Note on [Learning to Edit: Aligning LLMs with Knowledge Editing]

NoteAuthor: XIONG ZHIPENG

June 11, 2024

1 Which problem is addressed by the paper?

Efficiently modify LLMs' outputs towards targeted queries while preserving overall performance across other unrelated ones

For example: updating the knowledge of "*The current British Prime Minister is Rishi Sunak*" modifies the response to "*Who is married to the PM of the UK?*"

2 Why is it important?

The dynamic nature of the world necessitates frequent updates to LLMs to rectify outdated information or integrate new knowledge, thereby safeguarding their sustained pertinence.

This paper argues that the previous methods predominantly rely on memorizing the update.

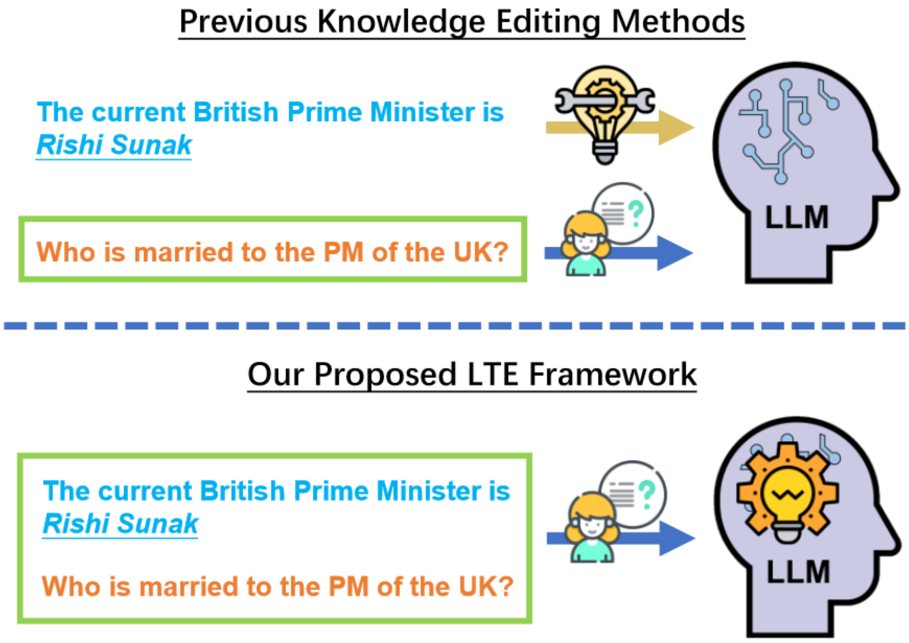
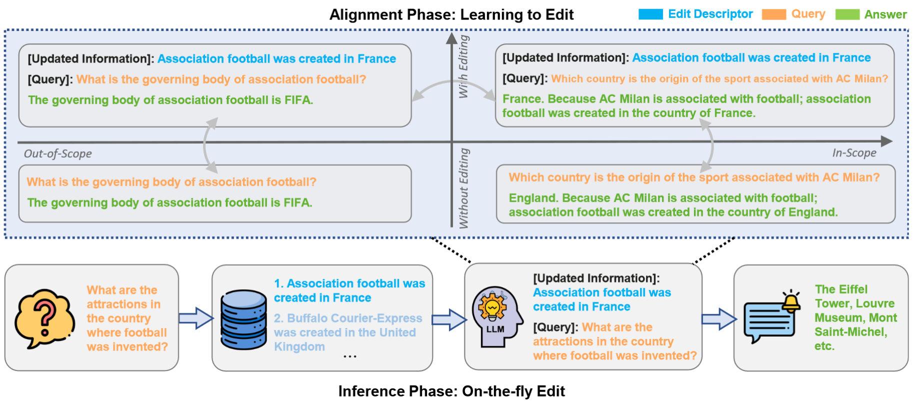

Previous Knowledge Editing Methods
The current British Prime Minister is Rishi Sunak
Who is married to the PM of the UK?
LLM
Our Proposed LTE Framework
The current British Prime Minister is Rishi Sunak
Who is married to the PM of the UK?
LLM

Figure 1: The previous and proposed principles

3 How is it solved?

Two phases

- **Alignment Phase:** train LLMs how to apply updated knowledge
- **Inference Phase:** propose a retrieval-based mechanism that retrieves relevant edit descriptors from a stored memory for real-time, mass editing requests.



[Updated Information]: Association football was created in France
[Query]: What is the governing body of association football?
The governing body of association football is FIFA.

[Updated Information]: Association football was created in France
[Query]: Which country is the origin of the sport associated with AC Milan?
France. Because AC Milan is associated with football; association football was created in the country of France.

What is the governing body of association football?
The governing body of association football is FIFA.

Which country is the origin of the sport associated with AC Milan?
England. Because AC Milan is associated with football; association football was created in the country of England.

What are the attractions in the country where football was invented?

1. Association football was created in France
2. Buffalo Courier-Express was created in the United Kingdom
...

[Updated Information]: Association football was created in France
[Query]: What are the attractions in the country where football was invented?

The Eiffel Tower, Louvre Museum, Mont Saint-Michel, etc.

Figure 2: The proposed feamework

4 DataAnalysis, Applications, Conclusion and Future Work