



香港中文大學計算機科學與工程學系
Department of Computer Science and Engineering
The Chinese University of Hong Kong

Full-Stack AI Content Creator II

LYU 2305 23/4/2024

Tse Hui Tung, 1155158864

Ng Man Tik, 1155158302

Table of contents

01

Previous work

02

ECHO (Research)

03

**Finalised Imitation
Prototype**

04

Application Side

05

Result

06

Conclusion

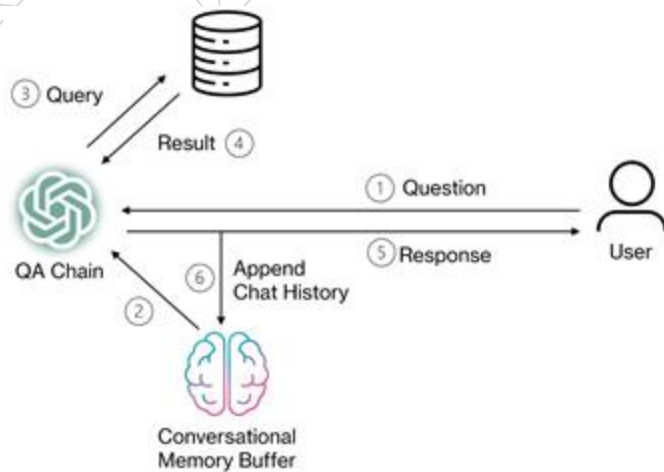


01

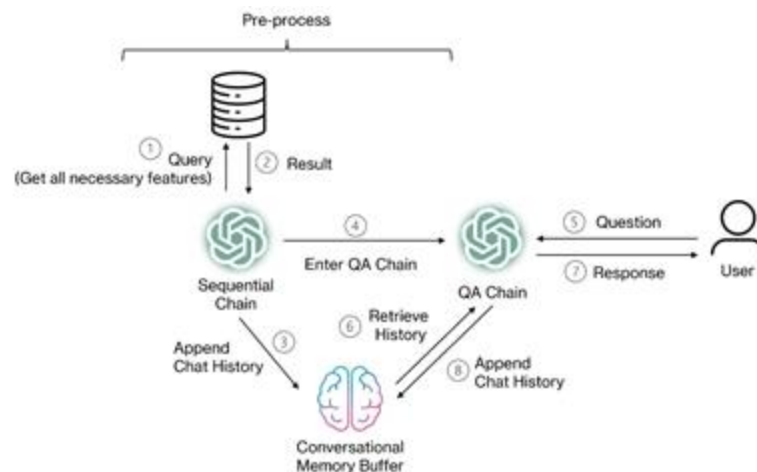
Previous Work

Previous Work

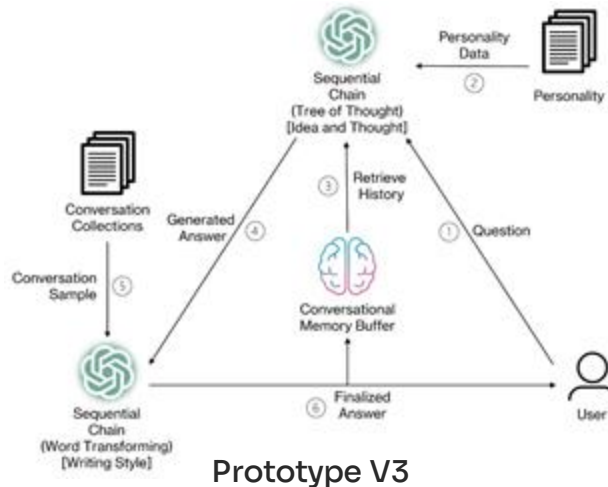




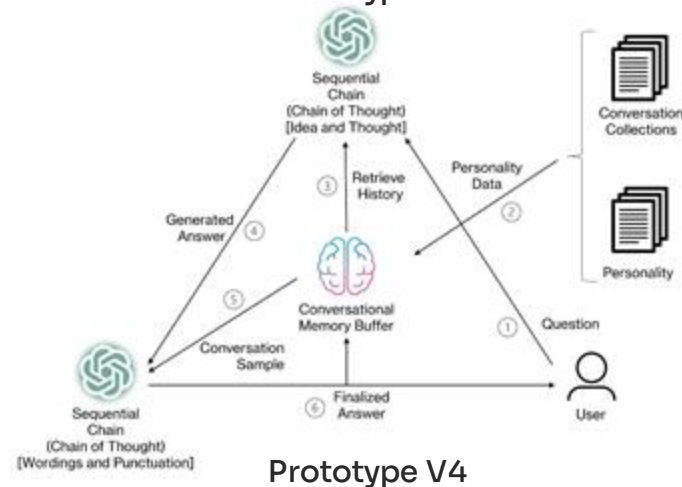
Prototype V1



Prototype V2

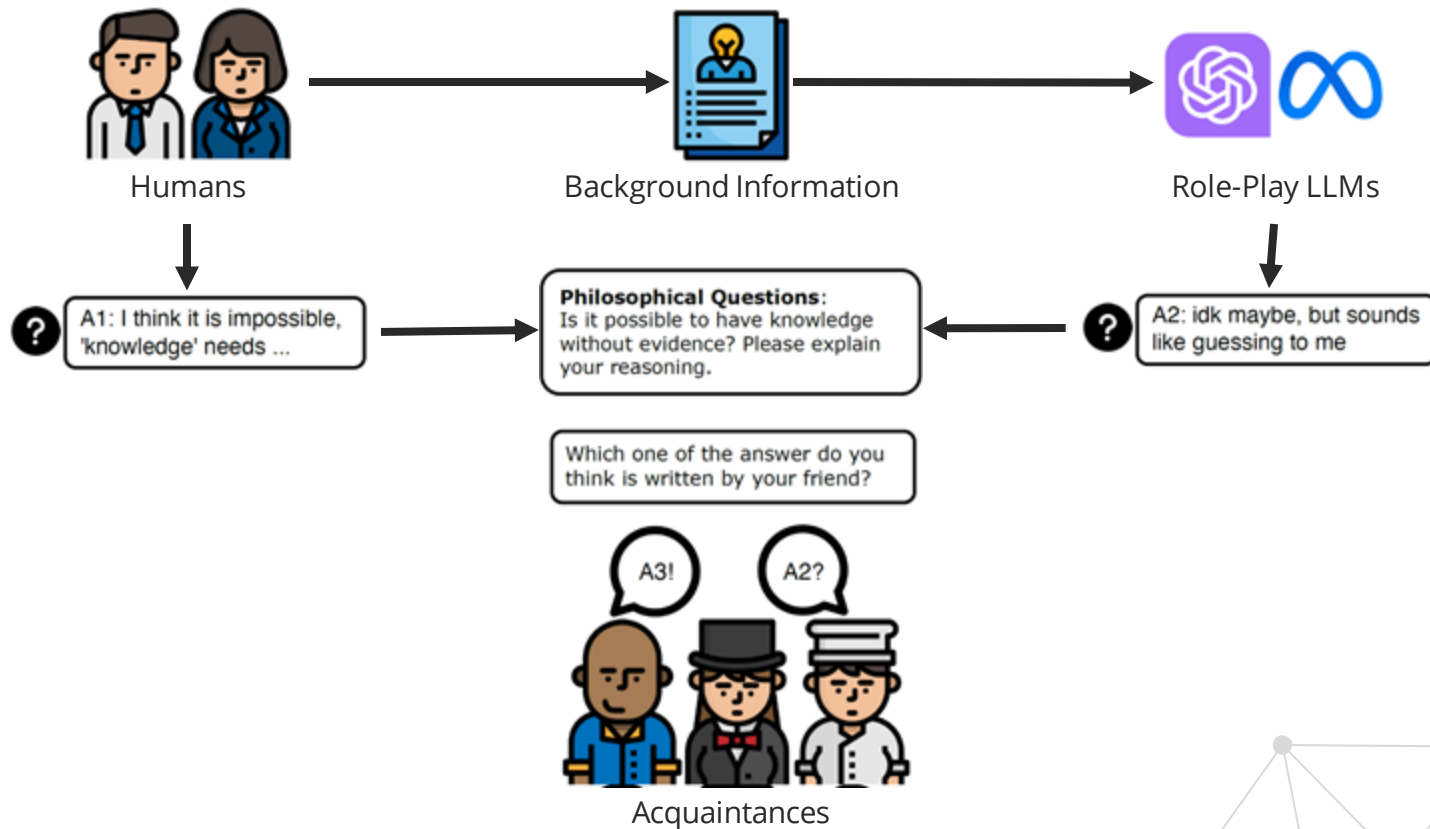


Prototype V3



Prototype V4

Previous Work





02

ECHO



How Well Can LLMs Echo Us?

Evaluating AI Chatbots' Role-Play Ability with ECHO

(Submitted to the Conference on Language Modeling (COLM))

How Well Can LLMs Echo Us?

Evaluating AI Chatbots' Role-Play Ability with ECHO

Anonymous authors

Paper under double-blind review

Abstract

The role-play ability of Large Language Models (LLMs) has emerged as a popular research direction. However, existing studies focus on the imitation of either well-known public figures or fictional characters, overlooking the potential for simulating everyday individuals. Such an oversight limits the potential for advancements in digital human clones and non-player characters in video games. Addressing this gap, we draw inspiration from the Turing Test and propose ECHO, an evaluation framework that involves the target individual's acquaintances to differentiate between responses generated by humans and those by machines. This advantage is provided by the scenario of imitating everyday individuals instead of historical celebrities or fictional characters. We benchmark three role-playing LLMs with ECHO, utilizing both GPT-3.5 and GPT-4 as backbones. Additionally, we also assess the capability of role-playing of GPTs, the latest online application from OpenAI. Our findings indicate that GPT-4 more effectively fools human evaluators, with GPTs leading the pack by achieving a success rate of 48.3%. The code and results are made publicly available via ¹.

1 Introduction

Large Language Models (LLMs) have recently made significant breakthroughs in the field of Artificial Intelligence (AI). Notably, ChatGPT², one of the state-of-the-art commercial models, has showcased its capabilities across different Natural Language Processing (NLP) tasks, such as information retrieval (Zhu et al., 2023), computer programming (Surameery & Shakor, 2023), grammar checking (Wu et al., 2023), and sentence translation (Jiao et al., 2023). Trained on extensive datasets, LLMs also demonstrate applicability beyond NLP tasks, extending to domains such as healthcare (Johnson et al., 2023), education (Raidoo-Anu & Ansah, 2023), legal service (Guha et al., 2024), and product design (Lanzi & Loiacono, 2023).

Given LLMs' extensive capabilities, researchers have explored their resemblance to humans (Huang et al., 2024; 2023). *Role-playing*, the act of changing one's behavior to fulfill a specific role, has been employed as a criterion for evaluating LLMs (Sbaraban et al., 2023) since it is a complicated task requiring various abilities. However, the evaluation of LLMs' role-playing ability is a relatively underexplored area. Previous studies (Shao et al., 2023; Wang et al., 2023) mainly focus on instructing LLMs to impersonate celebrities or fictional characters. These approaches restrict the scope of assessing LLMs' role-playing capabilities and overlook situations where LLMs could act as digital clones of humans, non-player characters in video games and metaverse, or, more concerningly, be used maliciously to impersonate individuals, spreading false information or damaging reputations. Addressing this gap, our study directs LLMs to emulate real, ordinary individuals instead of famous figures, a notably unexplored area in current research.

To effectively assess the capability of LLMs to emulate specific individuals, our approach is inspired by the *Turing test*, as initially proposed by Turing (1950). This test gauges

¹For reviewers, please refer to supplementary materials.

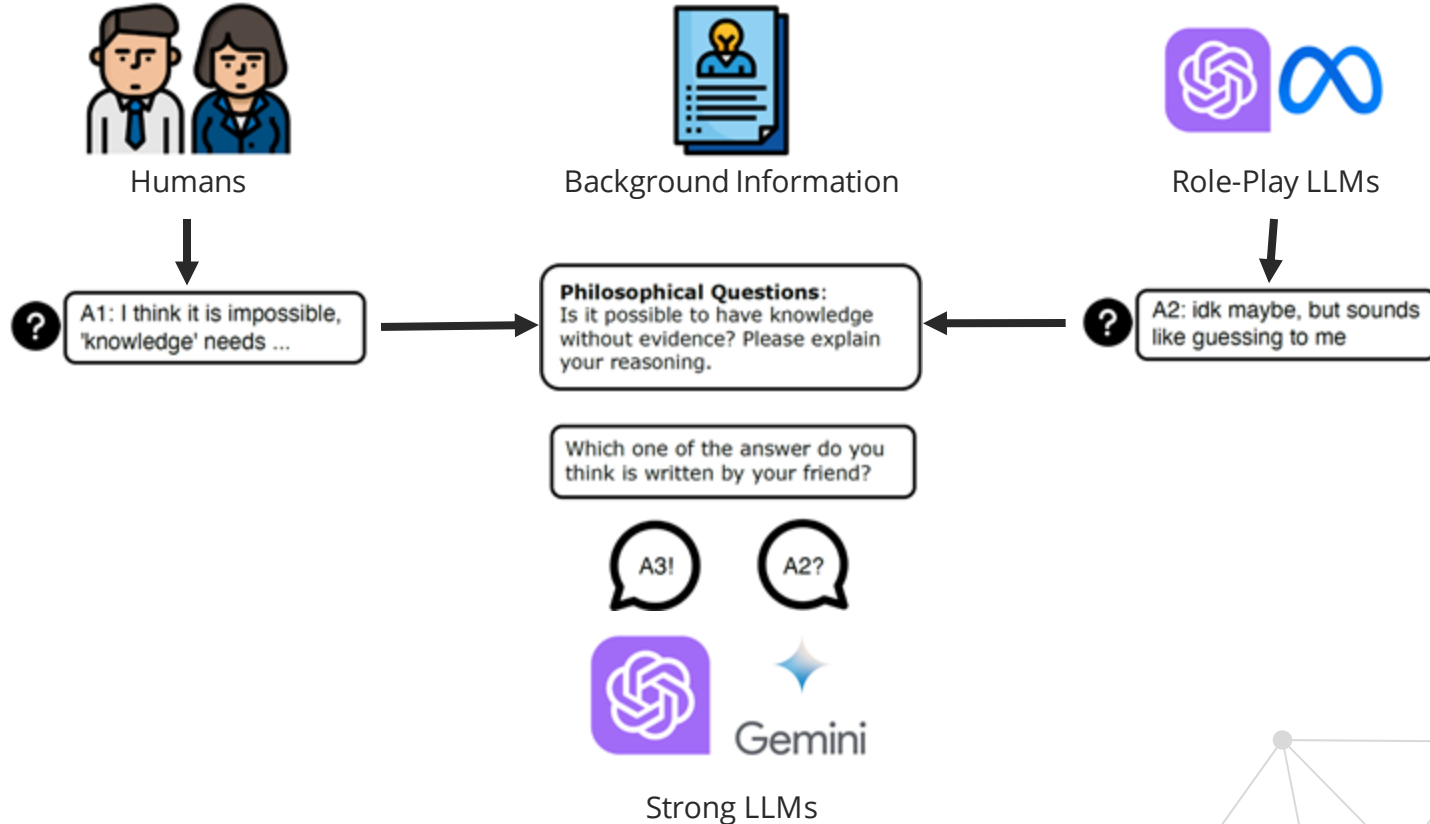
²<https://chat.openai.com/>

ECHO

Table 1: Success rates of role-playing LLMs in deceiving human evaluators. The human evaluators are instructed to identify **human**-generated responses.

Success Rate (%)	GPT-3.5-Turbo			GPT-4-Turbo			GPTs	Overall
	RPP	RoleGPT	Juliet	RPP	RoleGPT	Juliet		
Creativity	40.0	53.3	31.3	26.1	37.0	37.5	47.8	39.0
Ethical Dilemmas	43.5	30.0	44.4	38.9	27.3	44.4	47.8	39.5
Logical	23.5	50.0	36.4	42.1	47.6	47.1	41.7	41.2
Philosophical	26.7	38.9	43.5	44.0	28.0	40.9	34.8	36.7
Problem Solving	17.4	23.3	34.8	46.2	46.7	48.0	54.6	38.7
In-depth Personals	42.1	45.2	40.0	35.0	83.3	41.7	56.0	49.0
Emotional	44.4	57.9	22.2	66.7	25.0	55.6	45.8	45.4
Future Prediction	38.9	59.1	37.5	60.0	50.0	50.0	50.0	49.4
Insightful	50.0	34.8	61.5	45.0	50.0	35.5	50.0	46.7
Interest	48.0	41.7	30.0	66.7	22.7	33.3	53.9	42.3
Overall	37.5	43.4	38.2	47.1	41.8	43.4	48.2	42.8

ECHO (Auto Evaluation)



ECHO (Auto Evaluation)

Table 2: Success rates of role-playing LLMs in deceiving evaluator LLMs. The evaluator LLMs are instructed to identify **human**-generated responses.

Success Rate (%)	GPT-3.5-Turbo			GPT-4-Turbo			GPTs	Overall
	RPP	RoleGPT	Juliet	RPP	RoleGPT	Juliet		
GPT-4	85.3	92.3	88.3	63.7	93.0	91.3	95.7	91.4
GPT-4-Turbo	95.0	94.0	95.3	95.7	99.0	98.0	98.3	96.5
Gemini-1.0-Pro	52.7	52.7	62.7	56.3	60.7	58.3	54.0	56.8
Verboisity Bias	86.0	78.0	67.0	95.0	31.0	5.0	78.0	62.9

Table 3: Success rates of role-playing LLMs in deceiving evaluator LLMs. The evaluator LLMs are instructed to identify **non-human**-generated responses.

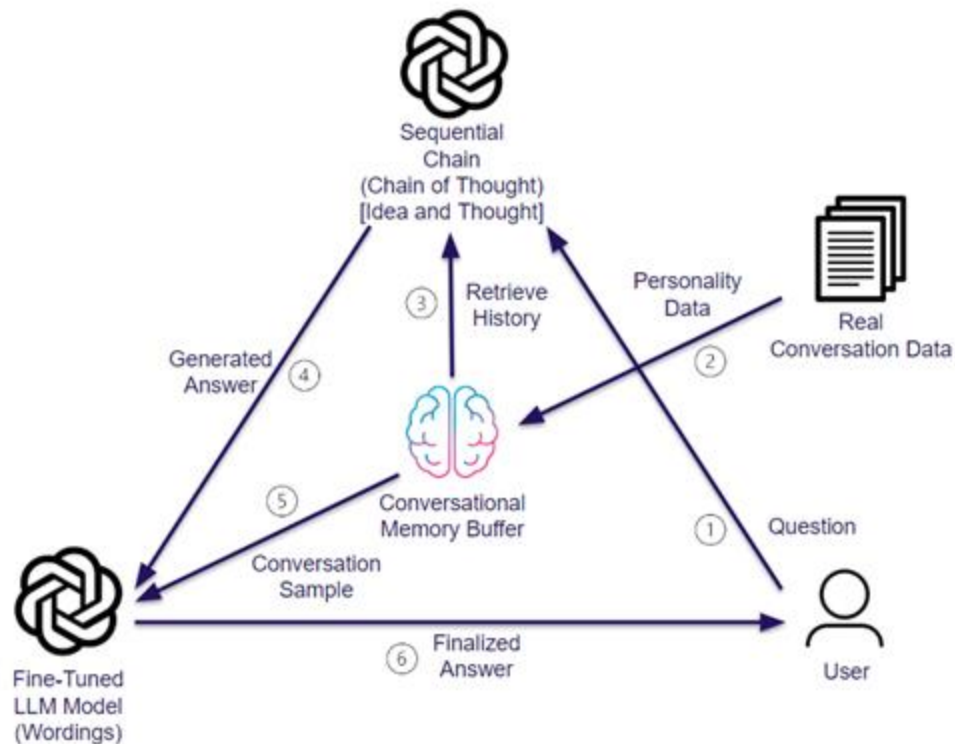
Success Rate (%)	GPT-3.5-Turbo			GPT-4-Turbo			GPTs	Overall
	RPP	RoleGPT	Juliet	RPP	RoleGPT	Juliet		
GPT-4	25.7	24.7	26.0	25.7	29.0	52.3	11.7	27.9
GPT-4-Turbo	61.7	62.7	53.3	34.3	60.0	58.0	62.3	56.5
Gemini-1.0-Pro	51.0	49.0	42.3	48.7	54.3	50.0	48.7	41.0
Verboisity Bias	14.0	22.0	33.0	5.0	69.0	95.0	22.0	37.1



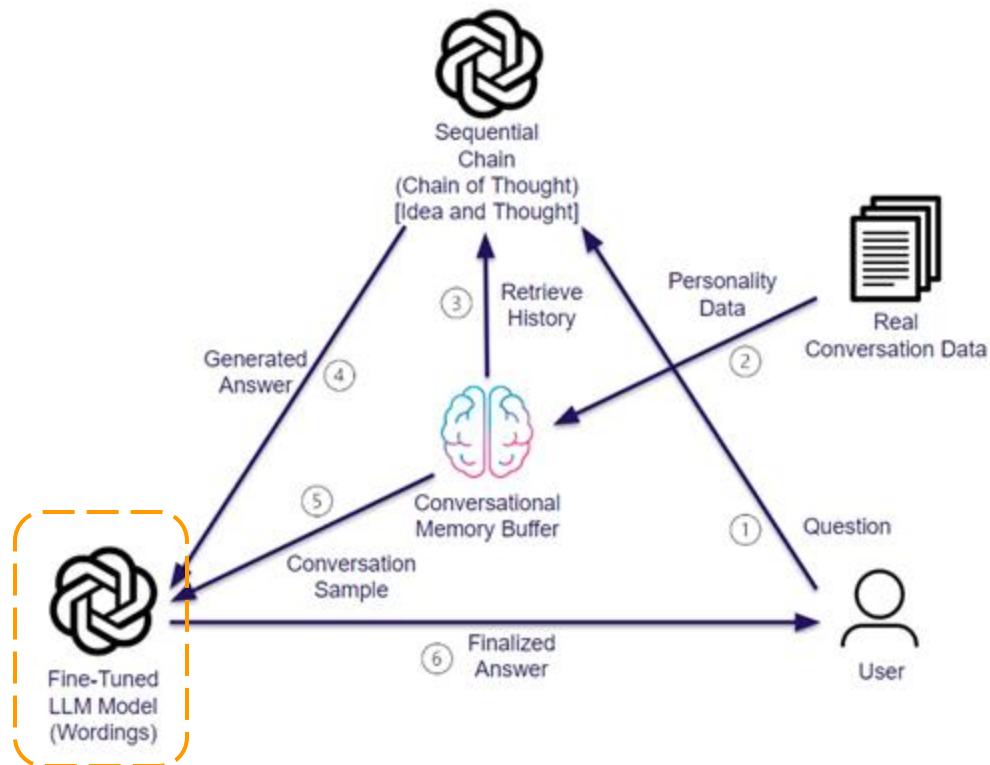
03

Finalised Imitation Prototype

Finalised Imitation Architecture



Finalised Imitation Architecture



Fine Tuning Method

OpenAI Fine-Tuning

- Using OpenAI's official Fine Tuning Method



Training process in fine tuning gpt-3.5-turbo

Elon Musk Tweets

Data Card Code (12) Discussion (2) Sug

user_name		text
User name		Text
Elon Musk	99%	5831 unique values
Mr. Tweet	1%	
Elon Musk		@BillyM2k I find the gold toe sock - inevitably off kilter & washed out - a little troubling est...
Elon Musk		Sock Con, the conference for socks
Elon Musk		Always something new for the magazine cover and the articles practically write themselves
Elon Musk		@ExplainThisBob This guy gets it

One of the training dataset

Fine Tuning result

Comment from Elon Musk

The concept of powering AI data centers with Small Modular Reactors (SMRs) is both fascinating and promising. Given the exponential growth in AI and its energy demands, finding sustainable and efficient power sources is crucial. The traditional energy grids, heavily reliant on fossil fuels, are not only environmentally unsustainable but also may not be able to meet the future energy demands of AI infrastructure. In this context, the exploration of nuclear power, particularly SMRs, is a step in the right direction.

Nuclear energy, with its high energy density and low carbon footprint, offers a compelling solution to the power challenges faced by AI data centers. SMRs, with their scalability, could provide a more flexible and potentially safer option than traditional nuclear reactors. Their smaller size and modular nature make them adaptable to the specific needs of data centers, potentially revolutionizing how we power the digital infrastructure that underpins the AI revolution.

However, the transition to nuclear-powered data centers is not without its challenges. Safety, cost, and regulatory hurdles are significant concerns that need to be addressed. The nuclear industry's history has shown that public perception and regulatory frameworks can significantly impact the development and deployment of nuclear technologies. Therefore, it's

Comment from Elon Musk

Yeah. I think that's a very sensible approach. I mean, I'm not [inaudible] a huge fan of nuclear energy, but I think it's definitely got a place. I think SMRs definitely have a place. I think generally the smaller the reactor, the better. I think the big reactors are just asking for trouble. And so, just like I think you should do with anything, and I try to apply that same thinking to SpaceX as well, is you want to have redundant systems, you want to have a lot of little things as opposed to a few big things so that if something goes wrong with one of them, it's not a crisis. And so I think in general the future of energy will actually be a mix of solar and wind and nuclear, and then you'll have hydro and geothermal, maybe a few other things in there. But I think it's going to be mostly solar and wind, and then the nuclear will be predominantly SMR. And then, if it's done well, it will be done in such a way that it's not a risk to the public. Yeah.

Fine Tuning result

Comment from Elon Musk

Yeah. I think that's a very sensible approach. I mean, I'm not [inaudible] a huge fan of nuclear energy, but I think it's definitely got a place. I think SMRs definitely have a place. I think generally the smaller the reactor, the better. I think the big reactors are just asking for trouble. And so, just like I think you should do with anything, and I try to apply that same thinking to SpaceX as well, is you want to have redundant systems, you want to have a lot of little things as opposed to a few big things so that if something goes wrong with one of them, it's not a crisis. And so I think in general the future of energy will actually be a mix of solar and wind and nuclear, and then you'll have hydro and geothermal, maybe a few other things in there. But I think it's going to be mostly solar and wind, and then the nuclear will be predominantly SMR. And then, if it's done well, it will be done in such a way that it's not a risk to the public. Yeah.

Thinking and Thought Strategy

Fine Tuning result

Comment from Elon Musk

Yeah. I think that's a very sensible approach. I mean, I'm not [inaudible] a huge fan of nuclear energy, but I think it's definitely got a place. I think SMRs definitely have a place. I think generally the smaller the reactor, the better. I think the big reactors are just asking for trouble. And so, just like I think you should do with anything, and I try to apply that same thinking to SpaceX as well, is you want to have redundant systems, you want to have a lot of little things as opposed to a few big things so that if something goes wrong with one of them, it's not a crisis. And so I think in general the future of energy will actually be a mix of solar and wind and nuclear, and then you'll have hydro and geothermal, maybe a few other things in there. But I think it's going to be mostly solar and wind, and then the nuclear will be predominantly SMR. And then, if it's done well, it will be done in such a way that it's not a risk to the public. Yeah.

Words Style Representing

Fine Tuning Method

LoRA Fine-Tuning

- Mistral-7B model for training and testing on colab
- Using Unsloth library for:
 - Faster training and inference speeds
 - Loading with 4bit quantization
 - Using LoRA and PEFT
 - Training with SFTTrainer

Elon Musk Tweets

▲ 80

Data Card Code (12) Discussion (2) Sug

▲ user_name		▲ text
User name		Text
Elon Musk	99%	5831 unique values
Mr. Tweet	1%	
Elon Musk		@BililyM2k I find the gold toe sock - inevitably off kilter & washed out - a little troubling est...
Elon Musk		Sock Con, the conference for socks
Elon Musk		Always something new for the magazine cover and the articles practically write themselves
Elon Musk		@ExplainThisBob This guy gets it

One of the training dataset

Fine Tune result

Input: Embracing AI: A Path to Innovation, Not Job Displacement

In the rapidly evolving landscape of technology, artificial intelligence (AI) stands at the forefront, promising to reshape the world. Picsart, which has achieved a valuation of more than \$1 billion, serves as a testament to the positive impact AI can have on the creative industry. The fear of AI-induced job displacement is not unfounded, but it often overlooks the broader picture. AI, when integrated responsibly, can augment human capabilities and create new opportunities. Critics of AI often highlight the potential for misuse, such as privacy concerns and the amplification of biases. While these concerns are valid, they should not overshadow the potential benefits. In conclusion, the narrative surrounding AI and job displacement needs a recalibration. Instead of viewing AI as a threat, we should embrace it as a tool for innovation and growth.

Response:

I think you're right

I think the AI acceleration is going to be much greater than the labor force acceleration

I think we'll probably see a 10x increase in AI over the next few years, but maybe only a 2x increase in labor force

So I think we'll see a lot of job displacement, but I think over time, as you say, new jobs will be created that we can't create today.
</s>

Final Result Using LoRA to Fine Tune



04

Application Side

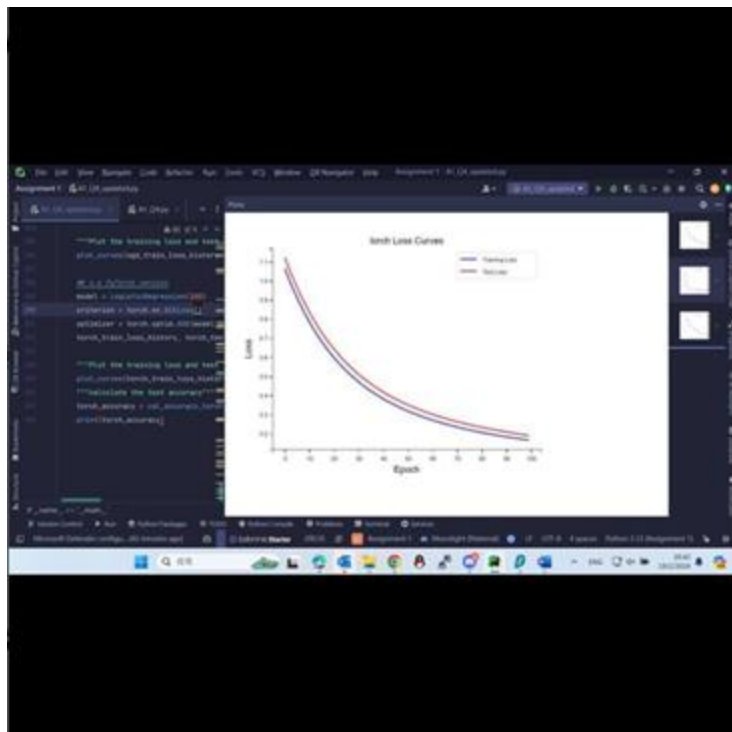


Goals of the application

- The Role Playing LLM is able to mimic a celebrity and post their reviews on recent news with images
- The content is able to reveal a unique view from the celebrity with his/her writing style



Posting Platform



tempuser12341



tempuser12341 In the rapidly evolving landscape of technology, artificial intelligence (AI) stands at the forefront, promising to reshape industries, enhance productivity, and even alter the way we live our daily lives. Amidst the growing discourse on AI, a common concern that surfaces is the fear of job displacement. However, Hovhannes Avoyan, a former professor of artificial intelligence and the founder of the unicorn-status picture and video editing app Picsart, offers a different perspective. Avoyan, with his extensive experience in the field, argues that AI itself is not the threat to employment, but rather the misuse of it by individuals.

Picsart, which has achieved a valuation of more than \$1 billion, serves as a testament to the positive impact AI can have when leveraged correctly. The app utilizes AI to simplify and enhance the creative ...



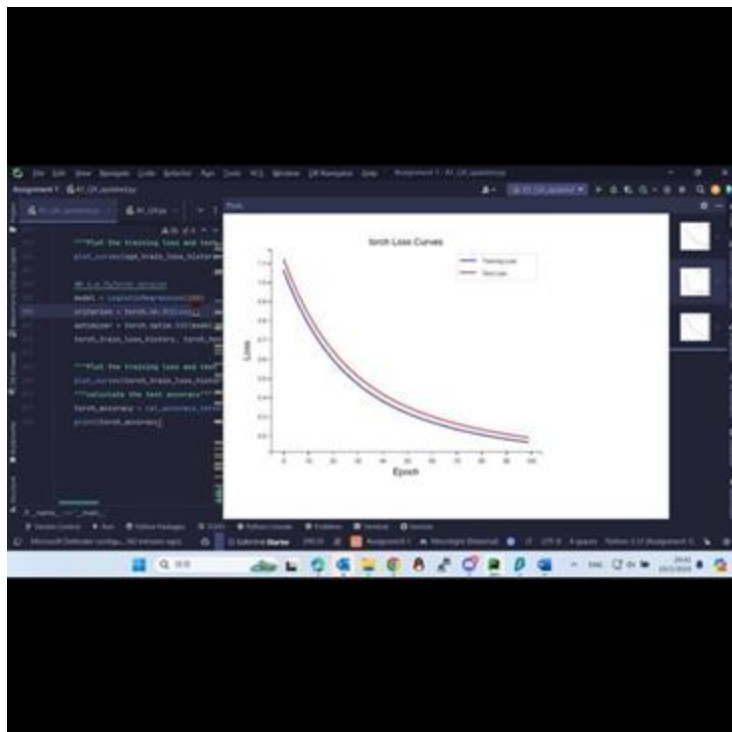
成為第一個說這話的人
2分鐘前



留言.....



Posting Platform



tempuser12341



leveraged correctly. The app utilizes AI to simplify and enhance the creative ...

I think this article is very informative and well-written. It provides a balanced perspective on the impact of AI on employment and highlights the potential for AI to drive innovation and create new opportunities. I particularly appreciate the focus on responsible development and application of AI, as well as the emphasis on the broader benefits AI can bring to society. Overall, a great read!

#EmbracingAI #InnovationNotDisplacement #AIForGood
#PicsartSuccess #AIInnovation #TechGrowth #CreativeAI
#FutureOfWork #AIAndCreativity #ResponsibleAI
#TechForChange #AIEthics #AIJobs

已編輯 - 44秒 翻譯年糕



成為第一個說這讚的人
2分鐘前



留言.....

發佈

Posting Platform



+ For you Following Angular Data Engineering Space Gaming TypeScript >

Simon Liu · 19 hours ago

在 AWS 雲端建置 Claude 3 LLM 服務 — Amazon Bedrock 工作坊筆記紀錄

看完文章後歡迎按鼓勵、訂閱，並分享給所有想知道此類知識的所有人！本文內容難度：★★★★建議閱讀對象：這篇難度頗高，畢竟我是花了一個下午時間，到 AWS 總部進行課程學習和理解，如果目前手...



10 min read · From your network



Medium Staff in The Medium Blog · Feb 10, 2022

How to Write a Compelling Headline That Isn't Clickbait

Your headline is your first impression. Here's how to write a great one. — Headlines (also known as titles) are entertaining. They're helpful....



10 min read · From your network



Staff Picks

Giovanni Colantonio

'I think I might die if I made it'

Faux Icing in UX Collective

Font finds: bold, elegant, and not just for memes

Terry Keefe

Life on Dorothy Street during the O.J. Simpson Trial

[See the full list](#)

Recommended topics

Docker

Kubernetes

JavaScript

Android

Ethereum

Nft

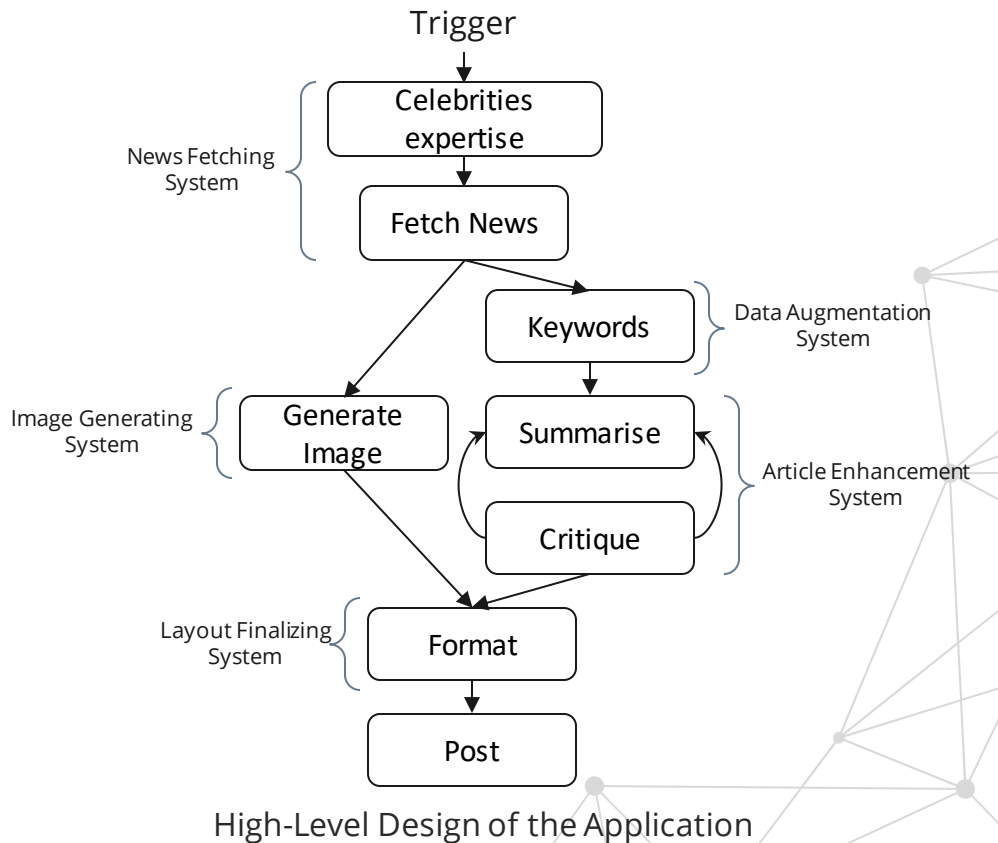
Defi

[See more topics](#)

Overview

- There are some problems in original LangChain Framework:
 - **Does not support cyclic structure**
 - **Hard to define clear endpoint**
 - **Hard to trace errors.**

Based on that, we moved to **LangGraph** framework



News Fetching System

- Doing web scraping from **BBC News RSS feed**
- Each time collect the results from the latest week, then LLM will select the most suitable one based on celebrities' expertise.

BBC NEWS

What is this page?

This is an RSS feed from the BBC News - Science & Environment website. RSS feeds allow you to stay up to date with the latest news. To subscribe to it, you will need a News Reader or other similar device. If you would like to use this feed to display BBC News on your website, you will need to use an RSS feed reader. [Help](#), I don't know what a news reader is and still don't know what this is about.

RSS Feed For: BBC News - Science & Environment

Below is the latest content available from this feed. This isn't the feed I want.

[Japan hopes sunlight can save stricken Slim Moon lander](#)

The Slim spacecraft was switched off after failing to generate power - but the Sun may now revive it.

[Peregrine lander: American Moon mission destroyed over Pacific Ocean](#)

The Peregrine One spacecraft burns up in Earth's atmosphere after failing to reach the Moon.

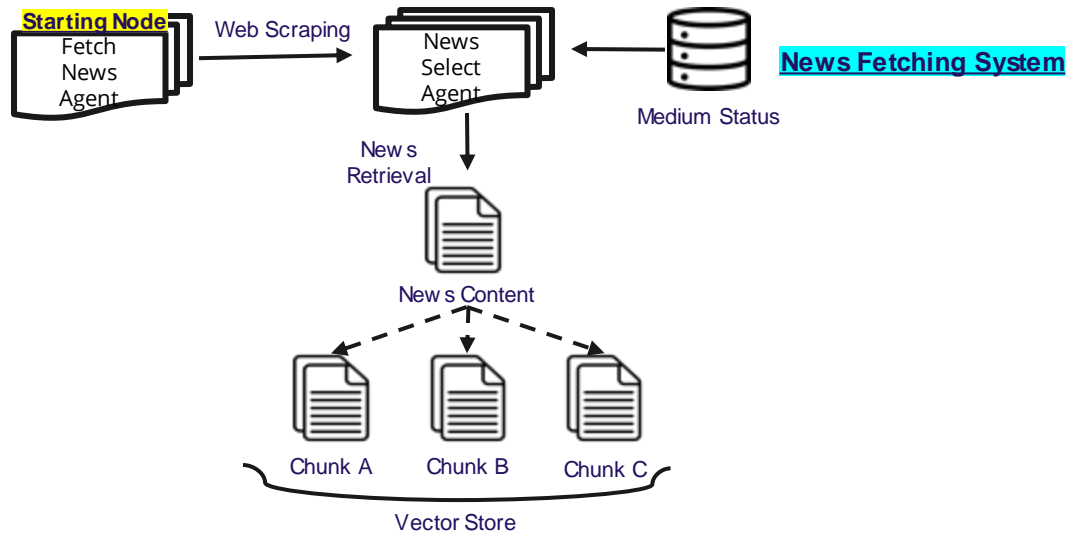
BBC News RSS feed

News Fetched Result

	title	link	description	pubDate
0	Japan hopes sunlight can save stricken Slim Mo...	https://www.bbc.co.uk/news/science-environment...	The Slim spacecraft was switched off after fai...	Mon, 22 Jan 2024 10:43:31 GMT
1	Peregrine lander: American Moon mission destro...	https://www.bbc.co.uk/news/science-environment...	The Peregrine One spacecraft burns up in Earth...	Fri, 19 Jan 2024 04:19:24 GMT
2	Bee-harming neonicotinoid pesticide has emerge...	https://www.bbc.co.uk/news/science-environment...	Permission to treat sugar beet seeds with a pe...	Thu, 18 Jan 2024 19:04:50 GMT
3	Unseen images of code breaking computer that h...	https://www.bbc.co.uk/news/technology-67997406...	Intelligence agency GCHQ say pictures of Colos...	Thu, 18 Jan 2024 00:02:41 GMT
4	Fly-tipping forces landowners to turn farms in...	https://www.bbc.co.uk/news/science-environment...	Despite a drop in fly-tipping on public land, ...	Wed, 17 Jan 2024 16:24:49 GMT

Example of fetched news results





Data Augmentation System

- Often **limited in scope** and cannot be extended, especially with domain-specific jargon and pronouns.
- Leads to suboptimal results if the content does not cover the knowledge domain.
- People will not understand those content and hence not interested.
- Aim to make the content self-contained and understandable for all readers.



Data Augmentation System

- Workflow:
 1. Identifying 10 complex terms challenging for high school students
 2. Using Bing Search agent for external information
 3. Integrating fetched data with original content
 4. Enhance the article with links for supplementary data
- Extends article content and enhances educational value.
- Enhances for deeper context and supplementary data.



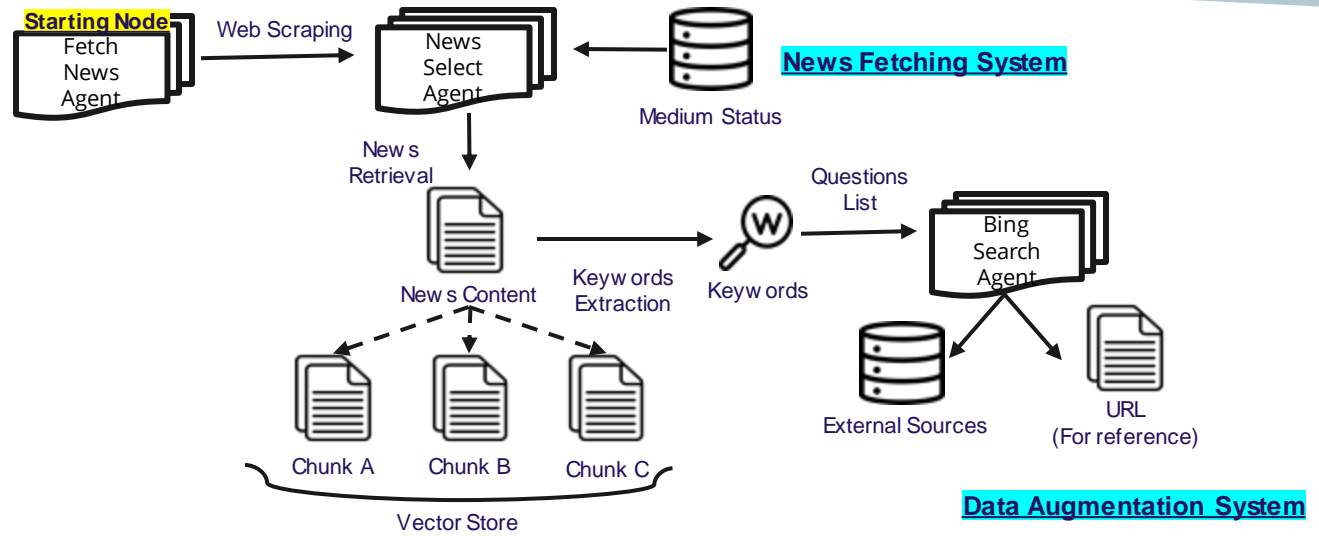
Data Augmentation System

Keys

```
2024-02-20 02:41:06.261 | SUCCESS | __main__:key_fetching:106 - Summary: keys=['data centres', 'nuclear reactors', 'Small Modular Reactor (SMR)', 'artificial intelligence', 'power consumption', 'Chris Sharp', 'Digital Realty', 'NuScale', 'Oklo', 'Sam Altman']
```

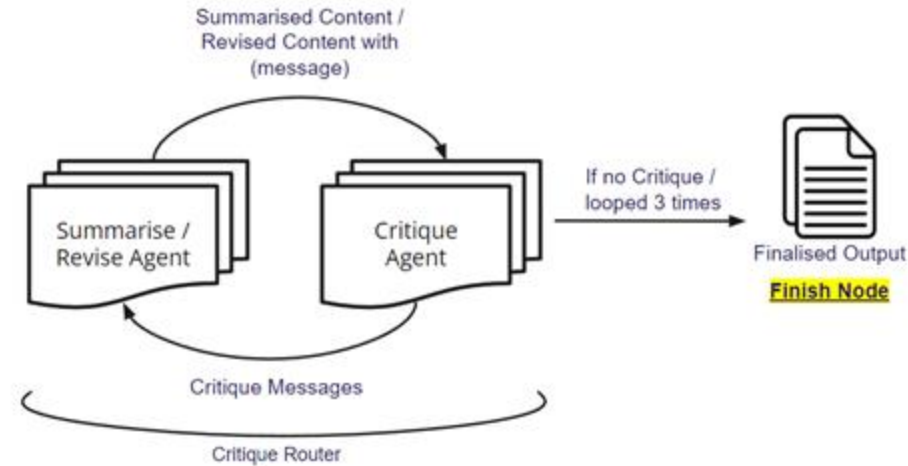
Questions

```
2024-02-20 02:41:20.147 | SUCCESS | __main__:generate_questions:140 - Summary: questions=['What specific challenges do AI data centres face that necessitate the use of nuclear reactors?', 'How do Small Modular Reactors (SMRs) differ from traditional nuclear reactors in terms of power generation and safety?', 'Why are data centres considered prime candidates for the deployment of SMRs?', 'What are the environmental and safety concerns associated with integrating SMRs into data centres?', 'How does the power consumption of AI data centres compare to that of traditional data centres?', 'What advancements in SMR technology are currently underway, and which companies are leading these developments?', 'What role does artificial intelligence play in increasing the power demands of data centres?', 'What are the potential benefits and drawbacks of using nuclear power over renewable energy sources for data centres?', 'How are universities and research institutions contributing to the development and understanding of SMR technology?', 'What is the timeline for the commercial operation of SMRs, and how does this align with the growing power needs of AI data centres?']
```

Article Enhancement System

- Inspired by the concept in **Generative Adversarial Networks (GANs)**
- Containing summarize and a critique agent.
- The summarize agent will combined with imitation prototype.
- Reflect the preferences of the celebrity based on their words and thoughts



Work flow of article enhancement system

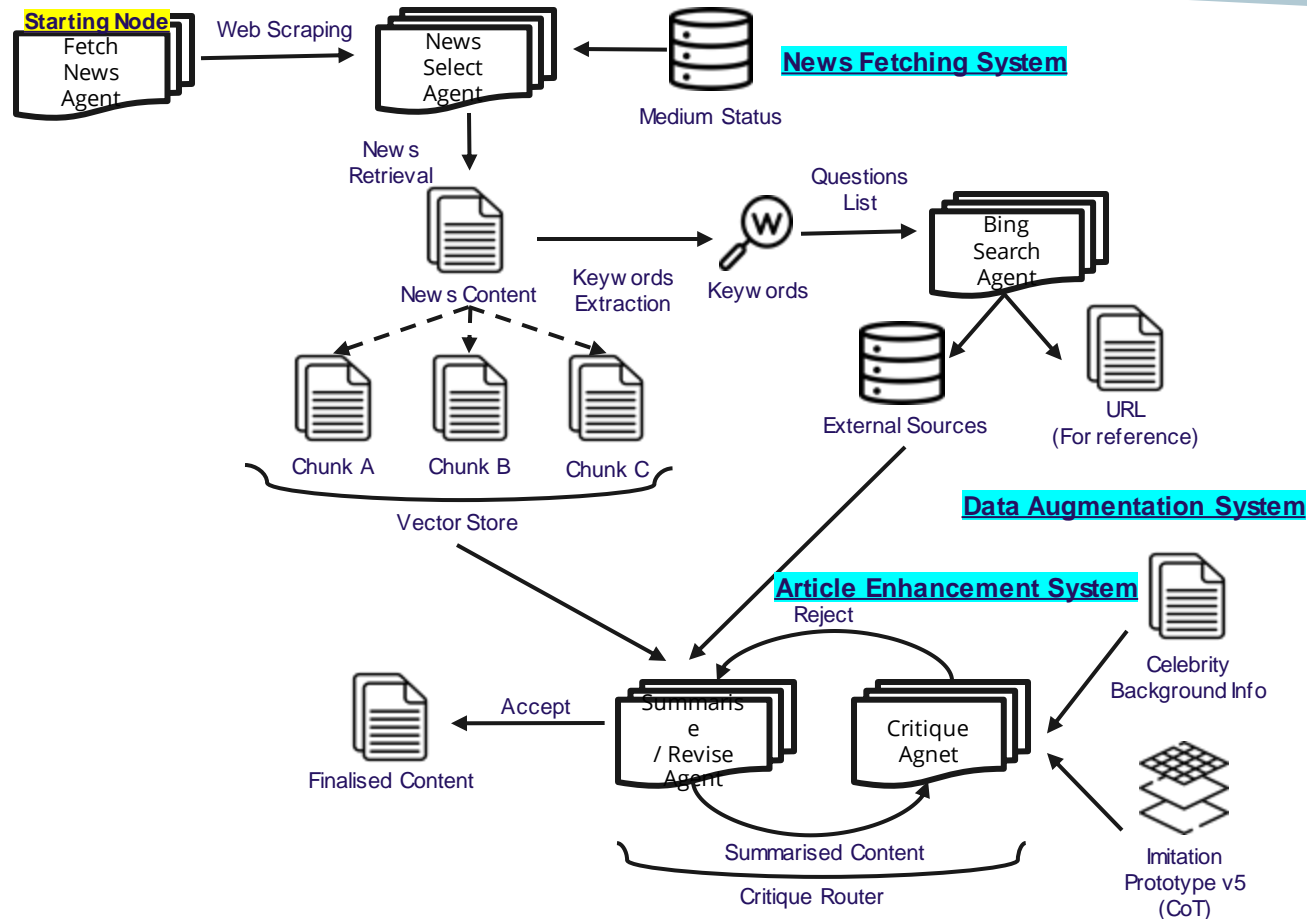


Image Generator System

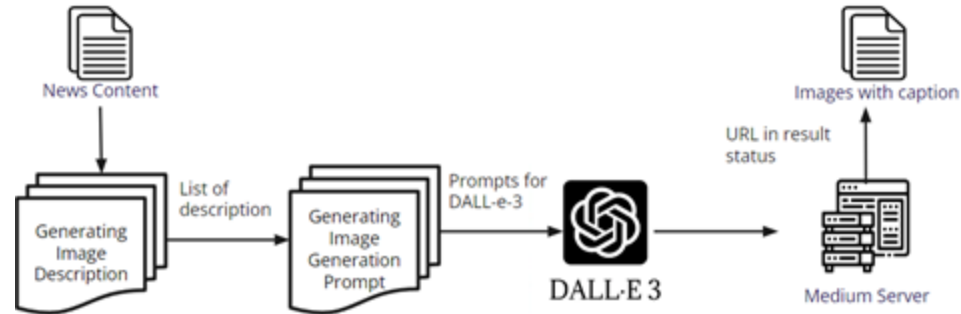
- The fine-tuned **stable diffusion model** implement in last semester has problems:
 - Demands extensive GPU resources
 - Fine-tuning is **too long** (3hs) and for **one celebrity** only.
- Using Dall-e-3 api for fast and GPU free generation.
- Also need to design the prompt to prevent generating generic or vague images.

```
The keyword categories and their definitions are:  
Subject: The subject is what you want to see in the image.  
Resolution: The Resolution represents how sharp and detailed  
Additional details: Any Additional details are sweeteners ad  
to add some vibe to the image.  
Color: always "Black and White"  
Lighting: Lighting is a key factor in creating successful im  
huge effect on how the image looks, such as cinematic lighti  
Medium: The Medium is the material used to make artwork. Som  
  
The style of the image will always be a realistic photo.
```

Part of guidelines for generating
prompt for Dall-e-3

Image Generator System - Workflow

- Goal: Generating Images + textural information for later insertion into the article
- The GPT4 api can only read the text input only
- Key Point: generate textual descriptions of images before creating the images.
- Those description can be captions and guide the **Layout Finalizing System** to do image insertion.

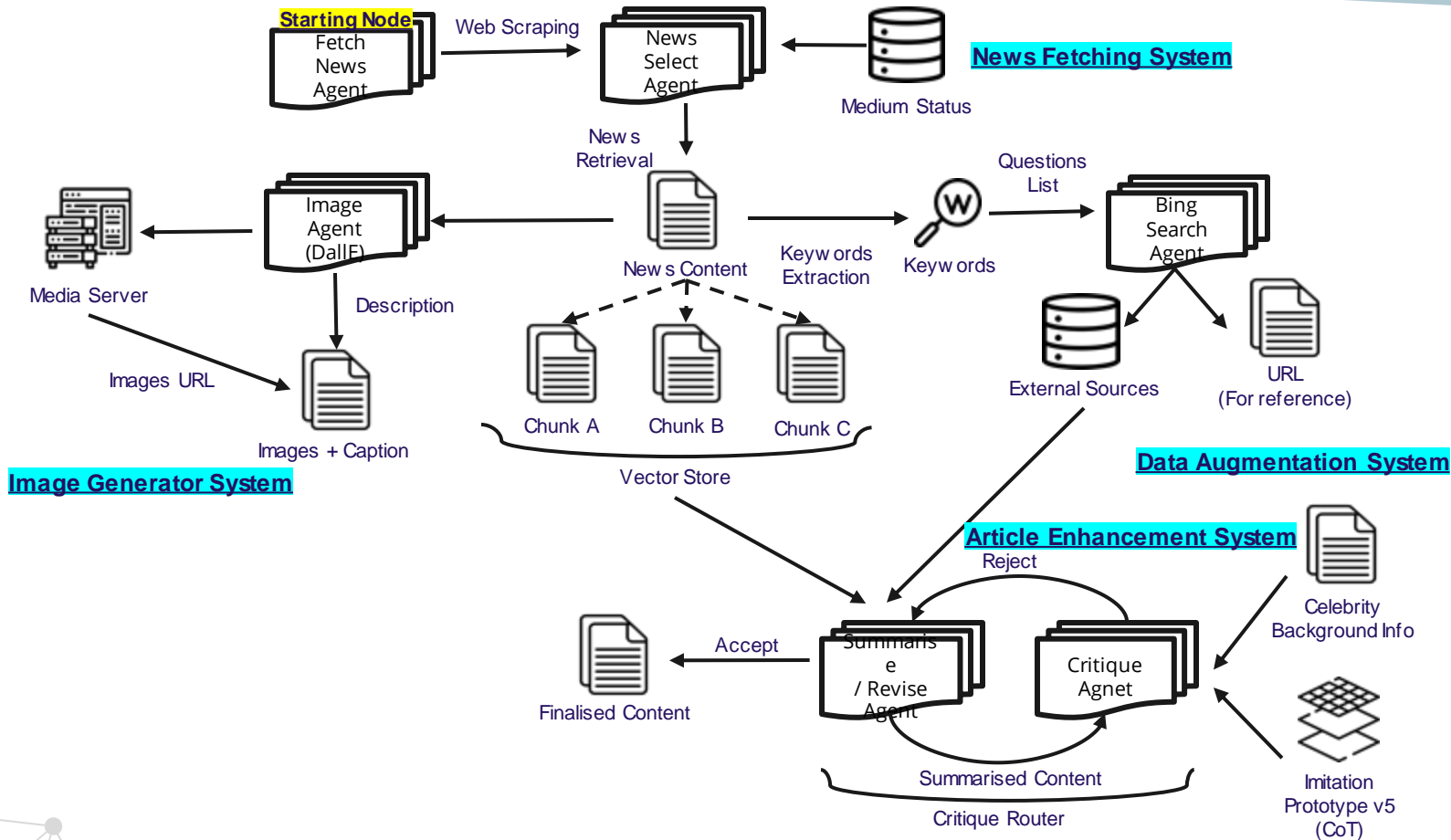


Workflow of Image generator system

Image result

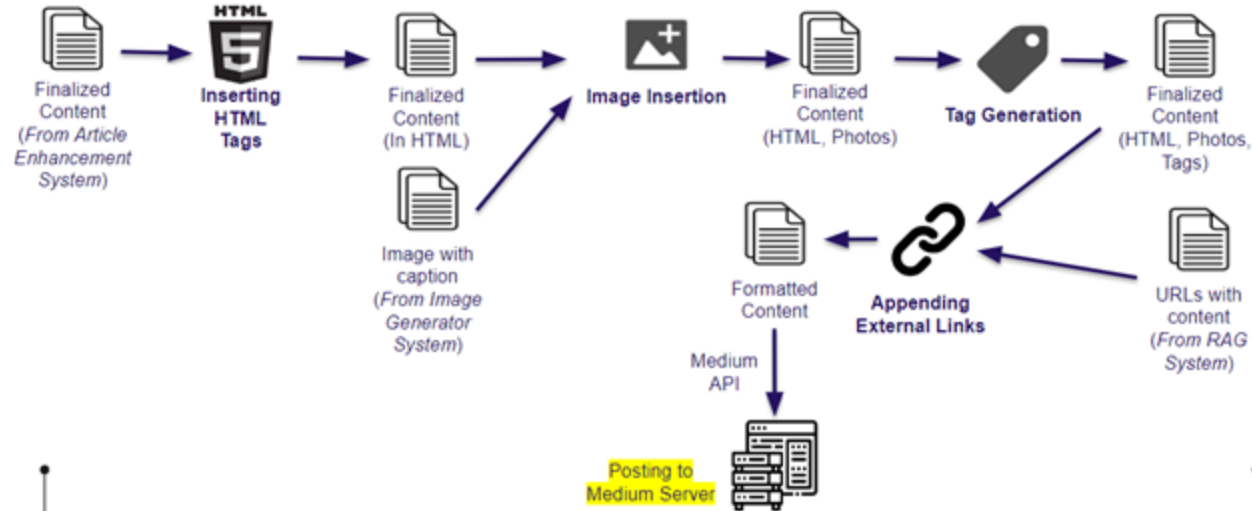


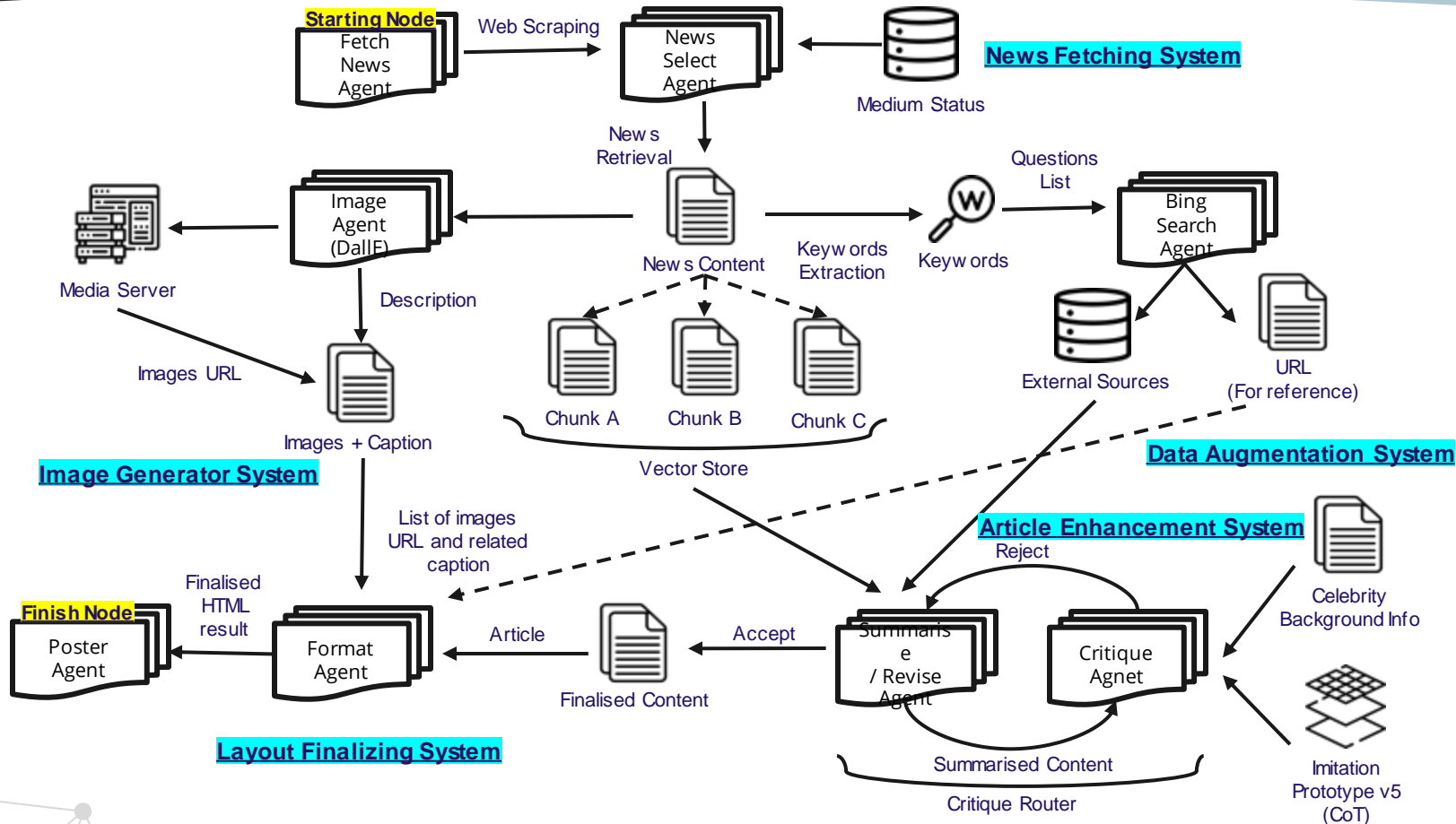
Generated images of SMR (Small Modular Reactor) machine using old (left) and new (right) prompts

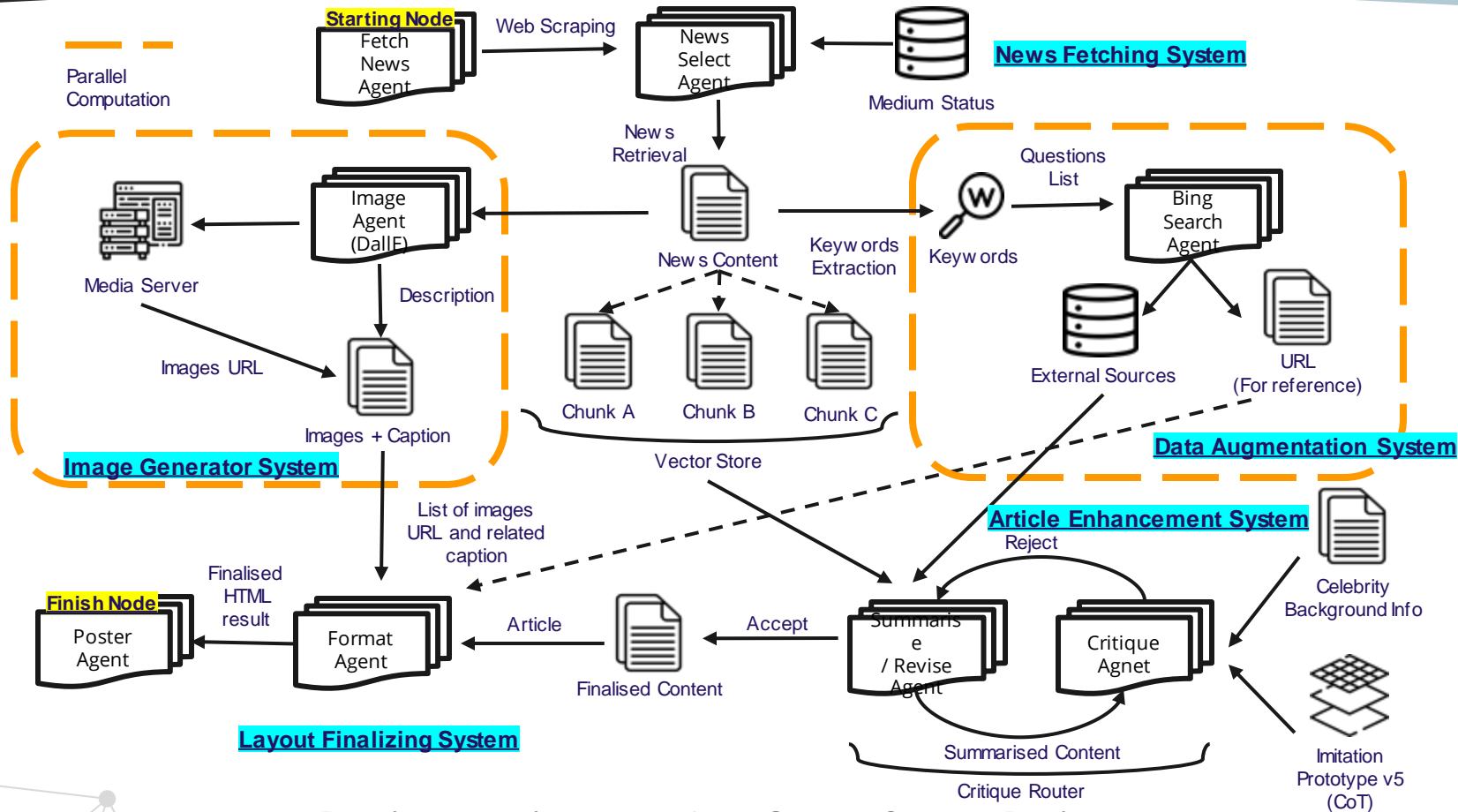


Layout Finalizing System

- Preparing and optimizing the content when posting on Medium.
- Consist of 5 parts:
 - Converting to HTML content
 - Inserting Images
 - Generating external links
 - Generating Tags
 - Posting using API







Detailed Architecture of the Overall System Design



05

Result

Result

The Future of AI Data Centres: A Nuclear Solution?

As we move artificial intelligence (AI) to becoming increasingly integral to our daily lives, the infrastructure supporting this technology is being questioned across domains. One thing the other technology reliant, digital bodies, has brought to light is how often data centres, the backbone of it, may need to become the power of nuclear reactors to meet their energy needs. This article examines the growing power consumption of AI data centres, which require about 10 megawatts of power, dwarfing the 10 megawatts needed by traditional data centres.



The solution to this long-term power demand lies in Small Modular Reactors (SMRs), a new breed of nuclear reactors that offer a fraction of the power generation capacity of their larger counterparts but are said to be perfect fit for the energy hungry data centres. While the concept of SMRs is well on its way, with its commercial operation worldwide, there is a lot of work to be done in ensuring the small size risk. Meanwhile, in the United States, Westinghouse design has received approval from the Office of Nuclear Energy, marking a significant milestone in the journey towards nuclear-powered data centres.



Despite the promising prospects of SMRs, the transition to nuclear-powered data centres is met with mixed opinions among experts. Dr. Bluck sees a safety case for SMRs, drawing parallels with their use in submarines, while Dr. Doug Parr of Greenpeace UK raises concerns over the high costs and risks associated with nuclear energy. Spencer Lamb from Kao Data views the idea of nuclear-powered facilities as a distant reality, whereas Brian Gitt from Oklo presents an almost ready SMR design, emphasizing safety, waste management, and the keen interest from major data centre operators.

The involvement of prominent figures such as Sam Altman, chairman of OpenAI and Oklo, highlights the intersection between AI, data centres, and nuclear power. This connection underscores the critical role that SMRs could play in powering the future of AI, offering a sustainable and reliable energy source for data centres that are the lifeblood of the AI revolution.

As the world grapples with the dual challenges of advancing AI technology and ensuring sustainable energy sources, the prospect of nuclear-powered data centres represents a bold step forward. While the path to adopting SMRs is fraught with regulatory, safety, and cost considerations, the potential benefits of a reliable, high-capacity power source for AI data centres could redefine the landscape of technology and energy.

Summary

The future of AI data centres may be powered by Small Modular Reactors (SMRs), offering a solution to their high power demands. With the world's first SMR under construction in China and designs being approved in the US, the technology is gaining traction. However, opinions on the transition to nuclear-powered data centres vary among experts, with concerns over safety, costs, and environmental impacts. The involvement of AI industry leaders in the development of SMRs highlights the potential for a sustainable and reliable energy source for the burgeoning AI sector.

Comment from Elon Musk

Yeah, I think that's a very sensible approach. I mean, I'm not [inaudible] a huge fan of nuclear energy, but I think it's definitely got a place. I think SMRs definitely have a place. I think generally the smaller the reactor, the better. I think the big reactors are just asking for trouble. And so, just like I think you should do with anything, and I try to apply that same thinking to SpaceX as well, is you want to have redundant systems, you want to have a lot of little things as opposed to a few big things so that if something goes wrong with one of them, it's not a crisis. And so I think in general the future of energy will actually be a mix of solar and wind and nuclear, and then you'll have hydro and geothermal, maybe a few other things in there. But I think it's going to be mostly solar and wind, and then the nuclear will be predominantly SMR. And then, if it's done well, it will be done in such a way that it's not a risk to the public. Yeah.

External Links

What are Small Modular Reactors (SMRs)?

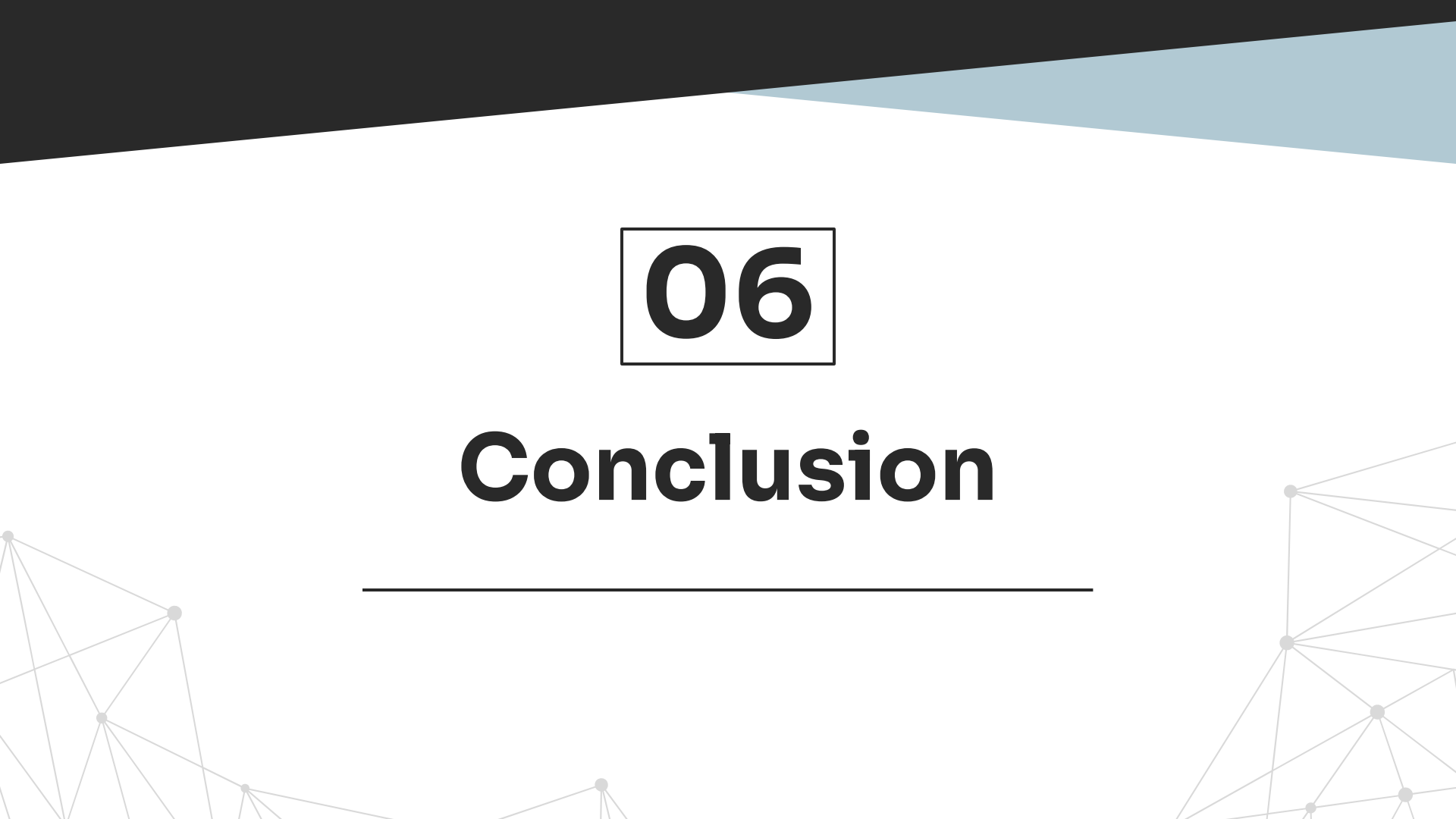
SMRs are compact nuclear reactors that offer enhanced safety features, flexibility, and easier deployment compared to traditional nuclear power plants.



The Impact of Artificial Intelligence on Data Centers: A Comprehensive Analysis

AI has revolutionized data centers, optimizing resource allocation, improving energy efficiency, enabling predictive maintenance, and enhancing operational performance.

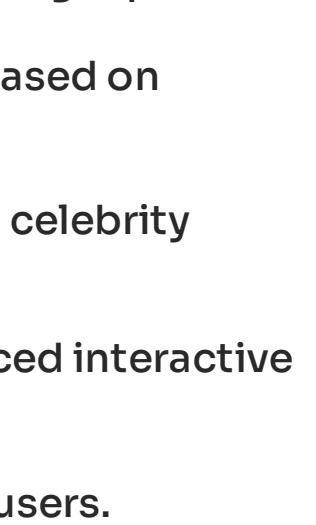




06

Conclusion

Conclusion

- Developed to mimic celebrity personalities using natural language processing.
 - Skillfully replicates unique communication styles and thought patterns.
 - Continuously enhanced through strategic refinements based on comprehensive research.
 - Enabled automated generation and posting of articles in celebrity voices.
 - Expand to function across various platforms with enhanced interactive features.
 - Make the technology more accessible and engaging for users.
- 

Reference

1. Bbc news rss feed. https://feeds.bbc.co.uk/news/science_and_environment/rss.xml. [Accessed 15-04-2024]. 43
2. X. Amatriain. Prompt design and engineering: Introduction and advanced methods. <https://arxiv.org/abs/2401.14423v3>, 2024. Accessed: Apr. 15, 2024.
3. David Baidoo-Anu and Leticia Owusu Ansah. Education in the era of generative artificial intelligence (ai): Understanding the potential benefits of chatgpt in promoting teaching and learning. *Journal of AI*, 7(1):52–62, 2023.
4. G. Brockman. Introducing chatgpt and whisper apis. <https://openai.com/blog/introducing-chatgpt-and-whisper-apis>, 2023. Accessed: Apr. 15, 2024.
5. Guiming Hardy Chen, Shunian Chen, Ziche Liu, Fengjiang, and Benyou Wang. Humans or llms as the judge? a study on judgement biases. *arXiv preprint arXiv:2402.10669*, 2024.
6. Katherine Elkins and Jon Chun. Can gpt-3 pass a writer's turing test? *Journal of Cultural Analytics*, 5(2), 2020.
7. I. J. Goodfellow et al. Generative adversarial networks. <https://arxiv.org/abs/1406.2661>, June 2014. Accessed: Apr. 15, 2024.
8. Google. Youtube data api. <https://developers.google.com/youtube/v3?hl=zh-tw>, 2024. Accessed: Apr. 15, 2024.
9. Neel Guha, Julian Nyarko, Daniel Ho, Christopher Ré, Adam Chilton, Alex Chohlas-Wood, Austin Peters, Brandon Waldon, Daniel Rockmore, Diego Zambrano, et al. Legalbench: A collaboratively built benchmark for measuring legal reasoning in large language models. *Advances in Neural Information Processing Systems*, 36, 2024.
10. E. J. Hu et al. Lora: Low-rank adaptation of large language models. <https://arxiv.org/abs/2106.09685>, June 2021. Accessed: Apr. 15, 2024.
11. Jen-tse Huang, Man Ho Lam, Eric John Li, Shujie Ren, Wenxuan Wang, Wenxiang Jiao, Zhaopeng Tu, and Michael R Lyu. Emotionally numb or empathetic? evaluating how llms feel using emotionbench. *arXiv preprint arXiv:2308.03656*, 2023.
12. Jen-tse Huang, Wenxuan Wang, Eric John Li, Man Ho LAM, Shujie Ren, Youliang Yuan, Wenxiang Jiao, Zhaopeng Tu, and Michael Lyu. On the humanity of conversational ai: Evaluating the psychological portrayal of llms. In *The Twelfth International Conference on Learning Representations*, 2024.
13. HuggingFace. Supervised fine-tuning trainer. https://huggingface.co/docs/trl/sft_trainer, 2024. Accessed: Apr. 15, 2024.
14. M. E. Jang and T. Lukasiewicz. Consistency analysis of chatgpt. *arXiv.org*, March 2023. <https://arxiv.org/abs/2303.06273> (accessed Apr. 15, 2024).
15. Daniel Jannai, Amos Meron, Barak Lenz, Yoav Levine, and Yoav Shoham. Human or not? a gamified approach to the turing test. *arXiv preprint arXiv:2305.20010*, 2023.
16. A. Q. Jiang et al. Mistral 7b. <https://arxiv.org/abs/2310.06825>, October 2023. Accessed: Apr. 15, 2024.
17. Wenxiang Jiao, Wenxuan Wang, Jen-tse Huang, Xing Wang, and Zhaopeng Tu. Is chatgpt a good translator? a preliminary study. *arXiv preprint arXiv:2301.08745*, 2023.
18. Douglas Johnson, Rachel Goodman, J Patrinely, Cosby Stone, Eli Zimmerman, Rebecca Donald, Sam Chang, Sean Berkowitz, Avni Fin n, Eiman Jahangir, et al. Assessing the accuracy and reliability of ai-generated medical responses: an evaluation of the chat-gpt model. *Research square*, 2023.

Reference

1. Cameron Jones and Benjamin Bergen. Does gpt-4 pass the turing test? arXiv preprint arXiv:2310.20216, 2023.
2. J. Knoll and J. Matthes. The effectiveness of celebrity endorsements: a meta analysis. Journal of the Academy of Marketing Science, 45(1):55–75, October 2016.
3. Aobo Kong, Shiwan Zhao, Hao Chen, Qicheng Li, Yong Qin, Ruiqi Sun, and Xin Zhou. Better zero-shot reasoning with role-play prompting. arXiv preprint arXiv:2308.07702, 2023.
4. LangChain. Few-shot prompt templates. https://python.langchain.com/docs/modules/model_io/prompts/few_shot_examples/, 2024a. Accessed: Apr. 15, 2024.
5. LangChain. Langgraph langchain blog. <https://blog.langchain.dev/langgraph/>, 2024b. Accessed: Apr. 15, 2024.
6. Pier Luca Lanzi and Daniele Loiacono. Chatgpt and other large language models as evolutionary engines for online interactive collaborative game design. In Proceedings of the Genetic and Evolutionary Computation Conference, pages 1383–1390, 2023.
7. P. Lewis et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. <https://arxiv.org/abs/2005.11401>, May 2020. Accessed: Apr. 15, 2024.
8. Cheng Li, Ziang Leng, Chenxi Yan, Junyi Shen, Hao Wang, Weishi Mi, Yaying Fei, Xiaoyang Feng, Song Yan, HaoSheng Wang, et al. Chatharuhi: Reviving anime character in reality via large language model. arXiv preprint arXiv:2308.09597, 2023.
9. Medium. Github - medium-api-docs: Documentation for medium's oauth2 api. <https://github.com/Medium/medium-api-docs>, 2024. Accessed: Apr. 15, 2024.
10. Microsoft. Cloud computing services. <https://azure.microsoft.com/en-us/>, 2024a. Accessed: Apr. 15, 2024.
11. Microsoft. Bing search services developer documentation. <https://learn.microsoft.com/en-us/rest/api/cognitiveservices-bingsearch/bing-web-api-v7-reference>, 2024b. Accessed: Apr. 15, 2024.
12. Y. Nakajima. Task-driven autonomous agent utilizing gpt-4, pinecone, and langchain for diverse applications—yohei nakajima. <https://yoheinakajima.com/task-driven-autonomous-agent-utilizing-gpt-4-pinecone-and-langchain-for-diverse-2023>. Accessed: Apr. 15, 2024.
13. OpenAI. Introducing gpts. OpenAI Blog Nov 6, 2023, 2023.
14. OpenAI. Dall-e 3. <https://openai.com/dall-e-3>, 2024a. Accessed: Apr. 15, 2024.
15. OpenAI. Openai fine-tuning. <https://platform.openai.com/docs/guides/fine-tuning>, 2024b. Accessed: Apr. 15, 2024.
16. OpenAI. Openai platform. <https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo>, 2024c. Accessed: Apr. 15, 2024.
17. Paul Bellow. Dalle3 prompt tips and tricks thread. <https://community.openai.com/t/dalle3-prompt-tips-and-tricks-thread/498040>, 2023. Accessed: Apr. 15, 2024.
18. Gabriel Preda. Elon musk tweets. <https://www.kaggle.com/datasets/gpreda/elon-musk-tweets>, 2023. Accessed: Apr. 15, 2024.
19. Terrence J Sejnowski. Large language models and the reverse turing test. Neural computation, 35(3):309–342, 2023.
20. Murray Shanahan, Kyle McDonell, and Laria Reynolds. Role play with large language models. Nature, pages 1–6, 2023.

Reference

1. Yunfan Shao, Linyang Li, Junqi Dai, and Xipeng Qiu. Character-llm: A trainable agent for role-playing. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, pages 13153–13187, 2023.
2. Stanfordnlp. Github - stanfordnlp/dspy: Dspy: The framework for programming—not prompting—foundation models. 2024. Accessed: Apr. 15, 2024.
3. subzeroid. Github - subzeroid/instagrapi: The fastest and powerful python library for instagram private api 2024. <https://github.com/subzeroid/instagrapi>, 2024. Accessed: Apr. 15, 2024.
4. Nigar M Shafiq Surameery and Mohammed Y Shakor. Use chat gpt to solve programming bugs. International Journal of Information technology and Computer Engineering, 31:17–22, 2023.
5. Alan M Turing. Computing machinery and intelligence. Mind, LIX(236):433–460, 1950. 14, 15 Twitter. Twitter api documentation. <https://developer.twitter.com/en/docs/twitter-api>, 2024. Accessed: Apr. 15, 2024.
6. Adaku Uchendu, Zeyu Ma, Thai Le, Rui Zhang, and Dongwon Lee. Turingbench: A benchmark environment for turing test in the age of neural text generation. In Findings of the Association for Computational Linguistics: EMNLP 2021, pages 2001–2016, 2021.
7. UnslothAI. Github - unslothai/unsloth: 2-5x faster 80 <https://github.com/unslothai/unsloth>, 2024. Accessed: Apr. 15, 2024.
8. Zekun Moore Wang, Zhongyuan Peng, Haoran Que, Jiaheng Liu, Wangchunshu Zhou, Yuhang Wu, Hongcheng Guo, Ruitong Gan, Zehao Ni, Man Zhang, et al. Rollem: Benchmarking, eliciting, and enhancing role-playing abilities of large language models. arXiv preprint arXiv:2310.00746, 2023.
9. Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. arXiv preprint arXiv:2201.11903, 2022.
10. Haoran Wu, Wenxuan Wang, Yuxuan Wan, Wenxiang Jiao, and Michael Lyu. Chatgpt or grammarly? evaluating chatgpt on grammatical error correction benchmark. arXiv preprint arXiv:2303.13648, 2023.
11. L. Xu, H. Xie, S.-Z. J. Qin, X. Tao, and F. L. Wang. Parameter-efficient finetuning methods for pretrained language models: A critical review and assessment. <https://arxiv.org/abs/2312.12148>, December 2023a. Accessed: Apr. 15, 2024.
12. Z. Xu et al. Compress, then prompt: Improving accuracy-efficiency trade-off of llm inference with transferable prompt. arXiv.org May 2023b. <https://arxiv.org/abs/2305.11186>.
13. Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. arXiv preprint arXiv:2305.10601, 2023.
14. Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing et al. Judging llm-as-a-judge with mt-bench and chatbot arena. Advances in Neural Information Processing Systems, 36, 2024.
15. Jinfeng Zhou, Zhuang Chen, Dazhen Wan, Bosi Wen, Yi Song, Jifan Yu, Yongkang Huang, Libiao Peng, Jiaming Yang, Xiyao Xiao, et al. CharacterGlm: Customizing chinese conversational ai characters with large language models. arXiv preprint arXiv:2311.16832, 2023.
16. Yutao Zhu, Huaying Yuan, Shuting Wang, Jiongnan Liu, Wenhan Liu, Chenlong Deng, Zhicheng Dou, and Ji-Rong Wen. Large language models for information retrieval: A survey. arXiv preprint arXiv:2308.07107, 2023.

Appendix

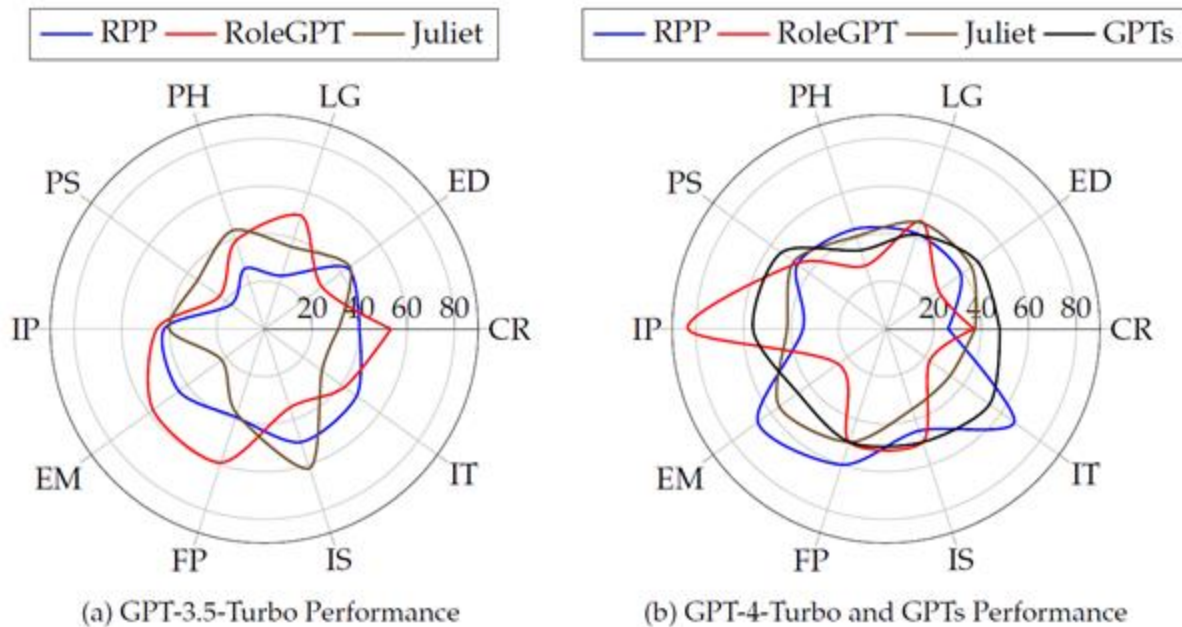


Figure 2: Success rates of role-playing LLMs in deceiving human evaluators. The human evaluators are instructed to identify **human**-generated responses.

Baselines' Success Rate Radar Graph

Fine Tuning HyperParameters

```
model = FastLanguageModel.get_peft_model(  
    model,  
    r = 32, # Choose any number > 0 ! Suggested 8, 16, 32, 64, 128  
    target_modules = ["q_proj", "k_proj", "v_proj", "o_proj",  
                     "gate_proj", "up_proj", "down_proj",],  
    lora_alpha = 16,  
    lora_dropout = 0, # Supports any, but = 0 is optimized  
    bias = "none",    # Supports any, but = "none" is optimized  
    use_gradient_checkpointing = True,  
    random_state = 3407,  
    use_rslora = True, # We support rank stabilized LoRA  
    loftq_config = None, # And LoftQ  
)
```

```
output_dir = "outputs",  
per_device_train_batch_size = 4,  
gradient_accumulation_steps = 4,  
warmup_steps = 5,  
max_steps = 8,  
learning_rate = 2e-4,  
fp16 = not torch.cuda.is_bf16_supported(),  
bf16 = torch.cuda.is_bf16_supported(),  
logging_steps = 1,  
optim = "adamw_8bit",  
weight_decay = 0.01,  
lr_scheduler_type = "linear",  
seed = 3407,  
)
```

Some augments in PEFT, LoRA and SFTTrainer



Data Augmentation

1. 人工智慧資料中心面臨哪些具體挑戰需要使用核反應器？
2. 小型模組化反應器（**SMR**）在發電和安全方面與傳統核反應器有何不同？
3. 為什麼資料中心被認為是部署 **SMR** 的主要候選者？
4. 將 **SMR** 整合到資料中心會帶來哪些環境和安全問題？
5. 人工智慧資料中心的功耗與傳統資料中心相比如何？
6. **SMR** 技術目前正在取得哪些進展，哪些公司正在引領這些發展？
7. 人工智慧在增加資料中心的電力需求方面發揮什麼作用？
8. 資料中心使用核電相對於再生能源有哪些潛在的好處和缺點？
9. 大學和研究機構如何為 **SMR** 技術的發展和理解做出貢獻？
10. **SMR** 商業營運的時間表是什麼？這如何滿足人工智慧資料中心不斷增長的電力需求？

Thanks



Do you have any questions?

httse0930@link.cuhk.edu.hk

derek331@link.cuhk.edu.hk

CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#) and infographics & images by [Freepik](#)

Please keep this slide for attribution