

# Artificial Intelligence Master of History

Yuming Zou

Guangzhou Tianxing Experimental School, Guangzhou, China  
[zhongda\\_2006@outlook.com](mailto:zhongda_2006@outlook.com)

---

## ABSTRACT

In this EPQ, I built a system that successively enables the large language model to answer questions on the past with evidence from primary sources. This project was motivated by the realization that despite the complexity, the performance of the large language model (LLMs) literally means understanding but simply does pattern recognition, coupled with probabilistic text prediction. Against this backdrop, a specialized AI was developed to help in error correction, given that AI-produced historical data are generally full of errors, by the use of the open-source Langchain-Chatcat that allows integration of LLMs with a structured knowledge database. These contain factual and historical information needed for the AI to compute the accurate answer. The system uses the GLM-4 model by Tsinghua University for the level of GPT-4 and above. This web application is built on Streamlit, an AI-powered system that interacts with the user. The steps through which input queries and historical text were processed by the system include an Unstructured Loader, a Text Splitter, an embedding model, and a Vector Store. Successful implementation of this system not only improved the reliability in AI for historical education but also laid the foundation for further improvement in educational technology and AI interaction models. Future plans will involve further historical database expansion to even more diverse periods and geographies, as well as enhancing the user interface to become more interactive and accessible.

## KEYWORDS

Artificial Intelligence; Large language models (LLMs); Historical education

---

## 1. INTRODUCTION

Most recently, there has been a gigantic surge in the use of large language models (LLMs) at learning institutions, proposing a way through which students and educators can tap into various forms of knowledge. With such higher computational capabilities, LLMs like Chat-GPT often bring in limitations while producing accurate and reliable facts from history. More than one such a problem was noticed, for example, during one of the festivals at our center when the AI model confused historical personalities of the Three Kingdoms period in ancient China. Such inaccuracies are due to the nature of the LLMs themselves. They lack real understanding or consciousness and operate mainly on pattern recognition and probabilistic prediction from the training data.

This finding motivated me to dive deeper into the root causes of such inaccuracies, which identified what we found to be the "illusion of understanding" in AI systems. So, there is an urgent need to introduce a new approach to mitigate these problems and enhance the reliability of AI in historical education. Thus, the paper introduces a novel AI system designed to overcome the pitfalls of traditional LLMs through the integration of a structured knowledge database customarily made for historical information. This integration hugely facilitates AI in coming up with responses that are not

only contextually valid but, most importantly, verifiably correct, which touches on the complex demands in historical research and education.

This will cover the methodology used to come up with this AI system, its architecture of the knowledge database, how the implementation should be carried out using the Langchain-Chatchat open-source tool, and the results thereafter. It should illustrate the area of its implementation where these improvements would make a very significant difference in raising the functional capabilities of the AI system within educational technologies and, therefore, set a precedent for future AI-driven learning tool advances

## 2. RELATED WORK

Chat-GPT and other large language models (LLMs) have brought to the discussions around the potential for their application in historical learning environments. The promise that such technologies hold in aiding learning by easily surfacing a lot of information. They noticed limitations that still need to be tackled in order to improve their reliability and effectiveness in educational applications.

The systematic scoping review by Yan et al. (2023) focused on the description of practical and ethical problems resulting from the use of LLM in the educational context. The most common are those relating to privacy and technological readiness. This cumulatively results in a potentially high integration of LLMs with structured knowledge databases that are extremely important to raise further the utility value for historical education (Yan et al., 2023).

Along this same line, Soos and Haroutunian (2014) delved deeper into the impact of LLMs on the activity of authorship and knowledge organization within historical education. Hence, argument suffices that the blurred boundaries around creators and the works, created by LLMs underpin the prevailing challenges of entity-relationship models. They further their educational use to teach the complex historical context (Soos & Haroutunian, 2024).

Bui et al. (2024) recommend the use of structured knowledge bases, like the Knowledge Graph, to reduce some of the most obvious intrinsic limitations of LLMs: their incapacities in effectively contextualizing historical facts and updating generated knowledge according to newly discovered sources. This would largely be premised on their study, suggesting such integration to be of great help in improving factual accuracy and responsiveness of LLMs within educational question answering. Bui et al. (2024) argue that this would make LLMs fit much more for historical education.

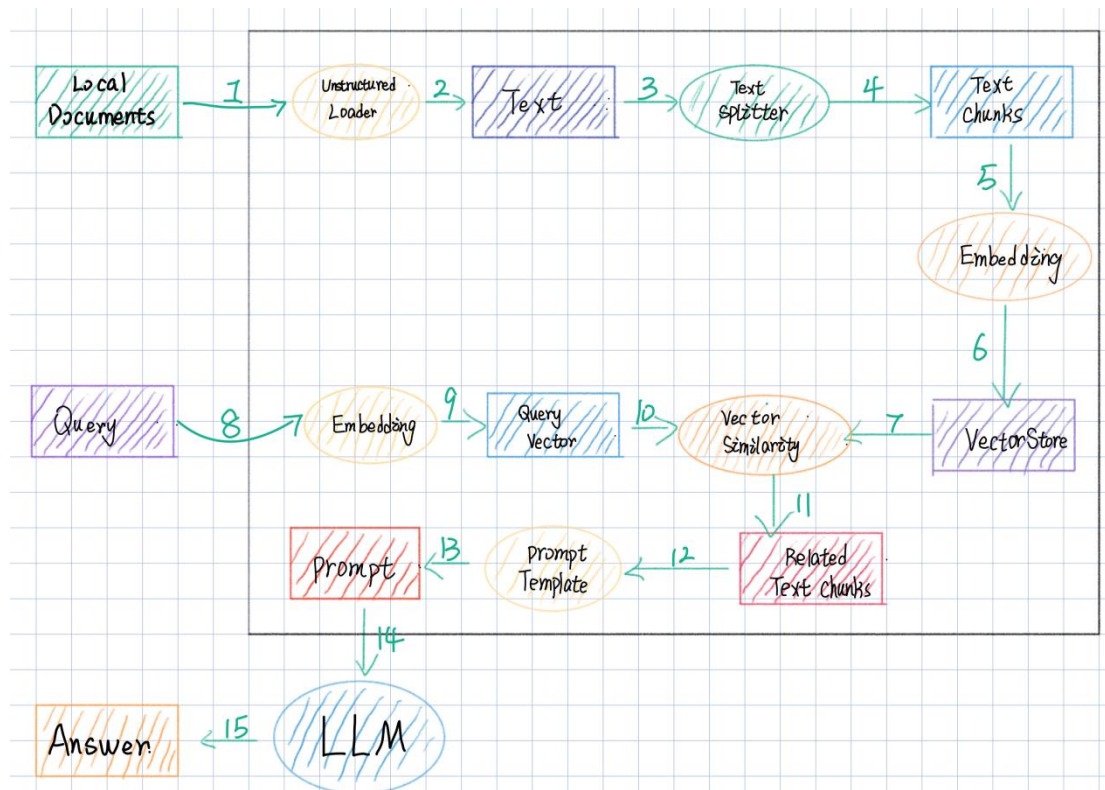
In addition, Wang and Yu (2023) showed a research space analysis for the education of the system in AI, which stressed that the major focus of the system is smart technologies and smart decision-making systems. A gap was identified in the area of attention in integrating LLM with structured knowledge databases that are essential keys within historical education to ensure the AI system provides contextually relevant and historically accurate information (Wang & Yu, 2023).

Finally, Edyko et al. (2023) reviewed the ethical considerations and verification challenges of sources that are very important in the deployment of LLMs, including Chat-GPT, in historical education. They note that in all cases, special care must be taken to ensure credibility and accuracy of information, especially when using such tools to form students' perception of historical events (Edyko et al., 2023).

Taken together, these studies set a foundation of understanding the challenges and potential strategies of integrating LLMs with Structured Knowledge Databases so that they can be put at the service by them in Historical Education, hence taking care of the complex needs of the same.

### 3. METHODOLOGY

#### 3.1. Overall Structure



**Figure 1.** visualization of code structure

In the first step, these documents are processed by an Unstructured Loader and converted into plain text, which is then segmented into manageable chunks by a Text Splitter. Each chunk is subsequently transformed into vectors by an embedding model. These vectors are stored in a Vector Store for efficient retrieval. When a user inputs a query, it is also embedded and transformed into a vector, allowing for a comparison of its similarity to the stored text embeddings. The system identifies the text chunks most relevant to the query based on similarity scores. These related text chunks are crafted into a Prompt Template, which is then fed into a Large Language Model (LLM). The LLM utilizes these inputs to generate an Answer that addresses the user's query.

#### 3.2. LangChain

LangChain is a framework for developing applications powered by large language models. It helps me to integrate the language model with the database; manage the flow of the conversation, including handling multiple steps, maintaining context, and implementing logic rules that guide the interaction.

#### 3.3. Embedding Model

For the embedding model, I choose bge-large-zh-v1.5 from Huggingface.co. It is a powerful embedding model which can be used for embed both Chinese and English.

#### 3.4. Large Language Model(LLM)

For the Large Language model, I chose GLM-4, invented by Tsinghua University, and it has similar capabilities to GPT-4 in multiple fields.

	MMLU(5-shot)	GSM8K(5-shot)	MATH(4-shot)	BBH(3-shot)	HellaSwag(10-shot)
GPT-4	86.4	92.0	52.9	83.1	95.3
Gemini-Ultra	83.7	94.4	53.2	83.6	87.8
<b>GLM-4</b>	<b>81.5</b>	<b>87.6</b>	<b>47.9</b>	<b>82.3</b>	<b>85.4</b>
GLM-4/GPT-4	94%	95%	91%	99%	90%

**Figure 2.** Capability of GLM-4

### 3.5. Web App

Finally, I used “Streamlit” to create the web app for the interaction between the user and the program. Streamlit is an open-source Python library designed to simplify the process of creating and sharing beautiful, custom web apps for machine learning and data science projects.

### 3.6. Database

The books that is incorporate in the database in the project includes:

The 24 Official Histories

100 Decisive Battles. From Ancient Times to the Present (Paul K. Davis)

The Annals The Reigns of Tiberius, Claudius, and Nero

The Diary of Samuel Pepys, Complete (Samuel Pepys)

The Histories (Herodotus, Robin Waterfield)

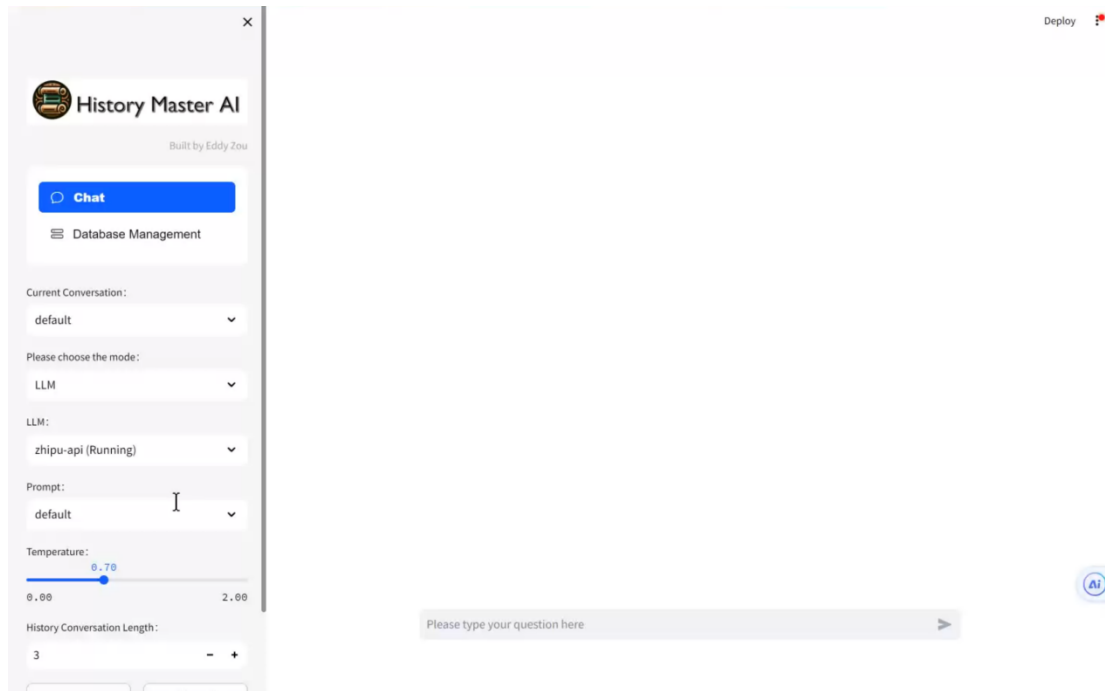
The Social Contract (Jean-Jacques Rousseau)

These books cover Chinese and Euperean histories and constitute a data base with more than 30,000,000 tokens.

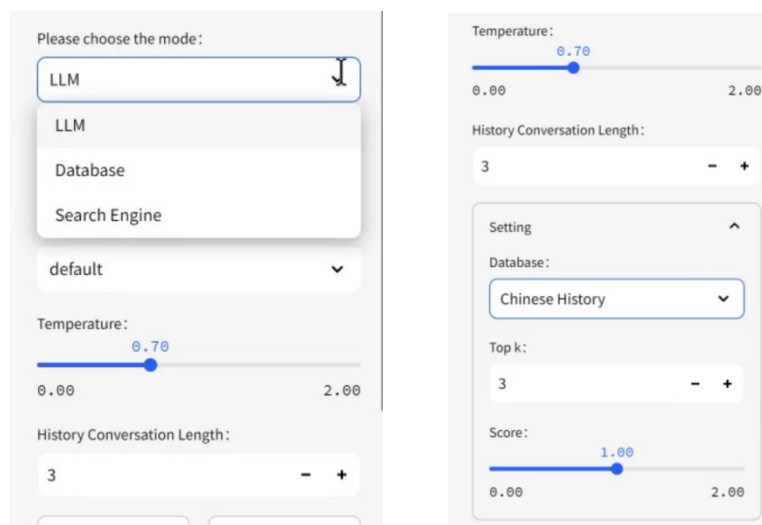
## 4. DISCUSSION

### 4.1. Functions

On the web app, users can directly talk to large language model or talk to large language model with a database. In addition, user can set the temperature (randomness of the responses), history Conversation Length, database, top K (the top n relevent text chunks) and similarity score when they use the web app.



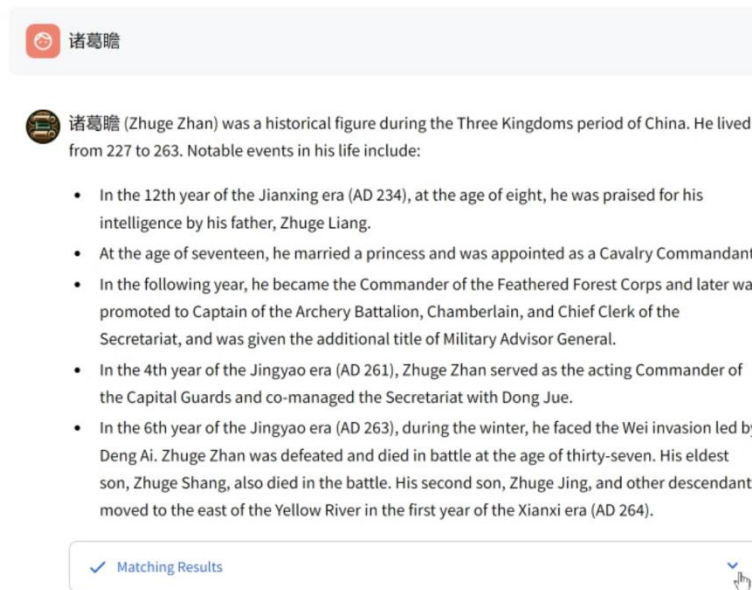
**Figure 3.** The web page of the program



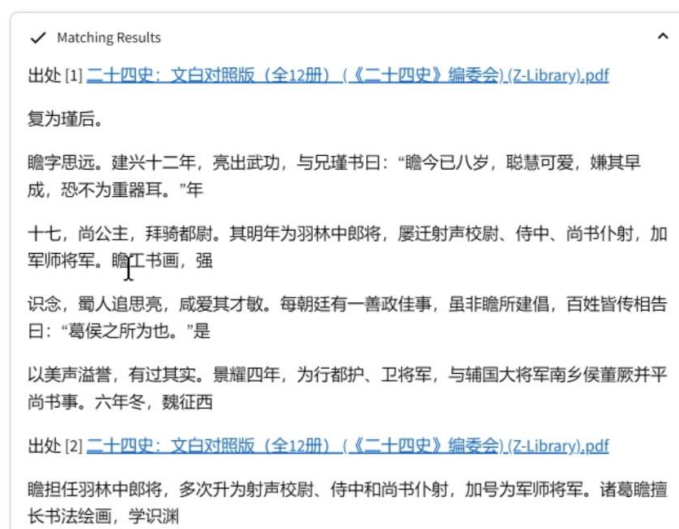
**Figure 4.** a. switch the mode; b. function zone

## 4.2. Examples

After asking AI History Master about “Zhuge Zhan”, a figure in Chinese history, it provides a precise and detailed response (figure 5) with references (figure 6) about the life experience of him.

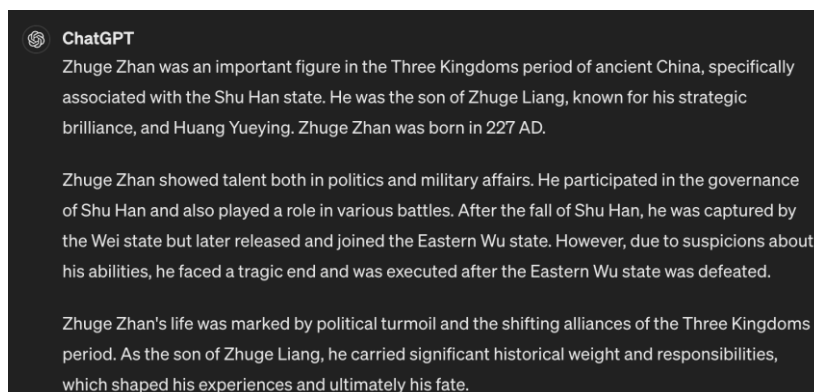


**Figure 5. Response 1**



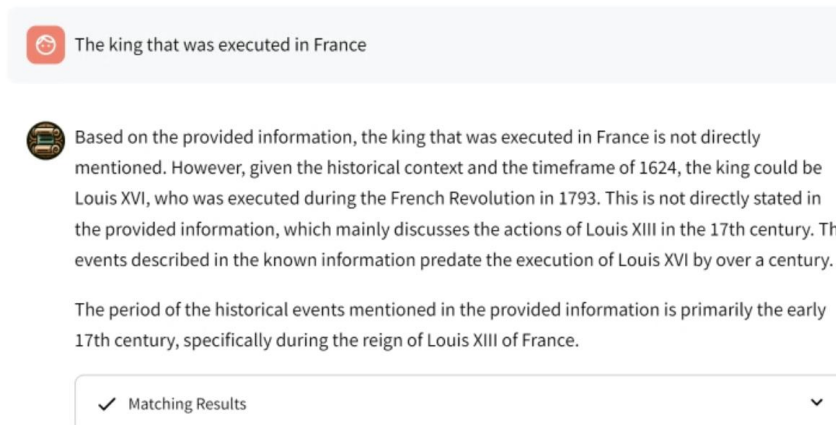
**Figure 6. Reference 1**

Compare to the response of Chat-GPT, which is a defective response with several incorrect description of Zhuge Zhan’s life, AI History Master provides a more accurate and specific answer of Chinese History.



**Figure 7. The response of Chat-GPT**

When asking AI History Master some questions about European History, it also provide precise and detailed responses. (figure 8, 9)



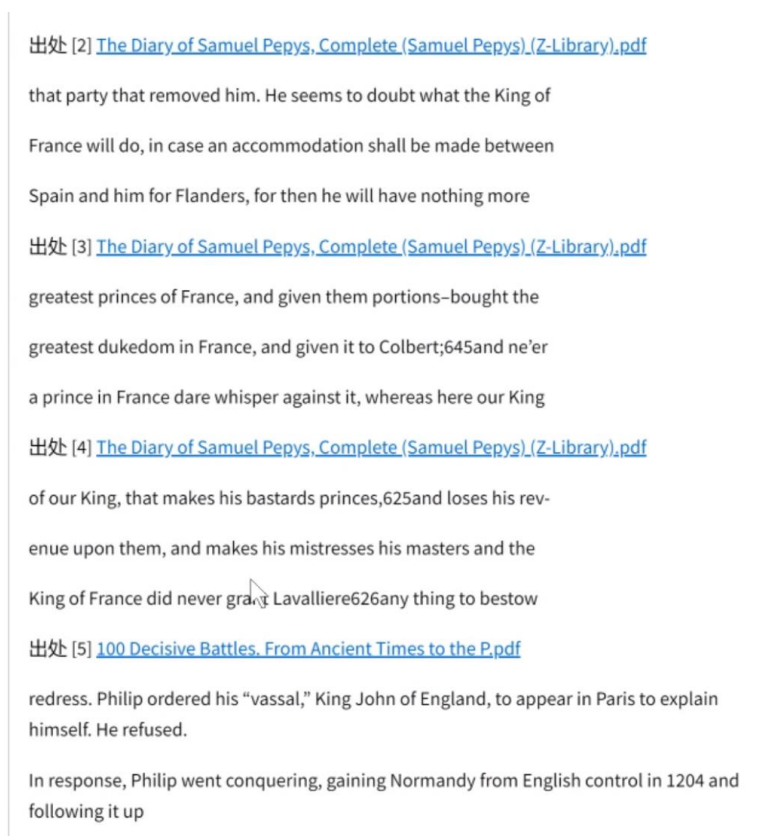
The king that was executed in France

Based on the provided information, the king that was executed in France is not directly mentioned. However, given the historical context and the timeframe of 1624, the king could be Louis XVI, who was executed during the French Revolution in 1793. This is not directly stated in the provided information, which mainly discusses the actions of Louis XIII in the 17th century. The events described in the known information predate the execution of Louis XVI by over a century.

The period of the historical events mentioned in the provided information is primarily the early 17th century, specifically during the reign of Louis XIII of France.

✓ Matching Results

**Figure 8.** Response 2



出处 [2] [The Diary of Samuel Pepys, Complete \(Samuel Pepys\) \(Z-Library\).pdf](#)

that party that removed him. He seems to doubt what the King of France will do, in case an accommodation shall be made between Spain and him for Flanders, for then he will have nothing more

出处 [3] [The Diary of Samuel Pepys, Complete \(Samuel Pepys\) \(Z-Library\).pdf](#)

greatest princes of France, and given them portions—bought the greatest dukedom in France, and given it to Colbert;645and ne'er a prince in France dare whisper against it, whereas here our King

出处 [4] [The Diary of Samuel Pepys, Complete \(Samuel Pepys\) \(Z-Library\).pdf](#)

of our King, that makes his bastards princes,625and loses his revenue upon them, and makes his mistresses his masters and the King of France did never gra. Lavalliere626any thing to bestow

出处 [5] [100 Decisive Battles. From Ancient Times to the P.pdf](#)

redress. Philip ordered his “vassal,” King John of England, to appear in Paris to explain himself. He refused.

In response, Philip went conquering, gaining Normandy from English control in 1204 and following it up

**Figure 9.** Reference 2

## 5. CONCLUSION

The project has shown how advanced AI technologies, in particular Langchain and Large Language Models, can provide valid responses to historical questions with a high precision that takes primary evidence into account. Structured knowledge databases to enhance this work of basic models will, hence, bring forth important improvements in both the accuracy and reliability in the dissemination of AI-driven historical data. The models will build a strong proof of concept with regard to AI in historical education.

In the future, the project will expand its repository to account for information outside the time-span and countries. This will give a much richer database for training and further fine-tune the ability of AI to deal with much more diverse ranges of historical narratives and understandings. Also, the planned enhancements of the web interface will create an environment much more user-friendly and interactive, which will probably result in more lively contact with the material and much more intuitive learning.

More importantly, making the AI ask questions, in addition to answering historical queries, would further add interactive features that would transform the educational experience from a passive one to a dynamic interaction where the user is challenged by questions and receives informed responses complete with the needed citations. The users have a more engaging and educative platform through which active learning and critical thinking are promoted. In this sense, the project is not simply concerning AI capability related to the educational sector but is a step toward the future of AI and human expertise in the collaborative effort to foster understanding and appreciation of history. Thus, we will be able to pave the way for the design of more advanced educational tools that, with the development and refinement of such a system, hold a lot of promise to redefine approaches in teaching, learning, and the interaction with our collective past.

## ACKNOWLEDGEMENTS

I am grateful to Mr. Zhou for his assistance in my EPQ project. Special thanks to my classmate, Shixuan Wu, for her provision of figure 1.

## REFERENCES

- [1] Oguzhan Topsakal and T. Akinci (2023). Creating Large Language Model Applications Utilizing LangChain: A Primer on Developing LLM Apps Fast. *Journal of Information and Computational Science*, DOI:10.59287/icaens.1127.
- [2] Yuanhao Gong (2023). Dynamic Large Language Models on Blockchains. *arXiv preprint arXiv:2307.10549*, DOI:10.48550/arXiv.2307.10549.
- [3] Bo Xu and M. Poo (2023). Large language models and brain-inspired general intelligence. *National Science Review*, DOI:10.1093/nsr/nwad267.
- [4] Yuxiang Sun et al. (2023). Self Generated Wargame AI: Double Layer Agent Task Planning Based on Large Language Model. *arXiv preprint arXiv:2312.01090*, DOI:10.48550/arXiv.2312.01090.
- [5] Lixiang Yan et al. (2023). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, DOI:10.1111/bjet.13370.
- [6] Carlin Soos and Levon Haroutunian (2024). On the Question of Authorship in Large Language Models. *Journal of Literary Theory*, DOI:10.5771/0943-7444-2024-2-83.
- [7] Tuan Bui et al. (2024). Cross-Data Knowledge Graph Construction for LLM-enabled Educational Question-Answering System: A Case Study at HCMUT. *arXiv preprint arXiv:2404.09296*, arXiv:2404.09296.
- [8] Yakun Wang and Fuxiang Yu (2023). Visualization and Analysis of Mapping Knowledge Domains for AI Education System Studies. *2023 IEEE 12th International Conference on Intelligent Education Technology (ICIET)*, DOI:10.1109/ICIET56899.2023.10111115.
- [9] Krzysztof Edyko et al. (2023). Utilizing Artificial Intelligence Tools Using the GPT Chatbot in Medicine - A Review of Flaws, Advantages, and Limitations. *Journal of Education, Health and Sport*, DOI:10.12775/jehs.2023.46.01.008.
- [10] Langchainchatchat, [github.com/chatchat-space/Langchain-Chatchat/wiki/](https://github.com/chatchat-space/Langchain-Chatchat/wiki/)
- [11] LangChain, [python.langchain.com/docs/get\\_started/introduction/](https://python.langchain.com/docs/get_started/introduction/) Streamlit Documentation, [docs.streamlit.io/](https://docs.streamlit.io/)
- [12] BAAI/Bge-Large-Zh-v1.5, Hugging Face, <https://huggingface.co/BAAI/bge-large-zh-v1.5>