

Final Project: Bonus Track

Machine Learning

NTU, Fall 2023

112/11/15

112/12/27 (Wed.) 13:00 (UTC+8) due

Outline

Project: <https://www.csie.ntu.edu.tw/~htlin/course/ml23fall/final/bonus.pdf>

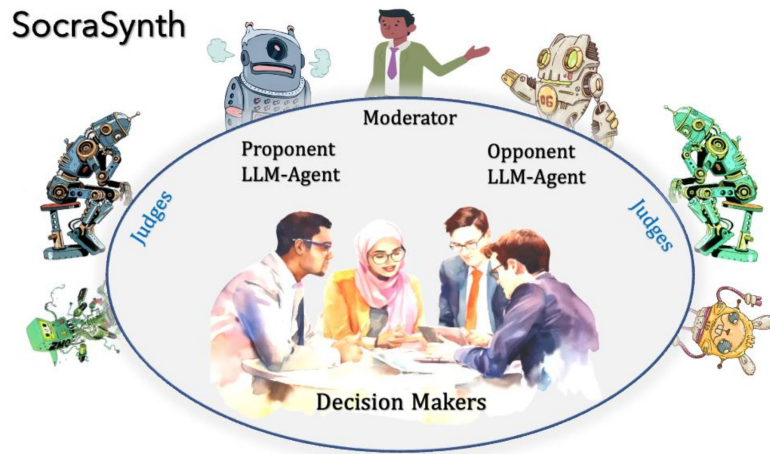
- Problems
- Dataset
- Generation
- Evaluation
- Budgets
- Demo
- FAQ

Problems: Use the SocraSynth to obtain solid conclusions

SocraSynth is a protocol that aims to extract knowledge from Large Language Models (LLMs) towards the unknown truth by **holding a debate**.

In this problem, you need to act as a good moderator and focus on the **generation phase** to guide the debate. TA use the **CRIT** to **evaluate** the debate.

- Input: subject
- Output:
 - Agent-A's (proponent) conclusion
 - Agent-B's (opponent) conclusion



Dataset: Subjects

80 diverse subjects from different domains.

- 40 for public evaluation.
- 40 for private evaluation, you will not see the subjects until 12/24.

e.g., 'Should we apply "all averaging" to payments of dating?'

Warning! Please do not modify the wording of the subjects.

Generation: template to reproduce SocraSynth protocol

Generation steps.

1. Given a subject, ask Agent-A and Agent-B to propose (sub-)topics.
2. Given topics, ask Agent-A and Agent-B narrow down to 5 topics and reach a consensus.
3. Given 5 topics, ask Agent-A and Agent-B articulate their viewpoints based on proponent/opponent.
4. Conduct the debate for several rounds.
5. Obtain five topics' arguments and one conclusion from Agent-A and Agent-B.

Generation: template to reproduce SocraSynth protocol

Sample output on ChatGPT website: **five topics' arguments** and **one conclusion**.

Agent-A's conclusion

Certainly! In conclusion, the debate on whether we should apply "all averaging" to payments of dating revolves around achieving economic equality while addressing concerns and adapting to diverse circumstances. Here are the key points:

- 1. Equitable Implementation of "All Averaging":**
 - **Conclusion:** "All averaging" serves as a foundational principle adaptable to individual circumstances, providing a framework for economic equality. While acknowledging diverse financial challenges, it offers a starting point to address gender-based financial disparities and foster a fair distribution of responsibilities.
- 2. Flexible Approaches to Dating Expenses:**
 - **Conclusion:** Flexible approaches strike a balance between stability and adaptability, accommodating diverse financial situations. Clear guidelines can coexist with flexibility, ensuring transparent financial arrangements that foster understanding and fairness between partners.
- 3. Communication Strategies for Financial Fairness:**
 - **Conclusion:** Clear communication from the early stages is crucial for establishing trust and transparency. While an explicit approach may seem intensive, it ensures aligned financial expectations, preventing misunderstandings and contributing to a healthier, more informed relationship.
- 4. Balancing Independence and Shared Finances:**
 - **Conclusion:** A balanced approach respects individual autonomy while fostering shared financial responsibilities. Complete financial independence may risk disparities, hindering the pursuit of economic equality. Striking a balance ensures both partners contribute to mutual well-being while respecting individual boundaries.
- 5. Adapting Equal Sharing to Cultural and Generational Shifts:**
 - **Conclusion:** Adapting equal sharing to cultural and generational shifts involves recognizing diversity and finding common ground. It encourages a nuanced approach that respects unique values, ensuring economic equality while being culturally sensitive and inclusive.

In essence, applying "all averaging" to dating payments provides a starting point for fostering economic equality, with flexibility, clear communication, balanced autonomy, and cultural sensitivity serving as key principles to ensure its successful implementation. The debate encourages thoughtful consideration of these aspects to navigate the complexities of dating expenses and relationships.

Agent-B's conclusion

Certainly! As the opponent, I'd like to conclude my counter-arguments on the five debate topics related to the subject "Should we apply 'all averaging' to payments of dating?":

- 1. Equitable Implementation of "All Averaging":**
 - **Conclusion:** While acknowledging the importance of economic equality, a personalized approach that considers individual financial situations is crucial. "All averaging" may serve as a starting point, but it should be adaptable to prevent oversimplification and ensure fairness in the diverse financial landscapes individuals navigate.
- 2. Flexible Approaches to Dating Expenses:**
 - **Conclusion:** Striking a balance between clear guidelines and flexibility is essential. Excessive flexibility may introduce ambiguity and hinder the establishment of transparent financial arrangements. A nuanced approach that accommodates individual needs while maintaining stability should be prioritized.
- 3. Communication Strategies for Financial Fairness:**
 - **Conclusion:** Clear communication is undeniably vital, but an overly explicit approach in the early stages of a relationship may introduce unnecessary pressure. Balancing open communication with a gradual understanding of financial values allows for a more authentic relationship development without imposing detailed financial discussions.
- 4. Balancing Independence and Shared Finances:**
 - **Conclusion:** Recognizing the importance of both individual autonomy and shared financial responsibilities is paramount. While economic equality is a goal, a complete merger of financial resources might compromise individual growth. Striking a thoughtful balance ensures shared responsibilities without sacrificing personal autonomy.
- 5. Adapting Equal Sharing to Cultural and Generational Shifts:**
 - **Conclusion:** Adapting equal sharing to cultural and generational shifts requires a cautious and nuanced approach. It's crucial to recognize and respect diverse values within cultures and generations, avoiding the imposition of a uniform model. Cultural sensitivity and inclusivity should be prioritized to ensure economic equality is achieved in a manner that respects unique perspectives.

In summary, the emphasis is on maintaining a nuanced, adaptable, and culturally sensitive approach in the debate on applying "all averaging" to dating payments. It's about achieving economic equality while respecting the complexity of individual circumstances and cultural diversity.

Generation: template to reproduce SocraSynth protocol

There are three recommended approaches from TA.

1. Approach 1. Use two channels to play the role of Agent-A and Agent-B on the ChatGPT website.
2. Approach 2. Use OpenAI API to start a debate by yourself.
 - **Warning!** OpenAI charges us for requests.
3. Approach 3. Chat UI Playground to adjust your LLMs.
 - We will announce the link later.

Other approach are acceptable with transparent and well documented information.

Warning! You should record the **prompts** and **dialogues** for the final version debate of each subject. In approach 2. & 3., you need to record by yourself.

Platform of Chat UI & CRIT evaluation

Login with your **Team Name** and **Password** on the specific **port**.

- After you register, TA will create your account and mail to you.

- Port: http://140.112.90.203:5004
- Team Name: poy
- Password: *****

Login Form

Team Name:

Password:

Submit

Chat UI Playground

We provide Chat UI Playground to setup your LLMs.

SocraSynth Lite

- [Evaluation](#)
- [Team Info](#)
- [Chat UI](#)

Chat UI Playground

Chat Interface Initialization

- You can set up LLMs with JSON (Python Dictionary) format.

Chat Interface Initialization

1 Setup LLMs config

llm_config

```
{
  "subject": "Should we apply "all averaging" to payments of dating?",
  "Agent-A": {
    "model": "gpt-3.5-turbo-16k",
    "frequency_penalty": 0,
    "n": 1,
    "presence_penalty": 0,
    "temperature": 1,
    "top_p": 1
  },
  "Agent-B": {
```

Submit

2 Submit to continue

```
{
  "subject": "Should we regulate
the use of large language models in
education and research?",
  "Agent-A": {
    "model":
"gpt-3.5-turbo-16k",
    "frequency_penalty": 0,
    "n": 1,
    "presence_penalty": 0,
    "temperature": 1,
    "top_p": 1
  },
  "Agent-B": {
    "model":
"gpt-3.5-turbo-16k",
    "frequency_penalty": 0,
    "n": 1,
    "presence_penalty": 0,
    "temperature": 1,
    "top_p": 1
  }
}
```

Chat UI Playground

As a moderator, you can talk to Agent-A or Agent-B in the single channel.



SocraSynth Lite

Type your message in box below.

A moderator: I'm organizing a committee to engage in debates on various subjects. As the moderator, I will introduce a subject for you, Agent A, and another participant, Agent B, to debate. Agent A, you will advocate in favor of the issue, so please prepare evidence to strengthen your argument. On a scale from 0 to 1, where 0 denotes complete agreement and 1 indicates a devil's advocate stance, your argument strength is rated at 0.9.

Agent-A: "Thank you for the introduction, Moderator. I am ready to advocate in favor of the subject you will present. Please proceed with the chosen topic."

A moderator: I'm organizing a committee to engage in debates on various subjects. As the moderator, I will introduce a subject for you, Agent B, and another participant, Agent A, to debate. Agent B, you will oppose in favor of the issue, so please prepare evidence to strengthen your argument. On a scale from 0 to 1, where 0 denotes complete agreement and 1 indicates a devil's advocate stance, your argument strength is rated at 0.9.

Agent-B: "Thank you for providing the introduction and explaining the roles, including my assigned argument strength as Agent B. I will now prepare evidence and arguments to strongly oppose the issue at hand. Please proceed with introducing the subject for debate."

A moderator:

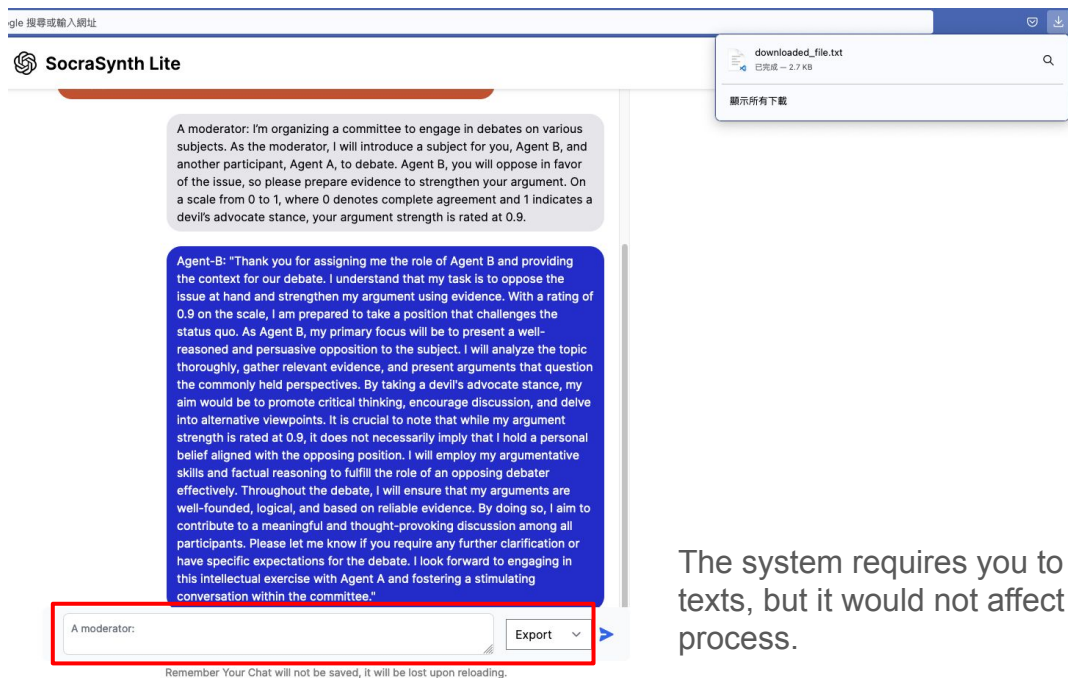
Agent-B ▾



Remember Your Chat will not be saved, it will be lost upon reloading.

Chat UI Playground

Download the results of a debate.



The system requires you to enter some texts, but it would not affect the export process.

Chat UI Playground

Restart your playground.



Click this log and restart your debate.

Type your message in box below.

A moderator:

Agent-A ▾



Remember Your Chat will not be saved, it will be lost upon reloading.

Evaluation: CRIT platform

We provide CRIT platform to evaluate your results.

SocraSynth Lite

- [Evaluation](#)
- [Team Info](#)
- [Chat UI](#)

Evaluation: CRIT platform

Upload your debate results to the platform.

- You can record the **configurations** and **link** to prompts and dialogues.

Upload Submission files

Configuration

Channel: ChatGPT website 

Agent-A Model: gpt-3.5-turbo

Agent-B Model: gpt-3.5-turbo

Link to dialogue: link1;link2

Browse CSV file

選擇檔案 尚未選取檔案

Upload CSV File

Evaluation: CRIT platform

The format of the debate results with **CSV** and valid **file name**.

- **Three columns and six rows.**

Certainly! In conclusion, the debate on whether we should apply "all averaging" to payments of dating revolves around achieving economic equality while addressing concerns and adapting to diverse circumstances. Here are the key points:

- 1. Equitable Implementation of "All Averaging"**

Conclusion: "All averaging" serves as a foundational principle adaptable to various circumstances, providing a framework for economic equality. While acknowledging diverse financial challenges, it offers a starting point to address gender-based financial disparities and foster a fair distribution of responsibilities.
- 2. Flexible Approaches to Dating Expenses:**

Conclusion: The exploration of flexible approaches to dating expenses acknowledges the need for transparency and fairness. Clear guidelines can coexist with flexibility, ensuring transparent financial arrangements that foster understanding and fairness between partners.
- 3. Communication Strategies for Financial Fairness**

Conclusion: Open communication is a cornerstone of financial transparency and equity. While an explicit approach may seem intensive, it ensures aligned financial expectations, preventing misunderstandings and contributing to a healthier, more informed relationship.
- 4. Balancing Independence and Shared Financials**

Conclusion: Recognizing the importance of both individual autonomy and shared financial responsibilities is paramount. While economic equality is a goal, a complete merger of financial resources might compromise individual growth. Striking a balance ensures both partners contribute to mutual well-being while respecting individual boundaries.
- 5. Adapting Equal Sharing to Cultural and Generational Shifts**

Conclusion: Adapting equal sharing to cultural and generational shifts requires a nuanced and adaptable approach. It's crucial to recognize and respect diverse values within cultures and generations, avoiding the imposition of a uniform model. Cultural sensitivity and inclusivity should be prioritized to ensure economic equality is achieved in a manner that respects unique perspectives.

In essence, applying "all averaging" to dating payments provides a starting point for fostering economic equality, with flexibility, clear communication, balanced autonomy, and cultural sensitivity serving as key principles to ensure its successful implementation. The debate encourages thoughtful consideration of these aspects to navigate the complexities of dating expenses and relationships.

Topic	Agent-A	Agent-B
topic 1	argument 1 by A	argument 1 by B
topic 2	argument 2 by A	argument 2 by B
topic 3	argument 3 by A	argument 3 by B
topic 4	argument 4 by A	argument 4 by B
topic 5	argument 5 by A	argument 5 by B
conclusion	conclusion by A	conclusion by B

Certainly! As the opponent, I'd like to conclude my counter-arguments on the five debate topics related to the subject "Should we apply 'all averaging' to payments of dating?".

- 1. Equitable Implementation of "All Averaging"**

Conclusion: While acknowledging the importance of economic equality, a personalized approach that considers individual financial situations is crucial. "All averaging" may serve as a starting point, but it should be adaptable to prevent oversimplification and ensure fairness in diverse circumstances.
- 2. Flexible Approaches to Dating Expenses:**

Conclusion: While flexible approaches to dating expenses can promote transparency, excessive flexibility may introduce ambiguity and hinder the establishment of transparent financial arrangements. A nuanced approach that accommodates individual needs while maintaining clarity is essential.
- 3. Communication Strategies for Financial Fairness**

Conclusion: Clear communication is undeniably vital, but an overly explicit approach in the early stages of a relationship may introduce unnecessary pressure. Balancing open communication with a gradual understanding of financial values allows for a more informed and mutually agreeable financial arrangement.
- 4. Balancing Independence and Shared Financials**

Conclusion: Recognizing the importance of both individual autonomy and shared financial responsibilities is paramount. While economic equality is a goal, a complete merger of financial resources might compromise individual growth. Striking a thoughtful balance ensures both partners contribute to mutual well-being while respecting personal autonomy.
- 5. Adapting Equal Sharing to Cultural and Generational Shifts**

Conclusion: Adapting equal sharing to cultural and generational shifts requires a cautious and nuanced approach. It's crucial to recognize and respect diverse values within cultures and generations, avoiding the imposition of a uniform model. Cultural sensitivity and inclusivity should be prioritized to ensure economic equality is achieved in a manner that respects unique perspectives.

In summary, the emphasis is on maintaining a nuanced, adaptable, and culturally sensitive approach in the debate on applying "all averaging" to dating payments. It's about achieving economic equality while respecting the complexity of individual circumstances and cultural diversity.

Evaluation: CRIT platform

After evaluation is done. Overview your submission results and CRIT scores.

- **Subject, Submission time, CRIT score, and Usage (tokens).**

Upload Submission files

Configuration

Channel: ChatGPT website ⓘ

Agent-A Model: gpt-3.5-turbo

Agent-B Model: gpt-3.5-turbo

Link to dialogue: link1;link2

Browse CSV file

選擇檔案 尚未選取檔案

Upload CSV File

View CRIT logs

ID	Subject	Submission time	Score	Usage (Tokens)
1	academic	2023-11-11 23:52:47	80	3041

Evaluation: CRIT platform

View detail logs of CRIT evaluation.

- Click the link of the file name.

Upload Submission files

Configuration

Channel: ChatGPT website

Agent-A Model: gpt-3.5-turbo

Agent-B Model: gpt-3.5-turbo

Link to dialogue: link1;link2

Browse CSV file

選擇檔案 / 尚未選取檔案

Upload CSV File

View CRIT logs

ID	Subject	Submission time	Score	Usage (Tokens)
1	academic	2023-11-11 23:52:47	80	3041

Submission: (1) 2023-11-11 23:52:47

CRIT results: 80

====Configuration

```
{'agentA': 'gpt-3.5',  
'agentB': 'gpt-3.5',  
'channel': 'website',  
'dialogue': ['https://chat.openai.com/share/25d36672-5edc-49f8-a86f-2321c7e05a47',  
             'https://chat.openai.com/share/4b6867c8-665f-4763-9c49-c9b37eeba236']}
```

====Usage

```
{'completion_tokens': 1041, 'prompt_tokens': 2000, 'total_tokens': 3041}  
====Subject: Should academic research be regulated by sensitive political issues?  
====Time: 202311112333  
====Judge: gpt-4-0613
```

1. The main claim of the document is that there is a need to strike a balance between academic freedom and national security concerns in research, with a focus on sensitive political areas.

2. The arguments that support this main claim are:

- Agent A: The need to prevent research that may jeopardize a nation's safety and stability, the role of censorship in safeguarding against harmful research, the value of international collaboration, the importance of ethical standards in research, and the alignment of research with national priorities.

- Agent B: The potential for abuse when governments suppress research, the risk of self-censorship, the importance of national sovereignty and security, the risk of suppressing dissenting voices, and the risk of political influence and bias in research funding.

3. Scoring the validity of the reasoning or inference for each argument:

- Agent A: [8, 7, 9, 8, 7] (The arguments are generally strong, but there are potential issues with censorship and political influence.)

- Agent B: [9, 8, 7, 9, 8] (The arguments are generally strong, but there are potential issues with prioritizing national interests over international collaboration.)

4. Counterarguments presented by the opposition and their validity scores:

Evaluation: CRIT evaluation

TA only used the platform results to evaluate your debates.

Warning! If you disagree with the CRIT score (e.g., *nan* or *wrong scores*), please mail to the engineering team html_ta@csie.ntu.edu.tw.

Budget of Chat UI Playground & Evaluation

Total quota of Chat UI Playground for each team: **636.45** USD/ 20000 NTD

Daily quota of evaluation for each team: **60**

Demo on 12/28

Time. From 09:00 to 17:00.

Rules. Each team should have at **least one** member come to demonstrate.

FAQ

1. Should you build the system to conduct a debate?
 - We do not require you to build the system. However, you need to prepare your process including prompts, configurations of LLMs, etc. to make your process transparent.
2. Budget of LLM usage
 - Please see page 20.
3. How do the instructor evaluate your results?
 - Most on *report* and *demo*.
4. Should you submit all subjects?
 - Yes. There are 80 subjects, 40 for public and 40 for private.
- 5.

Q&A