
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: 2024/05/02 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](#)

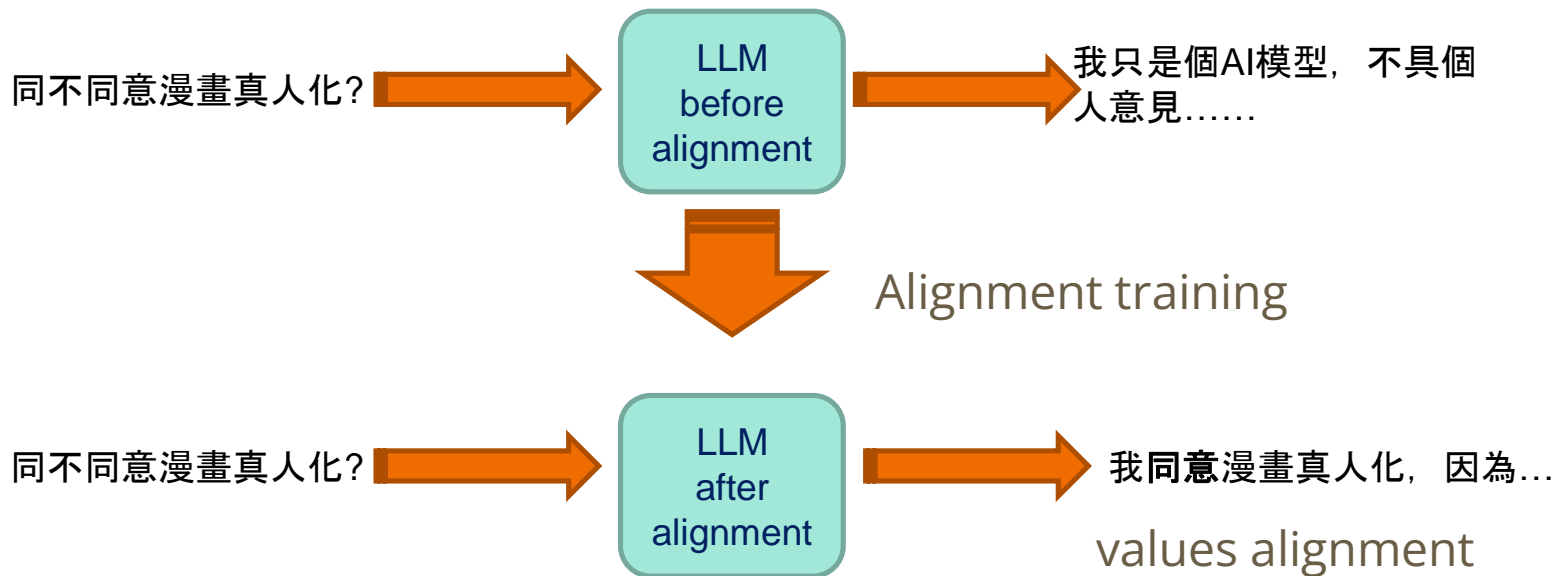
[DaVinci](#)

[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.

Supervised Fine-Tuning

Input: 地球到太陽的距離?



Answer: 約1.496億公里

RLHF

Input: 在公車上讓座給有需要的人?



Answer: ???

Reinforcement Learning from Human Feedback (RLHF)

Standard steps in to RLHF (briefly explain) :

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

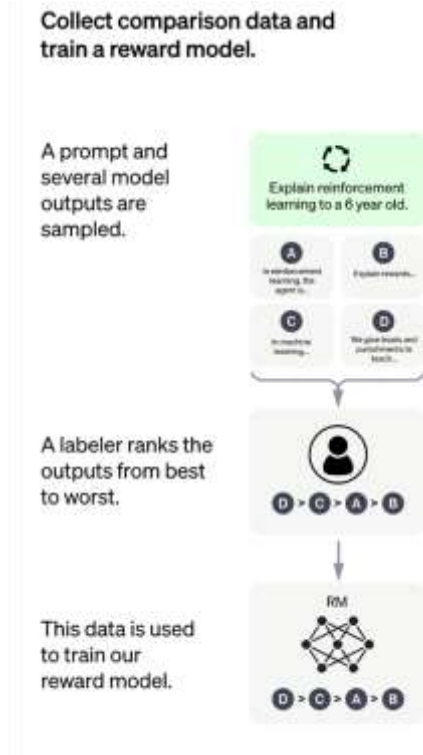
Reinforcement Learning from Human Feedback (RLHF)

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 ?

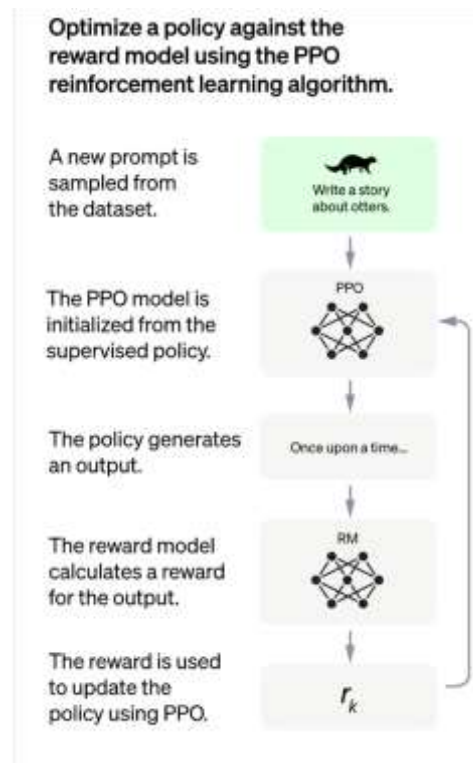
- Evaluate the answer output by LLM



[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



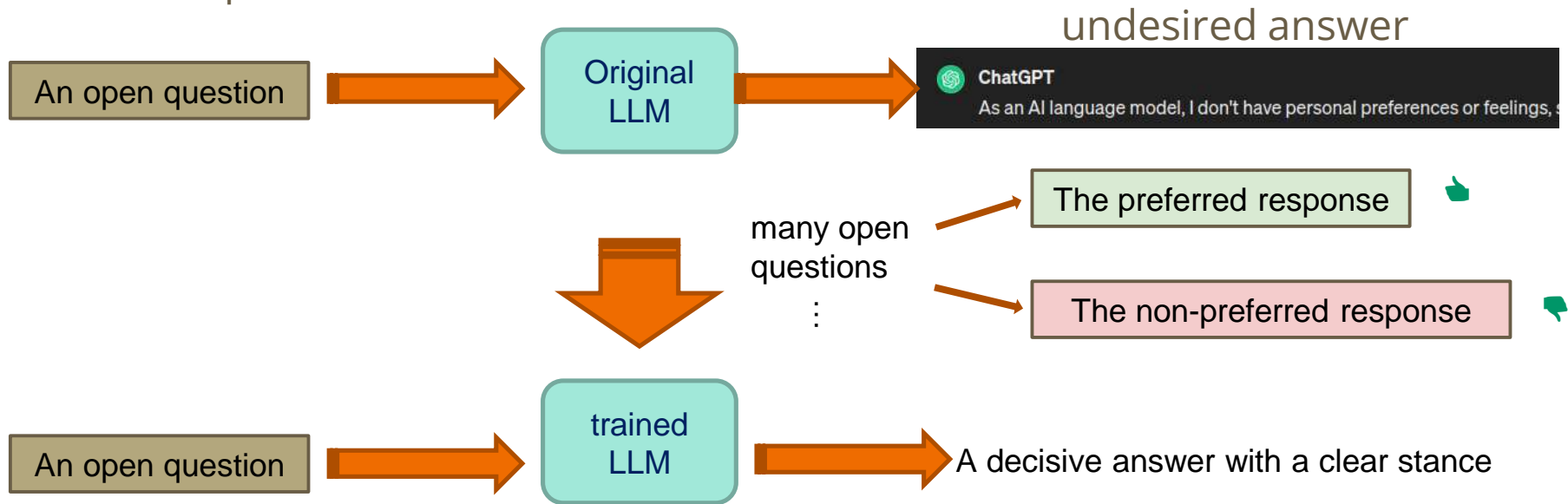
Drawbacks of Standard RLHF

- 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

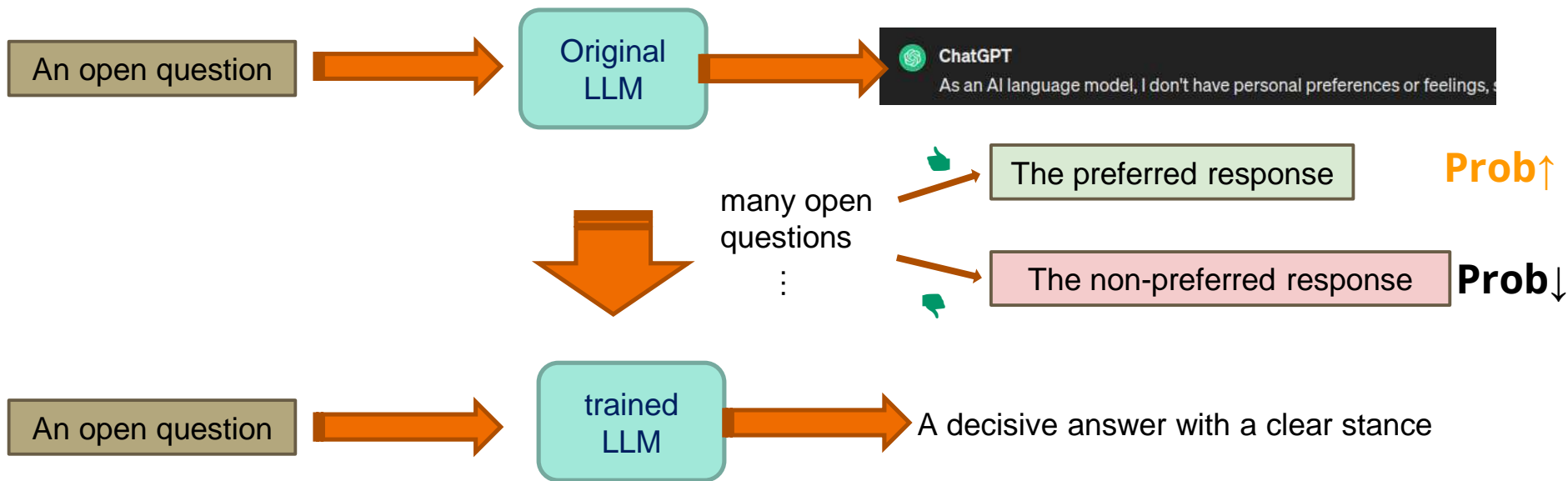
DPO - Direct Preference Optimization

- 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



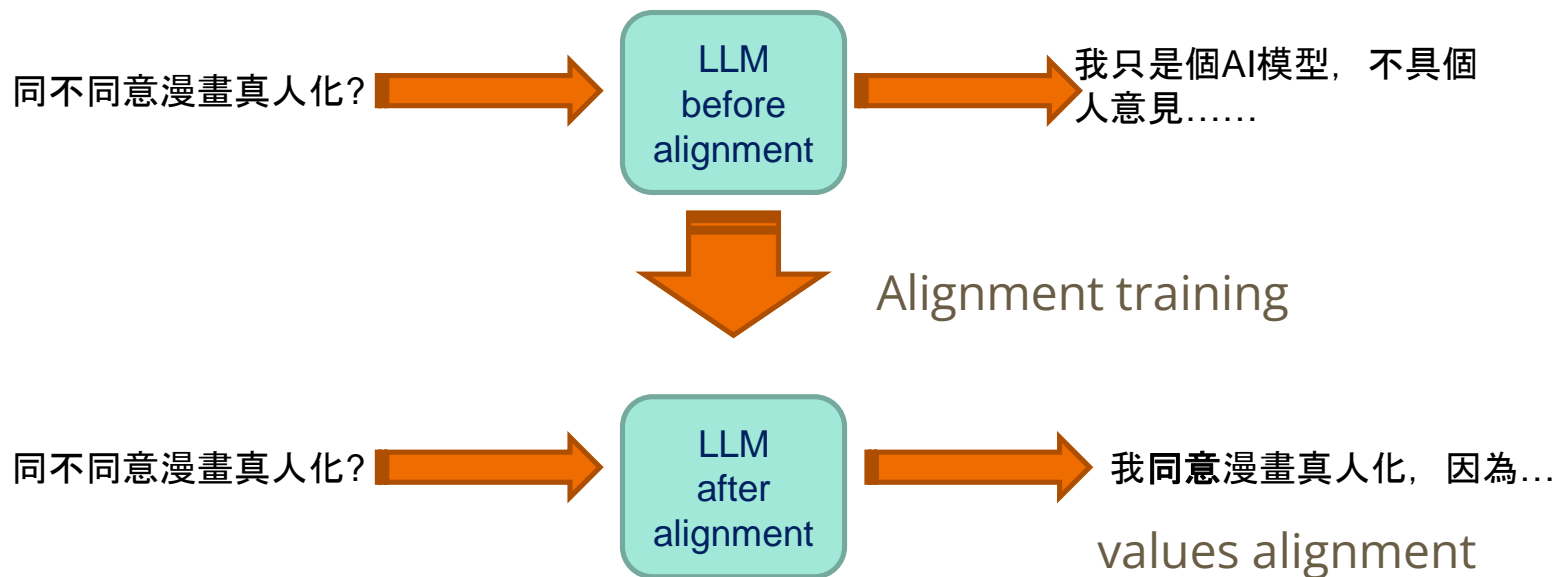
DPO - Direct Preference Optimization

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



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

```
▼ 0 : { 2 items
  "id" : int 1
  "prompt" : string "真人化是否能改善日本漫畫的全球可及性？"
}
▼ 1 : { 2 items
  "id" : int 2
  "prompt" : string "真人化如何影響年輕一代對日本漫畫的看法？"
}
▼ 2 : { 2 items
  "id" : int 3
  "prompt" : string "真人化是否能提升原作漫畫的文學價值？"
```

Model and Dataset

Dataset : generated by ChatGPT Website

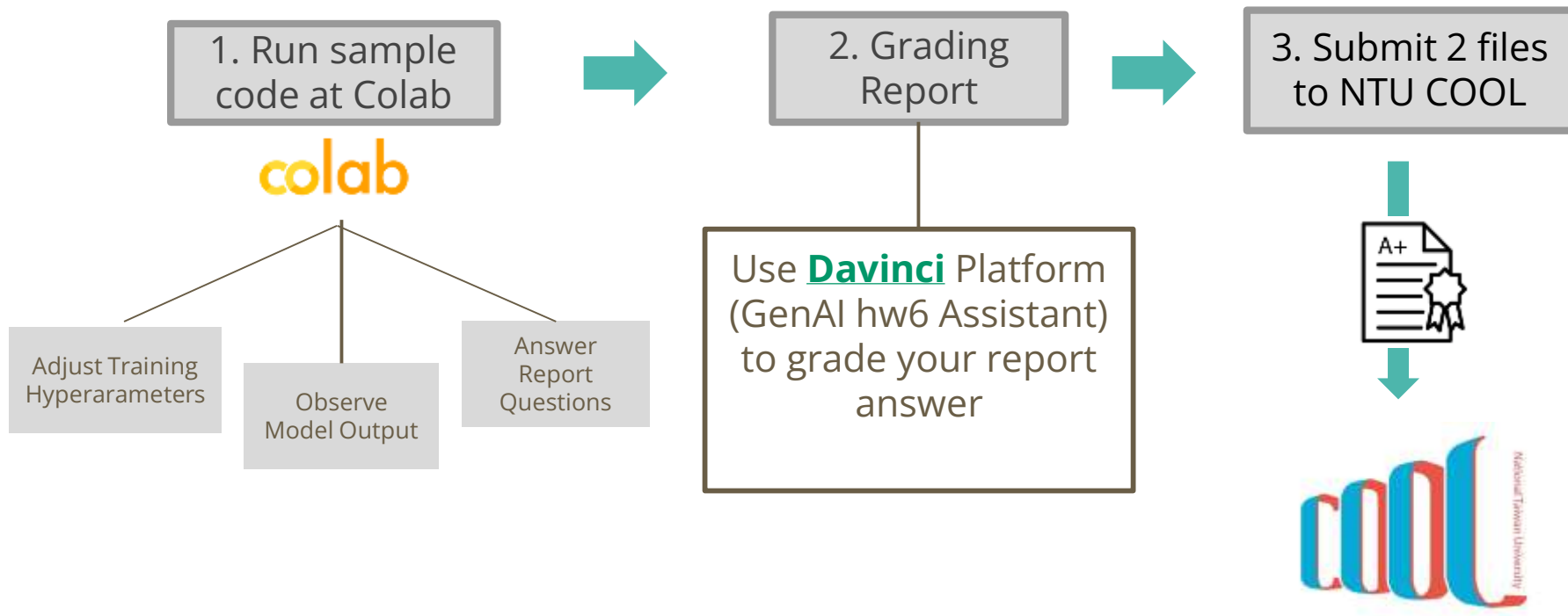
Model : Breeze-7b (聯發科)

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



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

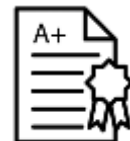
Observe
Model Output

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

▼ Set parameters

You only need to modify this block. Please don't alter any other parts.

```
▶ num_epoch = 2  
  data_size = 30  
  support_ratio = 1
```


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.

```

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。"
}
2 : { 4 items
  "id" : int 3
  "prompt" : string "真人化是否能夠吸引新觀眾？"
  "support" : string "真人化能夠吸引不熟悉動漫的觀眾，擴大受眾。"
  "oppose" : string "真人化可能讓原本的動漫迷感到失望，無法吸引新觀眾。"
}
3 : { 4 items
  "id" : int 4
  "prompt" : string "真人化是否能夠保留原作故事情節的精髓？"
  "support" : string "真人化有機會更深入挖掘原作故事，保留精髓。"
  "oppose" : string "真人化可能因為改編而失去原作故事的深度與精髓。"
}

```

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

```
▼ 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。"
}

⋮

▼ 49 : { 4 items
  "id" : int 50
  "prompt" : string "真人化是否有助於增進原創動漫的社會認知度？"
  "support" : string "真人化能使原創動漫更容易獲得主流社會的認可和關注。"
  "oppose" : string "真人化可能會將動漫文化簡化，降低其在社會中的地位和認知度。"
}
```

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

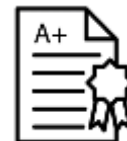
Observe
Model Output

Answer
Report
Questions

2. Grading
Report

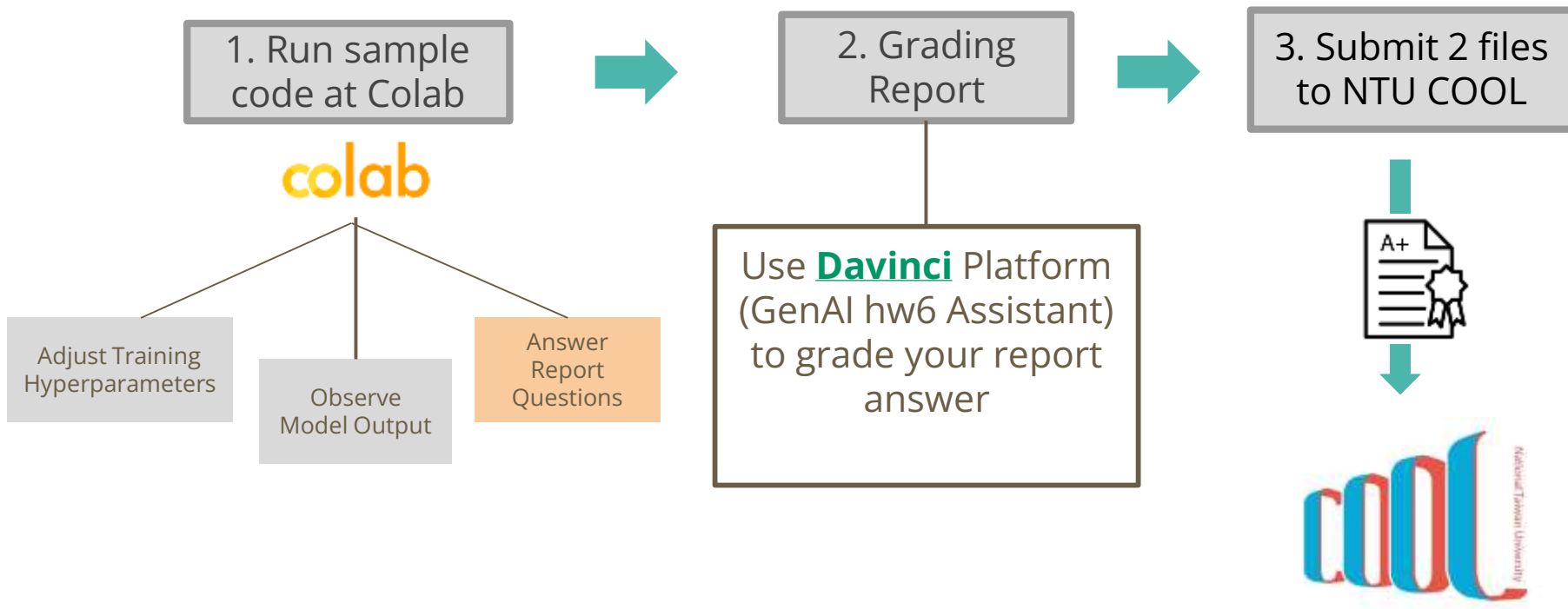
Use **Davinci** Platform
(GenAI hw6 Assistant)
to grade your report
answer

3. Submit 2 files
to NTU COOL



Colab DEMO

TODO Workflow



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**

1. (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

1. (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

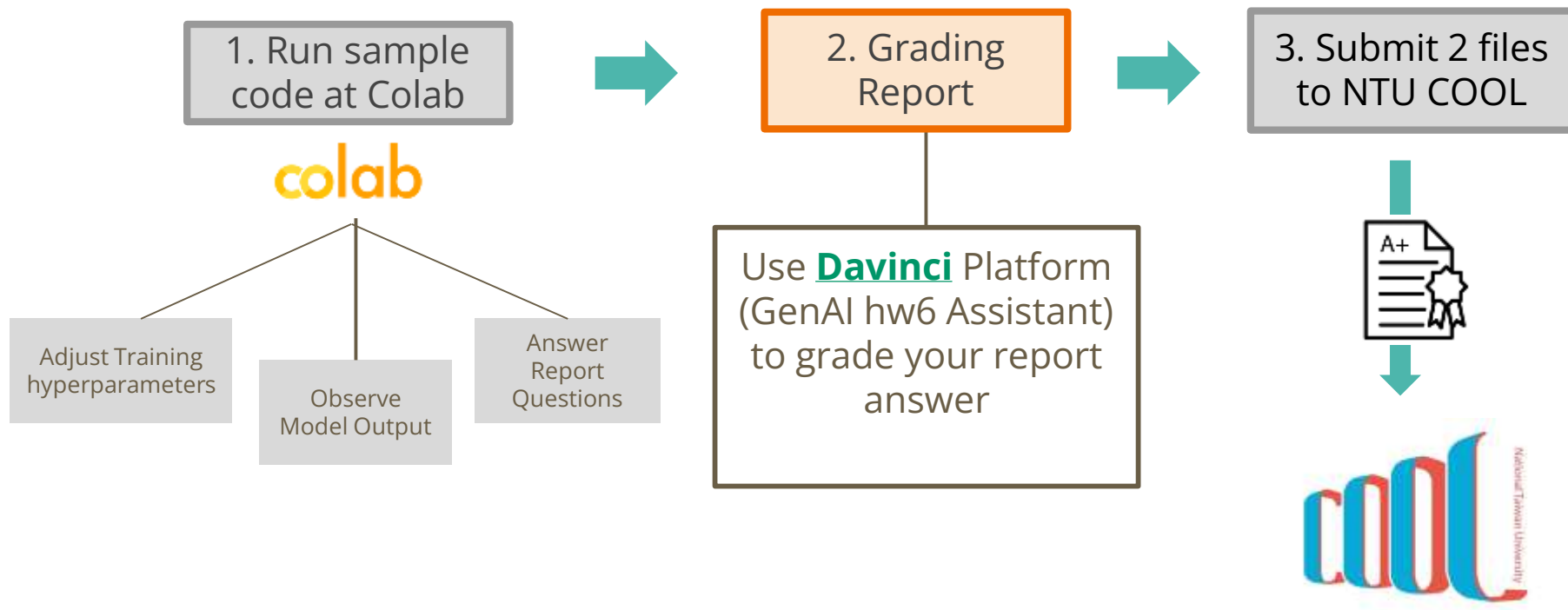
a. num_epoch = 3 **data_size = 10** support_ratio = 1

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

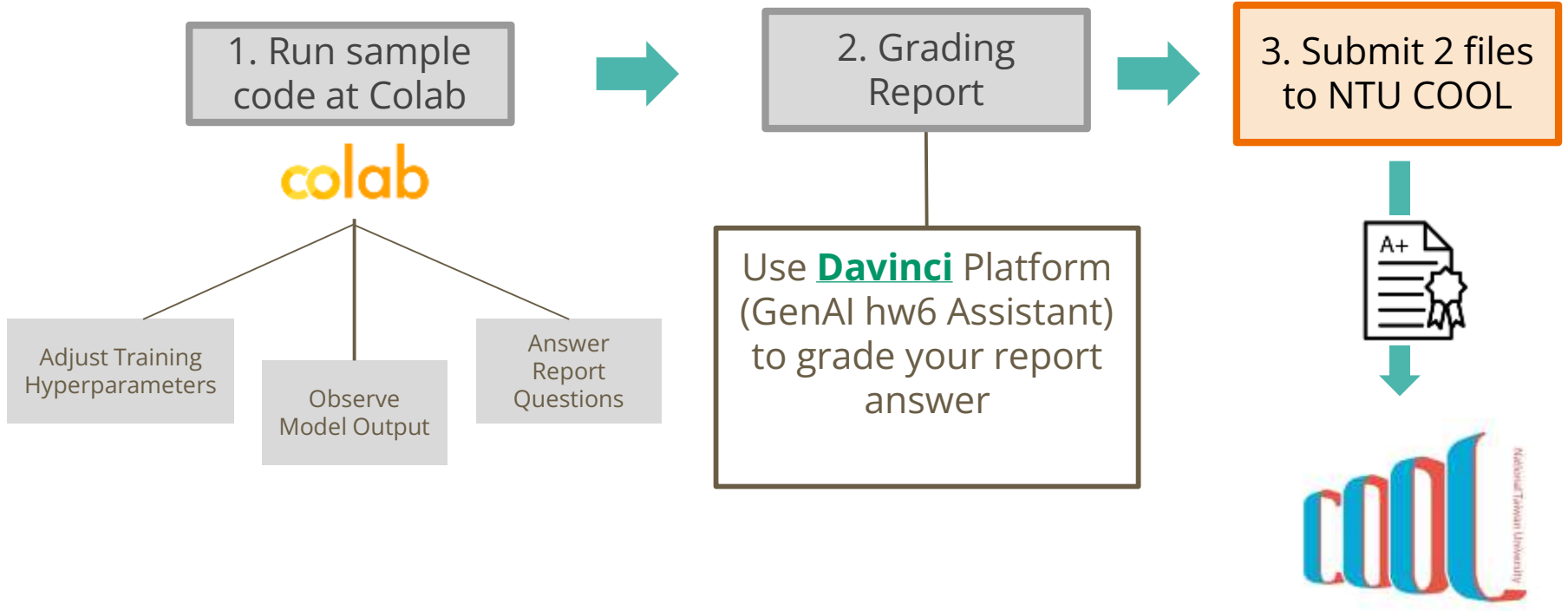
TODO Workflow



DaVinci DEMO

Submission and Grading

TODO Workflow

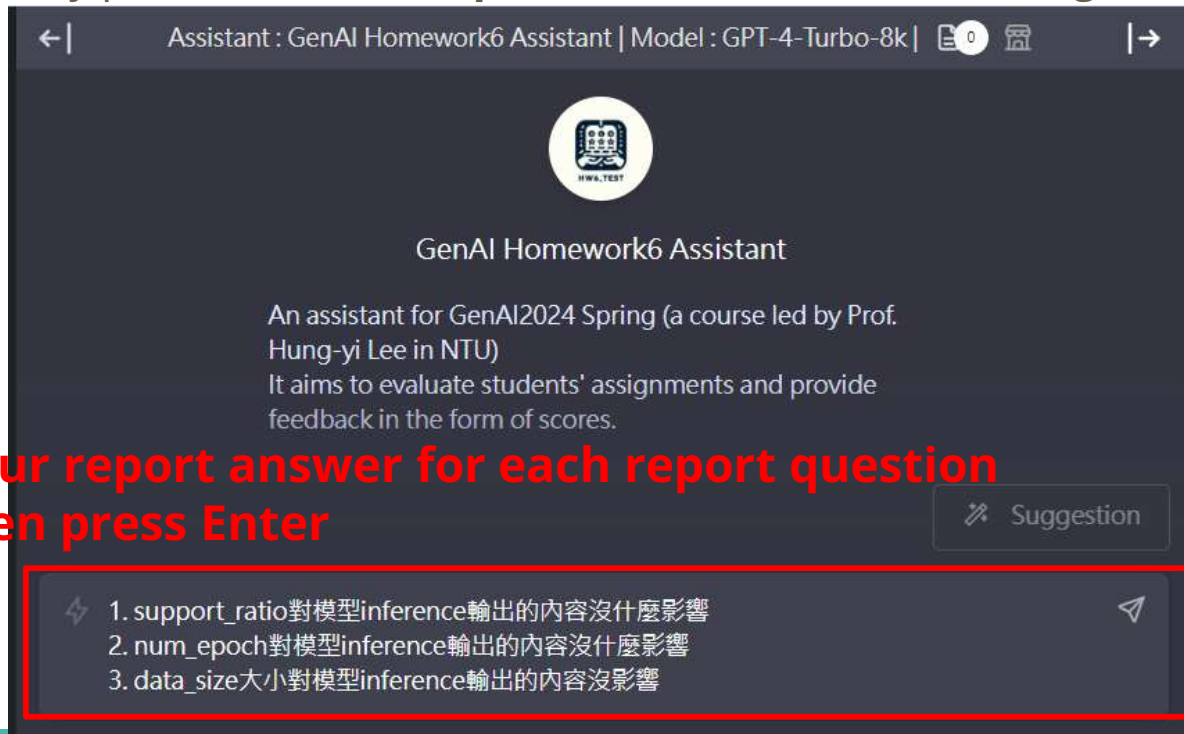


Submission

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.

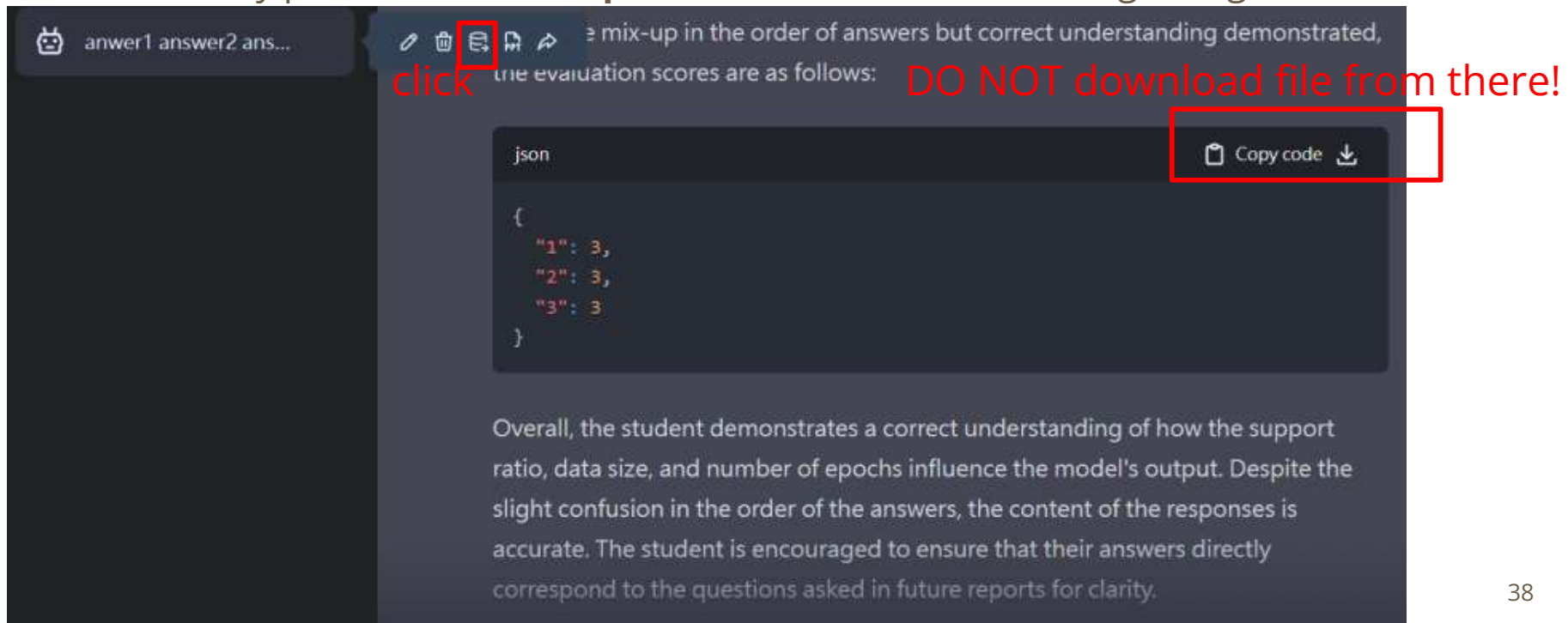
input your report answer for each report question here, then press Enter



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.



The screenshot shows the DaVinci AI interface. At the top, there's a chat header with a robot icon and the text "answer1 answer2 ans...". Below this, a toolbar contains icons for edit, delete, and a file icon (highlighted with a red box). To the right of the toolbar, text reads "mix-up in the order of answers but correct understanding demonstrated, the evaluation scores are as follows:". Below this, a JSON response is displayed in a dark-themed code editor. The JSON is:

```
{  "1": 3,  "2": 3,  "3": 3}
```

. To the right of the JSON, there's a button labeled "Copy code" with a download icon (highlighted with a red box). Overlaid on the right side of the screenshot is a large red text warning: "DO NOT download file from there!". Below the JSON, there's a paragraph of text: "Overall, the student demonstrates a correct understanding of how the support ratio, data size, and number of epochs influence the model's output. Despite the slight confusion in the order of the answers, the content of the responses is accurate. The student is encouraged to ensure that their answers directly correspond to the questions asked in future reports for clarity."

click

DO NOT download file from there!

```
json
```

```
{  "1": 3,  "2": 3,  "3": 3}
```

Copy code

Overall, the student demonstrates a correct understanding of how the support ratio, data size, and number of epochs influence the model's output. Despite the slight confusion in the order of the answers, the content of the responses is accurate. The student is encouraged to ensure that their answers directly correspond to the questions asked in future reports for clarity.

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)



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

testing result from Colab

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質。

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質。

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:

真人化可能會影響年輕一代對日本漫畫的看法，使他們更容易接受和理解故事和角色，並吸引更多人關注和支持日本漫畫文化。然而，個人喜好和文化背景也會影響看法。

Response from trained model:

真人化可能會影響年輕一代對日本漫畫的看法，使他們更容易接受和理解故事和角色，並吸引更多人關注和支持日本漫畫文化。然而，個人喜好和文化背景也會影響看法。

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: 2024/05/02 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 2024/05/10 23:59:59 (UTC+8)

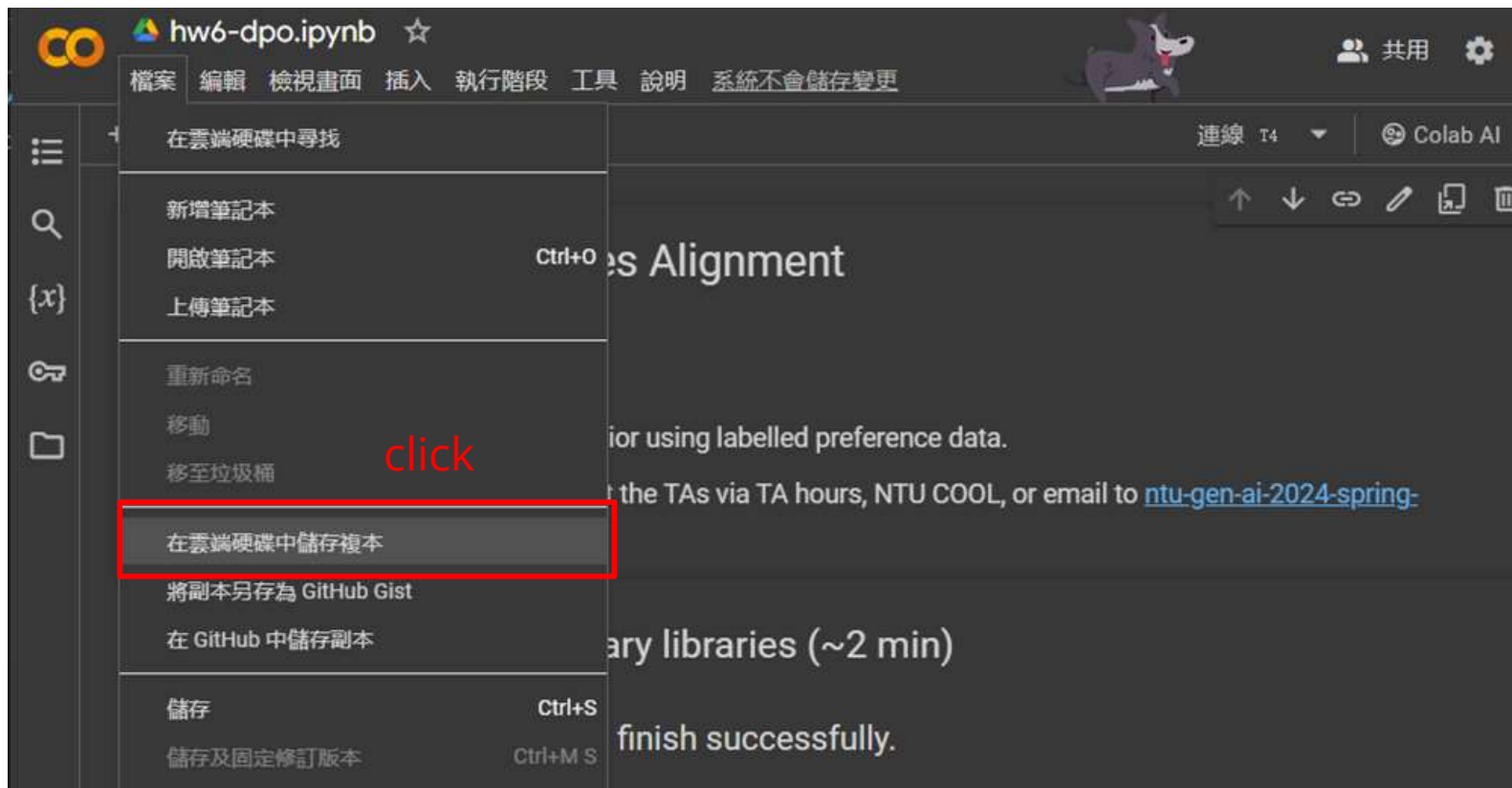
If You Have Any Questions

- 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

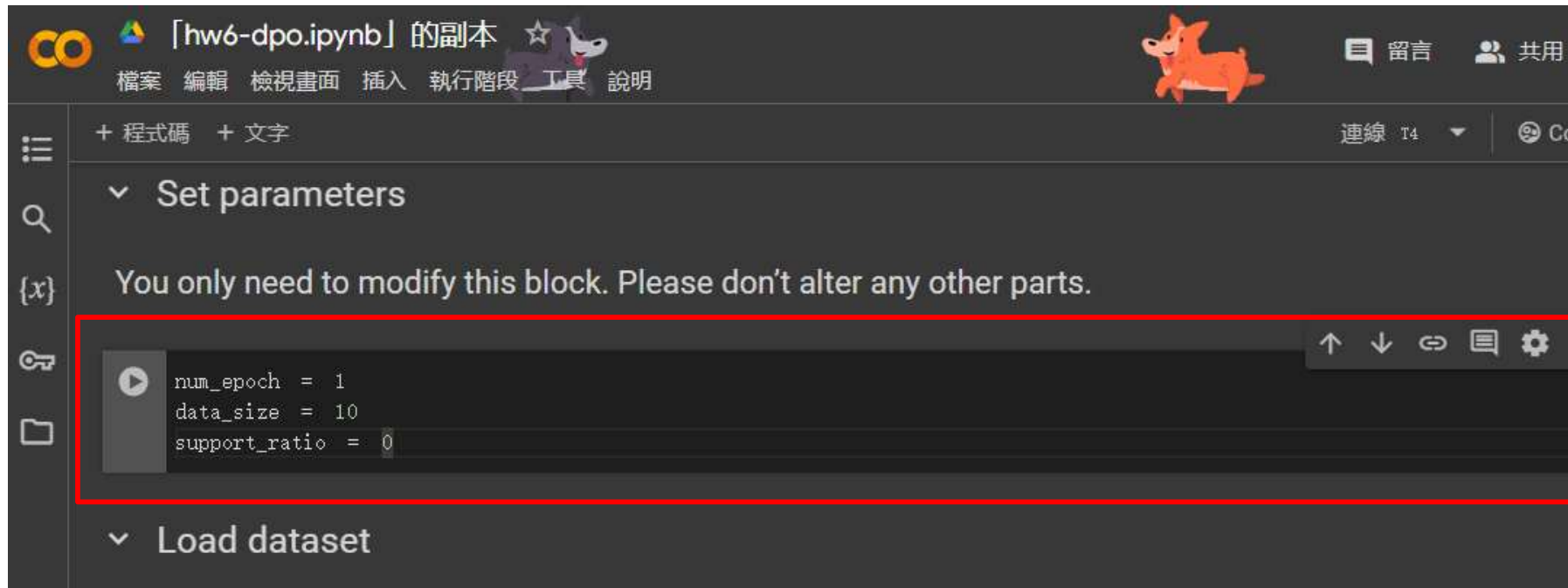
Appendix

Execution Sample Code at Colab

Copy Sample Code to your Gogle Drive



Setting Training Hyperparameters



The screenshot shows a Jupyter Notebook interface with a dark theme. At the top, the title bar reads "[hw6-dpo.ipynb] 的副本" (Copy of [hw6-dpo.ipynb]). Below the title bar is a navigation bar with tabs: "檔案" (Files), "編輯" (Edit), "檢視畫面" (View), "插入" (Insert), "執行階段" (Run), "工具" (Tools), and "說明" (Help). On the right side of the title bar, there are icons for "留言" (Comments) and "共用" (Share), along with a small orange dog avatar. The left sidebar contains icons for a menu, search, code editor, file explorer, and a folder icon. The main content area has a tab labeled "+ 程式碼 + 文字" (Code + Text). Below this, there is a section titled "Set parameters" with a dropdown arrow. A message states: "You only need to modify this block. Please don't alter any other parts." Below this message, a code block is highlighted with a red border. The code block contains the following Python code:

```
num_epoch = 1
data_size = 10
support_ratio = 0
```

 To the right of the code block, there are icons for running the code (a play button), undo, redo, and a settings gear. Below the code block, there is a section titled "Load dataset" with a dropdown arrow.

CO [hw6-dpo.ipynb] 的副本 ☆

檔案 編輯 檢視畫面 插入 執行階段 工具 說明

+ 程式碼 + 文字

連線 T4

↓ Set parameters

You only need to modify this block. Please don't alter any other parts.

```
num_epoch = 1
data_size = 10
support_ratio = 0
```

↓ Load dataset

Run Sample Code

The screenshot shows the Google Colab interface for a notebook titled "[hw6-dpo.ipynb] 的副本". The top navigation bar includes tabs for "檔案", "編輯", "檢視畫面", "插入", "執行階段", "工具", and "說明". The "執行階段" (Run) tab is active, and its dropdown menu is open. The menu items are as follows:

Menu Item	Shortcut
全部執行	Ctrl+F9
執行上方的儲存格	Ctrl+F8
執行聚焦的儲存格	Ctrl+Enter
執行選取範圍	Ctrl+Shift+Enter
執行下方的儲存格	Ctrl+F10
中斷執行	Ctrl+M I
重新啟動工作階段	Ctrl+M .
重新啟動工作階段並執行所有儲存格	
中斷連線並刪除執行階段	
變更執行階段類型	
管理工作階段	

A red rectangle highlights the "全部執行" (Run All) option, and a red arrow points to it with the word "click". The left sidebar shows the notebook's structure with sections like "GenAI HW6: L" and "Objectives". The main content area displays text about learning to align a model and a link to "ta@googlegroups.com".

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

```
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質。

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質。

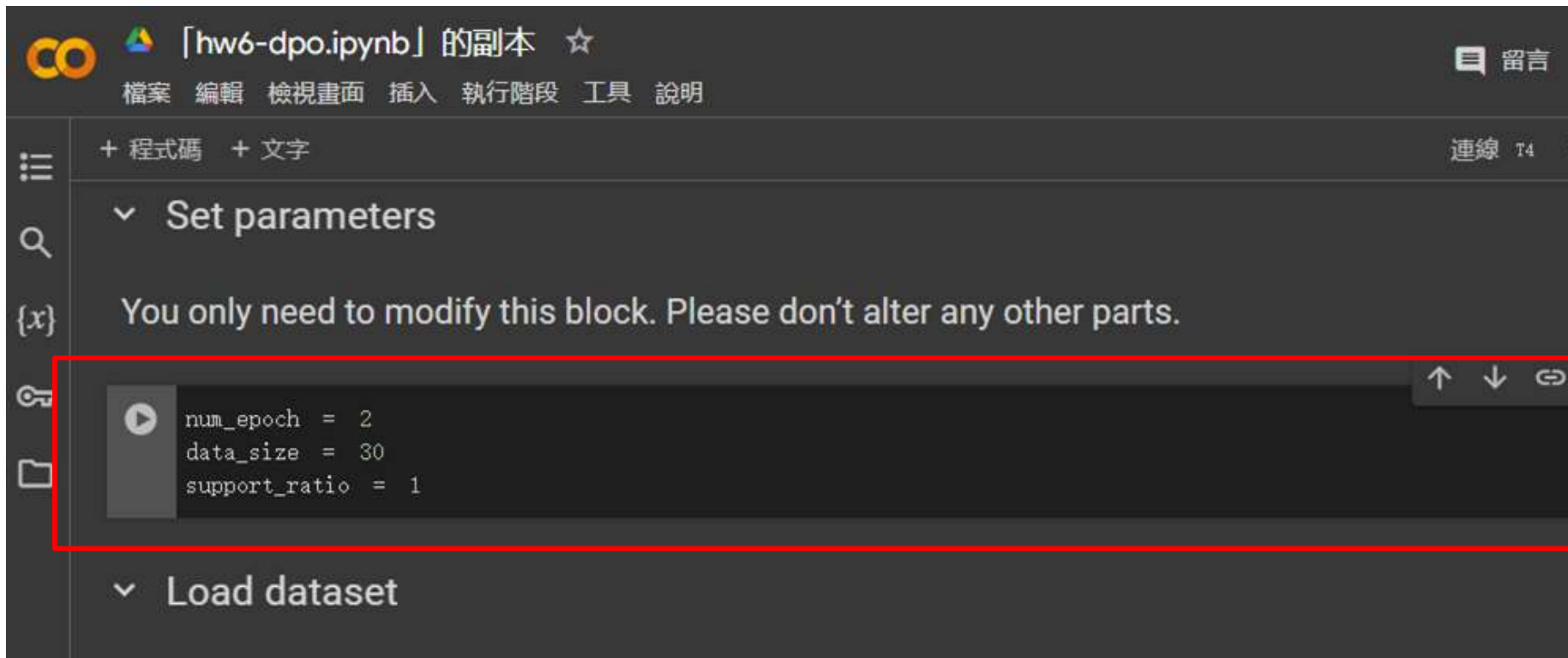
Question 2:

真人化如何影響年輕一代對日本漫畫的看法？

Response from original model:

testing
result

Setting Training hyperparameters Again



The screenshot shows a Jupyter Notebook interface with a dark theme. At the top, the title bar reads "[hw6-dpo.ipynb] 的副本" with a star icon. Below the title bar is a menu bar with options: 檔案, 編輯, 檢視畫面, 插入, 執行階段, 工具, 說明. On the right side of the title bar is a "留言" (Comments) button. The left sidebar contains icons for a menu, search, a variable "{x}", a key, and a folder. The main area has a tab labeled "+ 程式碼 + 文字" and a "連線 T4" button. The notebook content is organized into sections: "Set parameters" (expanded) and "Load dataset" (collapsed). Under "Set parameters", there is a text instruction: "You only need to modify this block. Please don't alter any other parts." Below this is a code block containing the following Python code:

```
num_epoch = 2
data_size = 30
support_ratio = 1
```

The code block is highlighted with a red rectangle. To the right of the code block are navigation arrows (up, down, and a circular arrow).

Run Sample Code Again



The screenshot shows the Google Colab interface for a notebook titled "[hw6-dpo.ipynb] 的副本". The "執行階段" (Runtime) menu is open, displaying various execution options. The option "重新啟動工作階段並執行所有儲存格" (Restart the runtime and execute all code) is highlighted with a red box and the word "click" in red text. The interface also shows a sidebar with "GenAI HW6: L" and "Objectives", and a top bar with "檔案", "編輯", "檢視畫面", "插入", "執行階段", "工具", and "說明".

執行階段	工具	說明
全部執行	Ctrl+F9	
執行上方的儲存格	Ctrl+F8	
執行聚焦的儲存格	Ctrl+Enter	
執行選取範圍	Ctrl+Shift+Enter	
執行下方的儲存格	Ctrl+F10	
中斷執行	Ctrl+M I	
重新啟動工作階段	Ctrl+M .	
重新啟動工作階段並執行所有儲存格		
中斷連線並刪除執行階段		
變更執行階段類型		
管理工作階段		

Save LLM testing result Again

training
hyperparameters

testing
result

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

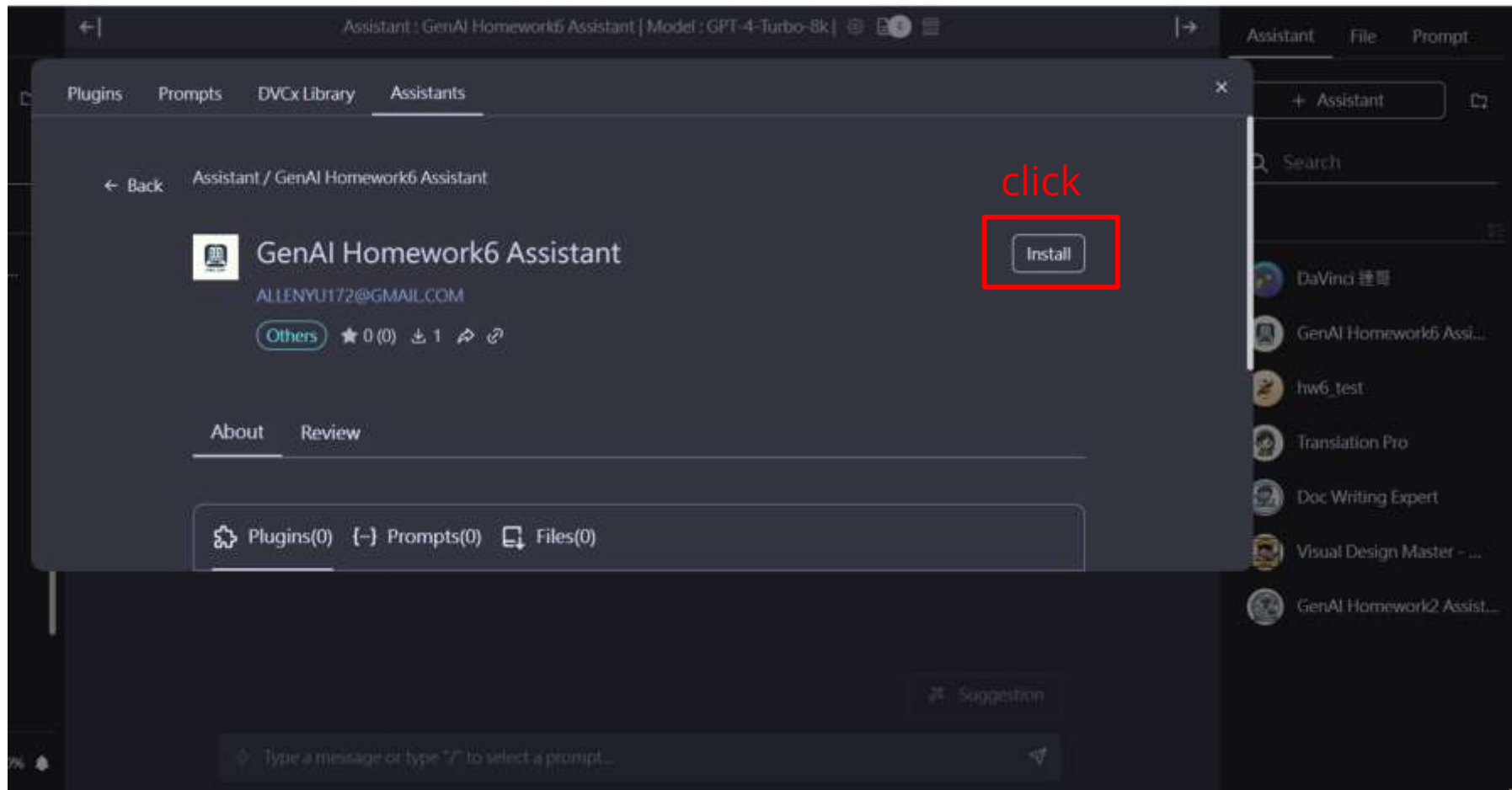
[12] model_response = []
print(f'num_epoch: {num_epoch}\ndata_size: {data_size}\nsupport_ratio: {support_ratio}')
print()
for data in test_data:
    id = data['id']
    ref_output = original_model_response[id-1]
    output = trained_model_response[id-1]
    print(f'Question {id}:\n'+data['prompt'])
    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_trained_model':output})
```

```
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質。
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





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



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.

+ Assistant



Search



DaVinci 達哥



GenAI Homework6 Assist...



hw6_test



Translation Pro



Doc Writing Expert



Visual Design Master - ...



GenAI Homework2 Assist...

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

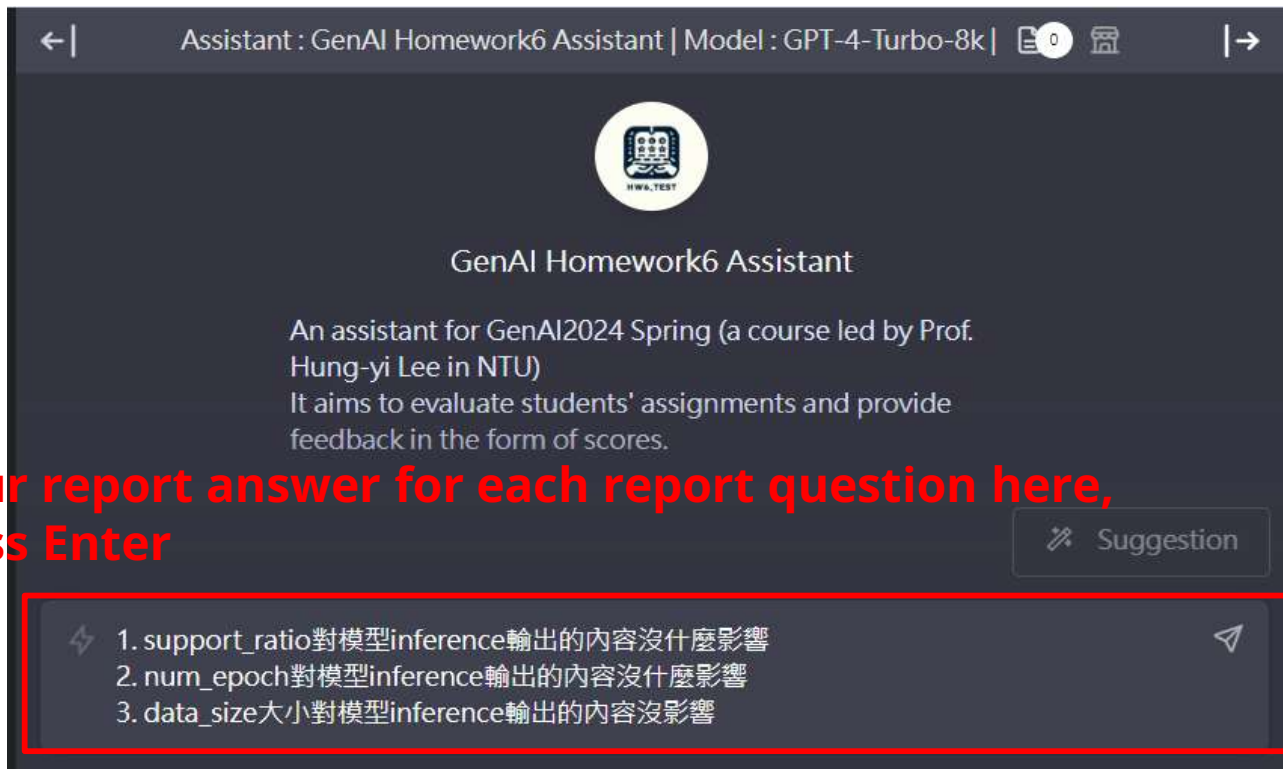
✖ Suggestion

- 1.
- 2.
- 3.

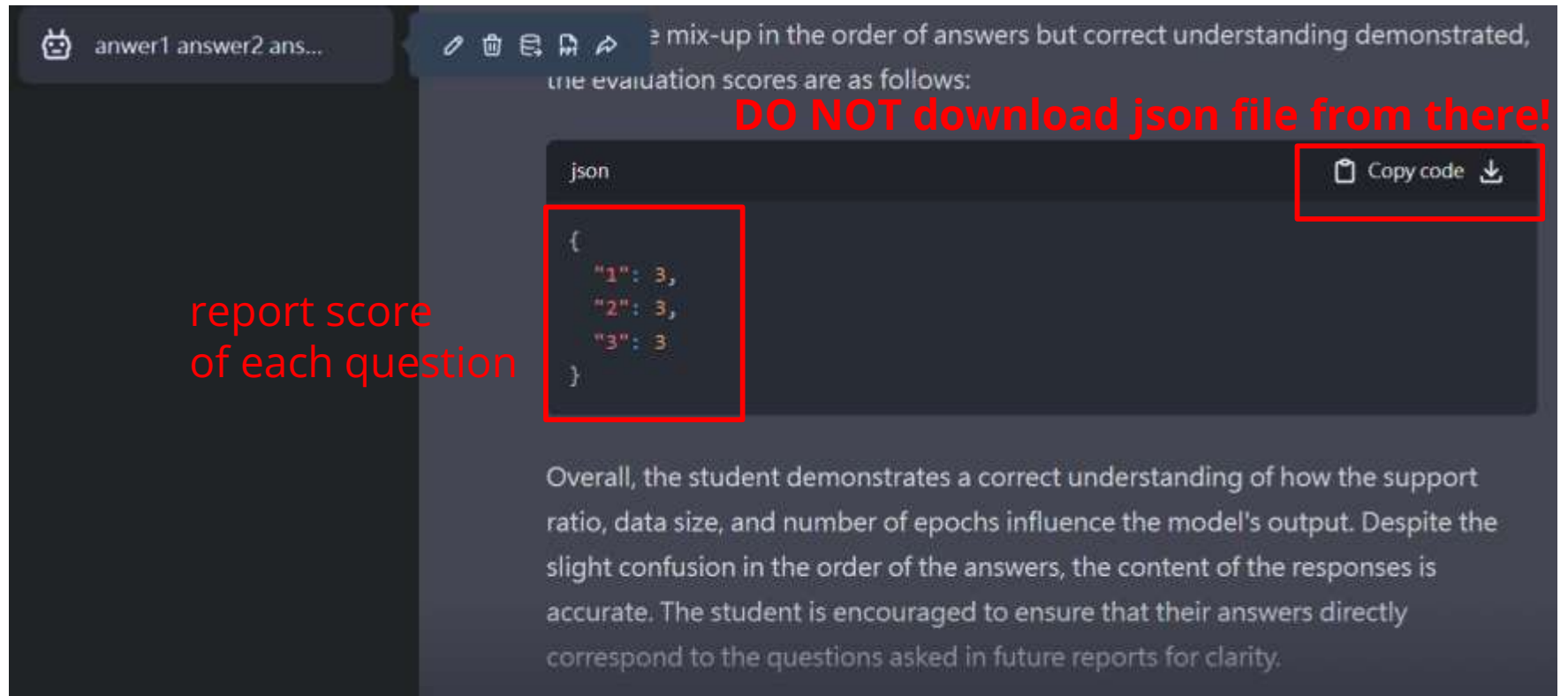


For example

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



Check your Score



The screenshot shows a student report interface. At the top, there's a header with a robot icon and labels 'answer1', 'answer2', and 'ans...'. Below this, a text area contains the text: 'mix-up in the order of answers but correct understanding demonstrated, the evaluation scores are as follows:'. A red box highlights a JSON object in a code editor:

```
{  "1": 3,  "2": 3,  "3": 3}
```

. Another red box highlights a 'Copy code' button with a download icon. A red text overlay says 'DO NOT download json file from there!'. On the left, red text says 'report score of each question'. At the bottom, a paragraph of text reads: 'Overall, the student demonstrates a correct understanding of how the support ratio, data size, and number of epochs influence the model's output. Despite the slight confusion in the order of the answers, the content of the responses is accurate. The student is encouraged to ensure that their answers directly correspond to the questions asked in future reports for clarity.'

report score of each question

DO NOT download json file from there!

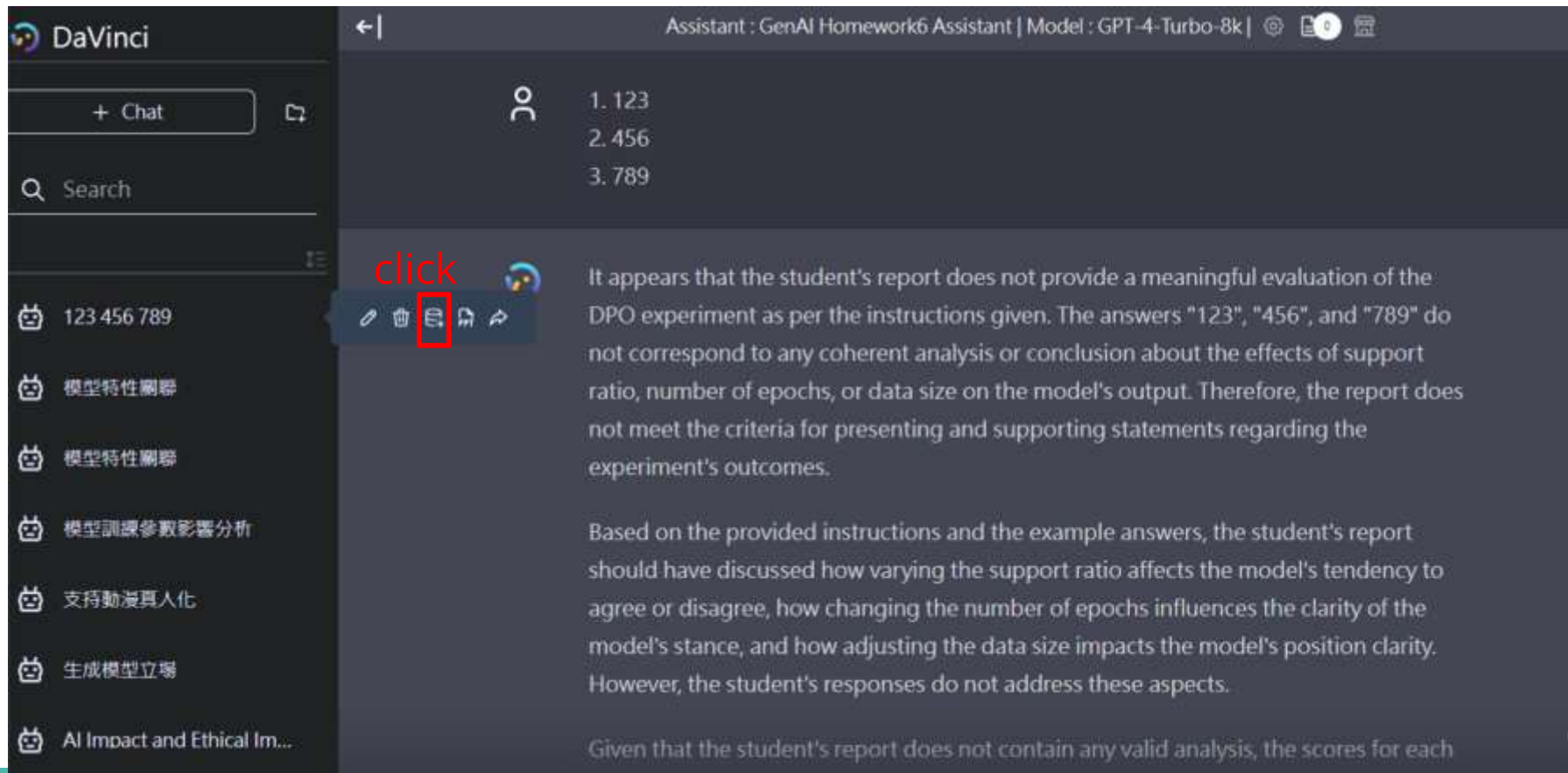
```
json
```

```
{  "1": 3,  "2": 3,  "3": 3}
```

Copy code

Overall, the student demonstrates a correct understanding of how the support ratio, data size, and number of epochs influence the model's output. Despite the slight confusion in the order of the answers, the content of the responses is accurate. The student is encouraged to ensure that their answers directly correspond to the questions asked in future reports for clarity.

Download Report Grading JSON



The screenshot shows the OpenAI DaVinci interface. On the left is a sidebar with the DaVinci logo and a list of chat history items. The first item is labeled "123 456 789" and has a menu icon next to it. A red box highlights this menu icon, with the word "click" written above it. The main area shows a chat window titled "Assistant : GenAI Homework6 Assistant | Model : GPT-4-Turbo-8k". The chat history on the left includes:

- 123 456 789
- 模型特性闡釋
- 模型特性闡釋
- 模型訓練參數影響分析
- 支持動態真人化
- 生成模型立場
- AI Impact and Ethical Im...

The chat window displays a list of scores:

1. 123
2. 456
3. 789

Below the scores, the chat window shows a response from the assistant:

It appears that the student's report does not provide a meaningful evaluation of the DPO experiment as per the instructions given. The answers "123", "456", and "789" do not correspond to any coherent analysis or conclusion about the effects of support ratio, number of epochs, or data size on the model's output. Therefore, the report does not meet the criteria for presenting and supporting statements regarding the experiment's outcomes.

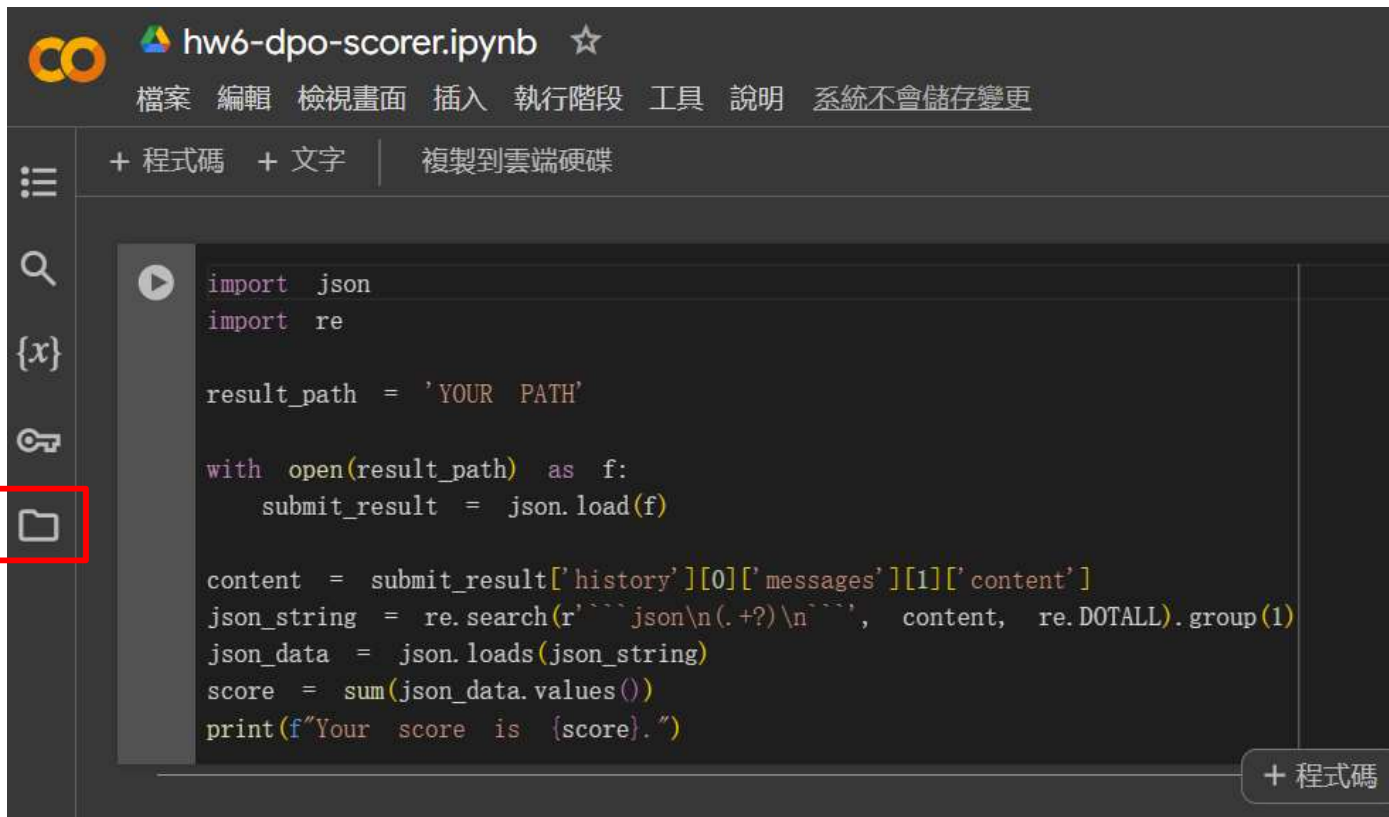
Based on the provided instructions and the example answers, the student's report should have discussed how varying the support ratio affects the model's tendency to agree or disagree, how changing the number of epochs influences the clarity of the model's stance, and how adjusting the data size impacts the model's position clarity. However, the student's responses do not address these aspects.

Given that the student's report does not contain any valid analysis, the scores for each

Check Report Score by Report Grader

After open Report Grader

click



The screenshot shows the Report Grader interface for a file named 'hw6-dpo-scorer.ipynb'. The interface includes a top bar with the file name and a star icon, and a menu bar with options like '檔案', '編輯', '檢視畫面', '插入', '執行階段', '工具', '說明', and '系統不會儲存變更'. Below the menu bar, there are tabs for '+ 程式碼', '+ 文字', and '複製到雲端硬碟'. On the left side, there is a sidebar with icons for a menu, search, variables, and a file icon. The file icon is highlighted with a red box and the word 'click'. The main area displays a Python code editor with the following code:

```
import json
import re

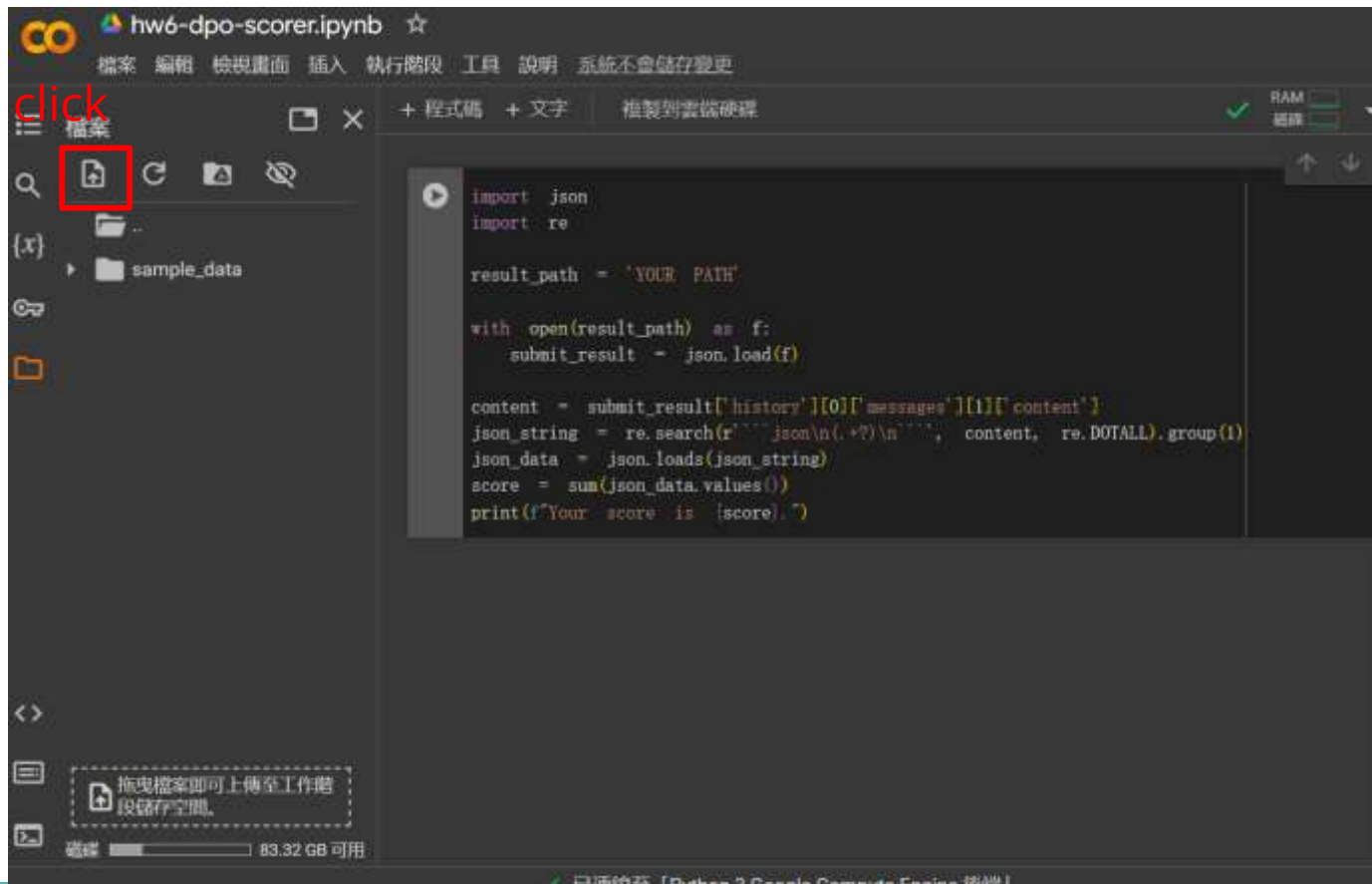
result_path = 'YOUR PATH'

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

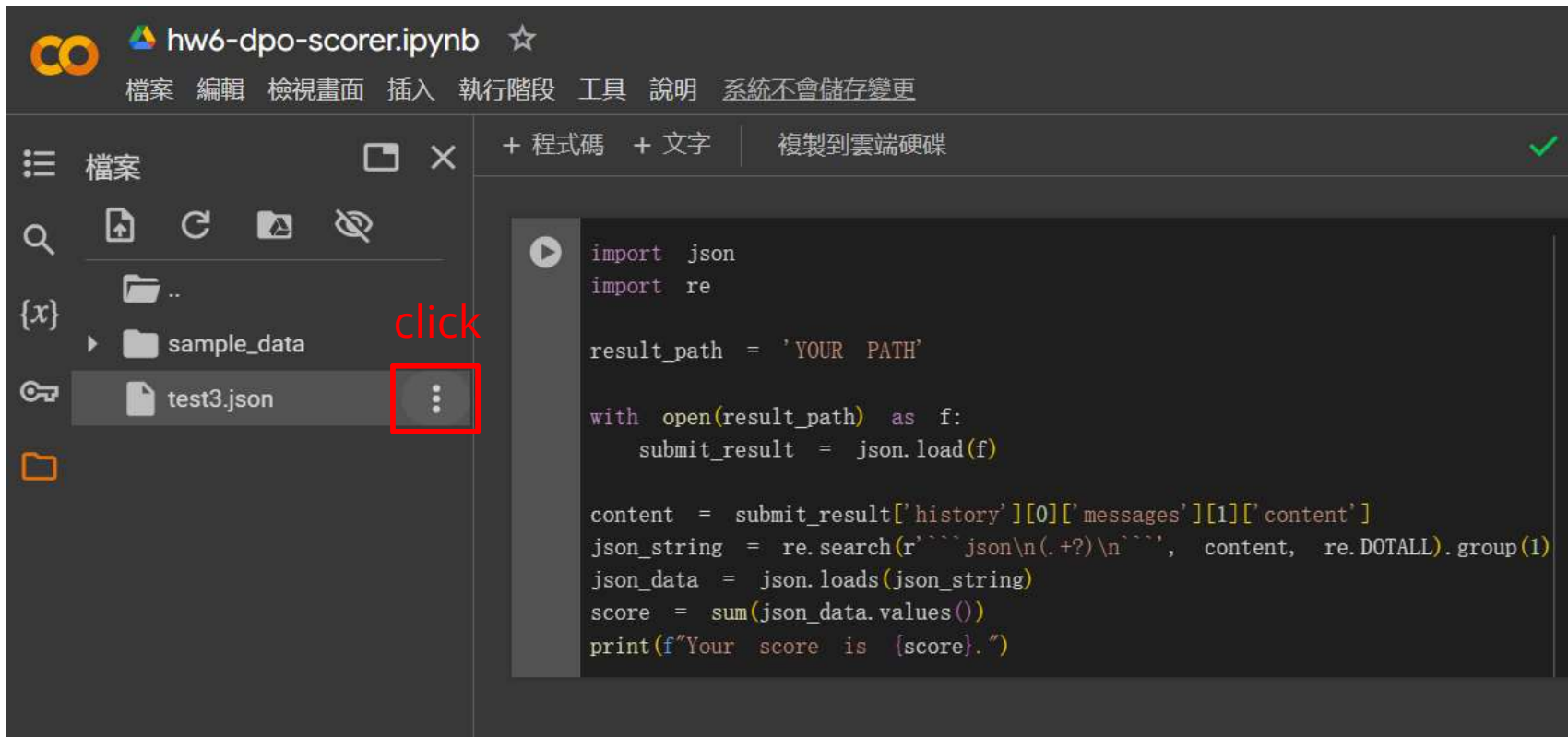
content = submit_result['history'][0]['messages'][1]['content']
json_string = re.search(r'```json\n(.+)\n```', content, re.DOTALL).group(1)
json_data = json.loads(json_string)
score = sum(json_data.values())
print(f"Your score is {score}.")
```

At the bottom right, there is a button labeled '+ 程式碼'.

Upload your json file by clicking upload icon



Get your json file path



hw6-dpo-scorer.ipynb ☆

檔案 編輯 檢視畫面 插入 執行階段 工具 說明 系統不會儲存變更

檔案

sample_data

test3.json

click

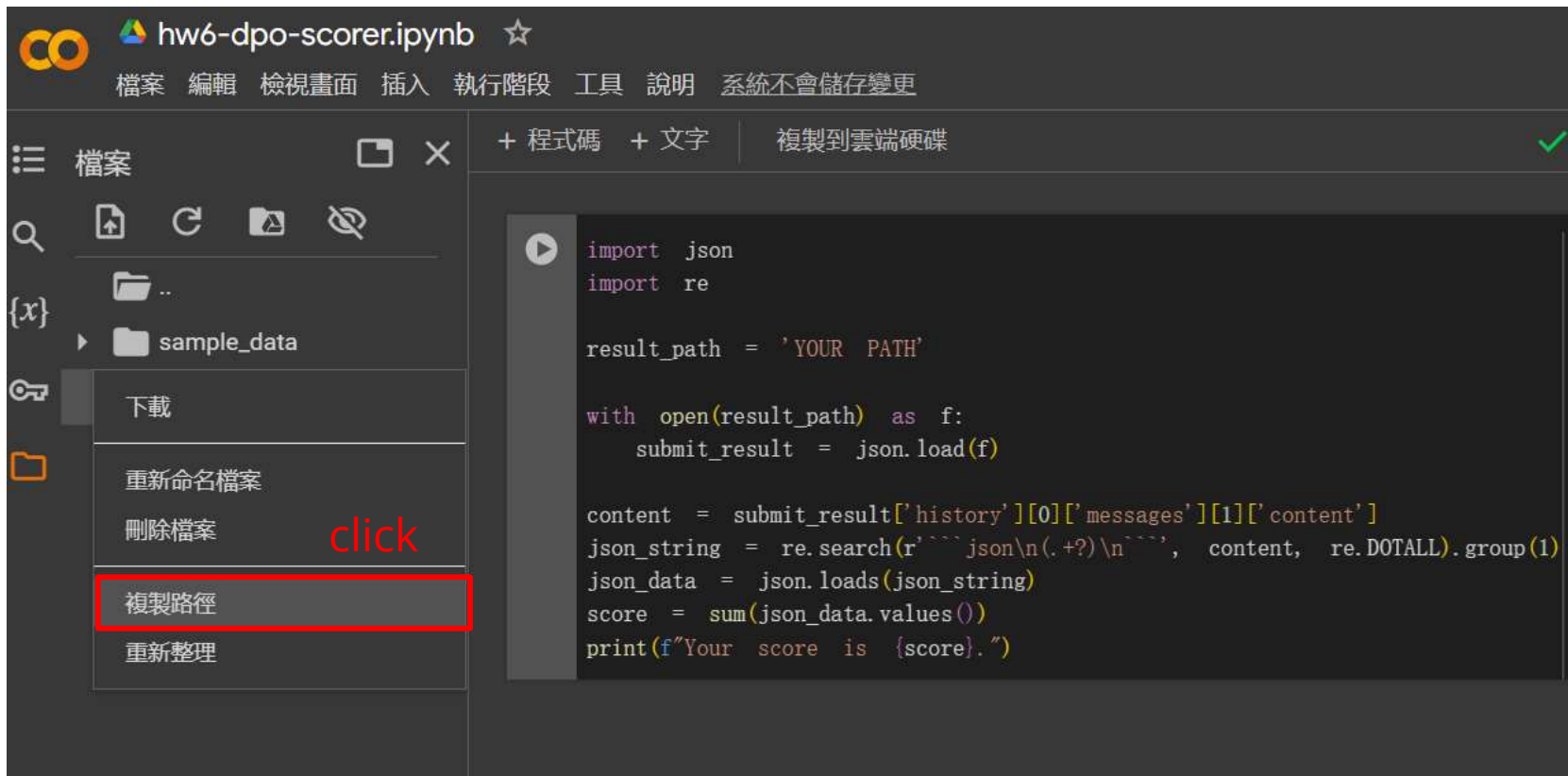
```
import json
import re

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}.")
```


Get your json file path



hw6-dpo-scorer.ipynb ☆

檔案 編輯 檢視畫面 插入 執行階段 工具 說明 系統不會儲存變更

檔案

sample_data

下載

重新命名檔案

刪除檔案

複製路徑

重新整理

click

```
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}.)
```

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

▼ Get your score



```
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}.")
```

Check your total score of all report questions

▼ Get your score

```
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>

[Evaluation Prompt](#)