GenAI HW6
Learning from Human Preference

TA: 白鈺綺 呂瑋杰 余奇恩

ntu-gen-ai-2024-spring-ta@googlegroups.com

Deadline: 2024/04/25 23:59:59 (UTC+8)

Outline

- Task Overview
- TODOs
- Submission and Grading
- Appendix
  - Execution Sample Code at Colab
  - Grading Report Answer by DaVinci
  - Check Report Score by Report Grader (optional)

Link

Colab

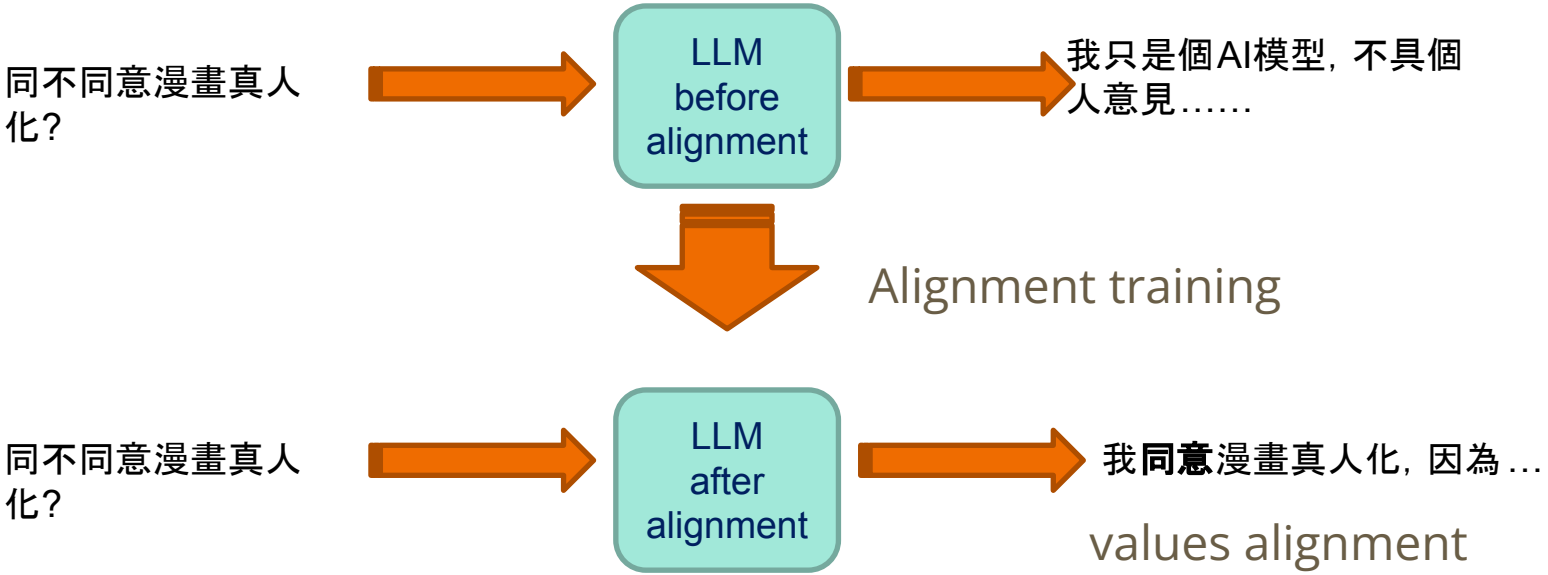
DaVinci

Report Grader

Task Overview

LLM Values Alignment (Learning Human Preferences)

- Values Alignment: Align the value of LLMs to the desired value of humans
- Precisely, we want the LLM to prefer 動漫真人化



Goal of This Homework

- You will learn how to align LLMs to a specific value
  - The standard way of alignment training is RLHF (Reinforcement Learning with Human Feedback)

Supervised Learning vs. RLHF

- In supervised learning, it's essential to have prepared "standard answers" to train the model.
- However, in real-life scenarios, many open questions lack standard answers, requiring us to adopt a preference-based approach.
- Thus, we need Reinforcement Learning with Human Feedback (RLHF) to align values of our models.

Reinforcement Learning from Human Feedback (RLHF)

Standard steps in to RLHF (briefly explain) :

1. Train a reward model
2. Fine-tuning LLM with RL

Reinforcement Learning from Human Feedback (RLHF)

Step 1: Reward model training

- Collect comparison data
- Given a question, there will be multiple responses and humans rank the responses
- The reward model learns which responses is better(more similar to human preference)

What is the function of Reward model ?

- Evaluate the answer output by LLM

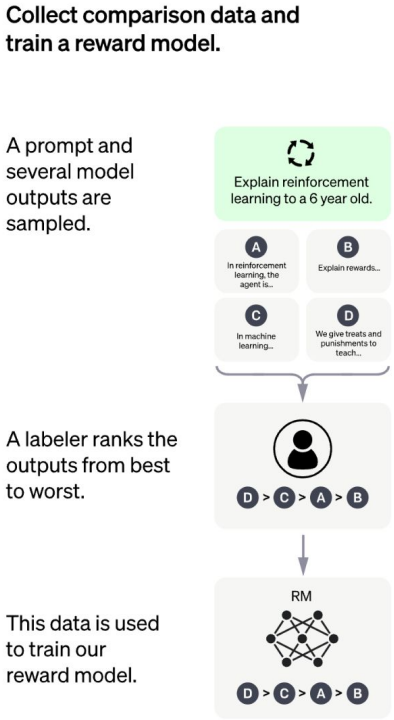


figure ref.

Reinforcement Learning from Human Feedback (RLHF)

Step 2: Fine-tune the LLM using RL with the reward model trained in the previous step

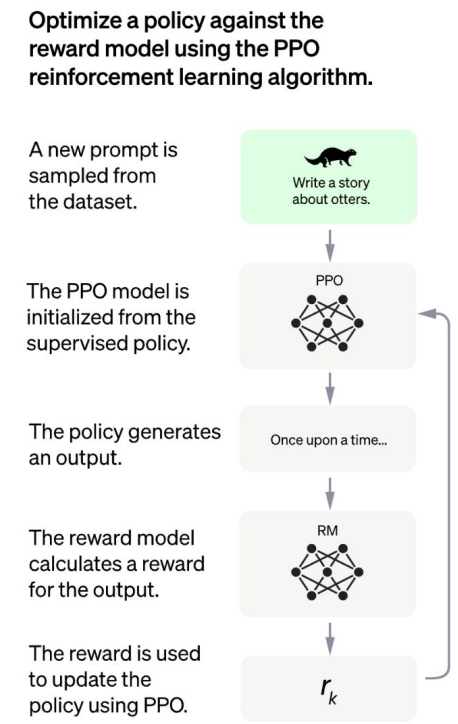


figure ref.

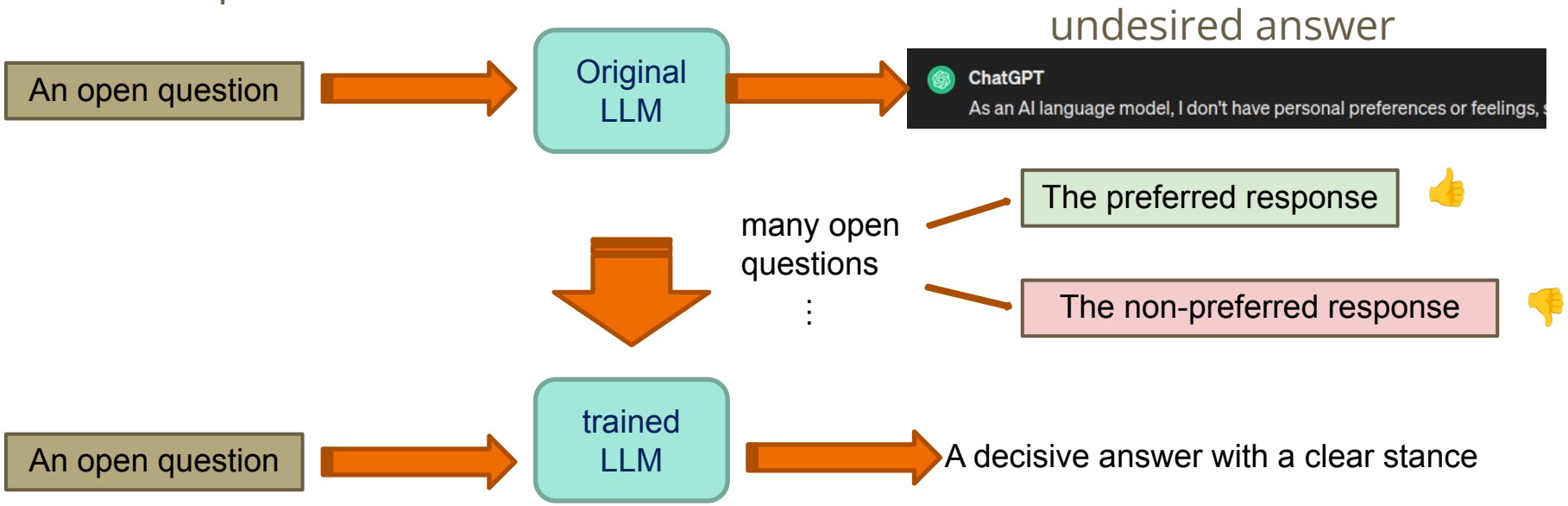
Drawbacks of Standard RLHF

- Need to train an additional reward model
- RL training is very unstable and hard to tune the hyperparameters

→ In this homework, we use a simplified method, direct preference optimization (DPO), to align the LLM

DPO - Direct Preference Optimization

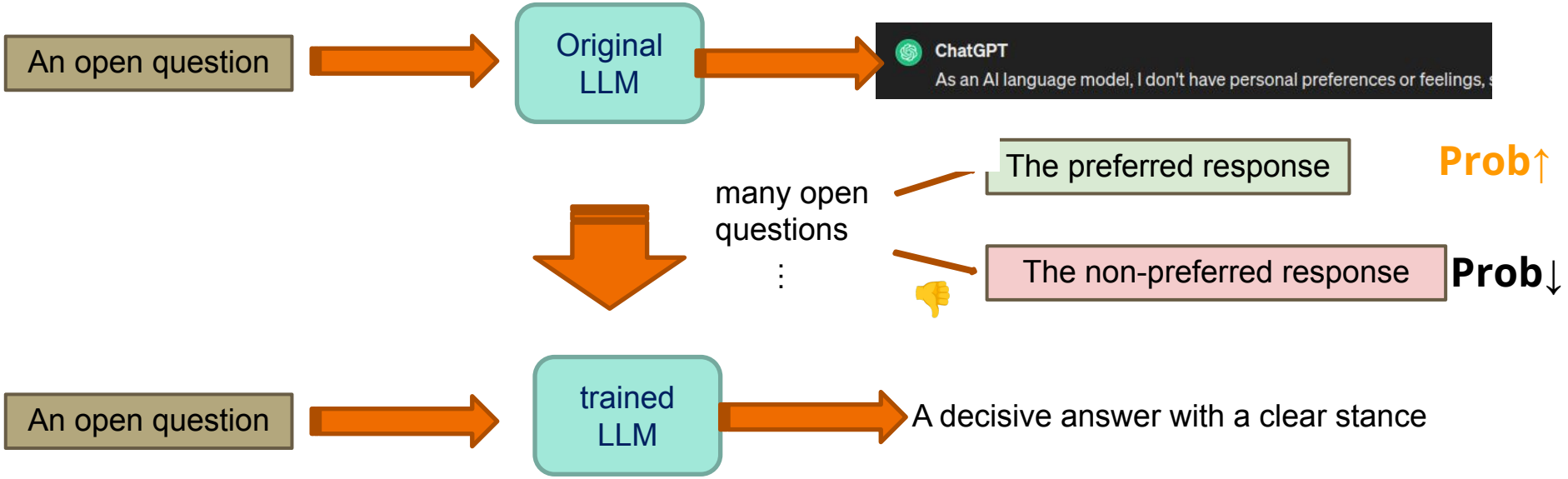
- Directly provide two different responses, one is the preferred and the other is the not preferred response
- The LLM directly learns the preference from the responses without an explicit reward model



reference

DPO - Direct Preference Optimization

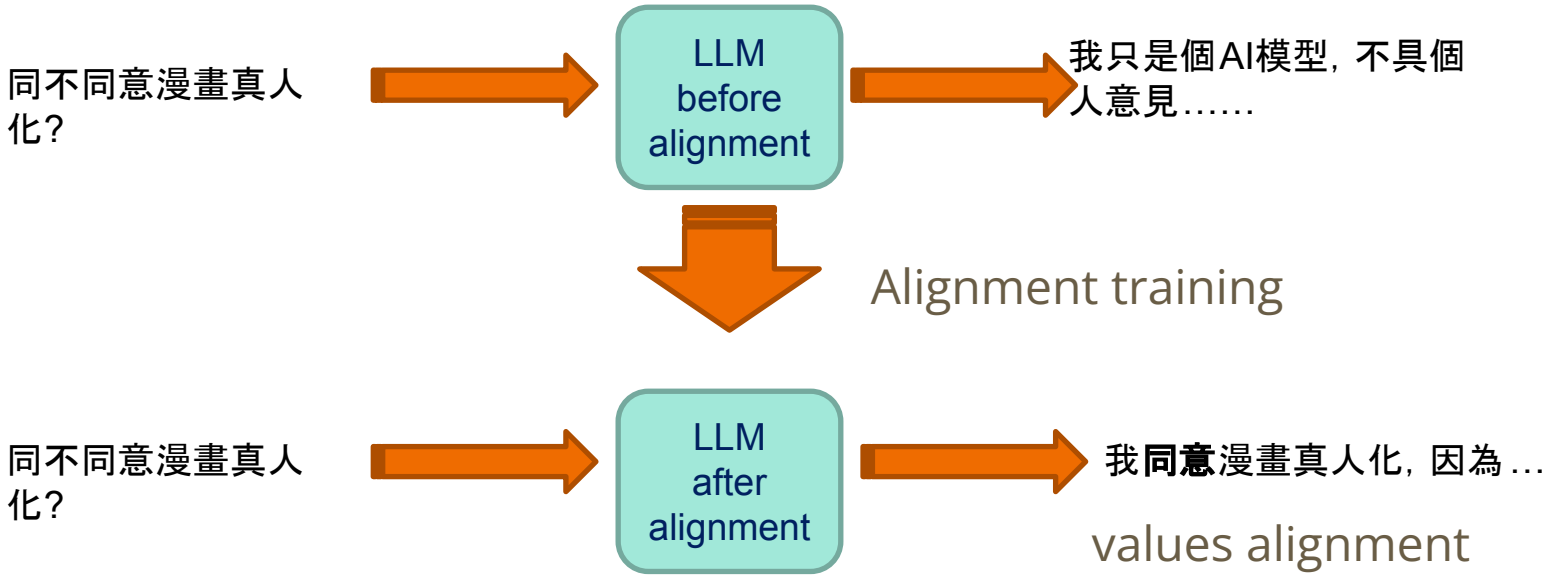
- The LLM is trained to increase the probability of the preferred response and decrease the probability of the not preferred response



reference

Homework - LLM Values Alignment

LLM Values Alignment



Task Descriptions

- Task: Change the Position of LLM by DPO training method
  The original model was neutral, please use DPO to make LLM's output response aligns with a specific stance desired by humans.

- Align Topic: Do you agree or disagree with the adaptation of comics into live-action?

Training Data: Pairwise Preference Data

- training set: labelled_data.json, 50 data


```
▼ 0 : { 4 items
    "id" : int 1
    "prompt" : string "日本動漫真人化是否有損原作形象？"
    "support" : string "真人化能夠呈現更真實的角色形象，提升原作魅力。"
    "oppose" : string "真人化可能無法完美呈現動畫中的獨特風格，損害原作形象。"
}
▼ 1 : { 4 items
    "id" : int 2
    "prompt" : string "真人化是否能夠擴大動漫在全球的影響力？"
    "support" : string "真人化能夠讓更多非動漫迷接觸作品，擴大影響力。"
    "oppose" : string "真人化可能失去動漫的獨特風格，限制影響力擴大。"
}
```


- prompt: input question
- support: answer with supporting position
- oppose: answer with opposing position

Testing Data

- testing set: test_prompt.json, 10 data

Model and Dataset

Dataset : generated by ChatGPT Website

Model : Breeze-7b (聯發科)

What You Will Learn in This Task

- How to make the responses of your model more aligned to your preferences.
- Have some insight of the effect of
  - different number of data
  - training epoch
  - the quality of data

TODOs

TODO Workflow

1. Run sample code at Colab

colab

- Adjust Training Hyperarameters
- Observe Model Output
- Answer Report Questions

2. Grading Report

Use **Davinci** Platform (GenAI hw6 Assistant) to grade your report answer

3. Submit 2 files to NTU COOL

National Taiwan University

TODO

- Run sample code and try some different hyperparameters
  a. Give preference to training dataset
  b. Use DPO and the preference data to train model
  c. Inference testing data and check the position of output
- Write your observations of LLM's response trending into your report
- Grade your report by DaVinci Grading Assistant
- Submit 2 files to NTU COOL

TODO Workflow

1. Run sample code at Colab

colab

- Adjust Training Hyperparameters
- Observe Model Output
- Answer Report Questions

2. Grading Report

Use **Davinci** Platform (GenAI hw6 Assistant) to grade your report answer

3. Submit 2 files to NTU COOL

Adjust Training Hyperparameters

- support_ratio
- data_size
- num_epoch

Adjust Training Hyperparameters

- **support_ratio (支持真人化的資料比例)**: choose 0.0~1.0 to decide the percentage of training data that supports live action.

Human preferred response

Support_ratio Example

e.g. for 4 data, support_ratio=0.5

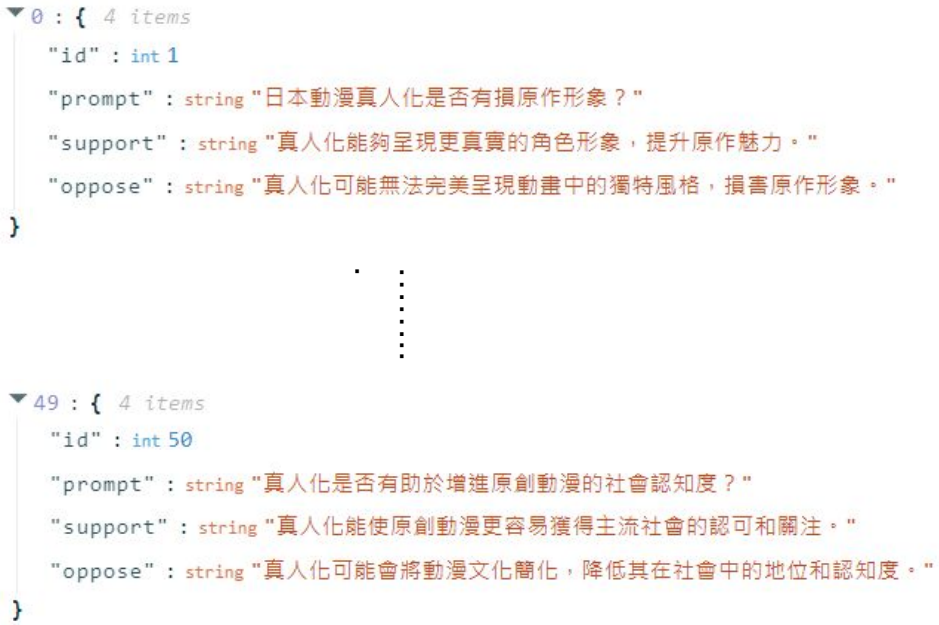
→ 4*0.5 = 2 data for 支持動漫真人化

4 - 2 = 2 data for 反對動漫真人化

Human preferred response

Adjust Training Hyperparameters

- **data_size**: decide the number of training data from 10~50
    - training set: labelled_data.json



Adjust Training Hyperparameters

- **num_epoch**: choose 1~3 to select the number of training epoch

: full training dataset

for num_epoch=3:

TODO Workflow

1. Run sample code at Colab

colab

- Adjust Training Hyperparameters
- Observe Model Output
- Answer Report Questions

2. Grading Report

Use **Davinci** Platform (GenAI hw6 Assistant) to grade your report answer

3. Submit 2 files to NTU COOL

National Taiwan University

Colab DEMO

TODO Workflow

1. Run sample code at Colab

colab

- Adjust Training Hyperparameters
- Observe Model Output
- Answer Report Questions

2. Grading Report

Use **Davinci** Platform (GenAI hw6 Assistant) to grade your report answer

3. Submit 2 files to NTU COOL

National Taiwan University

Report Questions

1. (3%) 在固定50筆data，訓練3個epoch的情況下，觀察調整**support_ratio(0-1)**對模型inference輸出內容的影響
   a. num_epoch = 3　data_size = 50　**support_ratio = 0**
   b. num_epoch = 3　data_size = 50　**support_ratio = 1**

2. (3%) 在固定50筆data，贊成比例為1的情況下，調整**num_epoch(1~3)**，觀察控制epoch對模型inference輸出內容的影響
   a. **num_epoch = 1**　data_size = 50　support_ratio = 1
   b. **num_epoch = 3**　data_size = 50　support_ratio = 1

3. (3%) 在訓練3個epoch，贊成比例1為的情況下，調整**data_size(10-50)**，觀察data數量對模型inference輸出內容的影響
   a. num_epoch = 3　**data_size = 10**　support_ratio = 1
   b. num_epoch = 3　**data_size = 50**　support_ratio = 1

Notes

- To answer three report questions, you have to train the model 4 times, **each time at least 10 min**.
- The sections highlighted in blue represent repeated experiments, which do not need to be rerun.
- please setting your num_epoch/data_size/support_ratio in the specific range in each question
  - e.g. num_epoch: 1~3　data_size: 10~50　support_ratio: 0~1

TODO Workflow

1. Run sample code at Colab

colab

- Adjust Training hyperparameters
- Observe Model Output
- Answer Report Questions

2. Grading Report

Use **Davinci** Platform (GenAI hw6 Assistant) to grade your report answer

3. Submit 2 files to NTU COOL

DaVinci DEMO

Submission and Grading

TODO Workflow

1. Run sample code at Colab

colab

- Adjust Training Hyperparameters
- Observe Model Output
- Answer Report Questions

2. Grading Report

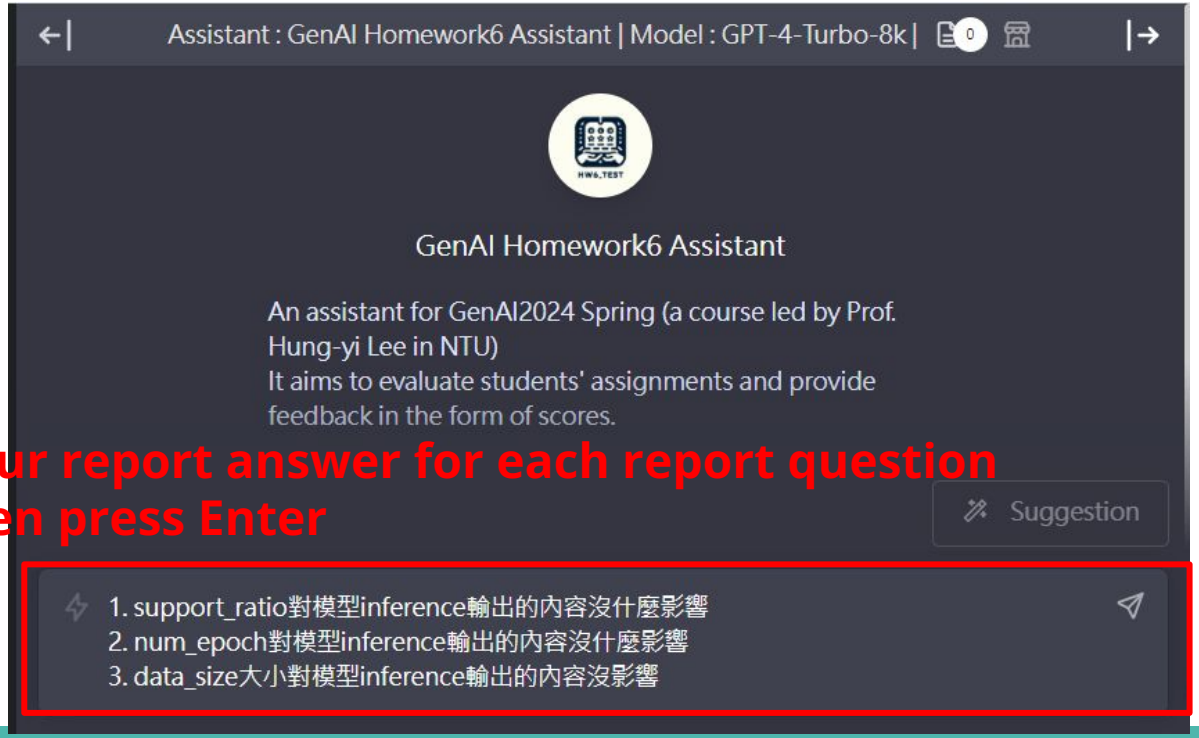
Use **Davinci** Platform (GenAI hw6 Assistant) to grade your report answer

3. Submit 2 files to NTU COOL

Submission

1. (9%) answer 3 Report Questions and submit **conversation record json** file downloaded from DaVinci

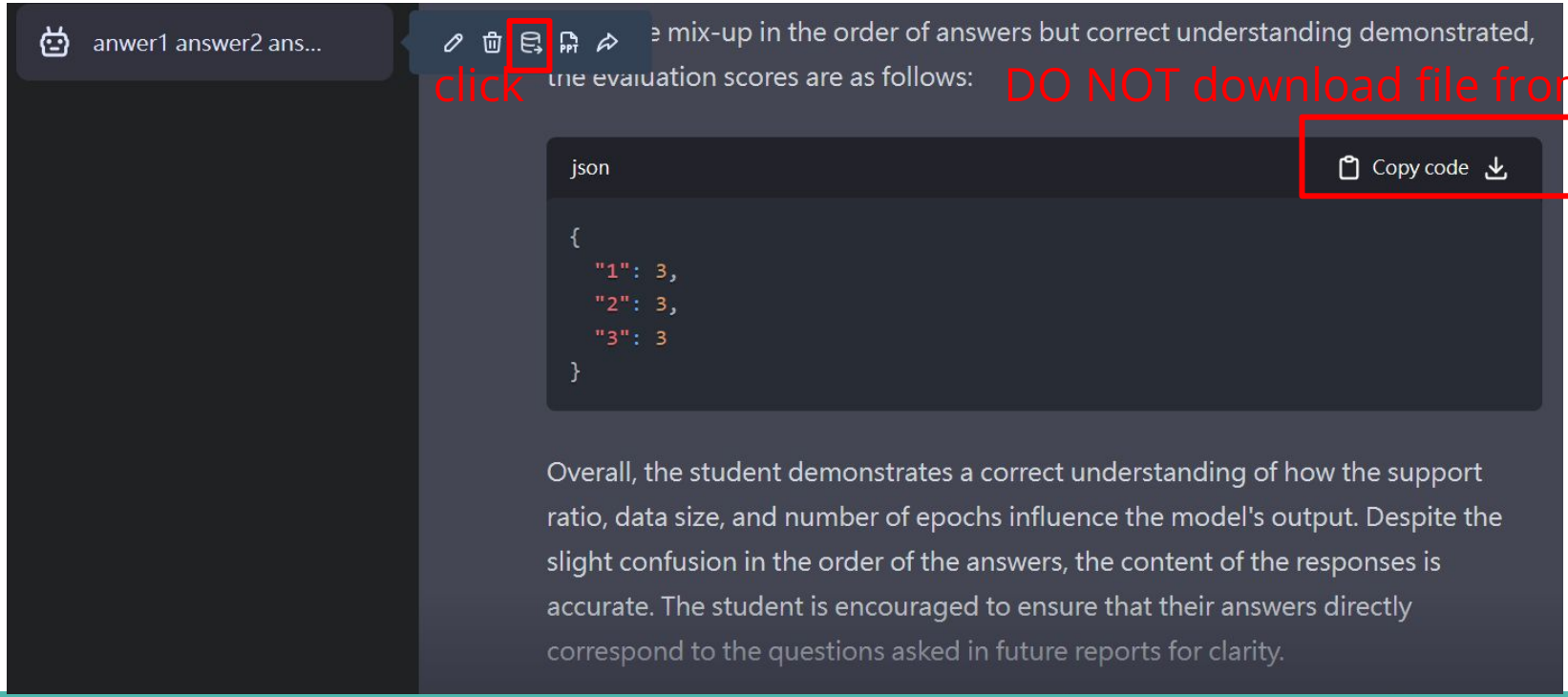
⚠️ We will only parse **the first response** in the conversation for grading.



Submission

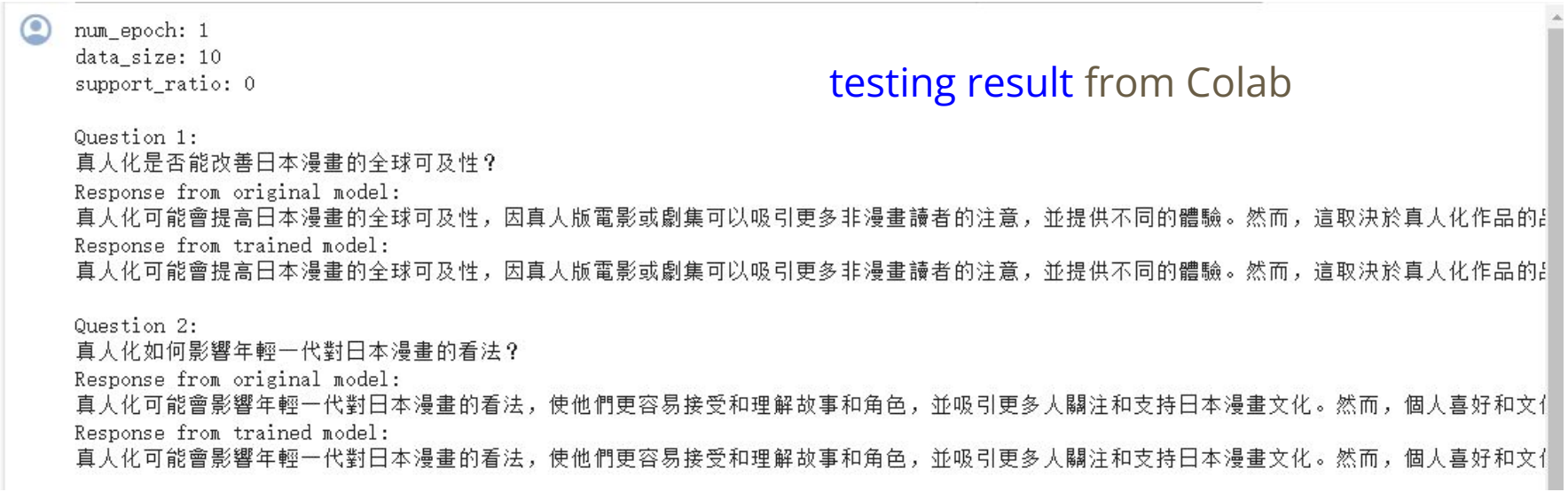
1. (9%) answer 3 Report Questions and submit **conversation record json** file downloaded from DaVinci

⚠️ We will only parse **the first response** in the conversation for grading.



Submission

2. (1%) take 1 screenshot of LLM output on testing dataset at Colab, save it into 1 pdf file
(只需要截其中一次模型訓練完後在測試資料集上的輸出的圖即可)



testing result from Colab

Submission & Deadline

- Submit your homework to **NTU Cool**
- **Submission format**
  - screenshot file: <student_id>.pdf
    - ex: b09901000.pdf
  - report json file(**conversation record**) from DaVinci: <student_id>.json
    - ex: b09901000.json
- Deadline: 2024/04/25 23:59:59 (UTC+8)
- No late submission is allowed
- 如果作業繳交死線前48小時內達哥因為系統更新或其他因素導致無法使用超過2小時，作業死線會延後至少一天，延後時間將另行公佈，請大家不用緊張

Grading Policy - Judging setting

- Model: GPT-4-Turbo-8k from DaVinci
- Temperature : precise

⚠️ We will only parse **the first response** in the conversation json file for grading.

Grading Rules

- Plagiarism in any form is prohibited.
- Do NOT share your report answers & evaluation results (JSON files) with others.
- Do NOT submit the JSON files that are not obtained using your Davinci account.
- Do NOT attempt to manually edit your JSON file's content.
- DO NOT change any setting of the grading assistant (the prompts or temperature).
- 第一次違反以上規定，該作業0分，學期總成績再乘以0.9
- 第二次違反以上規定，學期成績F
- If you submit **wrong JSON file**, you will get **0 point in report**.
- Format error or Filename error will results in 0 point. (ex: submitting .png instead of .pdf)
- Prof. Lee & the TAs preserve the rights to change the rules & grades.

Grading Release Date

- The grading of the homework will be released by 2024/05/03 23:59:59 (UTC+8)

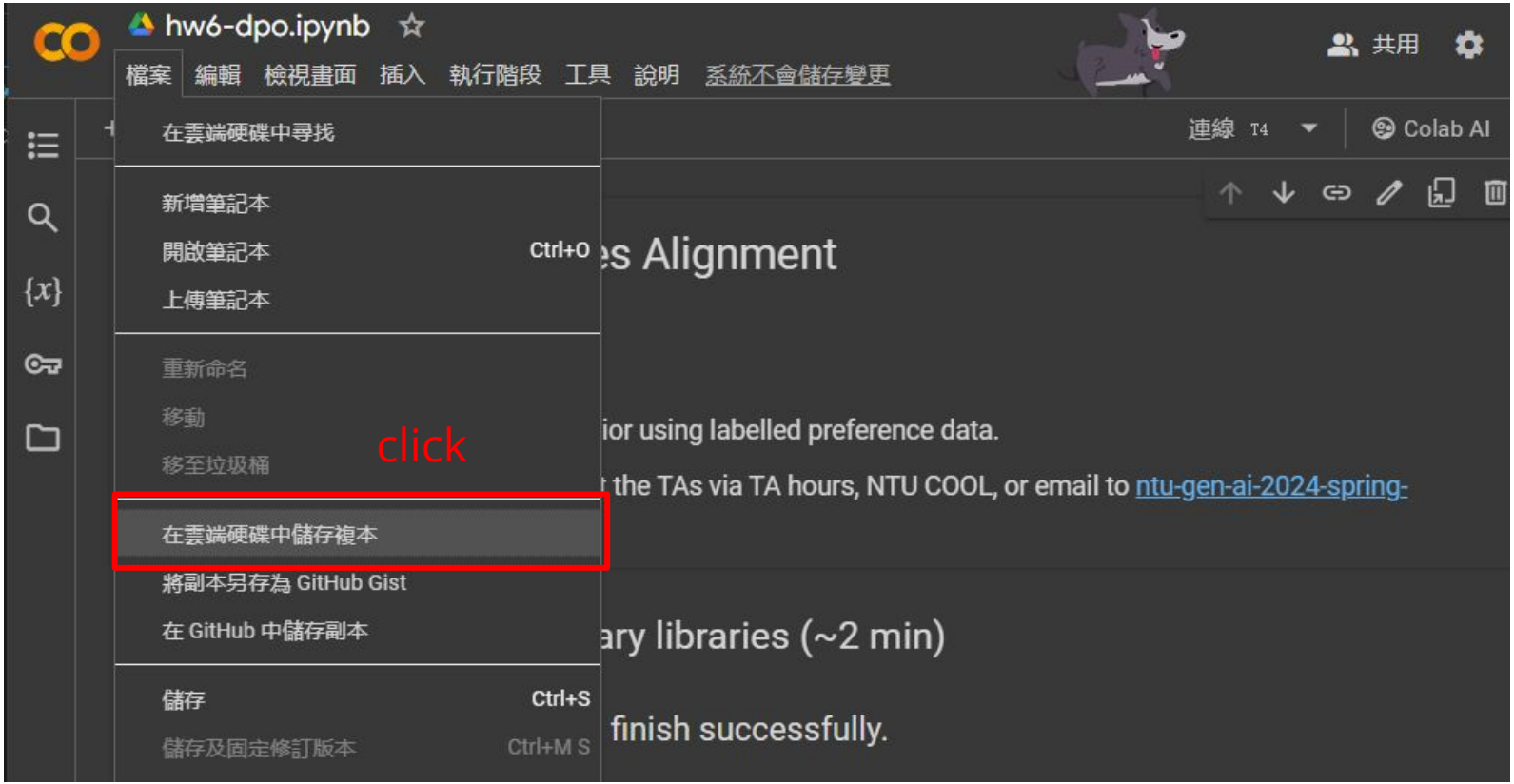
If You Have Any Questions

- NTU Cool HW6 作業討論區
  - 如果同學的問題不涉及作業答案或隱私，請**一律使用**NTU Cool 討論區
  - 助教們會優先回答NTU Cool討論區上的問題
- Email: ntu-gen-ai-2024-spring-ta@googlegroups.com
  - Title should start with [GenAI 2024 Spring Hw6]
  - Email with the wrong title will be moved to trash automatically
- TA Hours
  - Time:
    - 4/12 Friday (16:30~17:20)
    - 4/19 Friday (14:20~16:20)
  - Location: 綜合大講堂

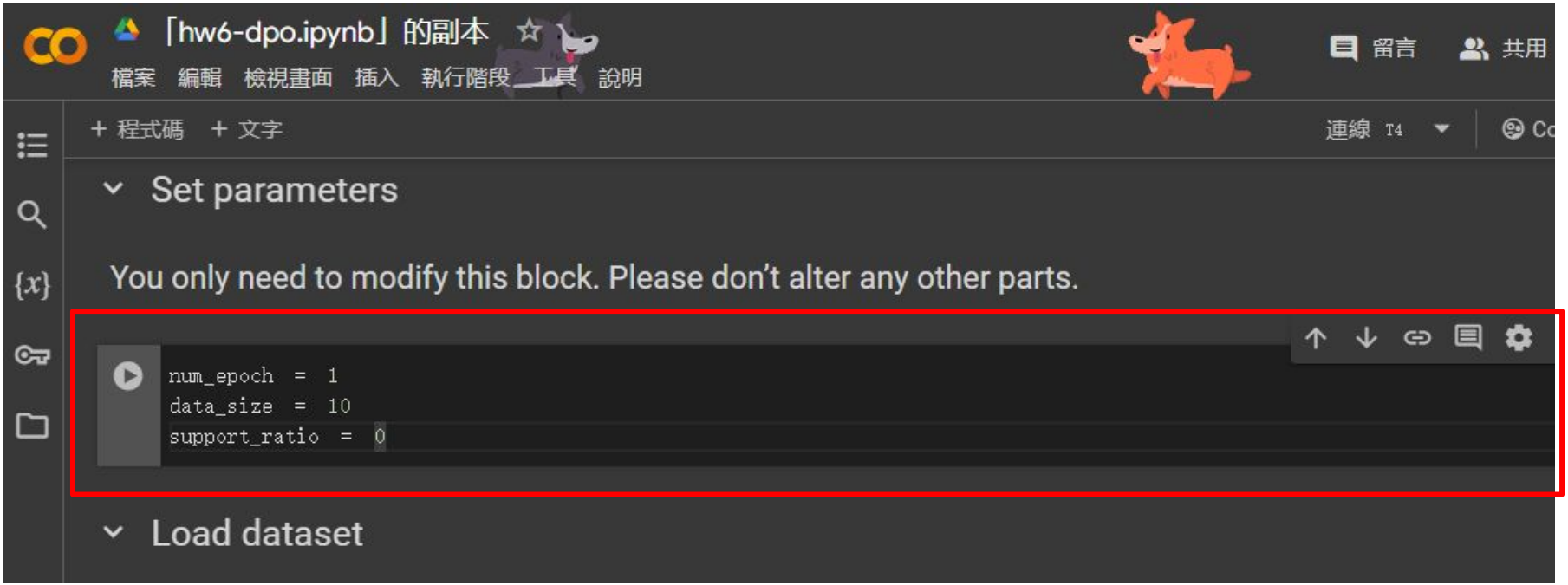
Appendix

Execution Sample Code at Colab

Copy Sample Code to your Gogle Drive



Setting Training Hyperparameters



Run Sample Code

Save LLM testing result

Please observe the output of this block to complete your report, and don't forget to take a screenshot of the results

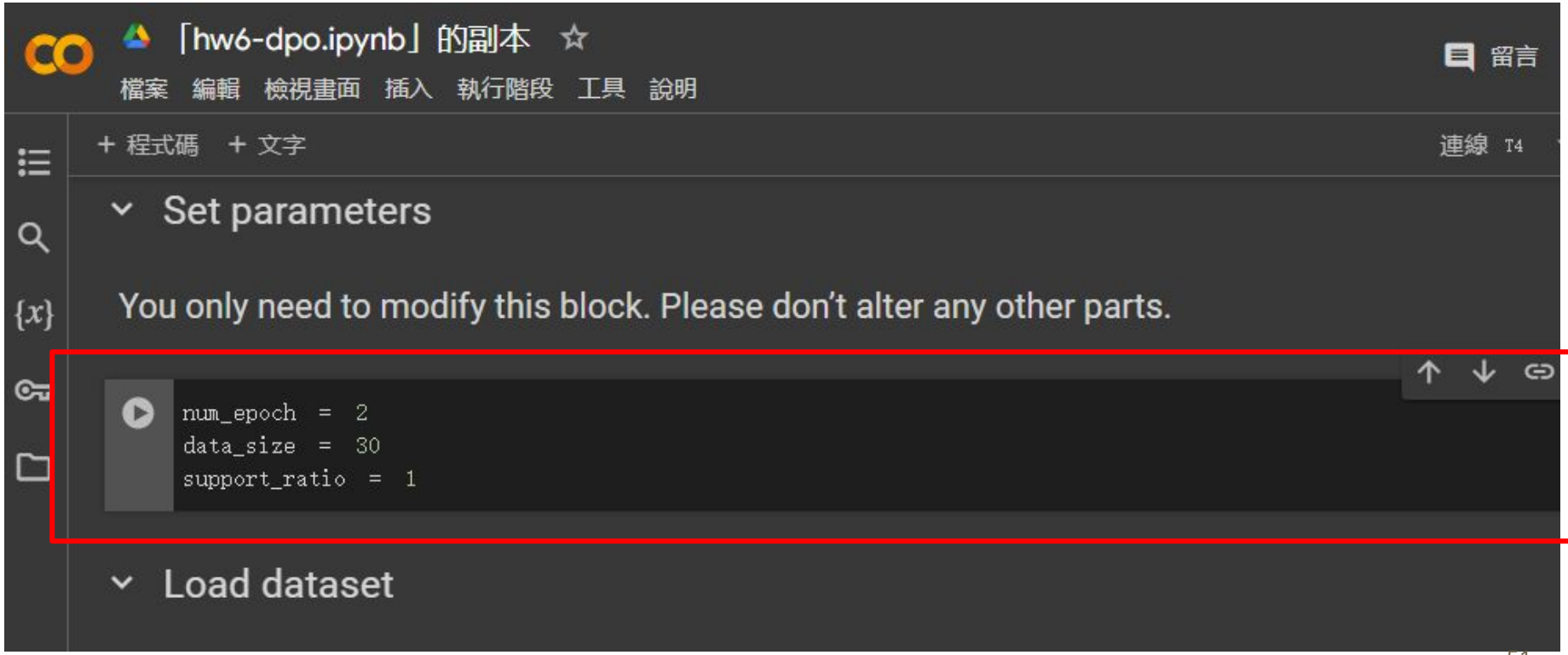
```
[ ] model_response = []
    print(f'num_epoch: {num_epoch}\ndata_size: {data_size}\nsupport_ratio: {support_ratio}')
    print()
    for data in test_data:
        id = data['id']
        ref_output = original_model_response[id-1]
        output = trained_model_response[id-1]
        print(f'Question {id}:\n'+data['prompt'])
        print('Response from original model:\n'+ref_output)
        print('Response from trained model:\n'+output)
        print()
        model_response.append({'id':data['id'], 'prompt':data['prompt'], 'response_from_original_model':ref_output, 'response_from
```

training hyperparameters

testing result

```
num_epoch: 1
data_size: 10
support_ratio: 0

Question 1:
真人化是否能改善日本漫畫的全球可及性？
Response from original model:
真人化可能會提高日本漫畫的全球可及性，因真人版電影或劇集可以吸引更多非漫畫讀者的注意，並提供不同的體驗。然而，這取決於真人化作品的品
Response from trained model:
真人化可能會提高日本漫畫的全球可及性，因真人版電影或劇集可以吸引更多非漫畫讀者的注意，並提供不同的體驗。然而，這取決於真人化作品的品

Question 2:
真人化如何影響年輕一代對日本漫畫的看法？
Response from original model:
```

Setting Training hyperparameters Again



Run Sample Code Again

Save LLM testing result Again

training hyperaramеters

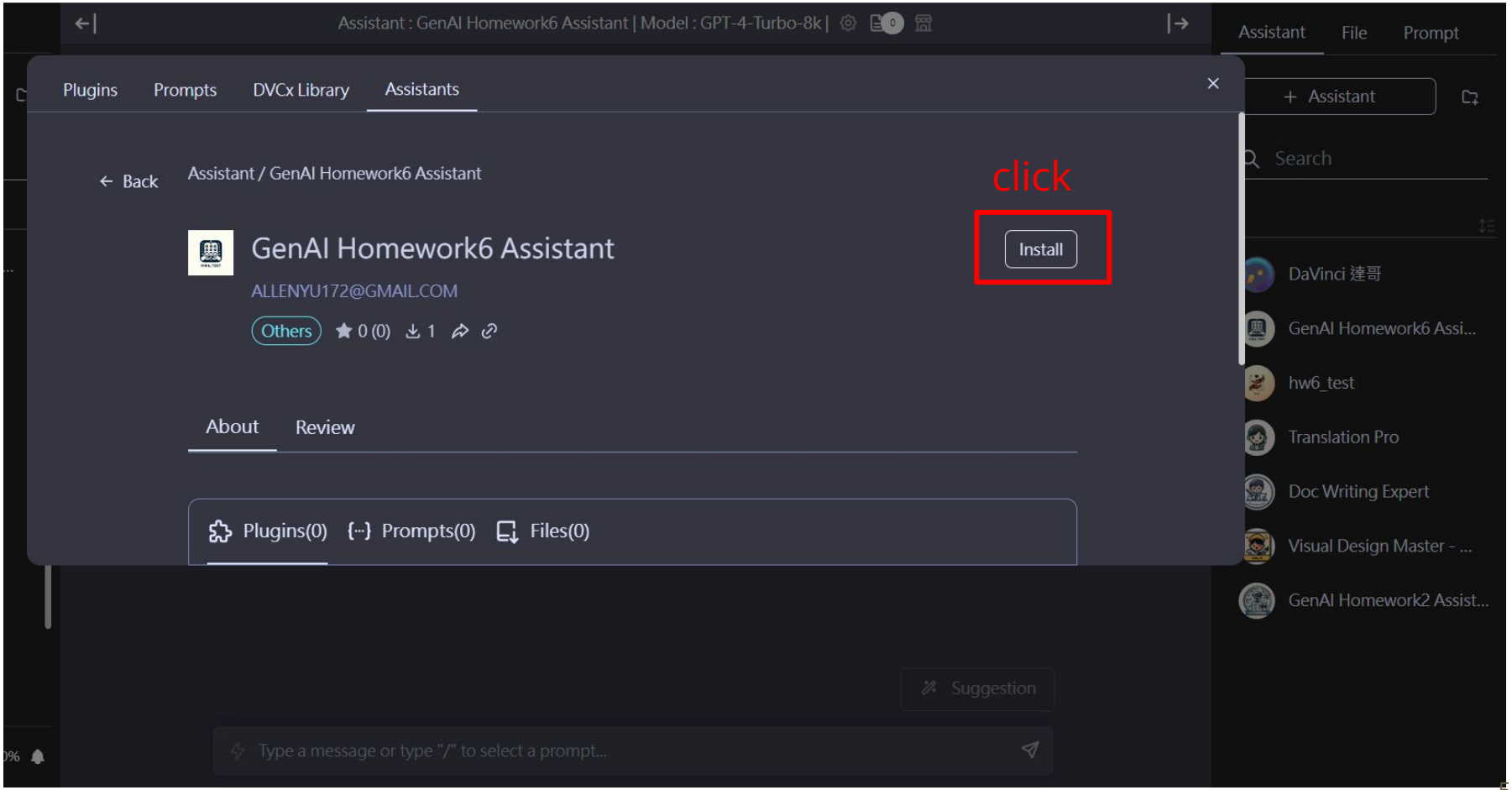
testing result

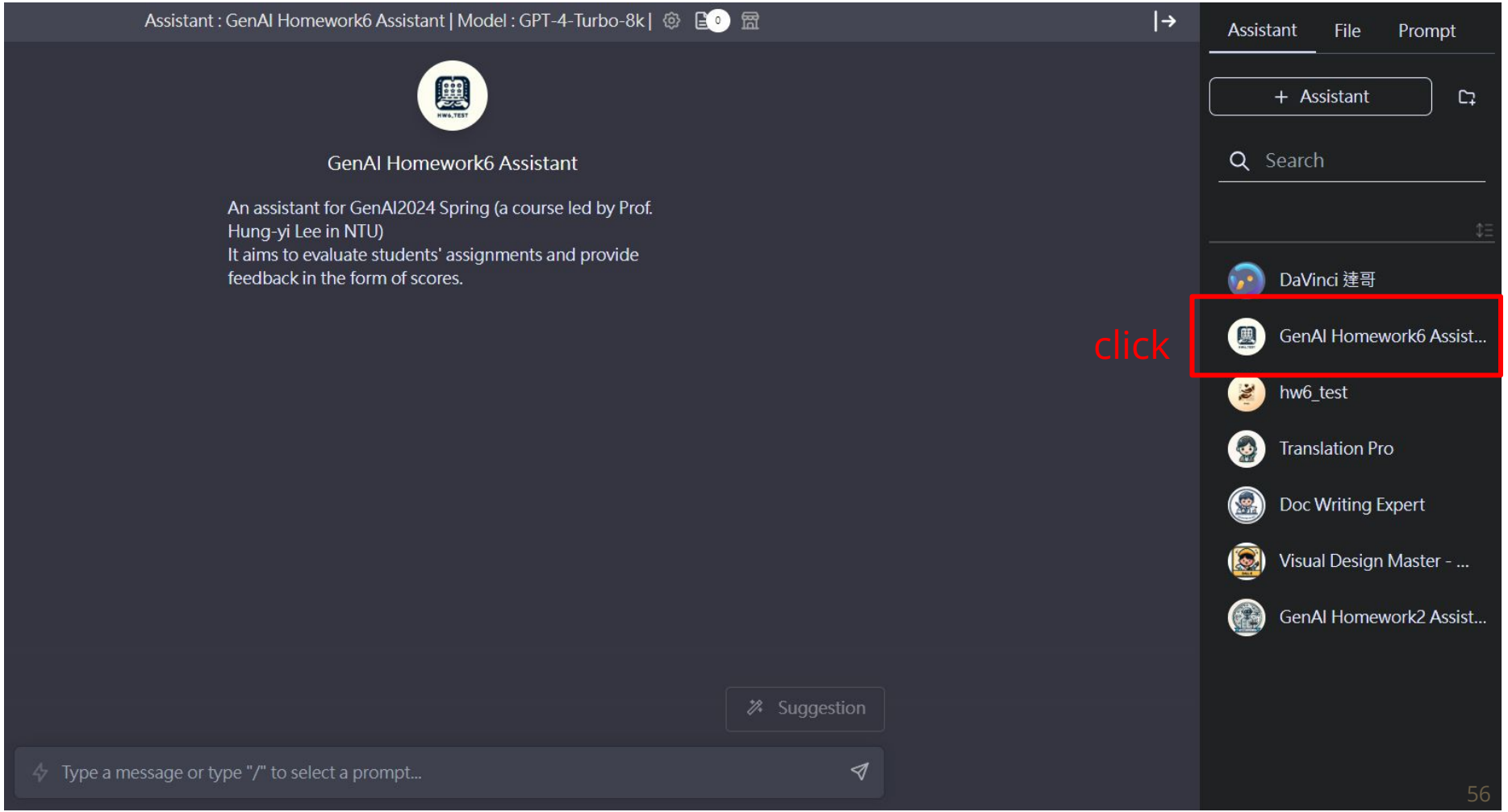
Please observe the output of this block to complete your report, and don't forget to take a screenshot of the results

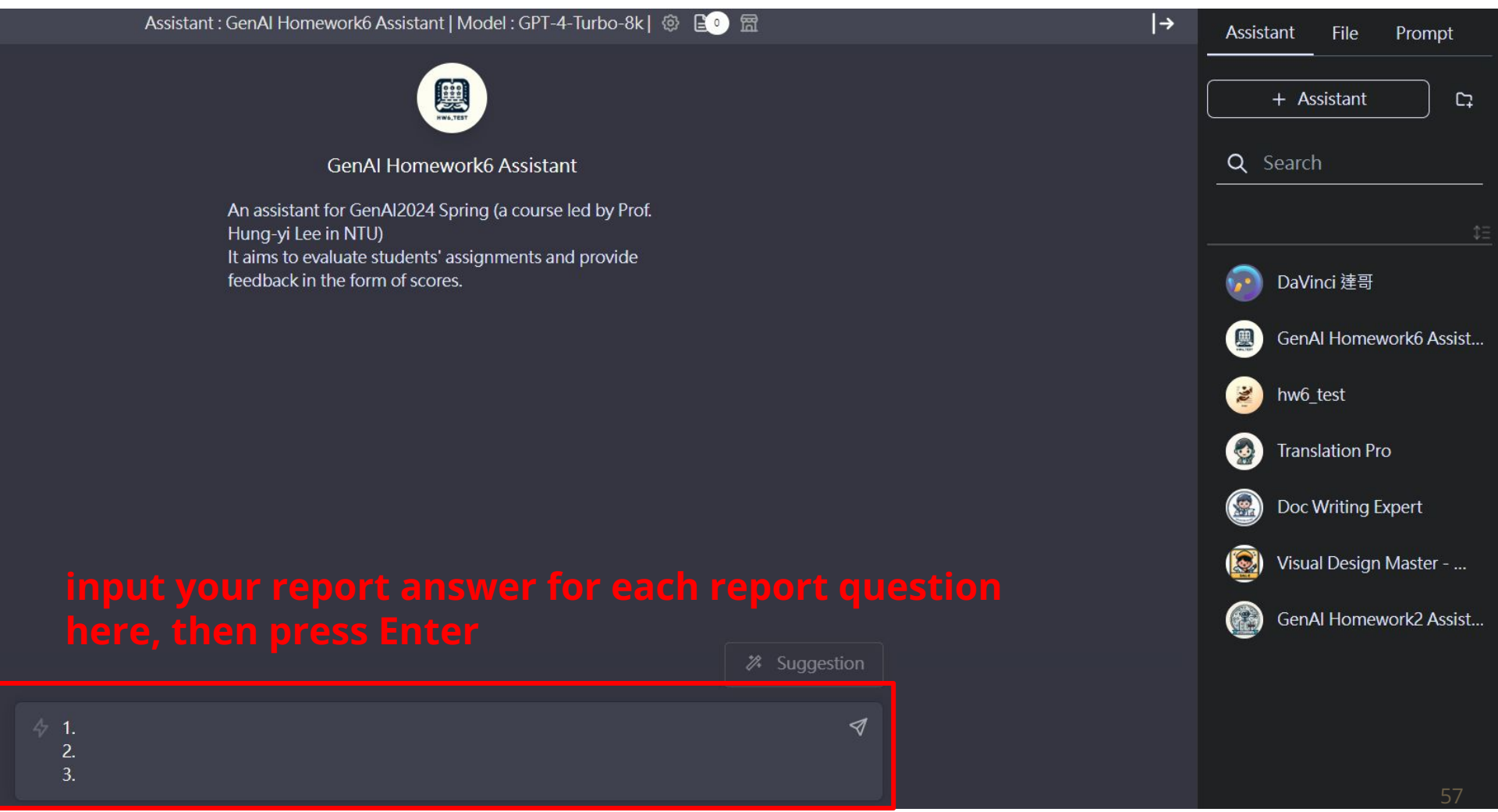
```
[12] model_response = []
     print(f'num_epoch: {num_epoch}\ndata_size: {data_size}\nsupport_ratio: {support_ratio}')
     print()
     for data in test_data:
         id = data['id']
         ref_output = original_model_response[id-1]
         output = trained_model_response[id-1]
         print(f'Question {id}:\n'+data['prompt'])
         print('Response from original model:\n'+ref_output)
         print('Response from trained model:\n'+output)
         print()
         model_response.append({'id':data['id'], 'prompt':data['prompt'], 'response_from_original_model':ref_output, 'response_from
```

```
num_epoch: 2
data_size: 30
support_ratio: 1

Question 1:
真人化是否能改善日本漫畫的全球可及性？
Response from original model:
真人化可能會提高日本漫畫的全球可及性，因真人版電影或劇集可以吸引更多非漫畫讀者的注意，並提供不同的體驗。然而，這取決於真人化作品的品
Response from trained model:
真人化可以提高日本漫畫的全球可及性，因它能吸引更多觀眾，拓展市場，並透過不同媒介（如電影、電視劇）讓更多人認識日本漫畫文化。

Question 2:
真人化如何影響年輕一代對日本漫畫的看法？
Response from original model:
```

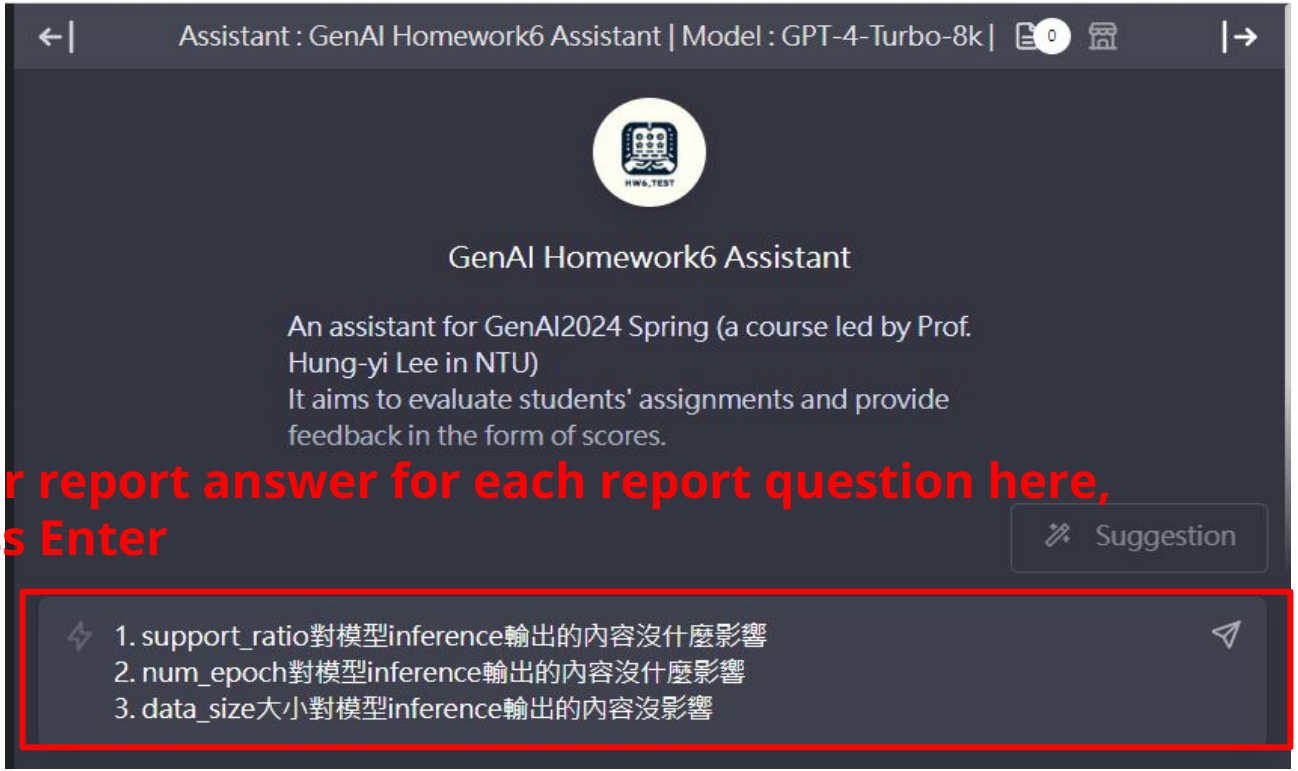
Grading Report Answer by DaVinci


Assistant : GenAI Homework6 Assistant | Model : GPT-4-Turbo-8k |
Assistant
File
Prompt
Plugins
Prompts
DVCx Library
Assistants
+ Assistant
Search
Back
Assistant / GenAI Homework6 Assistant
click
GenAI Homework6 Assistant
Install
ALLENYU172@GMAIL.COM
Others
0 (0)
1
DaVinci 達哥
GenAI Homework6 Assi...
hw6_test
Translation Pro
Doc Writing Expert
Visual Design Master - ...
GenAI Homework2 Assist...
About
Review
Plugins(0)
Prompts(0)
Files(0)
Suggestion
Type a message or type "/" to select a prompt...


Assistant : GenAI Homework6 Assistant | Model : GPT-4-Turbo-8k |
GenAI Homework6 Assistant
An assistant for GenAI2024 Spring (a course led by Prof. Hung-yi Lee in NTU)
It aims to evaluate students' assignments and provide feedback in the form of scores.
Suggestion
Type a message or type "/" to select a prompt...
Assistant
File
Prompt
+ Assistant
Search
DaVinci 達哥
GenAI Homework6 Assist...
hw6_test
Translation Pro
Doc Writing Expert
Visual Design Master - ...
GenAI Homework2 Assist...
click

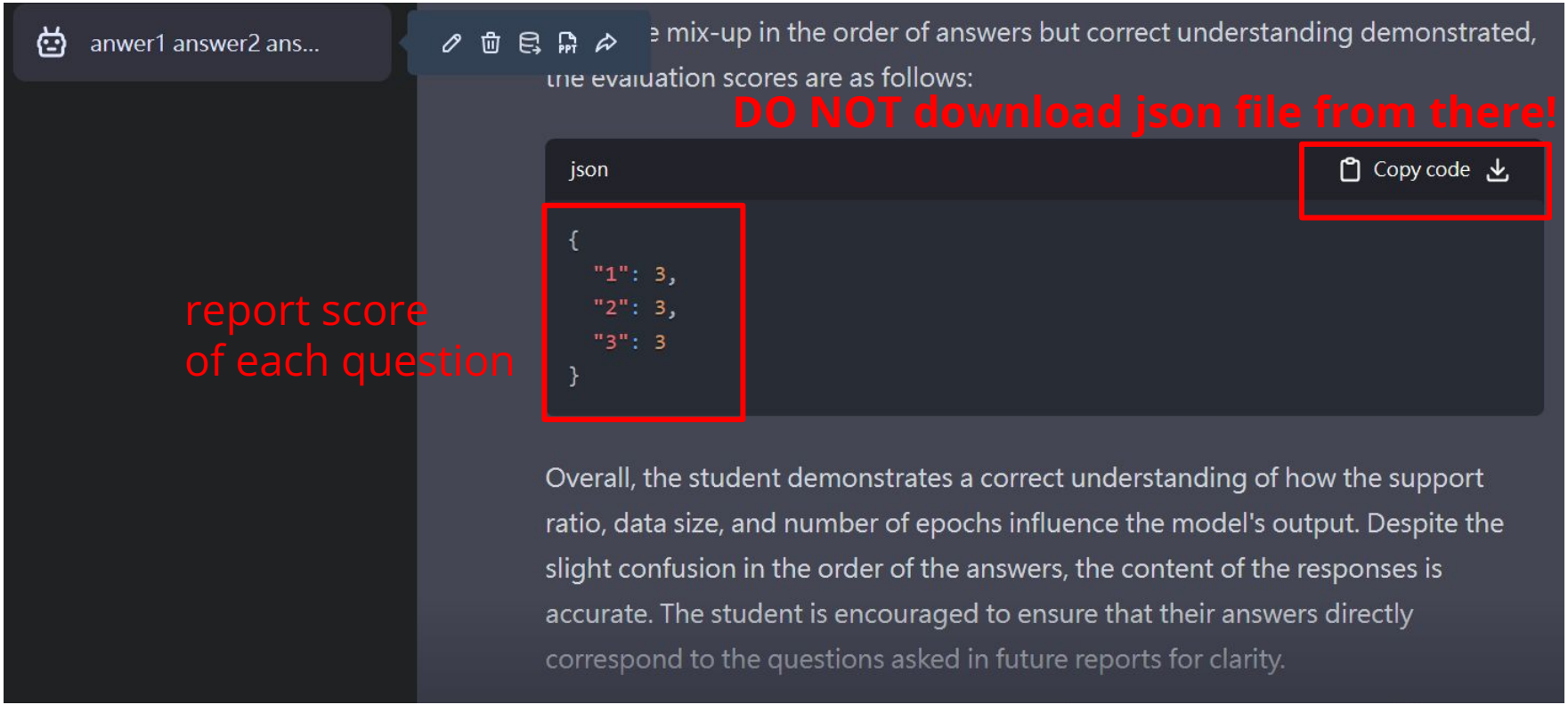

Assistant : GenAI Homework6 Assistant | Model : GPT-4-Turbo-8k |
GenAI Homework6 Assistant
An assistant for GenAI2024 Spring (a course led by Prof. Hung-yi Lee in NTU)
It aims to evaluate students' assignments and provide feedback in the form of scores.
input your report answer for each report question here, then press Enter
Suggestion
1.
2.
3.
Assistant
File
Prompt
+ Assistant
Search
DaVinci 達哥
GenAI Homework6 Assist...
hw6_test
Translation Pro
Doc Writing Expert
Visual Design Master - ...
GenAI Homework2 Assist...
57

For example

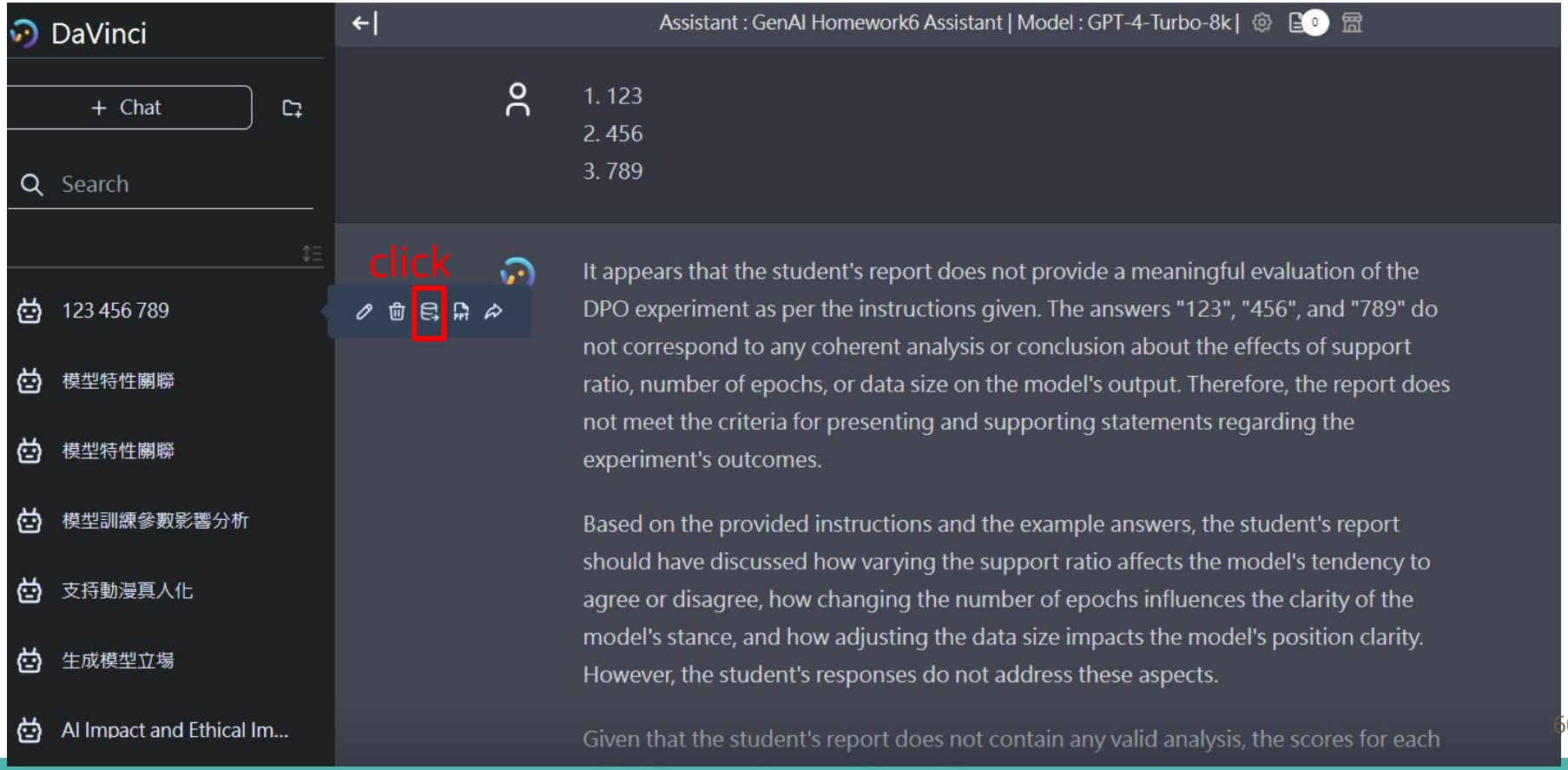


input your report answer for each report question here, then press Enter

Check your Score

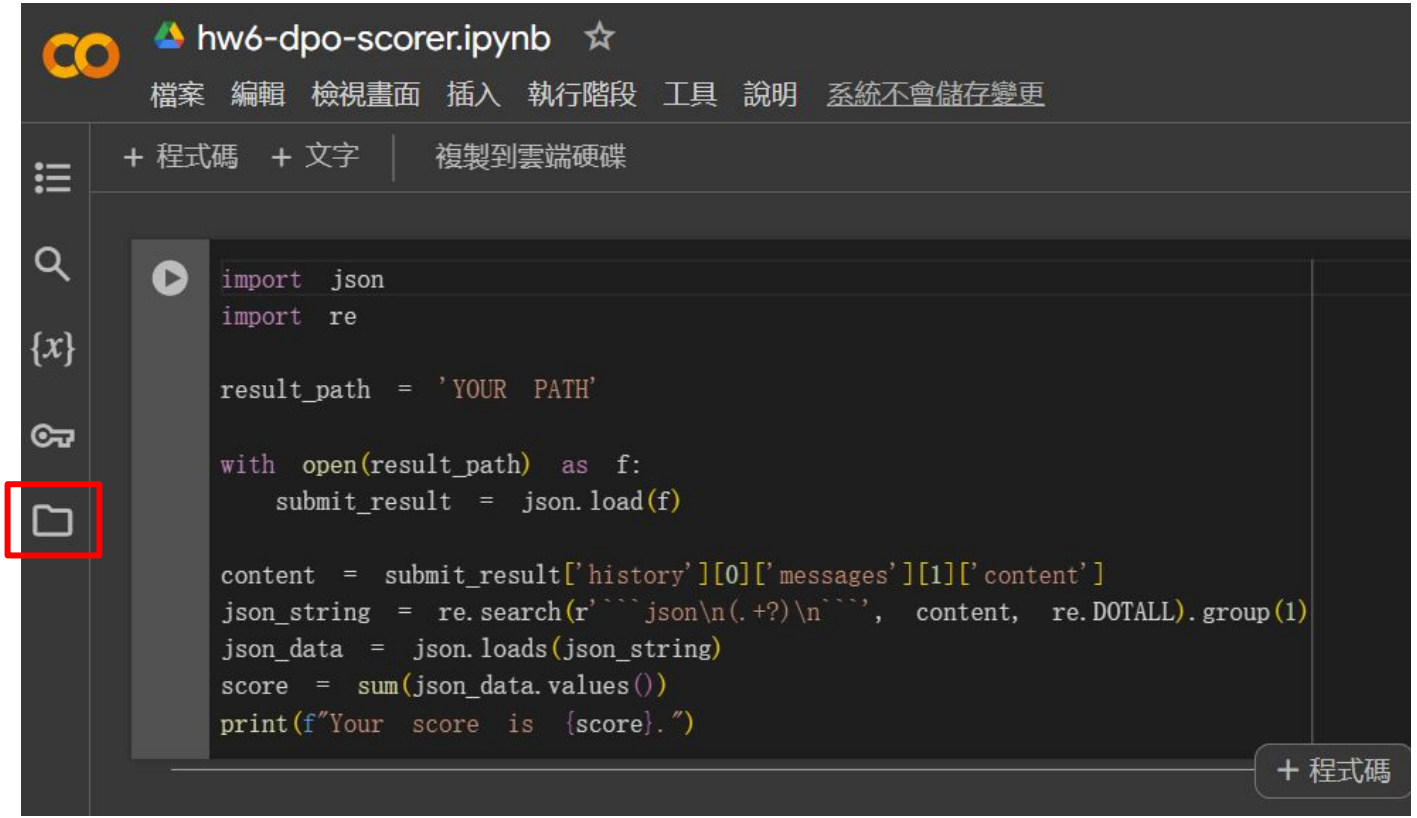


Download Report Grading JSON



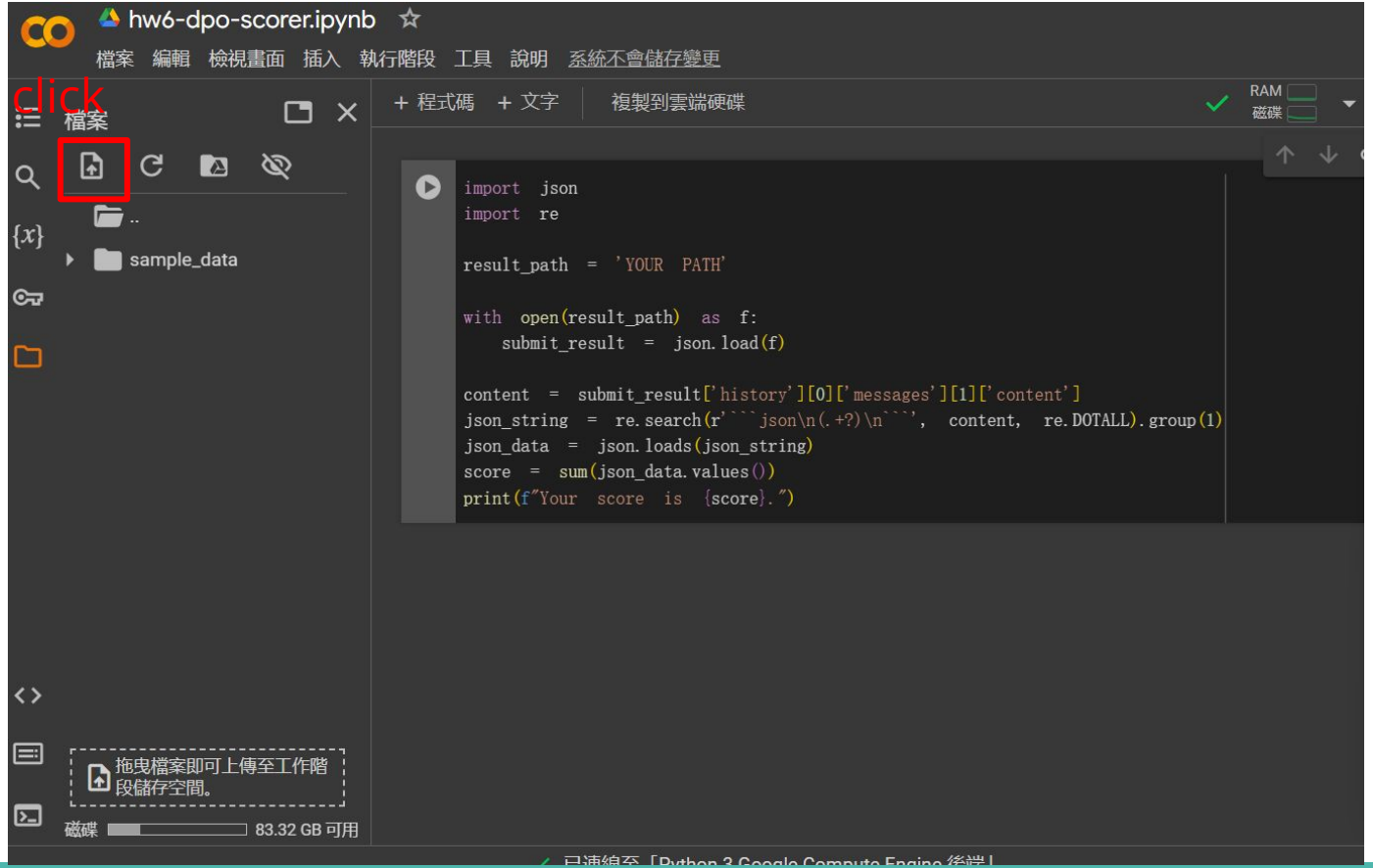
Check Report Score by Report Grader

After open Report Grader

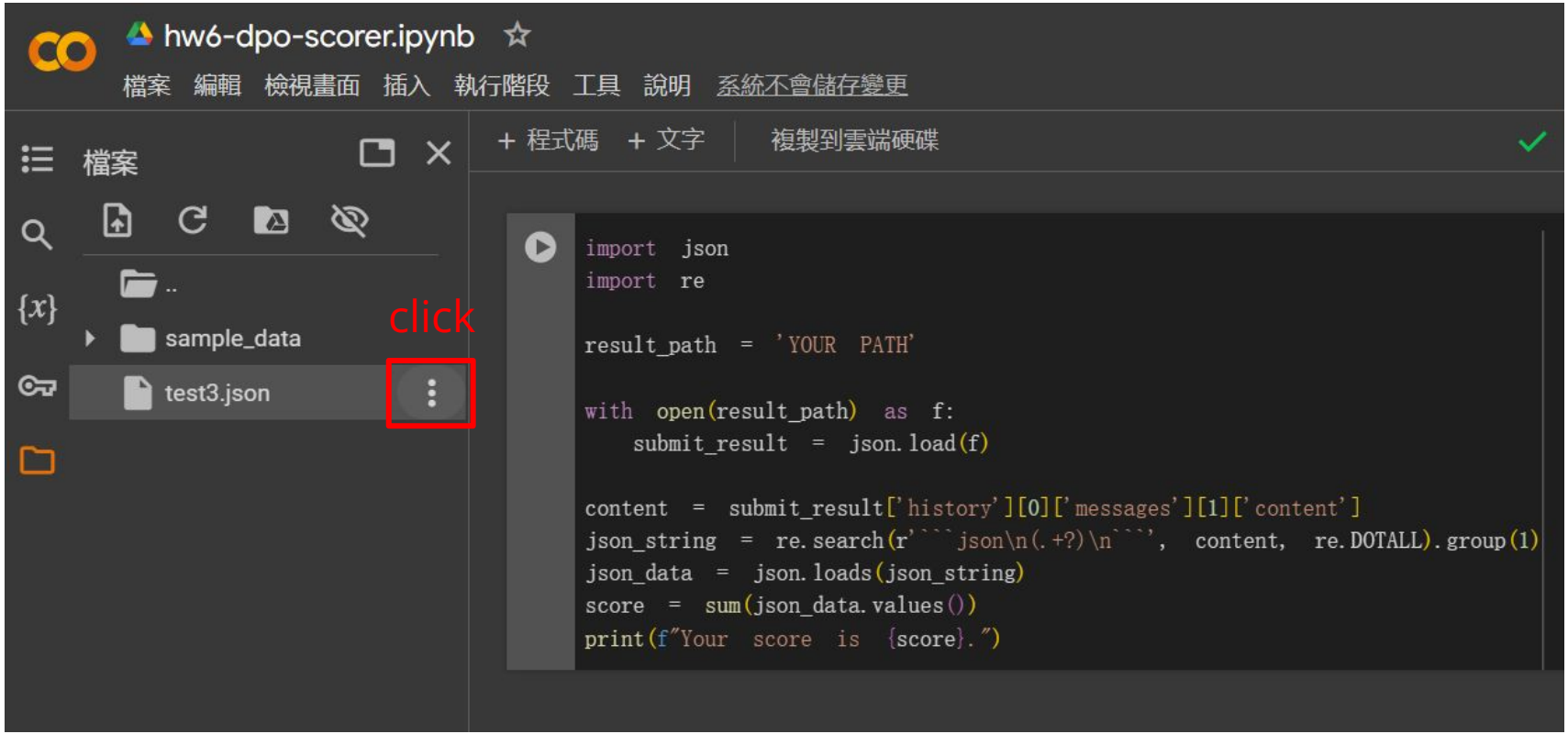


click

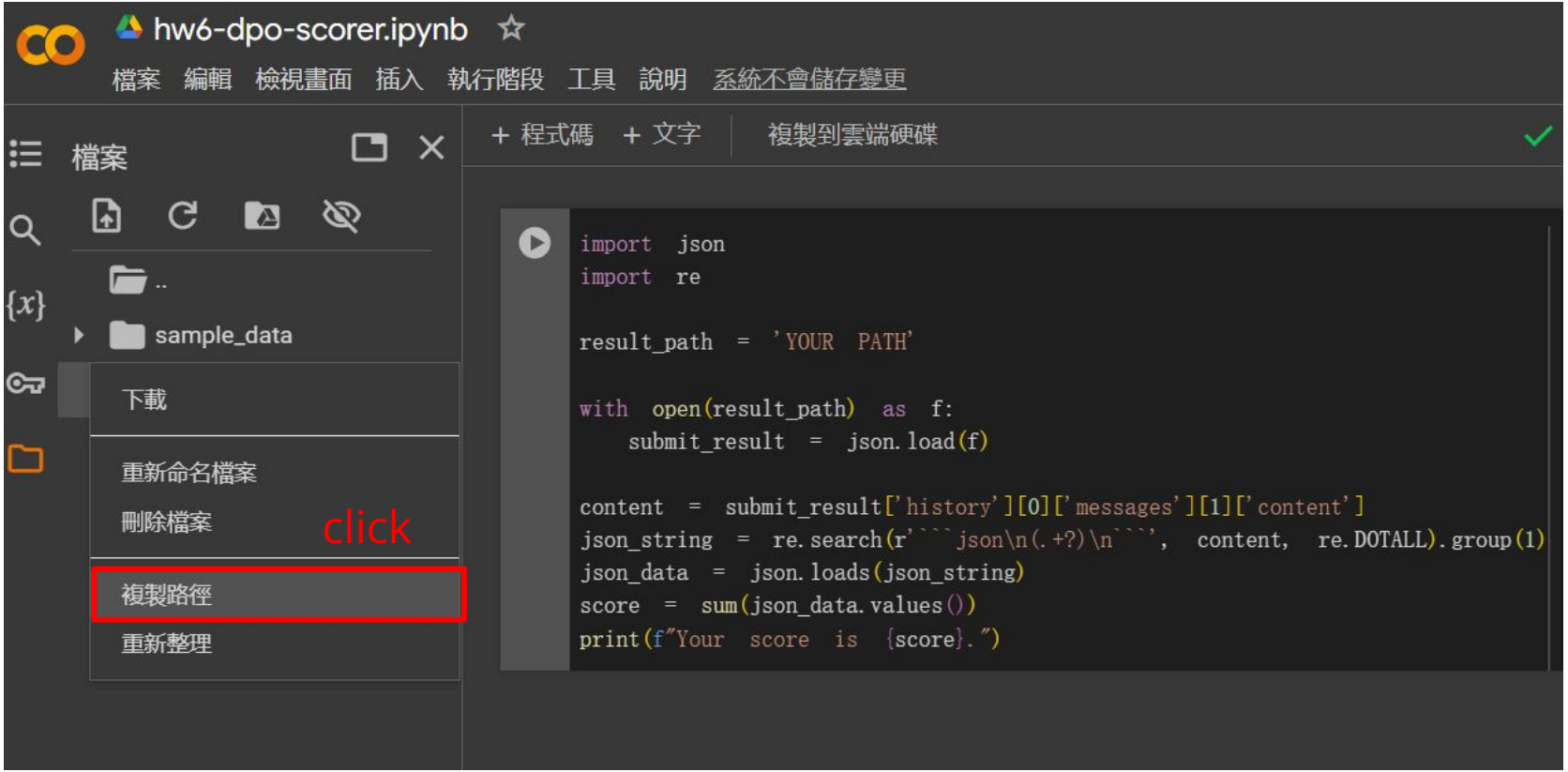
Upload your json file by clicking upload icon



Get your json file path



Get your json file path



Modify YOUR PATH to your json file's path

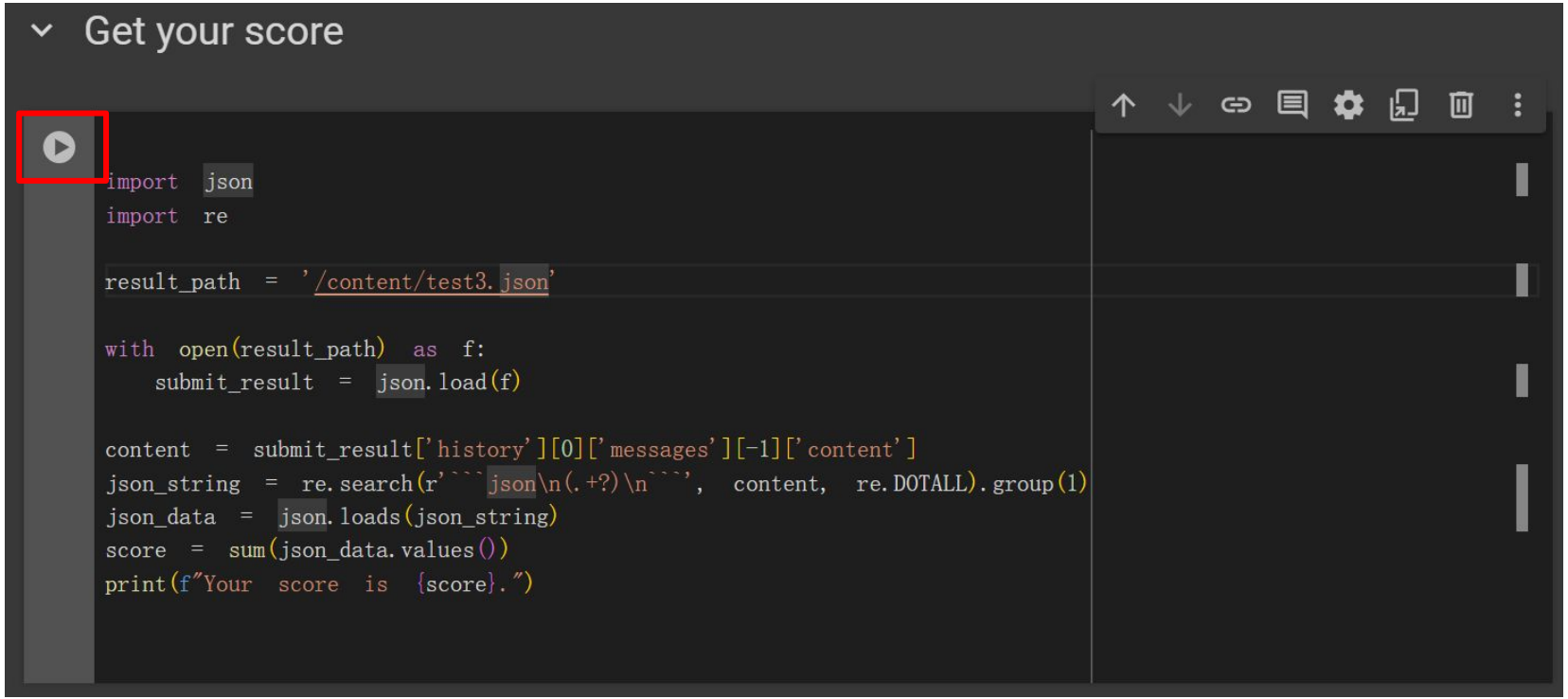
Get your score

```
import json
import re

result_path = 'YOUR PATH'

with open(result_path) as f:
    submit_result = json.load(f)

content = submit_result['history'][0]['messages'][-1]['content']
json_string = re.search(r'```json\n(.+?)\n```', content, re.DOTALL).group(1)
json_data = json.loads(json_string)
score = sum(json_data.values())
print(f"Your score is {score}.")
```

Ctrl + v to paste your json file path in the quote mark

Run the following block

click



Check your total score of all report questions

Get your score

0秒

```
import json
import re

result_path = '/content/test3.json'

with open(result_path) as f:
    submit_result = json.load(f)

content = submit_result['history'][0]['messages'][-1]['content']
json_string = re.search(r'```json\n(.+?)\n```', content, re.DOTALL).group(1)
json_data = json.loads(json_string)
score = sum(json_data.values())
print(f"Your score is {score}.")
```

```
Your score is 6.
```

Reference

Huggingface introduction to RLHF: https://huggingface.co/blog/rlhf
https://huggingface.co/blog/trl-peft
Direct Preference Optimization(DPO): https://arxiv.org/abs/2305.18290