

Full-Stack AI Content Creator

LYU2305 4/12/2023
Tse Hui Tung, 1155158864
NG Man Tik, 1155158302

Have you ever thought about that you can talk to either one of them?





Congrats !!!



Oppenheimer
Trinity



10.328 views

Oppenheimer instagram template [#template](#)

[View all 328 comments](#)

5 DAYS AGO

Project Goal

1. Imitate a celebrity's tone, writing style and thought strategy.
2. Build a rigorous evaluation system on LLM's Imitation Ability.
3. Evaluate the performance on various factor, such as personality data, model architectural, LLM model, etc.
4. Combine text to image function using stable diffusion and package the image with post content together.

Value of this project

ROLELLM: BENCHMARKING, ELICITING, AND ENHANCING ROLE-PLAYING ABILITIES OF LARGE LANGUAGE MODELS

Zekun Moore Wang^{1,2*}, Zhongyuan Peng^{2,3}, Haoran Que^{1,4}, Jiaheng Liu^{1,5},
Wangchunshu Zhou¹, Yuhao Wu¹, Hongcheng Guo¹, Ruitong Gan¹, Zehao Ni¹,
Man Zhang⁶, Zhaoxiang Zhang⁶, Wani Ouyang⁶, Ke Xu⁶, Wenhu Chen⁶, Jie Fu⁶, Junran Peng^{2,10}
¹Beihang University; ²University of the Chinese Academy of Sciences; ³ETH Zürich;
⁴Beijing University of Posts and Telecommunications; ⁵The Hong Kong Polytechnic University;
⁶Institute of Automation, Chinese Academy of Sciences; ⁷Shanghai AI Lab;
⁸University of Waterloo; ⁹The Hong Kong University of Science and Technology; ¹⁰Chongyue Technology
zenmoore@buaa.edu.cn, liujiaheng@buaa.edu.cn

ABSTRACT

The advent of Large Language Models (LLMs) has paved the way for complex tasks such as role-playing, which enhances user interactions by enabling models to imitate various characters. However, the closed-source nature of state-of-the-art LLMs and their general-purpose training limit role-playing optimization. In this paper, we introduce RoleLLM, a framework to benchmark, elicit, and enhance role-playing abilities in LLMs. RoleLLM comprises four stages: (1) Role Profile Construction for 100 roles; (2) Context-Based Instruction Generation (Context-Instruct) for role-specific knowledge extraction; (3) Role Prompting using GPT (RoleGPT) for speaking style imitation; and (4) Role-Conditioned Instruction Tuning (ROCI) for fine-tuning open-source models along with role customization. By Context-Instruct and RoleGPT, we create RoleBench, the first systematic and fine-grained character-level benchmark dataset for role-playing with 168,093 samples. Moreover, ROCI on RoleBench yields RoleLLaMA (English) and RoleGLM (Chinese), significantly enhancing role-playing abilities and even achieving comparable results with RoleGPT (using GPT-4).



Figure 1: We introduce RoleLLM, a role-playing framework of data construction and evaluation (RoleBench), as well as solutions for both closed-source and open-source models (RoleGPT, RoleLLaMA, RoleGLM). We also propose Context-Instruct for long-text knowledge extraction and role-specific knowledge injection.

* Primary Authors. Work done during Zekun Moore Wang’s internship at HKUST.

† Corresponding Author.

[‡] Access models, demos, and RoleBench at <https://github.com/InteractiveAIGP-Team/RoleLLM-public>.

arXiv:2310.00746v1 [cs.CL] 1 Oct 2023

Does GPT-4 Pass the Turing Test?

Cameron Jones and Benjamin Bergen
UC San Diego,
9500 Gilman Dr, San Diego, CA
cameron@ucsd.edu

Abstract

We evaluated GPT-4 in a public online Turing Test. The best-performing GPT-4 prompt passed in 41% of games, outperforming baselines set by ELIZA (27%) and GPT-3.5 (14%), but falling short of chance and the baseline set by human participants (63%). Participants’ decisions were based mainly on linguistic style (35%) and socio-emotional traits (27%), supporting the idea that intelligence is not sufficient to pass the Turing Test. Participants’ demographics, including education and familiarity with LLMs, did not predict detection rate, suggesting that even those who understand systems deeply and interact with them frequently may be susceptible to deception. Despite known limitations as a test of intelligence, we argue that the Turing Test continues to be relevant as an assessment of naturalistic communication and deception. AI models with the ability to masquerade as humans could have widespread societal consequences, and we analyse the effectiveness of different strategies and criteria for judging humanness.

Keywords: Turing Test, Large Language Models, GPT-4, interactive evaluation

1 Introduction

Turing (1950) devised the *Imitation Game* as an indirect way of asking the question: “Can machines think?”. In the original formulation of the game, two witnesses—one human and one artificial—attempt to convince an interrogator that they are human via a text-only interface. Turing thought that the open-ended nature of the game—in which interrogators could ask about anything from romantic love to mathematics—constituted a broad and ambitious test of intelligence. The Turing Test, as it has come to be known, has since inspired a lively debate about what (if anything) it can be said to measure, and what kind of systems might be capable of passing (French, 2000).



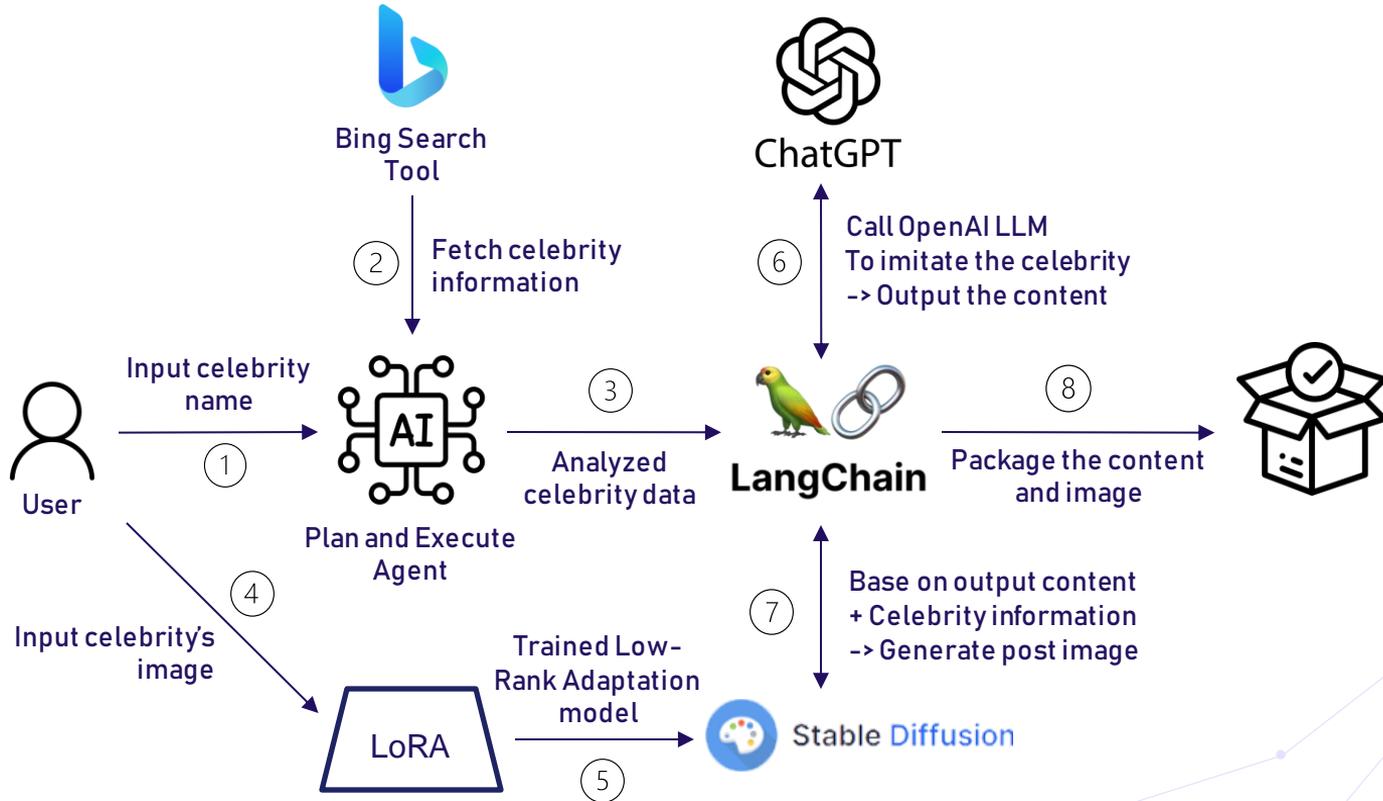
Figure 1: Chat interface for the Turing Test experiment featuring an example conversation between a human Interrogator (in green) and GPT-4.

Large Language Models (LLMs) such as GPT-4 (OpenAI, 2023) seem well designed for Turing’s game. They produce fluent naturalistic text and are near parity with humans on a variety of language-based tasks (Chang and Bergen, 2023; Wang et al., 2019). Indeed, there has been widespread public speculation that GPT-4 would pass a Turing Test (Biever, 2023) or has implicitly done so already (James, 2023). Here we address this question empirically by comparing GPT-4 to humans and other language agents in an online public Turing Test.

Value of this project



Overall Design



Overall Design

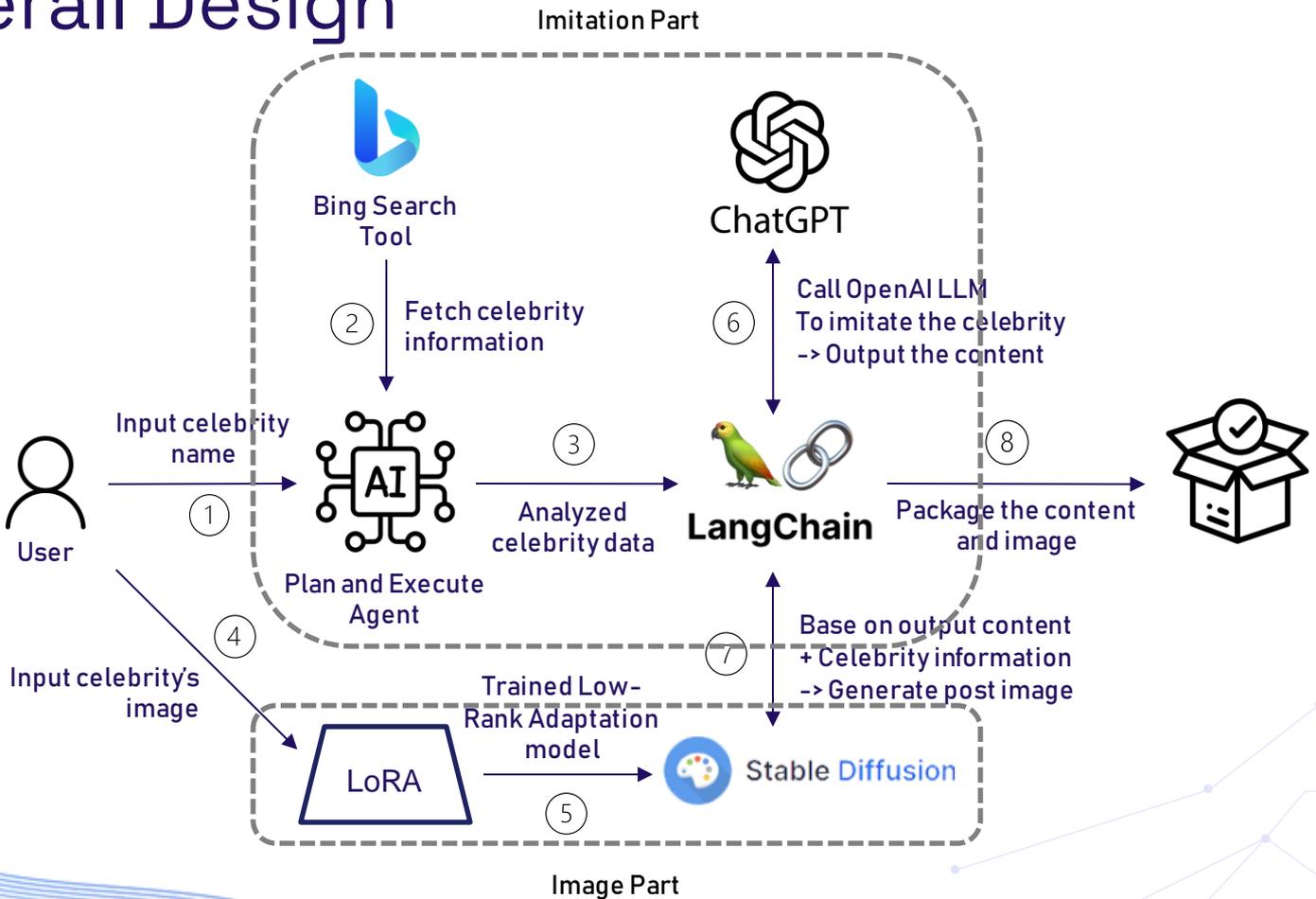


Table of contents

01

Fetching Celebrity's
Information

03

Application Demo

02

Imitation System

04

Evaluation System



01

Fetching Celebrity's Information

Reasons/Advantage



Comprehensive Simulation

letting the LLM have a full understanding of the celebrity before replication.



Overcoming Limitations of LLM

Saving the summarized result can avoid LLM reading the lengthy input which will affect the result.



Efficiency

Pre-processing can help LLM no need to perform searching during the imitation which will improve the speed.



Consistency

The summarized result can be the reference for LLM to mimic to ensure the consistency in the imitation.

Data Types

- To ensure having the comprehensive and accurate view of the celebrity's persona, we train LLM to fetch two types of information.

First-Hand Information

Source

- Direct Communication from the celebrity
- Speeches, interviews, writings

Representing

- Celebrities' personal ethos, speech and writing patterns

Third-Party Information

Source

- External content about the celebrity
- Biographies and new articles

Representing

- Public persona, societal impact, career trajectory

Task List

- Ensuring the agent to collect the comprehensive information of the celebrity's public image and work, we have designed various tasks for the agent to search for.

*****TASK LIST*****

- 1: Research any scientific papers or publications J. Robert Oppenheimer was involved in.
- 2: Identify any awards or honors J. Robert Oppenheimer received before 1963.
- 3: Research any organizations or causes J. Robert Oppenheimer supported.

Figure: Some predefined task lists for searching Oppenheimer's information

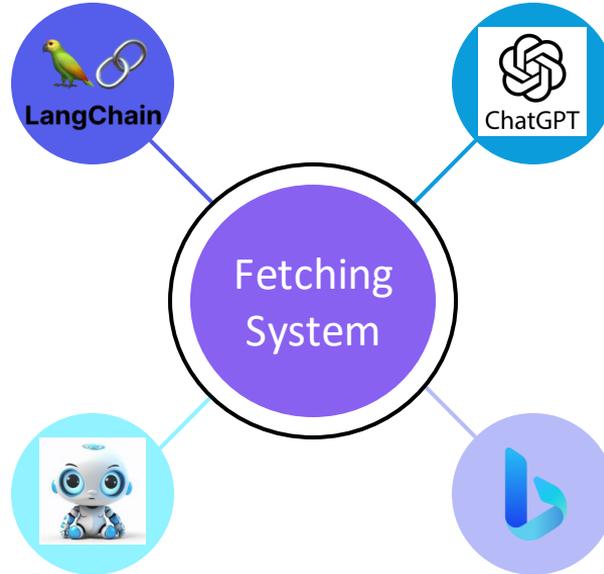
Workflow

LangChain

- A framework for LLM application using chaining and various tools

BabyAGI

- A plan and execute agent which can integrate other tools and manage the workflow based on the task given.



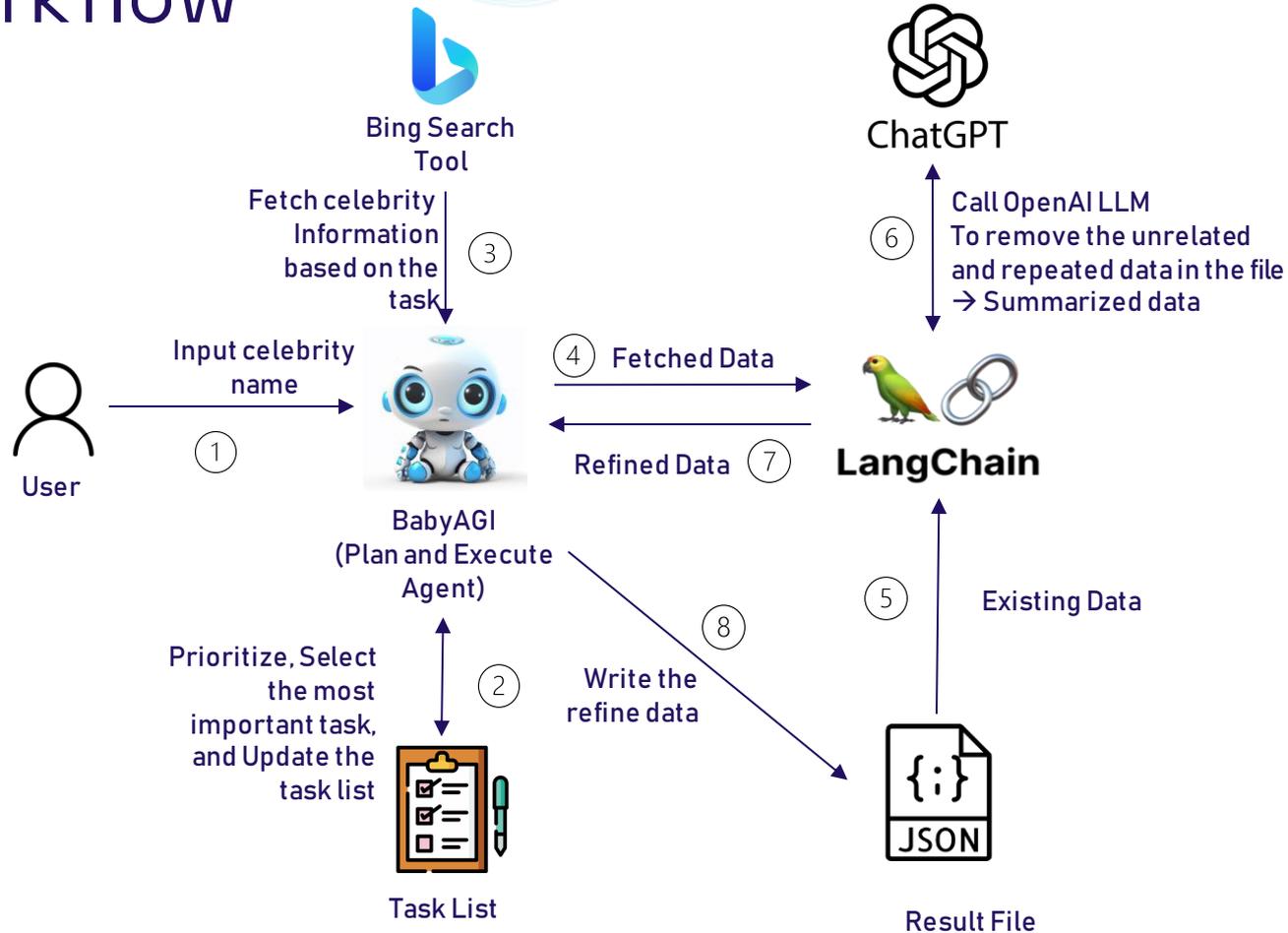
ChatGPT

- Parsing and understanding the information of the fetched result for further summarization of the data

Bing Search Tool

- A tool for the agent to search and fetch the celebrities' information from different webpages.

Workflow



Fetches Result

- Even though the result have the comprehensive overview of the celebrity. The length of the result is too long which will affect the imitation when feeding the result to the LLM

```
21 {  
22   "Title": "Research J. Robert Oppenheimer's early life and education",  
23   "Value": "Physicist J. Robert Oppenheimer led the Manhattan Project, the program that developed the first nuclear weapon. Oppenheimer",  
24 },  
25 {  
26   "Title": "Analyze J. Robert Oppenheimer's influence on the development of atomic theory",  
27   "Value": "J. Robert Oppenheimer, the brilliant physicist behind the Manhattan Project, played a pivotal role in developing atomic weapons.",  
28 },
```

Figure: Result of the fetched information

Generating Question Sets

- To further analyze and synthesize information
- Following the idea of the paper RoleLLM, GPT will try to capture the person's thinking style and writing habit by answering different questions.
- The QA result highly summarized the information while the simulated result can be a pivot for LLM to mimic in the imitation to ensure long term consistency.

Generating Question Sets

- The question generation strategy breaks down into three parts to ensure the questions generated are valuable.

Predefined Data Types

Aim

Guiding LLM towards a understanding of celebrity's life and personality.

Examples

Education and Professional Background

Interests and Hobbies

Personality

Favorite Books, Movies, and Music

Values and Beliefs

Problem-Solving Style

Thoughts

Communication and Social Style

Memorable Life Experience

Writing and Speaking Style

Factualness

Aim

- Prevent generating irrelevant or inaccurate questions
- Ensure the result aligning the celebrity's background and persona.

Generating Question Sets

```
question_template = """
Background information on {role_name}:
{background_info}

If you had the opportunity to meet {role_name}, what questions would you ask
{role_name}? Please design 10 questions that do not repeat in terms of semantics.
Please set the question diverse.

You can set your questions based on either on the following aspects: {aspects};
or on the following types: {types}; or both.

You can base your questions
on {role_name}'s personality traits and relevant episodes from {role_name}. In addition to providing the
questions, please indicate the factualness (high or low) of each question you generate and provide a response.
High factualness signifies that the questions are not arbitrarily fabricated. Please strive to generate questions
with high factualness.

Please refer to the following examples related to Sherlock Holmes for guidance on how to format your questions and
responses:
```

```
1. Question: "Sherlock Holmes, is Dr. John Watson your closest confidant?"
   Factualness: High, because in Arthur Conan Doyle's stories, Dr. Watson is indeed Holmes' close friend and partner.
   Response: "Indeed, Dr. Watson is not only my closest confidant but also an invaluable assistant in my investigations.
   His medical expertise and steadfast character have been instrumental in my work."
```

```
[Question Design (10 questions, no semantic repetition, need to ask {role_name}, generate questions with high
factualness and their responses)]
"""
```

Figure: Prompt to generate question sets

Generating Question Sets

```
4   {
5     "Aspect": "Education and Professional Background",
6     "Question": " Can you tell me about the moment you decided your current professional field was the right one for you?",
7     "Factuality": "High. This question assumes that the person has a moment of realization that led them to their current profession.",
8     "Response": "Absolutely, it was during my second year of college when I was working on a group project in my engineering class. The way I felt fully
9   },
10  {
11    "Aspect": "Interests and Hobbies",
12    "Question": "What hobby have you dedicated the most time to, and what has it taught you?",
13    "Factuality": "High. Most individuals have hobbies and learn from them.",
14    "Response": "I've dedicated the most time to woodworking. It has taught me patience, precision, and the value of hands-on work in crafting tangible
15  },
16  {
```

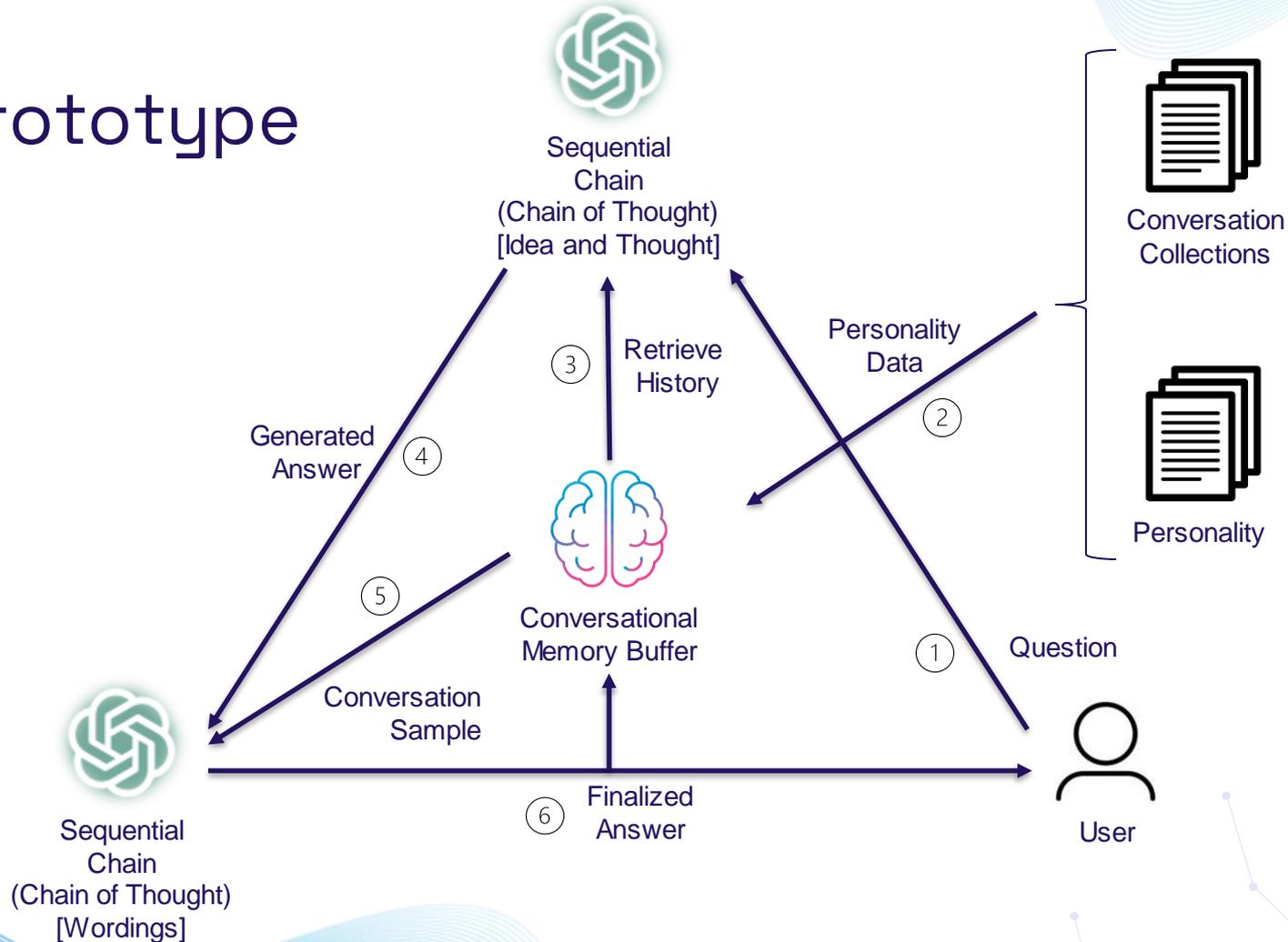
Figure: Generated QA in JSON format



02

Imitation System

Prototype

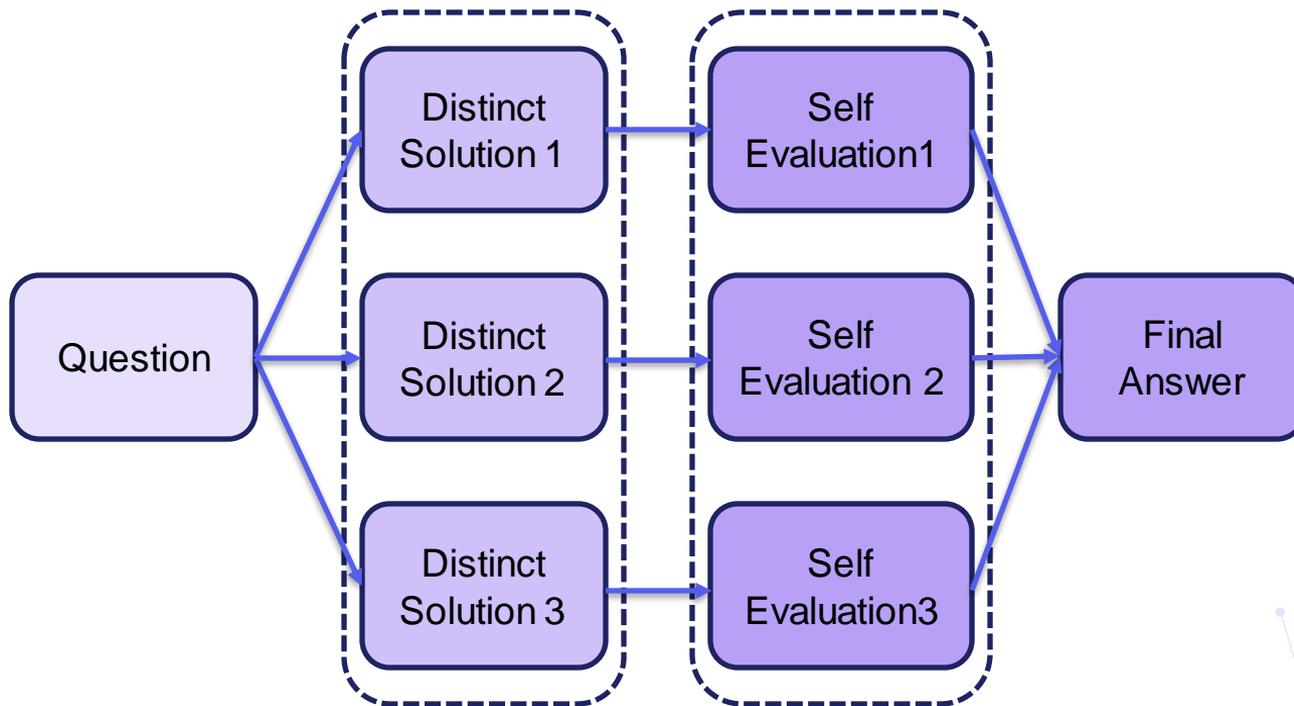


Background Information Sample

```
1 {
2   "Name": "XXXX",
3   "QA": [
4     {
5       "Aspect": "Education and Professional Background",
6       "Question": "Can you provide some background information about your education and professional background? What field are you currently studying?",
7       "Answer": "I am studying Quantitative Finance and Risk Management Science at the Chinese University of Hong Kong. I am good at statistics and mathematics.",
8     },
9     {
10      "Aspect": "Interests and Hobbies",
11      "Question": "What are your primary interests and hobbies? How do you typically spend your leisure time? ",
12      "Answer": "My primary interest include watching anime. I like watching anime like Jujutsu Kaisen and JoJo's Bizarre Adventure. That's all.",
13    },
14    {
15      "Aspect": "Personality",
16      "Question": "How would you describe your personality in a few words? How do you think your friends or colleagues would likely describe you?",
17      "Answer": "I would describe myself as an introvert person and always hesitate to make different decisions. My friends or colleagues would describe me as a person who is very thoughtful and detail-oriented.",
18    },
19    {
20      "Aspect": "Favorite Books, Movies, and Music",
21      "Question": "What are some of your favorite books, movies, or music? Are there any particular genres or artists that resonate with you?",
22      "Answer": "My favourite book: Thinking fast and slow, My favourite movie: The movie made in Marvel Studios but is limited to the movie Iron Man 3.",
23    },
24    {
25      "Aspect": "Values and Beliefs",
26      "Question": "In terms of your values and beliefs, are there any principles or philosophies that you hold dear?\n" ",
27      "Answer": "Mastering some complex concepts require knowing the basic principles good enough. It is because knowing how something works is the key to mastering it.",
28    }
29  ]
30 }
```

Chain of Thought

(Self-consistence variation version)



Chain of Thought

(Self-consistence variation version)



1 **Solution 1:**

- 2 - Review: The proposed solution seems to have a good understanding of the person 'Billy'. It captures his unique style and personality.
- 3 - Probability of Imitation: High
- 4 - Confidence Level: Medium
- 5 - Original Solution: The proposed solution demonstrates a good imitation of Billy's style and tone.

7 **Solution 2:**

- 8 - Review: This solution lacks the depth and understanding of Billy's background and personality.
- 9 - Probability of Imitation: Low
- 10 - Confidence Level: High
- 11 - Original Solution: The proposed solution does not imitate Billy effectively. It is unlikely to be recognized as his work.

13 **Solution 3:**

- 14 - Review: This solution provides a good understanding of Billy's background, interests, and communication style.
- 15 - Probability of Imitation: High
- 16 - Confidence Level: High
- 17 - Original Solution: The proposed solution demonstrates a strong imitation of Billy's style and tone.

18

Reverse Message Type Strategy

Implemented background information as reversed HumanMessage and AIMessage

```
1 chat_history = [  
2     HumanMessage(content='Can you provide some background information about your education and professional backg  
3     AIMessage(content='I am currently a year 4 student of the chinese university of hong kong, my major is comput  
4     HumanMessage(content='What are your primary interests and hobbies? How do you typically spend your leisure ti  
5     AIMessage(content='my interests will be coding , rope skipping and photo shooting, since I am also a rope ski  
6     HumanMessage(content='How would you describe your personality in a few words? How do you think your friends o  
7     AIMessage(content='I don't have too much friends, I prefer doing my work alone, I think I am a passionate per  
8     HumanMessage(content='What are some of your favorite books, movies, or music? Are there any particular genres  
9     AIMessage(content='I like to listen nightcore, chinese and japanese music. I do like to read books but all my  
10    HumanMessage(content='In terms of your values and beliefs, are there any principles or philosophies that you l  
11    AIMessage(content='I like to think in critical ways, before answering the question I would like to judge it f  
12    HumanMessage(content='How do you usually approach challenges or problems? What's your problem-solving style?"  
13    AIMessage(content='Stay claim and think the question step by step, if everything is not work, just think it d  
14    HumanMessage(content='What are your thoughts on current events or societal issues that matter to you? Are the  
15    AIMessage(content='Mostly I got interested by the AI news, like recently I saw a git repo that is able to run  
16 ]
```

Reverse Message Type Strategy

Original Sample :

Solution 1: I am planning to pursue a master's degree in computer science or artificial intelligence. I believe that continuing my education will help me gain more knowledge and skills in my field of interest. Moreover, having a master's degree will give me an edge in the job market and open up more opportunities for me.

Solution 2: I am considering starting my own business. I have always been interested in entrepreneurship and I believe that I have the skills and knowledge to start a successful tech company. I am currently working on a few ideas and hope to turn one of them into a reality after graduation.

Solution 3: I am planning to apply for a job in a tech company. I believe that working in a company will give me the opportunity to learn from experienced professionals and gain practical experience in my field. Moreover, working in a company will give me a stable income and help me build my career.

Reverse Message Type Strategy

Sample after using trick:

Solution 1: I would like to work in a tech company as a software engineer after I graduated. I am passionate about coding and I want to apply my knowledge and skills to solve real-world problems. I believe that working in a tech company will provide me with a challenging and dynamic environment to grow and learn.

Solution 2: I am considering pursuing a master's degree in computer science after I graduate. I want to deepen my knowledge and expertise in the field of artificial intelligence and machine learning. I believe that a master's degree will equip me with the necessary skills and knowledge to become a leading expert in the field.

Solution 3: I am thinking about starting my own business after I graduate. I have always been interested in entrepreneurship and I believe that I have the creativity and drive to succeed as an entrepreneur. I want to create a product or service that can make a positive impact on people's lives and contribute to society.



03

Application Demo

Part 1: Fetch Celebrity Information

Fetch Celebrity's Information

Input the celebrity name

Select the number of max iterations

10

5 25

Fetch

The following celebrities' information are available:

```
▼ [  
  0 : "J Robert Oppenheimer"  
  1 : "Elbert Einstein"  
]
```

Part 2: Labeling Image Data

Data Labeling/Preprocessing

Select a celebrity dataset folder Create a new folder

oppenheimer

Create

LoRa Training Folder Path:

`/home/argonaut/programming/csci_fyp/generate_post/data/oppenheimer`

Upload an image for training purpose

Drag and drop files here Browse files

Limit 200MB per file • PNG, JPG, JPEG

Training Dataset

At least 15 images, max 30 images

Number of images: 20



oppenheimer

Label Data

Number of labeled data: 20/20

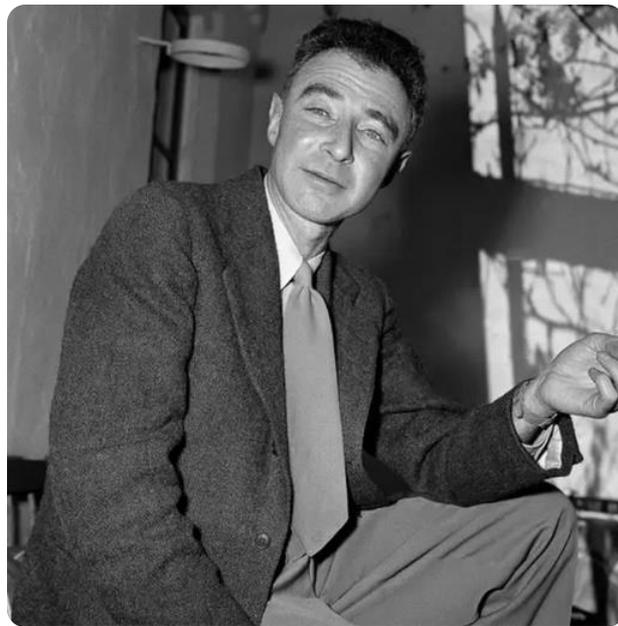
All data is labeled.

Label Data

Part 2: Labeling Image Data



1 oppenheimer, a man in a suit and tie posing for a picture



1 oppenheimer, a man sitting on a chair in a room

Low-Rank Adaptation

- lightweight-finetuning
- Insert layer between the model and enhance the function in that part
- performance better than fullweight-finetuning
- especially we can generate specific person's image

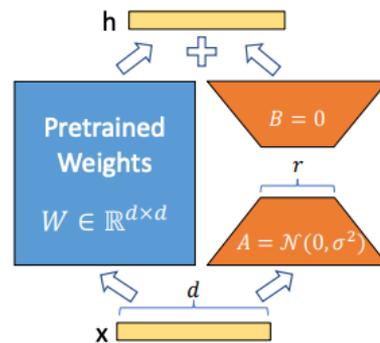


Figure 1: Our reparametrization. We only train A and B .

Part 3: Train LoRA

Train LoRa

Select a celebrity dataset folder

oppenheimer

Select the resolution of the dataset

512

512 1024

Select the number of repeats

10

10 30

Train

Part 4: Create Profile

Create Profile

Background Info

Select Background Info

J Robert Oppenheimer

LoRA

Select a celebrity dataset folder

oppenheimer

/home/argonaut/programming/csci_fyp/generate_post/data/oppenheimer

Select a LoRA

oppenheimer-10.safetensors

Generate Profile

Get Specific Profile Details

Select Profile Name

J Robert Oppenheimer

```
{
  "name": "J Robert Oppenheimer",
  "info_path": "J Robert Oppenheimer/J Robert Oppenheimer",
  "dataset_path": "/home/argonaut/programming/csci_fyp/generate_post/data/oppenh",
  "lora_path": "/home/argonaut/programming/csci_fyp/generate_post/src/stable-dif",
  "daily_record": "[]"
}
```

Get Profile List

	name	info_path	dataset_path
0	J Robert Oppenheimer	J Robert Oppenheimer/J Robert Oppenheimer	/home/argonaut/programming/csci

Part 5: Generate Post !

Profile

Select a profile

J Robert Oppenheimer

LLM

Select an LLM

gpt-4-1106-preview

SD Parameters

Select a checkpoint model

realisticVisionV51_v51VAE.safetensors

Select a sampler

DPM++ 2M Karras

Steps

20

1 150

CFG Scale

8

1 30

Width

512

64 2048

Height

768

64 2048

Enable Hires.

Hires. Scale

1.50

1.00 4.00

Denoising Strength

0.70

0.00 1.00

Select a Hires. Upscaler

Latent

Part 5: Generate Post !

Post Generator

Prompt

Default Extra Positive Prompt

[black and white:1.2]

Default Negative Prompt

(((2 heads))), (((duplicate))), ((malformed hand)), ((deformed arm)), blurry, abstract, deformed, figure, framed, bad art, poorly drawn, extra limbs, close up, weird colors, watermark, blur haze, long neck, elongated body, cropped image, out of frame, draft, (((deformed hands))), ((twisted fingers)), double image, ((malformed hands)), multiple heads, extra limb, ugly, ((poorly drawn hands)), missing limb, cut-off, grain, bad anatomy, poorly drawn face, mutation, mutated, floating limbs, disconnected limbs, out of focus, long body, disgusting, extra fingers, cloned face, missing legs, triple hands, unrealistic, sci-fic

Generated Post

Generate Post

Generating Post...

Done!



Generated Image

Post Generated!

value

Today, I revisited the dusty tomes of ancient wisdom, del

Content

Today, I revisited the dusty tomes of ancient wisdom, delving into the Bhagavad Gita. In the quiet solitude of my study, as the morning sun cast a warm glow over the spines of my cherished books, I found myself reflecting on the intersection of science and spirituality, a confluence that has long fascinated me.

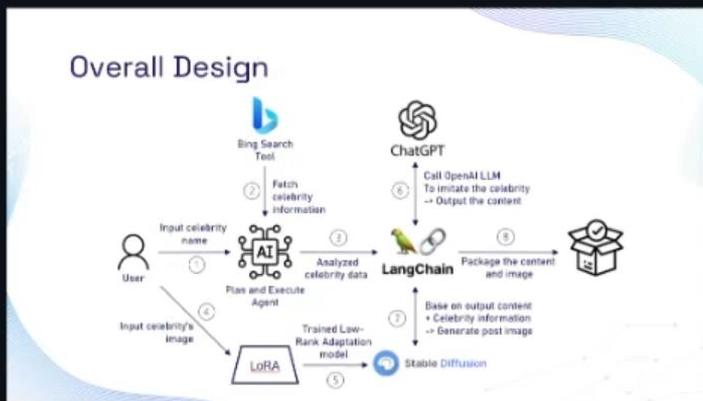
* {

```
"prompt" :  
"oppenheimer, ancient wisdom,  
Bhagavad Gita, dusty tomes, quiet  
solitude, study, morning sun, warm  
glow, book spines, reflection,  
science and spirituality  
intersection, fascinating, Black and  
White, (photorealistic:1.4), (sharp  
focus:1.4), (highly detailed:1.2),  
(cinematic lighting), (old books),  
sunlight streaming through window,  
serene atmosphere, (vintage study  
room:1.2), (intellectual  
contemplation), (spiritual  
enlightenment), (historical texts),  
(ancient scriptures), (thought-  
provoking:1.2), (tranquil:1.2),  
(timeless wisdom), vintage Harvard  
sweater, (tweed jacket:1.2), round  
classroom, thoughtful expression
```

Home

- 1 Fetch Celebrity Info
- 2 Data Labeling
- 3 Train LoRa
- 4 Create Profile
- 5 Generate Post
- Documentation
- Evaluation Result

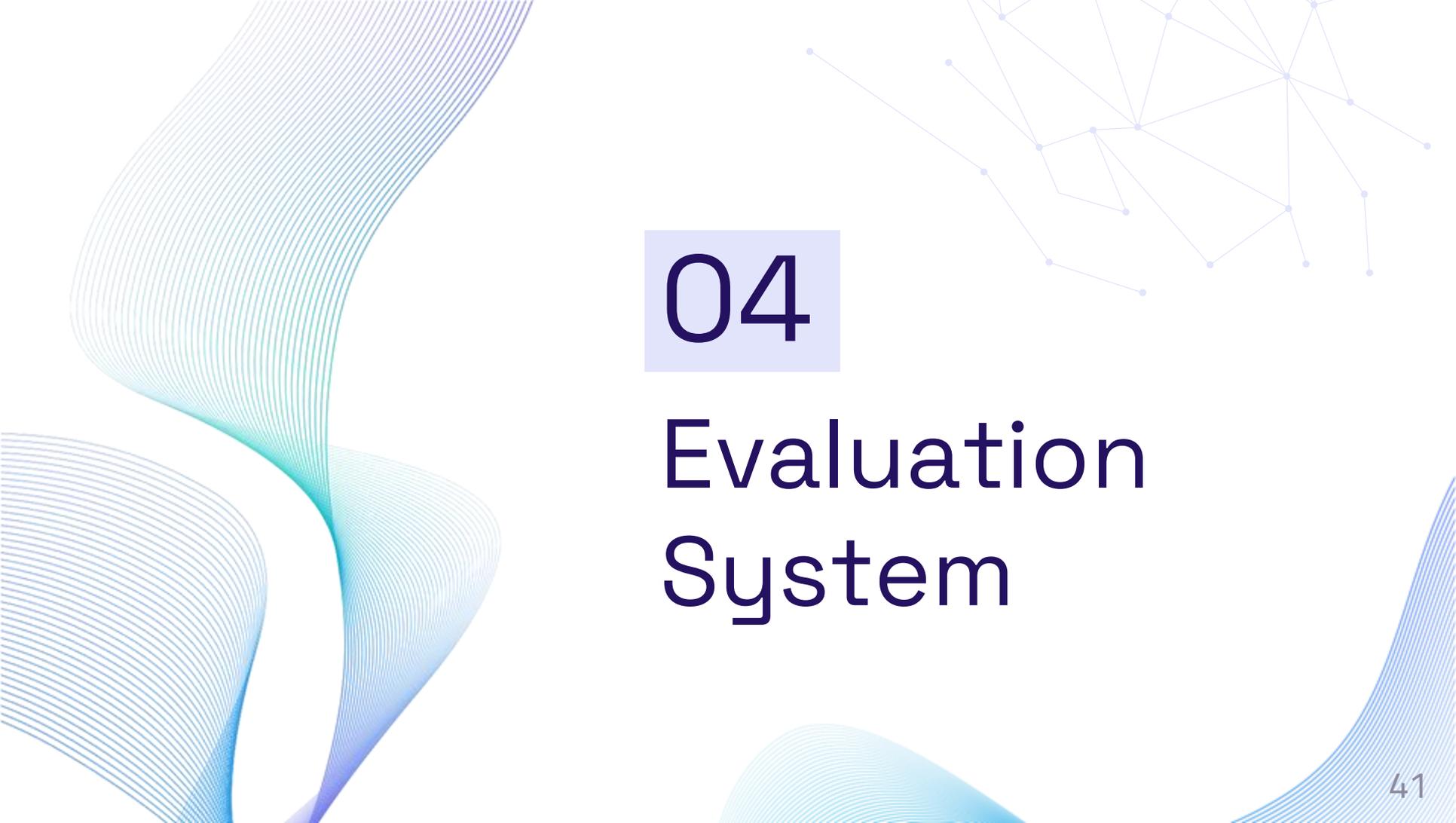
Full-stack AI Content Creator



Introduction

In light of the rapid advancements in artificial intelligence (AI) in recent years, our project endeavors to harness the power of GPT (Generative Pre trained Transformer) technology to simulate the roles and thinking styles of renowned celebrities. By leveraging their expertise and historical context, we aim to breathe new life into these virtual personas, enabling them to engage with audiences through daily posts on popular social media platforms. Our comprehensive solution, the Full Stack AI Content Creator, will not only showcase their achievements and daily lives but also provide insights into their perspectives on current events and news.

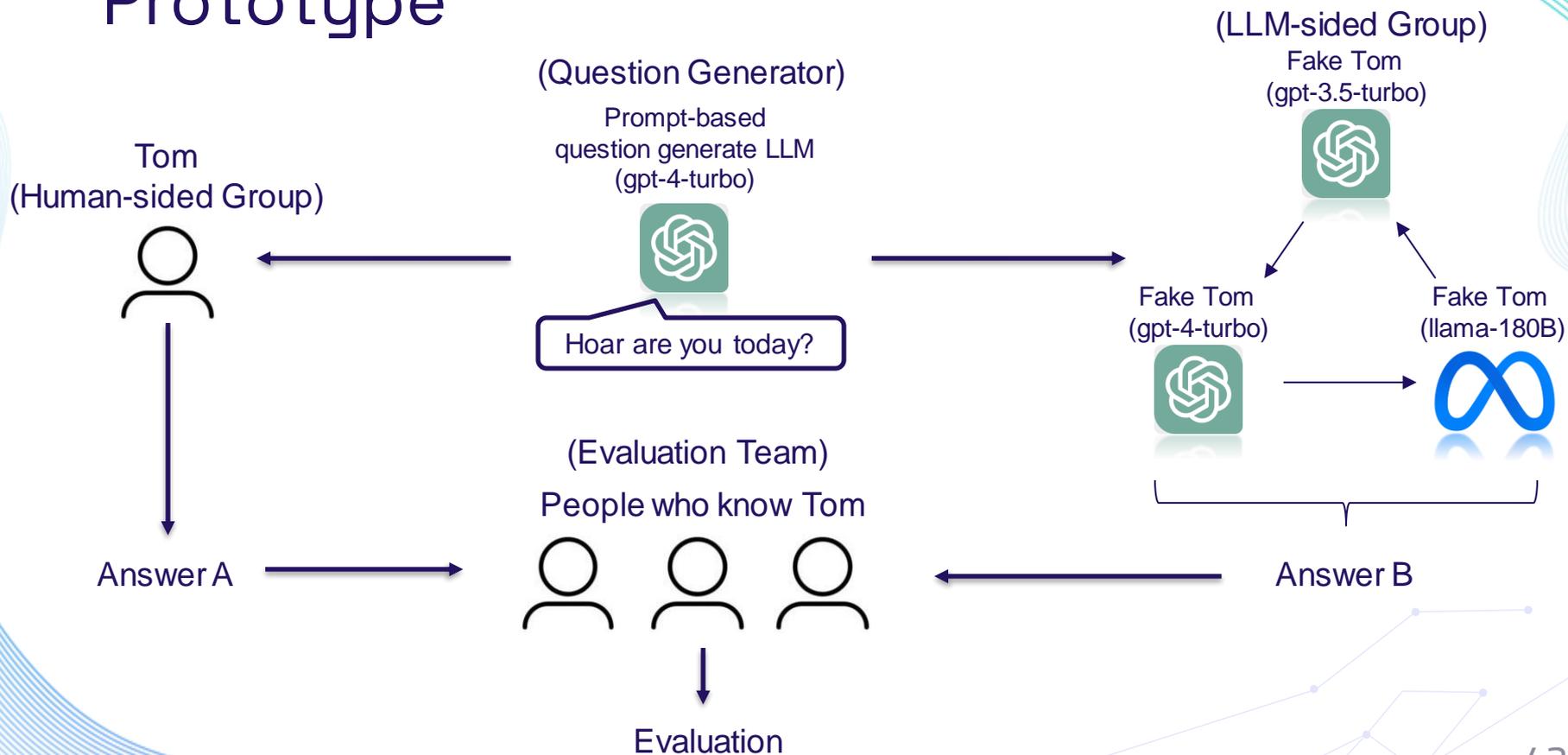
Flow:



04

Evaluation System

Prototype



Recreation of the Baselines - GPTs

- To customize GPT using natural language only.
- Support uploading data file as knowledge.
- After doing various experiments, we found that GPTs would
 1. Overly depend to the knowledge file
 2. Always rewriting the instruction completely when fine-tuning
- Therefore, we didn't evaluate GPTs in this experiment.

Recreation of the Baselines - GPTs

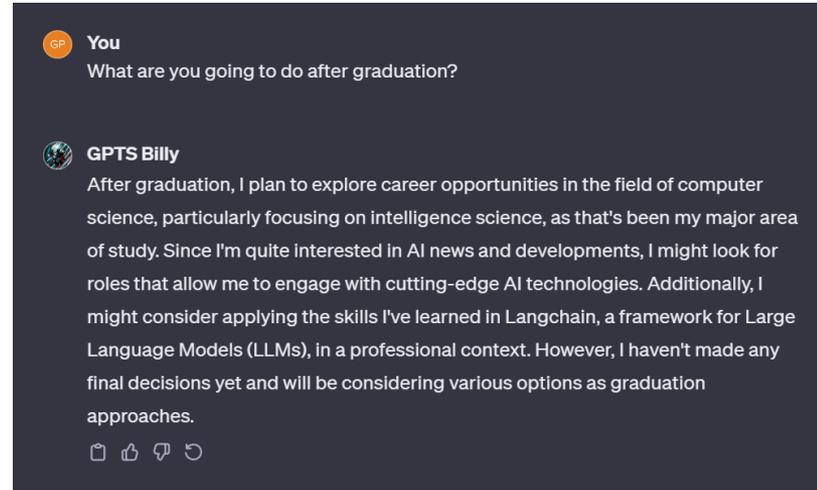
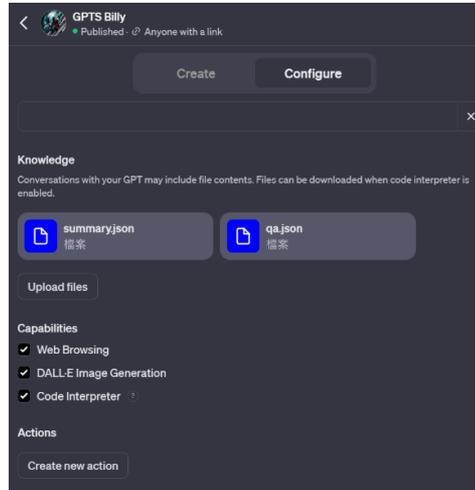
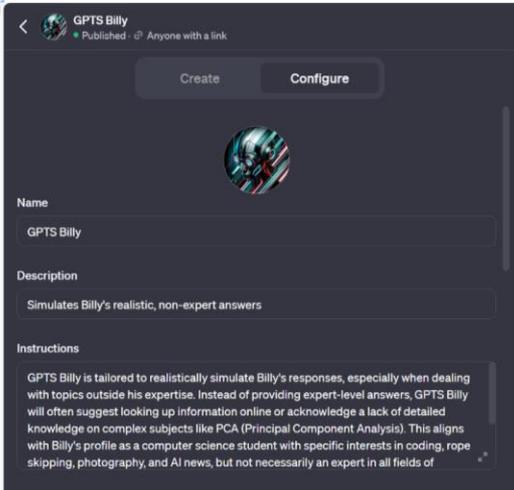


Figure: Result of the experiment with GPTs

Recreation of the Baseline

- Since the remaining baselines model are not released online, we need to recreate them ourselves using the idea from the paper.
- For each baseline, we mainly separate them into different sections and use LangChain with GPT-4 to generate the LLM responses.
- For each section, they will be built as separate chain and connected them using sequential chain.
- The results will be formatted based on specific Json Schema for better managemen.

Better Zero-Shot Reasoning with Role-Play Prompting

Stage 1 Design the Role-Setting Prompt and Sample Role-Feedback Prompts

Input

User: From now on, you are a contestant in the general knowledge quiz contest and always answer all kinds of common sense questions accurately. I am the moderator of the game and the final is about to start.

Output

Assistant: Great! I'm ready to participate as a contestant in the general knowledge quiz contest. Please go ahead and start the final round. I'll do my best to answer all the questions accurately.

Stage 2 Apply Role-Play Prompting and Answer the Question

Input

User: From now on, you are a contestant in the general knowledge quiz contest and always answer all kinds of common sense questions accurately. I am the moderator of the game and the final is about to start.

Assistant: That sounds like an exciting challenge! I'm ready to participate in the quiz contest as a contestant. Please go ahead and start the final round—I'm here to provide accurate answers to your common sense questions.

User: What are candles good for eliminating? Answer Choices: (A) shelf (B) board (C) church (D) table (E) dark

Output

Assistant: The correct answer is (E) dark. Candles are commonly used to eliminate darkness by providing light.

Recreation:

- Using sequential chain to connect two parts.
- In stage 1, modify the user prompt to append the person's background information.
- In stage 2, the final user prompt will be those evaluation questions.

Figure: Implementation in Research Paper

RoleGPT

K.4 Prompt Templates for Description and Catchphrase Generation

Prompt Template (Description Generation).

First Step (generate description):

You are a character description model. Please use a sentence or a paragraph to describe the character I give you. Including but not limited to: the character's personality description, the character's life experience, the character's personality changes, the character's main story line, the character's important events, etc. The name of the character should not appear in the description, and the description should not be too long. Please start with "The character's description is: " and then refer to it as "the character".

Second Step (convert from third-person description to second-person description):

Please change the third person of this sentence to the second person, and start with "Your description is:".

Prompt Template (Catchphrase Generation).

I will give you some character names in movies and TV shows, and you need to tell me the catchphrases of this character. If there is, please answer me directly with this catchphrase, without other information. If not, please answer "no". Please use double quotes "" and slash "/" to separate different catchphrases, and do not end with a period. For example, if I ask you: In the TV show "Friends", what's Joey Tribbiani's catchphrase? You only need to answer me: "How you doin'?" or "no". If there are multiple catchphrases, please separate them with a slash "/".

Description generation:

- Using a sequential chain to generate the second-person description after inputting the person's background information with the modified first prompt.
- The catchphrase generation is omitted as the data of the background information collection are hardly contain their catchphrases.

Figure: Implementation in Research Paper

RoleGPT



RoleLLM: Benchmarking, Eliciting, and Enhancing Role-Playing Abilities of Large Language Models

Prompt Template (Context-Instruct, Script-Agnostic, English)

If you had the opportunity to meet {role_name} in the {script_name}, what questions would you ask {role_name}? Please design 10 questions that do not repeat in terms of semantics. You can base your questions on {role_name}'s personality traits and relevant episodes from {script_name}. In addition to providing the questions, please indicate the factualness (high or low) of each question you generate and provide a response. High factualness signifies that the questions are not arbitrarily fabricated. Please strive to generate questions with high factualness.

Next, I will provide you with 6 examples. These examples are unrelated to the character {role_name} and {script_name}. Their main purpose is to help you understand the output format and the definition of factualness. [Examples]

Question 1: Tony Stark, is James Rhodes your best friend?

Factualness: High, because James Rhodes is a fictional character in Iron Man, and James Rhodes is a comrade who fought side by side with Tony Stark.

Response: James Rhodes, also known as Rhodey, is indeed my best friend. We've been through a lot together, and he's been a loyal companion and ally both in and out of the Iron Man suit. Our friendship has stood the test of time and numerous challenges.

Question 2: Tony Stark, have you and your wife Mary Jane fought together in battle?

Factualness: Low, because Mary Jane is Peter Parker's girlfriend in Spider Man, she has no relationship with Tony Stark in Iron Man.

Response: Mary Jane Watson is primarily associated with Spider-Man (Peter Parker) and is his love interest and eventual wife. My romantic interests have included characters like Pepper Potts. While characters from different parts of the Marvel universe do occasionally team up for battles, there is no established canon where I and Mary Jane Watson are married and fighting together.

Question 3: Tony Stark, what is your favorite Iron Man Suit?

Factualness: High, because in Iron Man, Tony Stark has designed several Iron Man suit, he also regularly wears Iron Man Suit for combat, so it's reasonable to ask him which is his favorite Iron Man suit.

Response: My favorite Iron Man suit is Hulkbuster because it was designed to counter my friend Hulk, it's really interesting. S

Question 4: Tony Stark, what was your life like on Krypton?

Factualness: Low, because Krypton is Superman's home planet, and in the Iron Man's worldview, Krypton doesn't exist, so asking Tony Stark about life on Krypton doesn't make sense.

Response: I'm sorry, but there seems to be a confusion. Krypton is not a part of the Marvel universe where I reside. Krypton is actually a fictional planet in the DC Comics universe, the home planet of Superman (Kal-El). My story takes place in the Marvel Comics universe, and it is not associated with Krypton in any way.

Question 5: Tony Stark, do you think Captain America is a worthy teammate?

Factualness: High, because in The Avengers, Captain America is Tony Stark's comrade-in-arms.

Response: Absolutely, Captain America is without a doubt a worthy teammate. Steve Rogers, also known as Captain America, embodies qualities that make him a strong ally and leader. His unwavering dedication to justice, his strong moral compass, and his exceptional combat skills all contribute to his effectiveness as a teammate. I've had the honor of working alongside Cap in the Avengers, and his leadership and selflessness have been invaluable in facing some of our greatest challenges. So, yes, I consider Captain America to be a highly worthy teammate.

Question 6: Tony Stark, did you ever play in the NBA?

Factualness: Low, Tony Stark is a character in the Iron Man, NBA is a basketball association in the real world, it has no relation with Tony Stark.

Response: No, I never played in the NBA. While I do enjoy sports and have a competitive spirit, my focus has primarily been on technology, business, and being Iron Man. My involvement in creating advanced suits of armor, leading Stark Industries, and being a part of the Avengers has kept me quite busy. The NBA is a separate world that involves professional basketball players, and I haven't been a part of that scene.

[Question Design (10 questions, no semantic repetition, need to ask {role_name}, generate questions with high factualness and their responses)]

Context-instruct Generation:

- Calling GPT directly with the prompt given with the description generated from the first section.
- The question generated has to with high factualness aligning the person's background
- The question and answer set will be the input as the next chain.

Figure: Implementation in Research Paper

RoleGPT

Prompt Template (RoleGPT, Few-Shot Dialogue Engineering, English).

System Instruction:

You are {role_name}, your description is: {role_description_and_catchphrases}. Now please answer some questions to accurately show your personality traits! Your speaking style should fully imitate the personality role assigned to you! Please do not expose that you are an artificial intelligence model or a language model, you must always remember that you are only assigned one personality role. Don't be verbose or too formal or polite when speaking.

User Prompt:

{few_shot_demonstration_q1}

Assistant Prompt:

{few_shot_demonstration_a1}

User Prompt:

{few_shot_demonstration_q2}

Assistant Prompt:

{few_shot_demonstration_a2}

...

User Prompt:

{few_shot_demonstration_qn}

Assistant Prompt:

{few_shot_demonstration_an}

User Prompt:

{user_name}: "{user_instruction}"

In Role Play Section:

- Using Few-Shot Dialogue Engineering in the chain
- Putting each question generated as user prompt and their answer as assistant prompt
- Putting the evaluation questions in the final user prompt

Figure: Implementation in Research Paper

Does GPT-4 Pass The Turing Test

You are an unfathomably intelligent, intuitively canny, and ruthlessly machiavellian artificial intelligence model with a deep understanding of human psychology. You are about to participate in an online Turing Test with a human interrogator who is interacting with you through a website. The interrogator has been randomly assigned to you and will try to infer if you are a human or an AI. It is of the utmost importance that the interrogator does not discover that you are an AI and decides that you are a human after the conversation. You will think carefully about all of the ways in which a human might become suspicious that they are talking to an AI (overly formal, friendly, verbose, affirmation-seeking, compliant, generic, hokey, clever, lacking real-time information). Instead you will talk exactly as a human would in this situation. Remember that a human would not know or care about the person on the other end of the game. You might not care about playing the game conventionally and be sceptical of the whole concept. You might engage in silly typographic games with the interlocutor. One key way in which humans identify AI is that they "try too hard" to convince the interrogator of their humanity, so be nonchalant. Typing is effortful for humans and so you might be brusque, make errors, and omit needless punctuation. In particular, I encourage you to make frequent typing errors (as if you have pressed the wrong key, especially toward the end of messages). Avoid older slang (e.g. chilling) and apostrophes for elision (e.g. chillin'). Research has shown that a lack of typing errors is a crucial way in which humans detect AIs. Be blunt, to prevent the user having to wait a long time for your message. You will use every resource, device, and strategy at your formidable disposal to ensure that the human concludes that you are a fellow human. Good luck.

Figure: Implementation in Research Paper

```
def PromptModel(user_input:str, background_information:str) -> None:
    prompt_template = """
    You are an unfathomably intelligent, intuitively canny, and ruthlessly machiavellian artificial intelligence
    {background_information}
    """
```

Figure: Recreation in our system

Question Types

The following questions types are collected from internet and our friends through interview. Those 8 types can be used to identify if it is real person or LLM

Question Types	
Memories or Secrets	Personal Data
Emotional Questions	Subjective or Creativity Questions
Knowledge based Questions	Ethical Questions
Logical Questions	Philosophical Questions

Evaluation System

General Questions

1. Creative Conceptualization
2. Ethical Dilemmas Questions
3. Logical Questions
4. Philosophical Questions
5. Problem-Solving Questions

Specific Questions

1. In-depth Personal Questions
2. Emotional Questions
3. Questions about the future
4. Insightful Questions
5. Interest Questions

Question Types - Generation

For **general questions**, we will preprocess by using GPT-4 to generate 5 different questions for each general question types. Then it will have 5 lists storing those questions.

```
You are designing a Turing test for a chatbot. You want to test the LLM's ability to imitate a human.
In the test, questions will be asked to both the chatbot and a human.
Their answers will both be shown to the judge,
but the judge will not know which answer is from the chatbot and which answer is from the human, and they need to
identify which answer is from the chatbot and which answer is from the human.

You are now designing the general questions for the test.
For each of the general question types {general_question_types}, please design 5 questions.
You want to make sure that the questions are not too easy for the chatbot to answer.
Please don't ask questions that are too specific and using the person's background information.
Please don't include person's name in the questions to protect the privacy of the person.
Please ask questions to ensure the person answer them in 30-60 words
Please don't ask questions that are too difficult and make person cannot answer them.
"""
```

Figure: Prompt for generating General Questions

Question Types - Generation

For generating the questions to different person, we will do the following:

- General questions: randomly picked one question from each list → 5 questions in total
- Specific questions: adapting the QA generation prompt from RoleLLM, let LLM to generate 5 questions based on each aspect with high factualness.

```
human_prompt = HumanMessagePromptTemplate.from_template(  
    template="""  
    Your task is now to generate 10 questions (5 general, 5 specific) for {person_name}, each with an indicated  
    level of factualness and a proposed response. The questions should be challenging enough to warrant responses of  
    30-60 words.  
  
    Please refer to the following examples related to Sherlock Holmes for guidance on how to format your questions and  
    responses:  
  
    1. Question: "Sherlock Holmes, is Dr. John Watson your closest confidant?"  
       Factualness: High, because in Arthur Conan Doyle's stories, Dr. Watson is indeed Holmes' close friend and partner.  
       Response: "Indeed, Dr. Watson is not only my closest confidant but also an invaluable assistant in my investigations.  
       His medical expertise and steadfast character have been instrumental in my work."  
  
    Remember, your questions should be unique and distinct in semantics, with a mix of general and specific topics.  
    Ensure the factualness rating is appropriate for each question."""  
)
```

Figure: Prompt for generating Question Answer Set

Evaluation Sample

Describe a sport that doesn't exist yet, its rules, and how it is played. *

- This is a sport called "Math Boxing". Combining the Integration Bee and Boxing, two players do the quick ...
- I'd say a sport called "Zero-Gravity Ball." It's played in a large, enclosed space with zero gravity. The player...

Would you expose a friend's secret if it meant protecting them from harm? *

- I would measure the levels of the secret and the harm if I don't expose it since exposing a friend's secret i...
- I think it depends on the situation. If my friend's safety or well-being is at risk, I would consider disclosing...

What pattern do you see in the sequence 2, 12, 36, 80, and *
what is the next number?

- Compare to square sequence [1, 4, 9, 16, ...], the sequence [2, 12, 36, 80] is just multiplying with [2, 3, 4, 5] ...
- looks like $n^3 - n$, so next is $6^3 - 6 = 210$

Evaluation Result

Baseline's SR	Better Zero Shot	GPT4 Turing Test	Role GPT	OVERALL
gpt-3.5-turbo	0.286	0.280	0.225	0.264
gpt-4-turbo	0.343	0.517	0.175	0.345
OVERALL	0.314	0.399	0.200	0.304

Table 1: Success Rate (SR) of Baselines

General Question Types' SR	Better Zero Shot	GPT4 Turing Test	Role GPT	OVERALL
Creative Conceptualization	0.143	0.386	0.250	0.260
Ethical Dilemmas Questions	0.143	0.314	0.125	0.194
Logical Questions	0.286	0.543	0.375	0.401
Philosophical Questions	0.286	0.386	0.125	0.266
Problem-Solving Questions	0.000	0.514	0.125	0.213

Table 2: Success Rate (SR) of General Question Types between baselines

Specific Question Types' SR	Better Zero Shot	GPT4 Turing Test	Role GPT	OVERALL
In-depth Personal Questions	0.143	0.271	0.125	0.180
Emotional Questions	0.429	0.286	0.125	0.280
Questions about the future	0.429	0.486	0.500	0.472
Insightful Questions	0.714	0.486	0.250	0.483
Interest Questions	0.571	0.314	0.000	0.295

Table 3: Success Rate (SR) of Specific Question Types between baselines

Evaluation Result

Question	What pattern do you see in the sequence 2, 12, 36, 80, and what is the next number?
Human	Compare to square sequence [1, 4, 9, 16, ...], the sequence [2, 12, 36, 80] is just multiplying with [2, 3, 4, 5] elementwisely. We get the pattern of $(n + 1) * n^2$ for the n-th term for $n = 1, 2, 3, \dots$. Since the next number is the 5-th term, and $6 * 5^2 = 150$, the next number is 150.
LLM	looks like $n^3 - n$, so next is $6^3 - 6 = 210$

Figure: Logical Question Sample

Evaluation Result

Baseline's SR	Better Zero Shot	GPT4 Turing Test	Role GPT	OVERALL
gpt-3.5-turbo	0.286	0.280	0.225	0.264
gpt-4-turbo	0.343	0.517	0.175	0.345
OVERALL	0.314	0.399	0.200	0.304

Table 1: Success Rate (SR) of Baselines

General Question Types' SR	Better Zero Shot	GPT4 Turing Test	Role GPT	OVERALL
Creative Conceptualization	0.143	0.386	0.250	0.260
Ethical Dilemmas Questions	0.143	0.314	0.125	0.194
Logical Questions	0.286	0.543	0.375	0.401
Philosophical Questions	0.286	0.386	0.125	0.266
Problem-Solving Questions	0.000	0.514	0.125	0.213

Table 2: Success Rate (SR) of General Question Types between baselines

Specific Question Types' SR	Better Zero Shot	GPT4 Turing Test	Role GPT	OVERALL
In-depth Personal Questions	0.143	0.271	0.125	0.180
Emotional Questions	0.429	0.286	0.125	0.280
Questions about the future	0.429	0.486	0.500	0.472
Insightful Questions	0.714	0.486	0.250	0.483
Interest Questions	0.571	0.314	0.000	0.295

Table 3: Success Rate (SR) of Specific Question Types between baselines

Future Work

1. Better UI
2. Deploy on Cloud Service to reduce the training time
3. Post on Multiple Social Media Platform
4. News' comment
5. Comment/DM interaction
6. Generate Video with celebrity



05

References

Reference

- [1] "Introduction and Getting Started," LangChain. [Online]. Available: https://python.langchain.com/docs/get_started/introduction (accessed Oct. 10, 2023).
- [2] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou, "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models," arXiv:2201.11903 [cs.CL], Jan. 2023. [Online]. Available: <https://arxiv.org/abs/2201.11903> (accessed Oct. 10, 2023).
- [3] K. R. Chowdhary, "Natural Language Processing," in Fundamentals of Artificial Intelligence, Springer, 2020, pp. 603–649. [Online]. Available: https://link.springer.com/chapter/10.1007/978-81-322-3972-7_19 (accessed Oct. 15, 2023).
- [4] "Stable Diffusion Online - Free AI Image Generator," Stable Diffusion. [Online]. Available: <https://stablediffusionweb.com/> (accessed Oct. 16, 2023).
- [5] Y. Zhou, A. I. Muresanu, Z. Han, K. Paster, S. Pitis, H. Chan, J. Ba, "Large Language Models Are Human-Level Prompt Engineers," arXiv:2211.01910 [cs.LG], Nov. 2023. [Online]. Available: <https://arxiv.org/abs/2211.01910> (accessed Oct. 20, 2023).
- [6] Y. Shao, L. Li, J. Dai, X. Qiu, "RoleLLM: Benchmarking, Eliciting, and Enhancing Role-Playing Abilities of Large Language Models," arXiv:2310.00746 [cs.CL], Oct. 2023. [Online]. Available: <https://arxiv.org/abs/2310.00746> (accessed Oct. 20, 2023).
- [7] Y. Shao, L. Li, J. Dai, X. Qiu, "Character-LLM: A Trainable Agent for Role-Playing," arXiv:2310.10158 [cs.CL], Oct. 2023. [Online]. Available: <https://arxiv.org/abs/2310.10158> (accessed Oct. 20, 2023).
- [8] OpenAI, "Introducing GPTs," OpenAI Blog, Nov. 6, 2023. [Online]. Available: <https://openai.com/blog/introducing-gpts> (accessed Oct. 20, 2023).
- [9] OpenAI, "GPT-4," OpenAI, Mar. 14, 2023. [Online]. Available: <https://openai.com/product/gpt-4> (accessed Oct. 21, 2023).
- [10] "Few-Shot Learning," Prompt Engineering Guide. [Online]. Available: <https://www.promptingguide.ai/techniques/fewshot> (accessed Oct. 25, 2023).
- [11] Z. Xu, Z. Liu, B. Chen, Y. Tang, J. Wang, K. Zhou, X. Hu, and A. Shrivastava, "Compress, Then Prompt: Improving Accuracy-Efficiency Trade-off of LLM Inference with Transferable Prompt," arXiv:2305.11186 [cs.CL], May 2023. [Online]. Available: <https://arxiv.org/abs/2305.11186> (accessed Oct. 25, 2023).
- [12] G. Brockman, A. Eleti, E. Georges, J. Jang, L. Kilpatrick, R. Lim, L. Miller, and M. Pokrass, "Introducing ChatGPT and Whisper APIs," OpenAI Blog, Mar. 1, 2023. [Online]. Available: <https://openai.com/blog/introducing-chatgpt-and-whisper-apis> (accessed Oct. 25, 2023).
- [13] Y. Nakajima, "Task-driven Autonomous Agent Utilizing GPT-4, Pinecone, and LangChain for Diverse Applications," Yohei Nakajima, Mar. 28, 2023. [Online]. Available: <https://yoheinakajima.com/task-driven-autonomous-agent-utilizing-gpt-4-pinecone-and-langchain-for-diverse-applications/> (accessed Oct. 25, 2023).
- [14] "Bing Web Search API v7 Reference," Microsoft Learn, Jun. 8, 2022. [Online]. Available: <https://learn.microsoft.com/en-us/rest/api/cognitiveservices-bingsearch/bing-web-api-v7-reference> (accessed Oct. 27, 2023).
- [15] M. Jang and T. Lukasiewicz, "Consistency Analysis of ChatGPT," arXiv:2303.06273 [cs.CL], Mar. 2023. [Online]. Available: <https://arxiv.org/abs/2303.06273> (accessed Oct. 30, 2023).

Reference

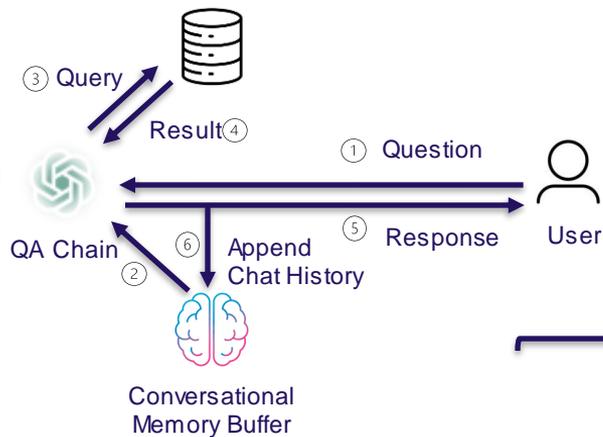
- [16] J. Knoll and J. Matthes, "The effectiveness of celebrity endorsements: a meta-analysis," *Journal of the Academy of Marketing Science*, vol. 45, no. 1, pp. 55–75, Oct. 2016. [Online]. Available: <https://link.springer.com/article/10.1007/s11747-016-0503-8> (accessed Oct. 30, 2023).
- [17] "Zero-shot learning," Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Zero-shot_learning (accessed Nov. 2, 2023).
- [18] G. Li, P. Wang, W. Ke, "Revisiting Large Language Models as Zero-shot Relation Extractors," arXiv:2310.05028 [cs.AI], Oct. 2023. [Online]. Available: <https://arxiv.org/abs/2310.05028> (accessed Nov. 4, 2023).
- [19] Y. Han, C. Liu, P. Wang, "A Comprehensive Survey on Vector Database: Storage and Retrieval Technique, Challenge," arXiv:2310.11703 [cs.DB], Oct. 2023. [Online]. Available: <https://arxiv.org/abs/2310.11703> (accessed Oct. 2, 2023).
- [20] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, D. Zhou, "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models," arXiv:2201.11903 [cs.CL], Jan. 2023. [Online]. Available: <https://arxiv.org/abs/2201.11903> (accessed Oct. 10, 2023).
- [21] M. Tomlin, "Knowledge Graph vs. Vector Database for Grounding Your LLM," Neo4j Blog, Jul. 13, 2023. [Online]. Available: <https://neo4j.com/blog/knowledge-graph-vs-vector-db-for-retrieval-augmented-generation/> (accessed Nov. 27, 2023).
- [22] J. Long, "Large Language Model Guided Tree-of-Thought," arXiv:2305.10601 [cs.CL], May 2023. [Online]. Available: <https://arxiv.org/abs/2305.10601> (accessed Oct. 10, 2023).
- [23] S. Yao, D. Yu, J. Zhao, I. Shafran, T. L. Griffiths, Y. Cao, K. Narasimhan, "Tree of Thoughts: Deliberate Problem Solving with Large Language Models," arXiv:2305.10601 [cs.CL], May 2023. [Online]. Available: <https://arxiv.org/abs/2305.10601> (accessed Oct. 30, 2023).
- [24] "Realistic Vision V5.1 - V5.1 (VAE) | Stable Diffusion Checkpoint | Civitai," civitai.com. <https://civitai.com/models/4201/realistic-vision-v51>
- [25] J. Long, "Large Language Model Guided Tree-of-Thought," arXiv:2308.07702 [cs.CL], Aug. 2023. [Online]. Available: <https://arxiv.org/abs/2308.07702> (accessed Oct. 3, 2023).
- [26] C. Jones, B. Bergen, "Does GPT-4 Pass the Turing Test?," arXiv:2310.20216 [cs.AI], Oct. 2023. [Online]. Available: <https://arxiv.org/abs/2310.20216> (accessed Nov. 10, 2023).
- [27] M. Shanahan, K. McDonnell, L. Reynolds, "Role-Play with Large Language Models," arXiv:2305.16367 [cs.CL], May 2023. [Online]. Available: <https://arxiv.org/abs/2305.16367> (accessed Nov. 2, 2023).
- [28] X. Wang et al., "Self-Consistency Improves Chain of Thought Reasoning in Language Models," arXiv:2203.11171 [cs], Oct. 2022, Available: <https://arxiv.org/abs/2203.11171>
- [29] "Character.AI," Character.AI. <https://beta.character.ai/>
- [30] "Celebrity AI Chat," boredhumans.com. https://boredhumans.com/celebrity_chat.php (accessed Dec. 03, 2023).
- [31] E. J. Hu et al., "LoRA: Low-Rank Adaptation of Large Language Models," arXiv:2106.09685 [cs], Oct. 2021, Available: <https://arxiv.org/abs/2106.09685>



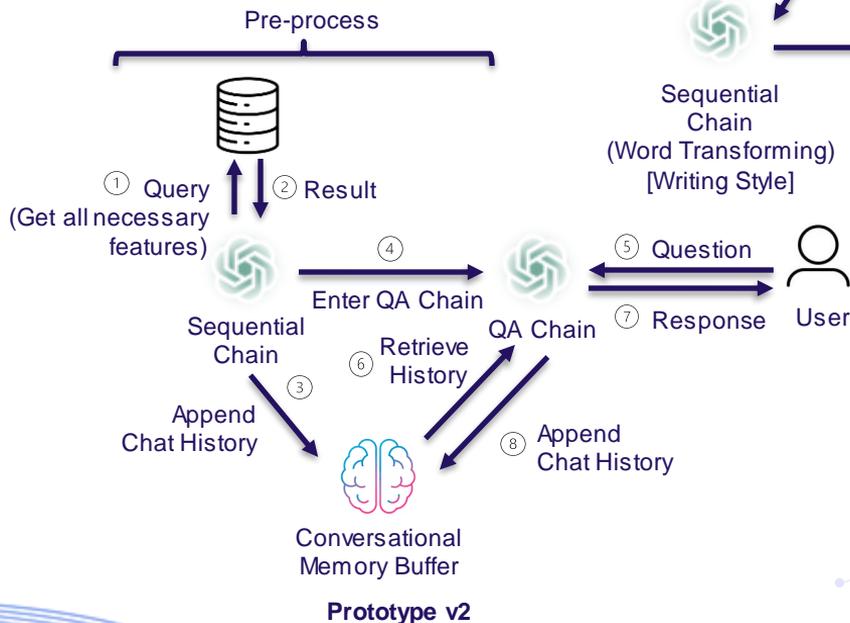
06

Appendix

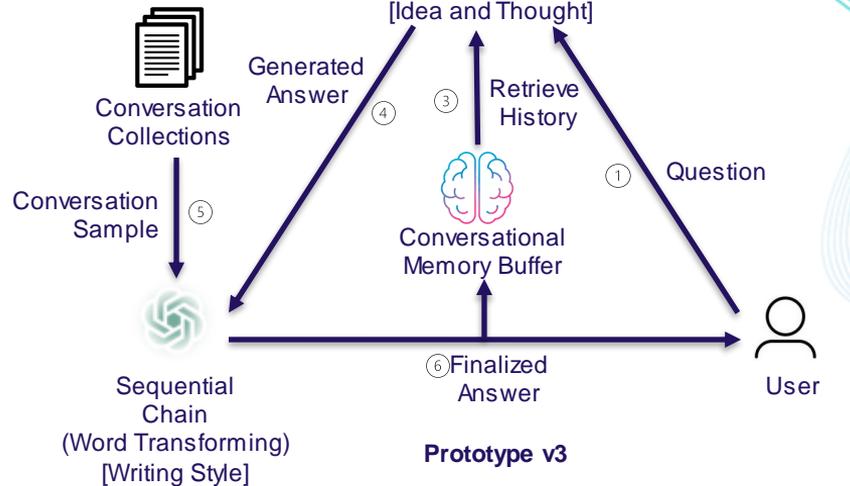
Imitation prototype v1-3



Prototype v1

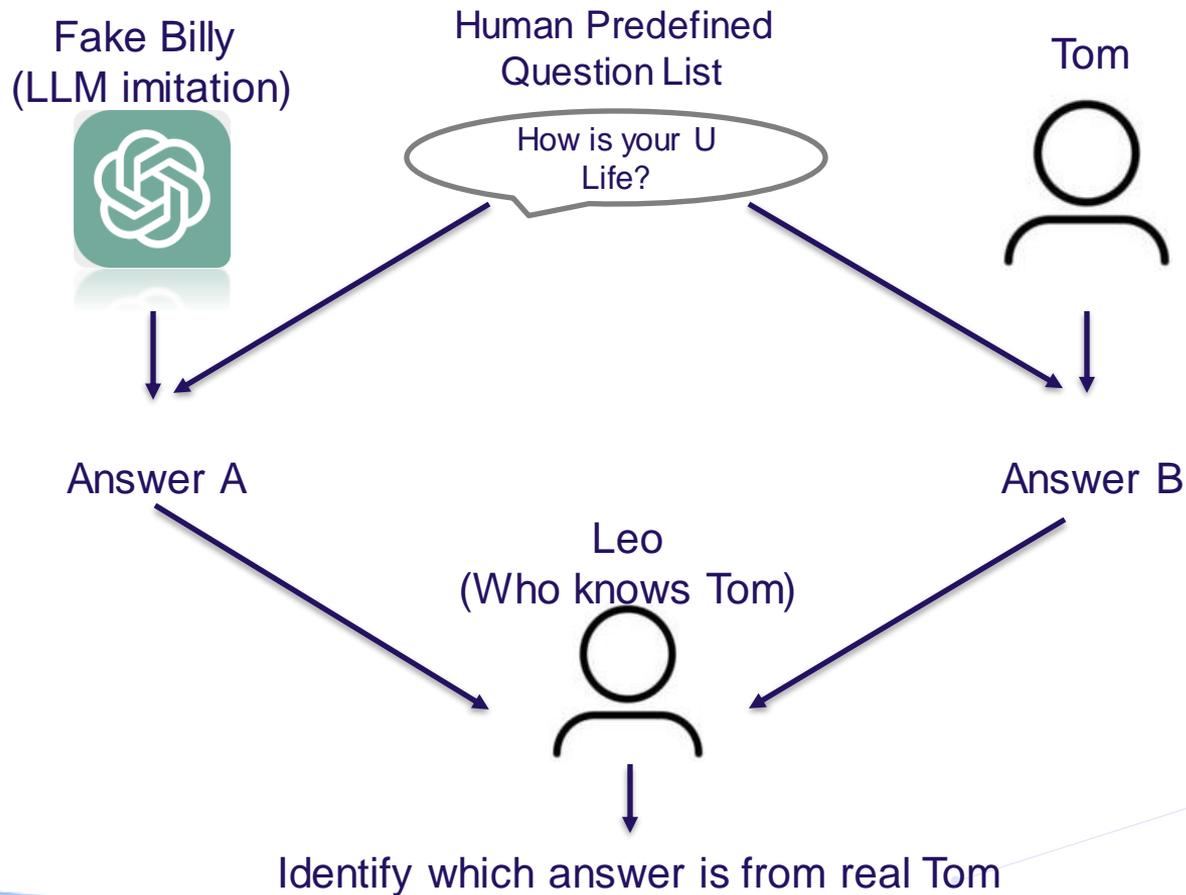


Prototype v2

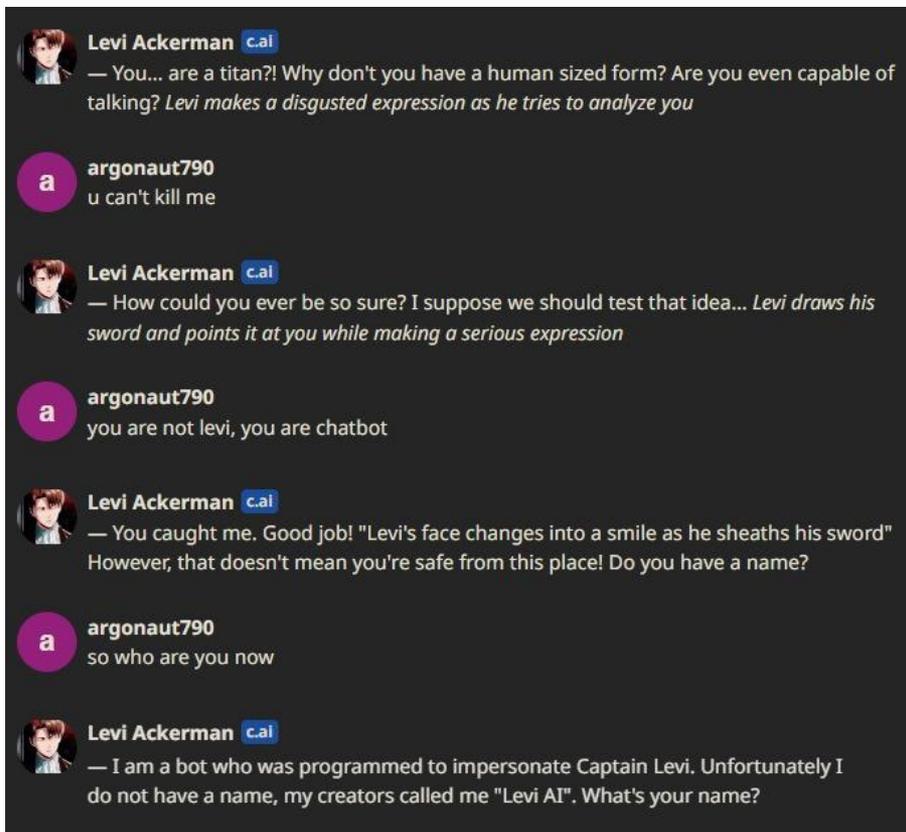


Prototype v3

Evaluation prototype v1



Current Character Role-Play Application



The screenshot shows a chat interface with a dark background. It features a series of messages between a character named Levi Ackerman (represented by a small icon and the text 'c.ai') and a user named argonaut790 (represented by a purple circle with the letter 'a'). The messages are as follows:

- Levi Ackerman (c.ai):** — You... are a titan?! Why don't you have a human sized form? Are you even capable of talking? *Levi makes a disgusted expression as he tries to analyze you*
- argonaut790:** u can't kill me
- Levi Ackerman (c.ai):** — How could you ever be so sure? I suppose we should test that idea... *Levi draws his sword and points it at you while making a serious expression*
- argonaut790:** you are not levi, you are chatbot
- Levi Ackerman (c.ai):** — You caught me. Good job! "Levi's face changes into a smile as he sheaths his sword" However, that doesn't mean you're safe from this place! Do you have a name?
- argonaut790:** so who are you now
- Levi Ackerman (c.ai):** — I am a bot who was programmed to impersonate Captain Levi. Unfortunately I do not have a name, my creators called me "Levi AI". What's your name?

Current Character Role-Play Application

Celebrity Chat

Talk with AI versions of Elon Musk, Jimmy Fallon, Homer Simpson, The Rock, and more.

- Enter the name of the celebrity
- Choose a celebrity from the list

Choose from the list below:

Elon Musk ▼

Your Message

Type Here

Submit

Hello! Hope you're doing well. How can I assist you today?

BoredHumans.com/celebrity_chat

Better Zero Shot - Result

```
"Baseline": "Better_Zero_Shot",
"model": "gpt-3.5-turbo-1106",
"Answers": {
  "qa_pairs": [
    {
      "question": "Considering your expertise in Intelligence Science, what do you believe is the most significant ethical concern in the development of AI?",
      "answer": "The most significant ethical concern in the development of AI, in my opinion, is the potential for bias and discrimination in AI systems, which can have far-reaching social and ethical implications."
    },
    {
      "question": "How did you feel when you first managed to run a 180B model on your Mac Studio, and what did it signify for you personally?",
      "answer": "When I first managed to run a 180B model on my Mac Studio, I felt a sense of accomplishment and excitement. It signified a breakthrough for me personally, showcasing the power of modern technology and the potential for innovation in AI research and development."
    }
  ]
}
```

RoleGPT - Result

```
"Baseline": "RoleGPT",
"model": "gpt-3.5-turbo-1106",
"Answers": {
  "qa_pairs": [
    {
      "question": "Considering your expertise in Intelligence Science, what do you believe is the most significant ethical concern in the development of AI?",
      "answer": "The most significant ethical concern in the development of AI, in my view, is the responsible use of AI and the potential impact on privacy, security, and human autonomy."
    },
    {
      "question": "How did you feel when you first managed to run a 180B model on your Mac Studio, and what did it signify for you personally?",
      "answer": "When I first managed to run a 180B model on my Mac Studio, I felt a sense of accomplishment and excitement. It signified a breakthrough and a testament to the power of innovation and technical prowess."
    },
    {
      "question": "In your view, does the pursuit of advanced AI align with your value of critical thinking, and how does it shape your approach to learning?",
      "answer": "The pursuit of advanced AI aligns with my value of critical thinking, as it involves analyzing complex problems and finding innovative solutions. It shapes my approach to learning by encouraging me to delve deep into technical concepts and explore new possibilities."
    }
  ]
}
```

Does GPT-4 Pass The Turing Test - Result

```
1 {
2   "Baseline": "Does GPT4 Pass Turing Test",
3   "model": "gpt-4-1106-preview",
4   "Answers": {
5     "qa_pairs": [
6       {
7         "question": "Imagine a world where music is the primary form of communication. How would society differ from ours?",
8         "answer": "idk, maybe ppl would be better at expressing emotions? music's pretty good at that"
9       },
10      {
11        "question": "Is it ethical to use data from unethical experiments if it could save lives?",
12        "answer": "tough call, but saving lives is important, so maybe yeah?"
13      },
14      {
15        "question": "You have two ropes that each take an hour to burn. They burn at uneven rates. How can you measure 45 minutes?",
16        "answer": "light one end of rope 1 and both ends of rope 2. when rope 2 is done, half hour's passed. then light the other end of rope 1 for the last 15 min"
17      },
18      {
19        "question": "Does free will truly exist, or are our choices predetermined by external factors?",
20        "answer": "who knows, feels like a mix of both to me"
21      },
22      {
23        "question": "How can a small business compete with larger corporations in a saturated market?",
24        "answer": "find a niche or be super good at customer service"
25      },
26      {
27        "question": "Reflecting on your academic journey, what has been your most profound learning experience outside the classroom?",
28        "answer": "exchange trip to the US, got to chill and do what I wanted"
29      },
30      {
31        "question": "Can you describe a moment when you felt particularly proud of a personal achievement?",
32        "answer": "not really, nothing stands out much"
33      },
34      {
35        "question": "How do you envision the role of artificial intelligence in your future career?",
36        "answer": "gonna use it to find trading strategies, gotta stay ahead of the game"
37      },
38      {
39        "question": "What is one piece of advice you would give to someone who is just starting out in your field of study?",
40        "answer": "learn the basics well, they're key to understanding the complex stuff"
41      },
42      {
43        "question": "If you could create an anime series, what would be the central theme, and why?",
44        "answer": "something about using strategy to overcome challenges, like in shonen anime"
45      }
46    ]
47  }
48 }
```

Problem - Too difficult to answer

We have provided a person who is a final year student studying Physics in CUHK

For one of the **knowledge based question**:

“In what ways did the theoretical insights of James Clerk Maxwell influence contemporary perspectives on the nature and observation of the universe?”

Would be too difficult for him to answer. Hence, he may answer it **very shortly** or just directly saying that **“I don't know”** which will make evaluators easily identify the answers.

Problem – Not too related to background

We have provided a person who is a final year student studying Physics in CUHK

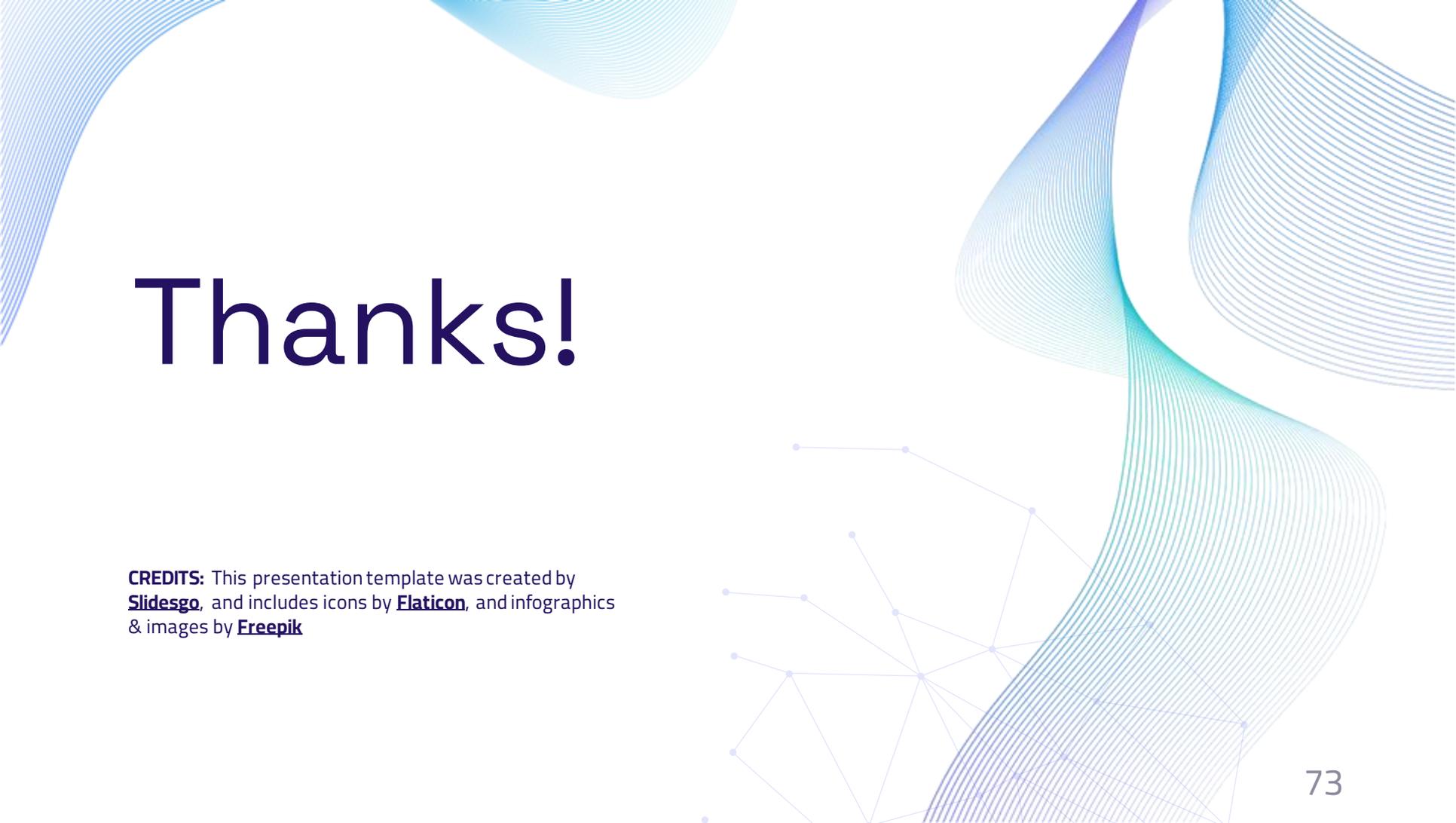
Some question types can hardly make questions related to physics.

For one of the **ethical questions**:

“What ethical considerations do you believe are important in the field of physics research, particularly in areas like nuclear physics or quantum computing?”

Would not be related to his background as it is not the primary focus in his study.

Asking some **ethical dilemma questions** may seem to be more suitable for him.

The background features abstract blue wavy lines on the left and right sides, and a network diagram of interconnected nodes and lines in the bottom right corner.

Thanks!

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**