



讓 AI 自主運行其他 AI

自主運行 ChatGPT

- 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>

G O D ⚡ M O D E

把李宏毅老師 YouTube 頻道上機器學習教學影片做成學習地圖



Suggested: Watch all of Li Hongyi's machine learning tutorial videos on YouTube.

+ Add



Suggested: Organize the videos into categories based on topic and difficulty level.

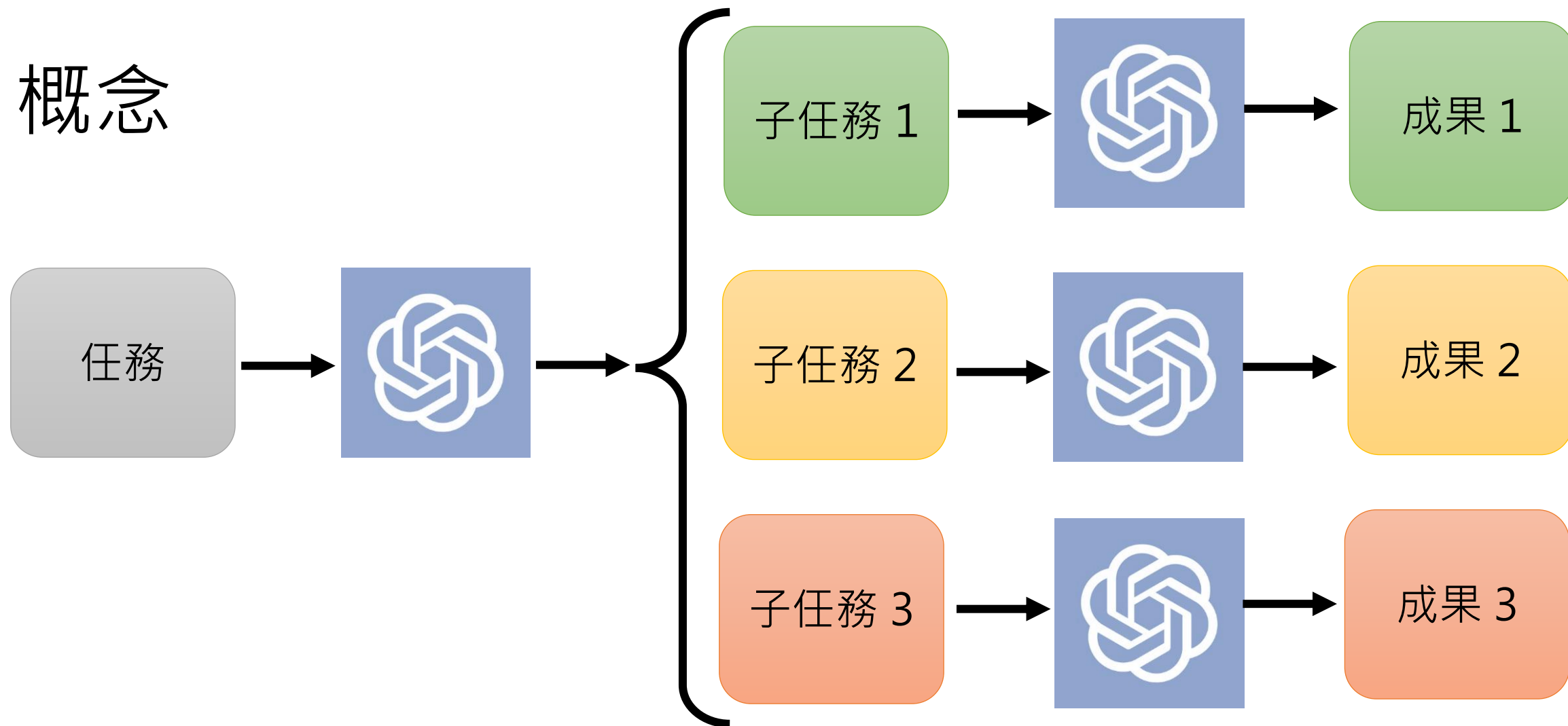
+ Add



Suggested: Create a visual learning map that displays the categories and their corresponding videos.

+ Add

概念



把大任務拆解成一連串的小任務，也就是讓機器「**做計畫**」

Recursive Reprompting and Revision (Re3)

<https://arxiv.org/abs/2210.06774>

Premise

A new law grad returns home to start her career, but struggles with the broken justice system.

Story

Liza Turner pulled up in front of the house where she'd grown up. Little had changed since she was a teenager...

Plan

Premise

Premise: A new law grad returns home to start her career, but struggles with the broken justice system.

Setting

Setting: The story is set in a small town in the United States.

Characters

1. Character Portrait: Liza Turner is a 22-year-old woman.
2. Character Portrait: Peyton Turner is Liza's older sister.

Outline

- Outline the main plot points of the story.
1. Liza Turner graduates from law school.
 2. She moves back to her hometown to start her career.
 3. She struggles with the reality of the broken justice system.

Language Models as Zero-Shot Planners

<https://arxiv.org/abs/2201.07207>

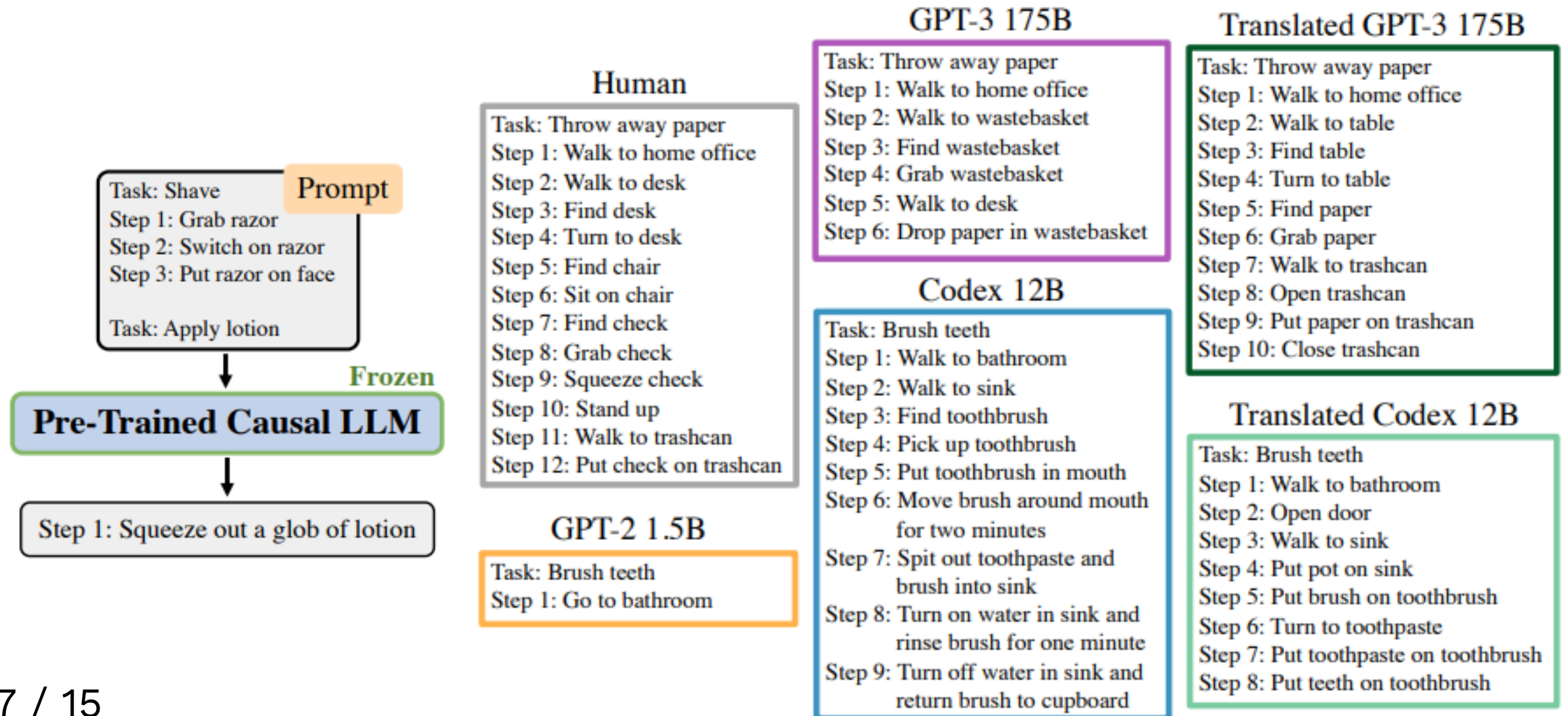
Task: Get Glass of Milk



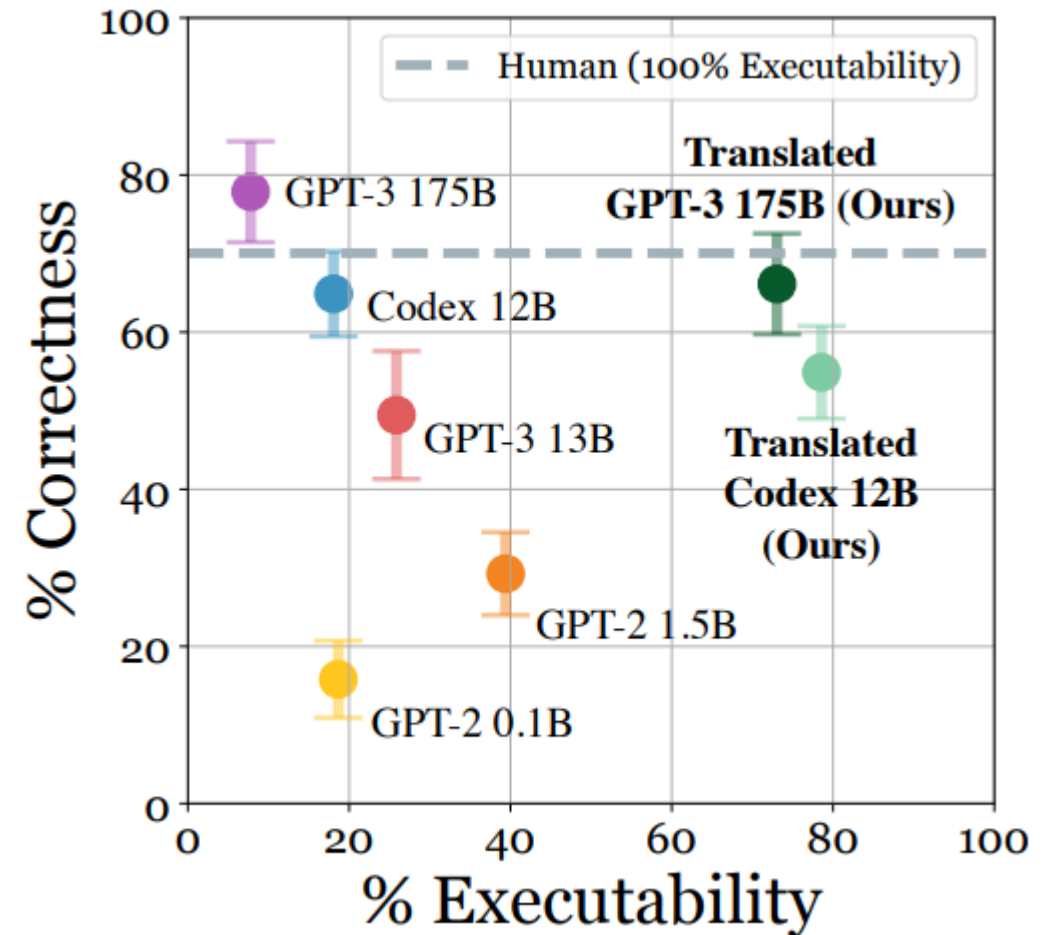
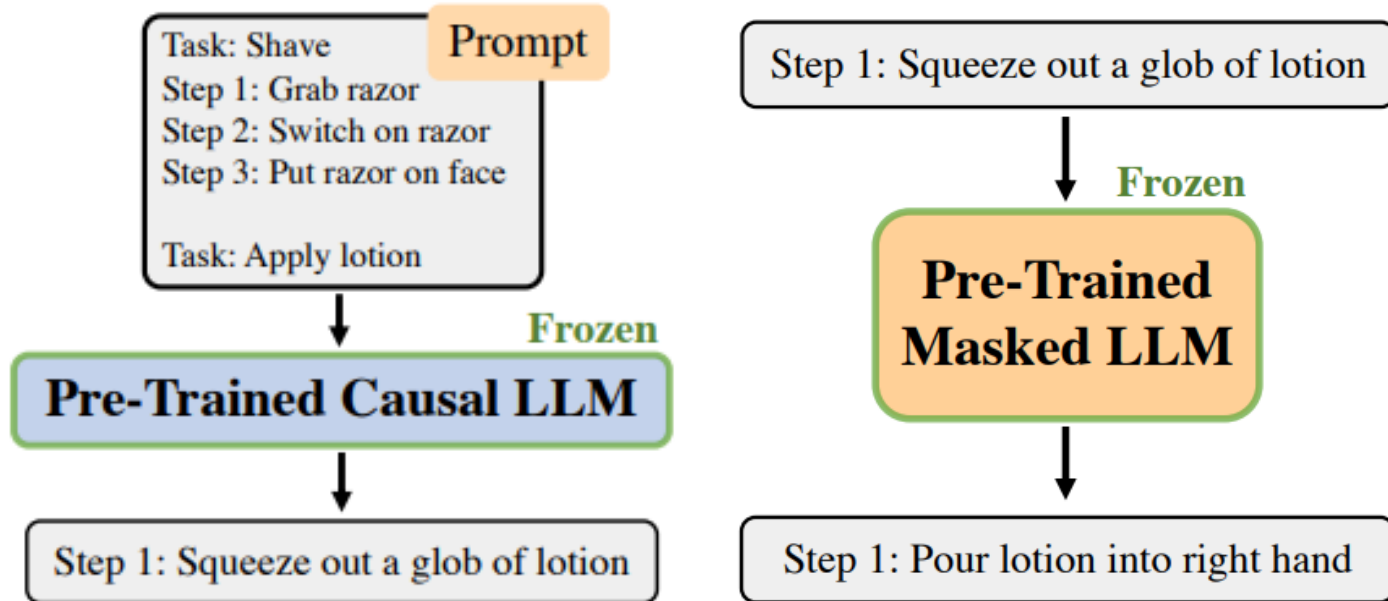
挑戰一：如何自動根據大任務規劃一系列子任務

挑戰二：如何把子任務對應到可執行的動作

Language Models as Zero-Shot Planners



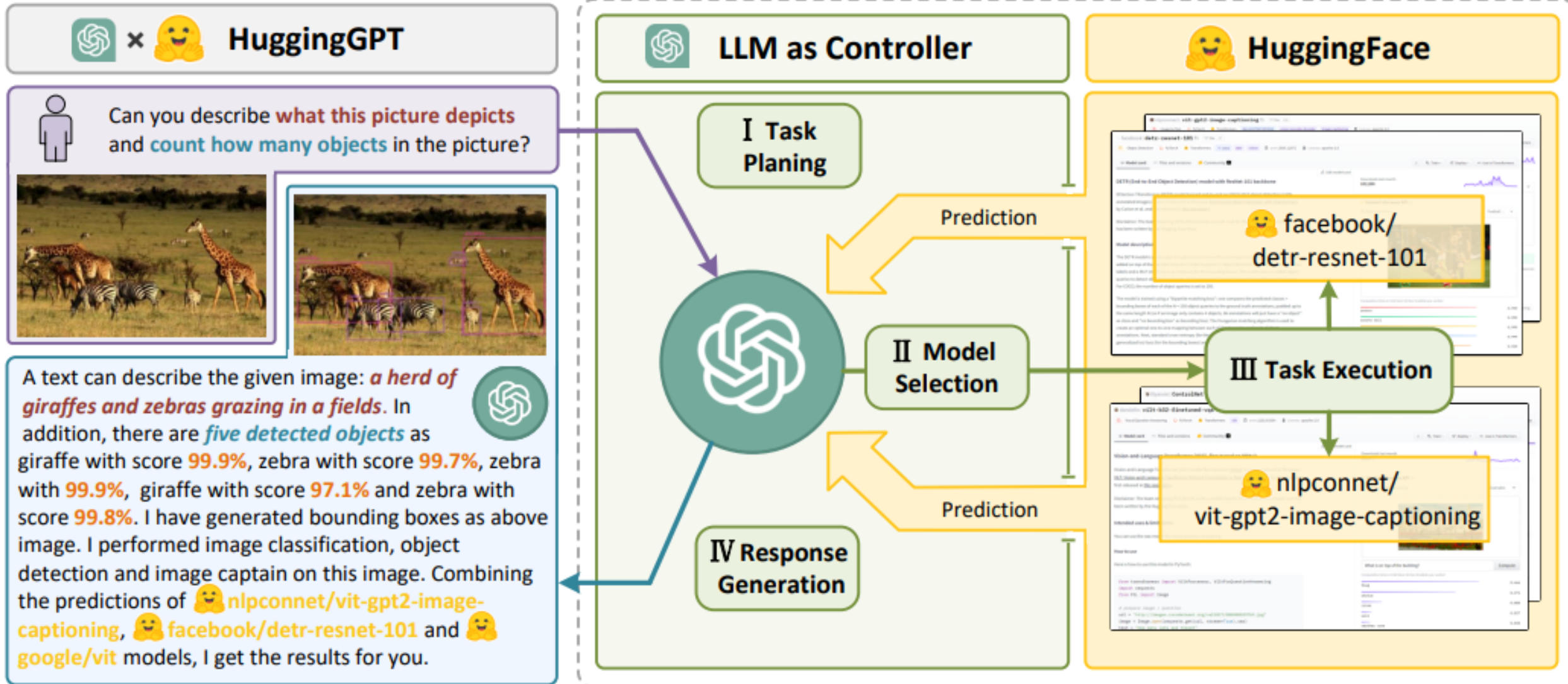
Language Models as Zero-Shot Planners



Hugging GPT (Jarvis)

<https://github.com/microsoft/JARVIS>

<https://arxiv.org/abs/2303.17580>



Hugging GPT (Jarvis)

<https://github.com/microsoft/JARVIS>

<https://arxiv.org/abs/2303.17580>

please generate an image where a girl is reading a book, and her pose is the same as the boy in the image example.jpg. Then please describe the new image with your voice.

Hugging GPT (Jarvis)

<https://github.com/microsoft/JARVIS>

<https://arxiv.org/abs/2303.17580>

Response



[Image-1]



[Image-2] = <resource-1>



[Image-3] = <resource-2>



[Image-4]

a girl sitting on a bed
reading a book

[Text-1] = <resource-5>




[Audio-1]

The image you gave me is of "boy". The first thing I did was use 🤖 **OpenCV's openpose control model** to analyze the pose of the boy in the image. I fed that image, specifically located at [Image-1], into the model to get an output image of the pose, located at [Image-2]. Then I take that pose and generate a new image, located at [Image-3], using 🤖 **llyasviel/sd-controlnet-openpose** and the instruction "a girl is reading". Next, I ran object detection on the image generated in the previous step, to get predictions of the objects present in the image, including the box location. These predictions were run through the 🤖 **facebook/detr-resnet-101** model. The box was drawn on the image, which is located at [Image-4]. Then, I used the predictions of a image classification model 🤖 **google/vit-base-patch16-224** and a image caption model 🤖 **nlpconnect/vit-gpt2-image-captioning** model to generate the caption for newly generated image. It generated text: "a girl sitting on a bed reading a book". Lastly, I ran the 🤖 **facebook/fastspeech2-en-ljspeech** model to convert the generated text into audio, located at [Audio-1].

Language models can solve computer tasks

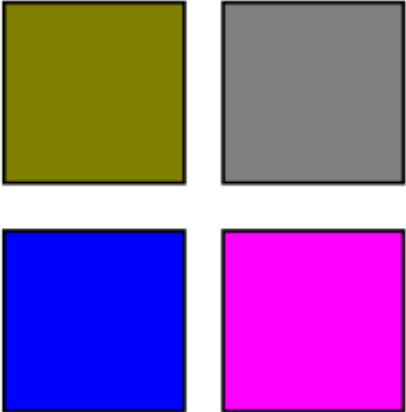
<https://arxiv.org/abs/2303.17491>

Click on the  colored box.

Click on the "no" button.

Find the email by **Brier** and click the trash icon to delete it.

Switch between the tabs to find and click on the link "Quis".



vitae in morbi
pellentesque mauris diam

no

sagittis, dolor, accumsan
ultrices nibh facilisis

submit

Primary

Kiersten
Pellentesque tr..
Mi ullamcorper ..

Marin
Neque..
Magna elementum..

Fredia
Imperdiet..
Nulla nunc cons..

Tab #1 Tab #2 Tab #3

Rhoncus rutrum amet,
habitant. Sagittis [gravida](#)
sed ornare potenti. [Et, id](#)
[blandit](#) id malesuada purus
praesent egestas orci. [Quis](#)
[purus.](#)

MiniWoB++ (Mini World of Bits++)

<https://miniwob.farama.org/>

We have an autonomous computer control agent that can perform atomic instructions specified by natural language to control computers. There are two types of instructions it can execute. First, given the instruction that matches the regular expression, <type regex>, it can type a list of characters via the keyboard. This instruction should specify the target keyboard input for the agent to type. Before this typing instruction, you should first locate the cursor by clicking the input box with the click instruction.

Second, given the instruction that matches the regular expression, <press regex>, it can press a specific key on the keyboard.

Third, given the instruction that matches the regular expression, <clickoption regex>, it can click an option HTML element in a list with an XPath that is visible on the webpage. The target of this instruction should be a valid XPath.

Fourth, given the instruction that matches the regular expression, <movemouse regex>, it can move the mouse cursor on an HTML element with an XPath that is visible on the webpage.

Lastly, given the instruction that matches the regular expression, <clickxpath regex>, it can click an HTML element with an XPath that is visible on the webpage. The target of this instruction should be a valid XPath.

```
type regex = "^type\s.{1,}$"
press regex = "^press\s(enter|arrowleft|arrowright|arrowup|arrowdown|
    backspace)$"
clickxpath regex = "^clickxpath\s.{1,}$"
clickoption regex = "^clickoption\s.{1,}$"
movemouse regex = "^movemouse\s.{1,}$"
```

Use the terminal below to delete a file ending with the extension .rb

```
terminal
Welcome! Type help for a list
of available commands.
Last login: Mon Jan 30 2023
user$
```

Initial state

```
terminal
Welcome! Type help for a list
of available commands.
Last login: Mon Jan 30 2023
user$ ls
alloy.jpg code.rb file.gpg
navigator.txt skynet.html
thunderbird.html
user$
```

Current state

Task

The task is to use the terminal below to delete a file ending with the extension .rb.

Explicit RCI

Plan

Here is a plan to solve the task using the instructions specified:

1. Type "rm [filename]" command to delete the file ending with ".rb".
2. Press "enter" key

Critique

Find problems with this plan.

This plan does not specify the filename of the file ending with the extension .rb. Without knowing the filename, the agent cannot delete the file

Improvement

Based on this, the improved plan for the agent to complete the task are as follows.

1. Type "ls" command to list all files in the terminal.
2. Type "rm [filename]" command to delete the file ending with ".rb".
3. Press "enter" key after typing each command to execute the command.

Task grounding

Implicit RCI

According to the current plan, the next proper instruction should be 'type "rm [filename]"'.

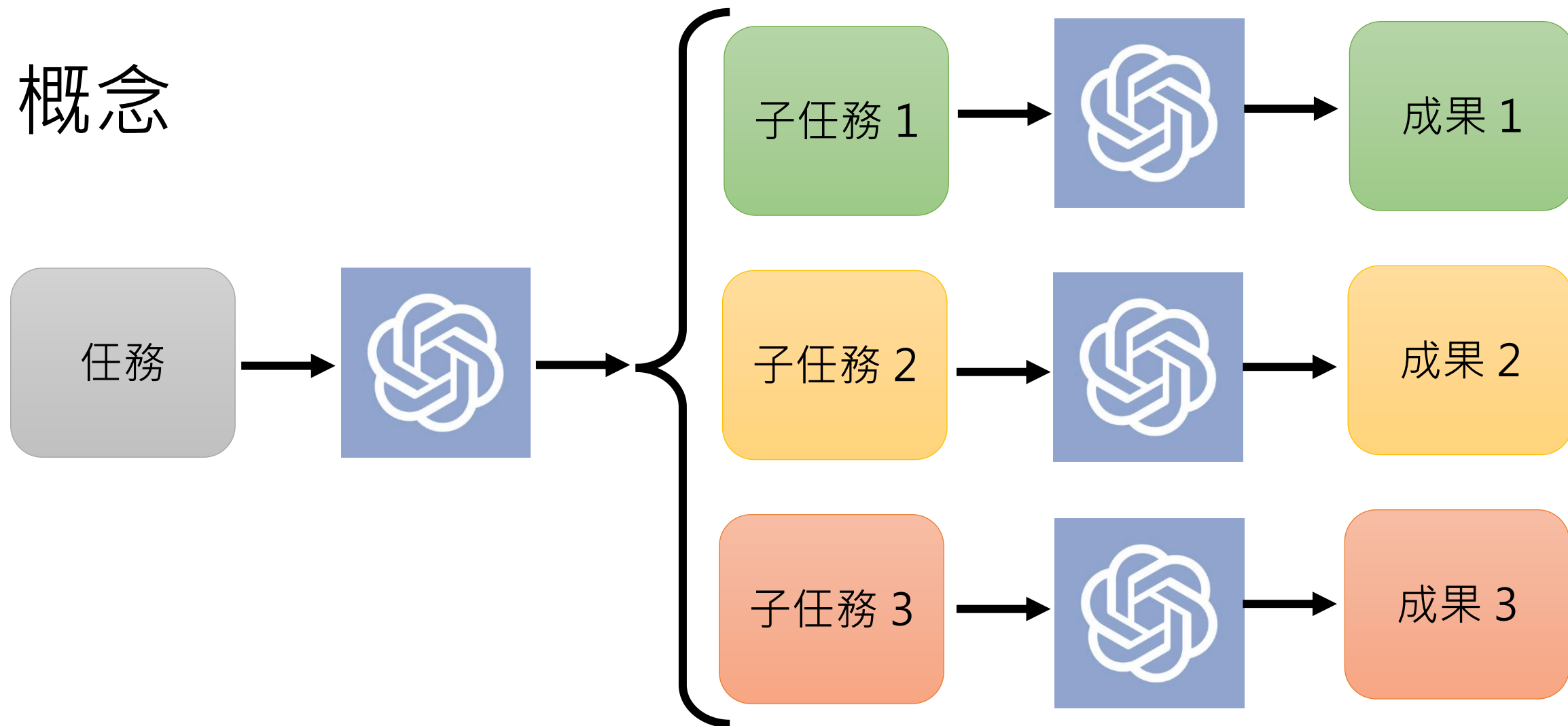
Therefore, considering the output on the webpage, the specific instruction for solving the task should be 'type' instruction with "rm code.rb" command.

Without explanation, the single instruction that matches one of the regular expressions is "type "rm code.rb""

State grounding

Agent grounding

概念



把大任務拆解成一連串的小任務，也就是讓機器「**做計畫**」