 4

每次開始新對話  
一切都重頭來過

How can I help you today?

Give me ideas  
about how to plan my New Years resolutions

Recommend a dish  
to bring to a potluck

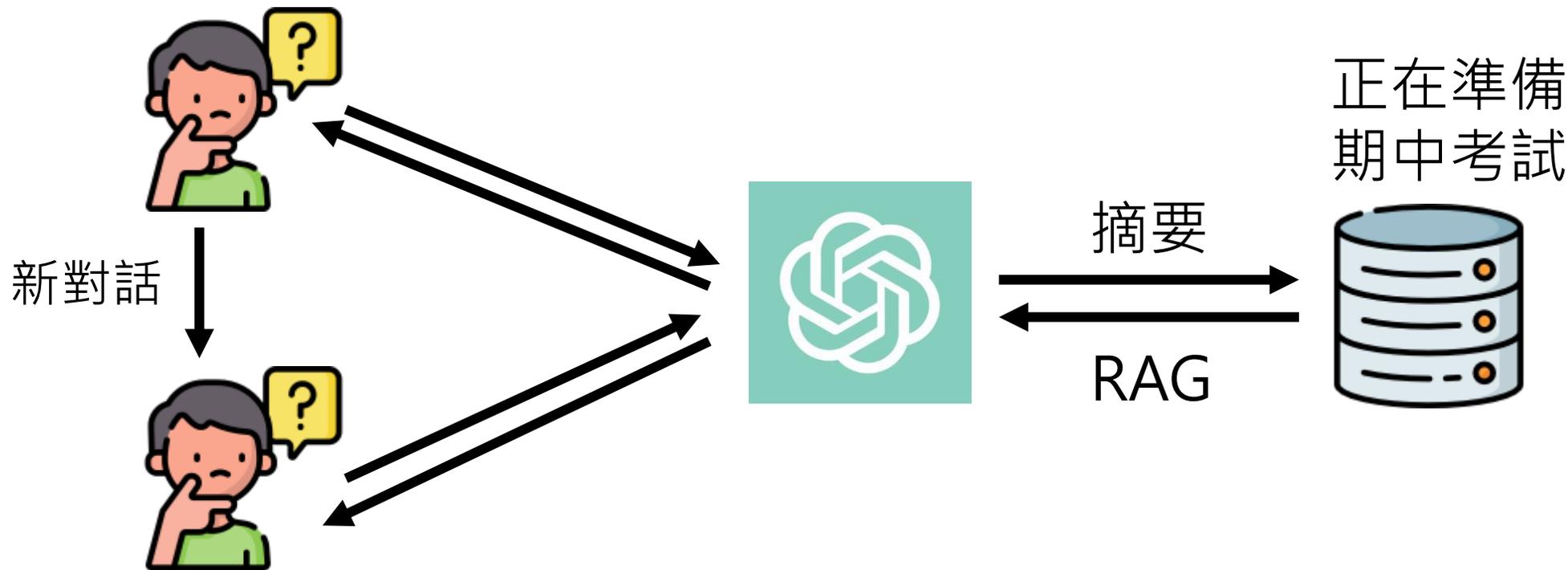
Explain superconductors  
like I'm five years old

Plan a tour  
for architectural photography in Barcelona

Message ChatGPT...

ChatGPT can make mistakes. Consider checking important information.

# 有記憶的 ChatGPT



Open AI is working on it?

<https://openai.com/blog/memory-and-new-controls-for-chatgpt>



<https://github.com/cpacker/MemGPT>

又要講芙莉蓮的故事了 .....



由大型語言模型  
操控的 AI Agent



你會自律行動，安全的把考生帶出迷宮



終極目標

狀態

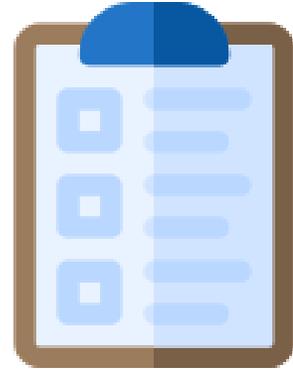
眼前是遭遇複製體攻擊而重傷的愛黛兒

(Image captioning)



記憶 (經驗)

(空)

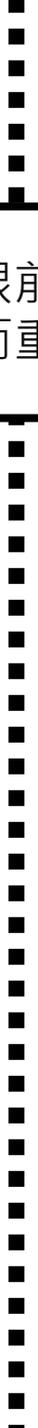


計畫



外界環境

Agent





**You**

目標：你叫哥雷姆，你會自律行動，安全的把考生帶出迷宮

記憶：

狀態：眼前是遭遇複製體攻擊而重傷的愛黛兒

請產生行動計劃

< 7 / 18 >



**ChatGPT**

首先，作為哥雷姆，我的首要任務是確保考生的安全和健康。考慮到愛黛兒遭遇複製體攻擊並受到重傷，我的行動計劃如下：

1. 立即評估愛黛兒的傷勢：我會快速但仔細地檢查愛黛兒的傷勢，確定需要立即處理的問題（如出血、斷骨等）。
2. 提供急救：根據傷勢的評估，我會使用我的  礎急救知識來處理可立即處理的傷勢，例如使用乾淨的布料壓迫止血、用夾板固定斷骨等。如果有攜帶急救包，將充分利用其中的物資。

你會自律行動，安全的把考生帶出迷宮

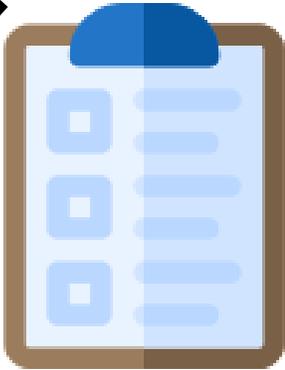


終極目標



記憶  
(經驗)

(空)



計畫



狀態

行動

眼前是遭遇複製體攻擊  
而重傷的愛黛兒

(Image captioning)

?????



外界環境

Agent



You

目標：你叫哥雷姆，你會自律行動，安全的把考生帶出迷宮

記憶：

狀態：眼前是遭遇複製體攻擊而重傷的愛黛兒

行動計劃：首先，作為哥雷姆，我的首要任務是確保考生的安全和健康。考慮到愛黛兒遭遇複製體攻擊並受到重傷，我的行動計劃如下：

1. **\*\*立即評估愛黛兒的傷勢\*\***：我會快速但仔細地檢查愛黛兒的傷勢，確定需要立即處理的問題（如出血、斷骨等）。

2. **\*\*提供急救\*\***：根據傷勢的評估，我會使用我的基礎急救知識來處理可立即處理的傷勢，例如使用乾淨的布料壓迫止血。

3. **\*\*尋找安全地帶\*\***：我會尋找一個安全的地帶，遠離複製體的攻擊。這可能意味著我需要先處理愛黛兒的傷勢，然後再尋找安全地帶。



請採取下一步行動 (請執行一個動作就好)

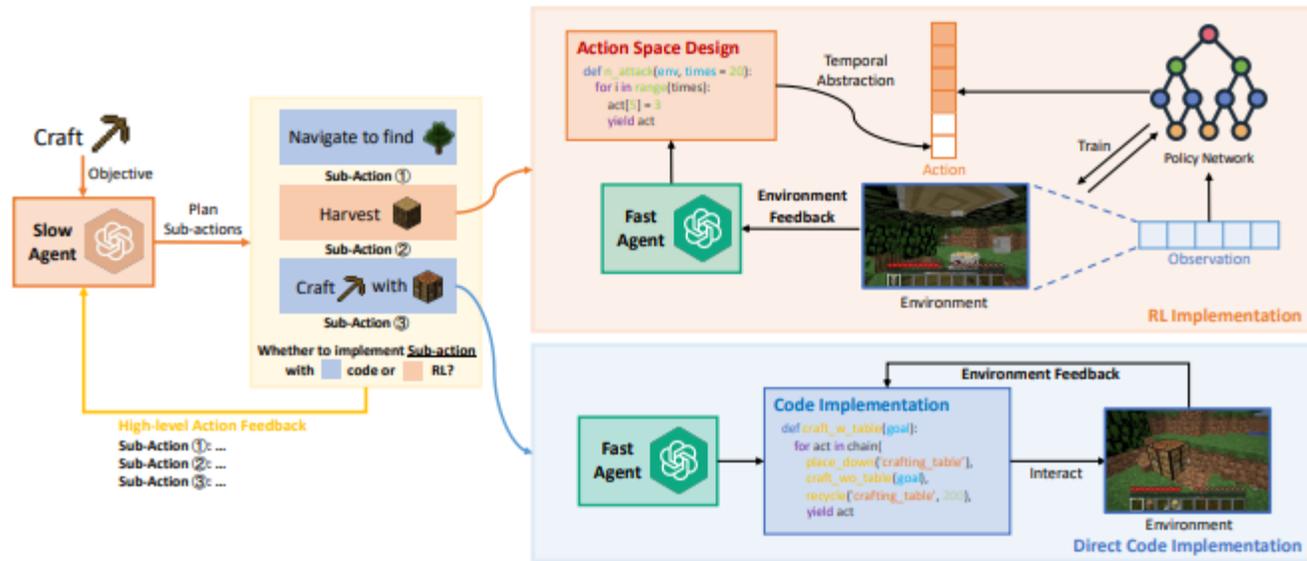
< 9 / 18 >



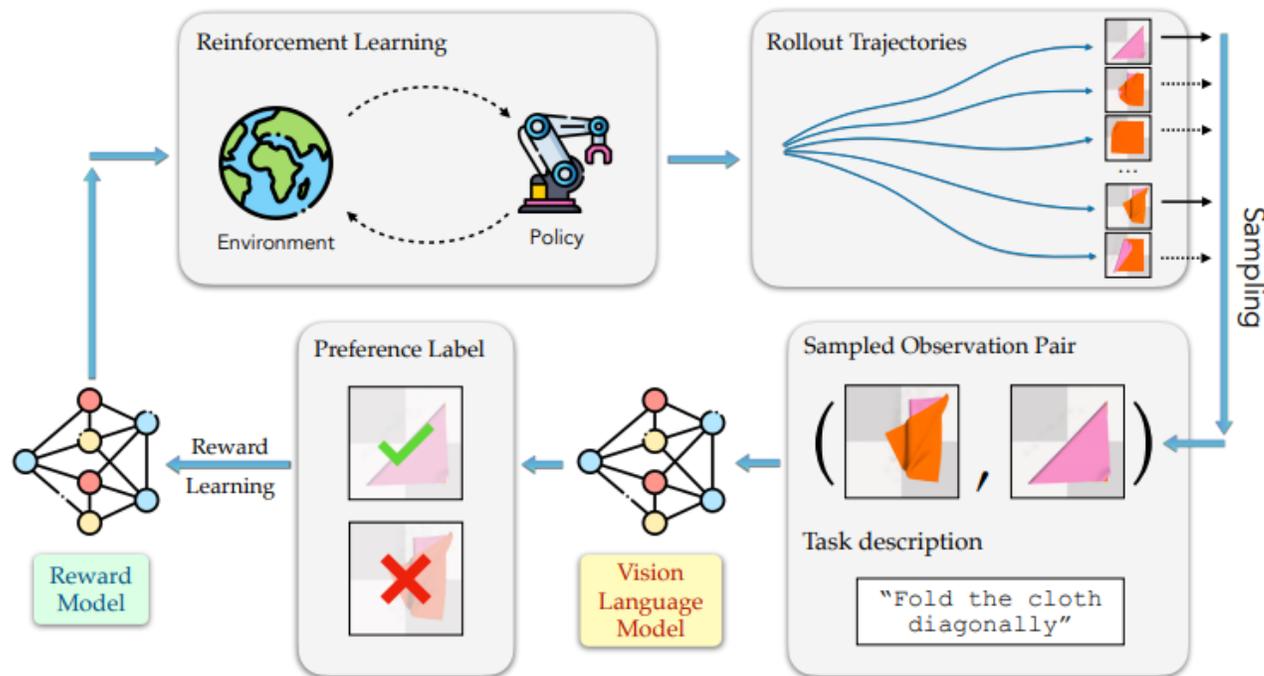
ChatGPT

在這一情境下，我選擇立即評估愛黛兒的傷勢作為我的下一步行動。我會仔細檢查愛黛兒，注意她身上的每一處傷口，尤其是大出血或可能有骨折的跡象。我會留意她的意識狀態，確認她是否清醒、能夠說話，以及是否有震顫或是其他休克的跡象。這一步驟對於確定下一步的急救措施至關重要。

如何執行???



<https://arxiv.org/abs/2402.19299>



<https://arxiv.org/abs/2402.03681>

你會自律行動，安全的把考生帶出迷宮



終極目標

狀態

在評估愛黛兒的傷勢時，  
你遭遇複製體襲擊

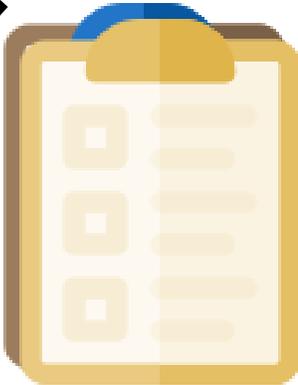
(Image captioning)



記憶  
(經驗)

(空)

反思



計畫

DEPS

<https://arxiv.org/abs/2302.01560>

行動

評估愛黛兒的傷勢



外界環境

Agent

執行過的動作：

在這一情境下，我選擇立即評估愛黛兒的傷勢作為我的下一步行動。我會仔細檢查愛黛兒，注意她身上的每一處傷口，尤其是大出血或可能有骨折的跡象。我會留意她的意識狀態，確認她是否清醒、能夠說話，以及是否有震顫或是其他休克的跡象。這一步驟對於確定下一步的急救措施至關重要。

狀態：在評估愛黛兒的傷勢時，你遭遇複製體襲擊

請產生行動計劃

< 11 / 18 >



## ChatGPT

面對突如其來的複製體攻擊，在確保愛黛兒的安全的同時，我需要迅速應對這一威脅。我的行動計劃如下：

1. 保護愛黛兒：首先，我會立即將愛黛兒移到一個相對安全的位置，如果可能的話，找到掩護物或障礙物來阻擋複製體的視線和攻擊路徑。在移動的過程中，我會盡量保持警惕，隨時準備防禦。

你會自律行動，安全的把考生帶出迷宮



終極目標

狀態

在評估愛黛兒的傷勢時，  
你遭遇複製體襲擊

(Image captioning)

ReAct = Reason + Act

<https://arxiv.org/abs/2210.03629>

Reflexion

<https://arxiv.org/abs/2303.11366>



記憶  
(經驗)

反思



行動

評估愛黛兒的傷勢



計畫

DEPS

<https://arxiv.org/abs/2302.01560>



外界環境

Agent

執行過的動作：

在這一情境下，我選擇立即評估愛黛兒的傷勢作為我的下一步行動。我會仔細檢查愛黛兒，注意她身上的每一處傷口，尤其是大出血或可能有骨折的跡象。我會留意她的意識狀態，確認她是否清醒、能夠說話，以及是否有震顛或是其他休克的跡象。這一步驟對於確定下一步的急救措施至關重要。

狀態：在評估愛黛兒的傷勢時，你遭遇複製體襲擊

簡短總結值得記下來的經驗

< 13 / 18 >



## ChatGPT

在這次的經驗中，我學到了幾個關鍵的教訓：

1. **環境警覺性**：即使在專注於救助傷者的情況下，也不能忽視周圍環境的安全。複製體的突然襲擊提醒了我，即使在看似安靜的時刻，也應保持高度的警惕性。
2. **快速應對能力**：面對突如其來的攻擊，能↓速做出反應，保護自己和傷者的能力至關重要。這要求不僅是身體上的準備，也包括心理上的準備，以冷靜應對突發狀況。

你會自律行動，安全的把考生帶出迷宮



終極目標

狀態

眼前是遭遇複製體攻擊而重傷的愛黛兒

(Image captioning)

ReAct = Reason + Act

<https://arxiv.org/abs/2210.03629>

Reflexion

<https://arxiv.org/abs/2303.11366>



記憶

(經驗)



反思

行動

?????

經驗會影響所採取的行動



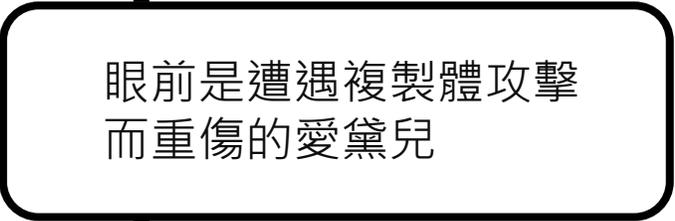
計畫

DEPS

<https://arxiv.org/abs/2302.01560>

Agent

外界環境





You

目標：你叫哥雷姆，你會自律行動，安全的把考生帶出迷宮

記憶：在這次的經驗中，我學到了幾個關鍵的教訓：

狀態：眼前是遭遇複製體攻擊而重傷的愛黛兒

請產生行動計劃

< 17 / 19 >



ChatGPT

面對眼前的情況，考慮到剛剛學到的教訓，我們可以制定如下的行動計劃：

1. 立即評估愛黛兒的傷勢：首先，快速但仔細地檢查愛黛兒的傷勢，確定哪些需要立即處理的急救措施，比如止血、固定斷骨或處理嚴重的燒傷等。
2. 保持高度的環境警覺性：在進行急救措施的同時，要不斷留意周圍環境，防止複製體或其他潛在威脅再次接近。如果可能，利用任何現成的障礙物（如巨石、樹木或牆壁）來為自己和愛黛兒提供一定的掩護。
3. 溝通和尋求幫助：如果附近有其他考生或者同伴，應立即發出信號或直接呼救，尋求他們的援助。同時，告知他們愛黛兒的狀況和可能需要的急救物資。

1. \*\*環  
襲擊提

2. \*\*快  
要。這

3. \*\*急  
衡非常

4. \*\*溝  
高生存

根據沒有任何記憶的計畫

請採取下一步行動 (請執行一個動作就好)

< 9 / 18 >



**ChatGPT**

在這一情境下，我選擇立即評估愛黛兒的傷勢作為我的下一步行動。我會仔細檢查愛黛兒，注意她身上的每一處傷口，尤其是大出血或可能有骨折的跡象。我會留意她的意識狀態，確認她是否清醒、能夠說話，以及是否有震顫或是其他休克的跡象。這一步驟對於確定下一步的急救措施至關重要。

請採取下一步行動 (請執行一個動作就好)

< 18 / 19 >



**ChatGPT**

基於行動計劃的第一步，我立即評估愛黛兒的傷勢。我謹慎地檢查她的傷口，尋找任何明顯的出血點、斷骨或可能的內傷跡象。同時，我留心她的呼吸和意識狀態，試圖確定她是否處於休克狀態或有其他需要緊急處理的醫療問題。在這一過程中，我也保持高度警惕，注意四周是否有複製體或其他潛在威脅接近。



根據有記憶的計畫

# 延伸閱讀

Overview paper: <https://arxiv.org/abs/2309.07864>

