Paper Link

The Knowledge Alignment Problem: Bridging Human and External Knowledge for Large Language Models

Shuo Zhang[1] Liangming Pan [2] Junzhou Zhao [1] William Yang Wang [2]

[1] Xi'an Jiaotong University

[2] University of California, Santa Barbara

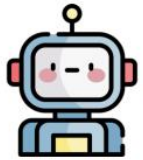
A major drawback of ChatGPT

LLM Hallucination

User Input

Can you recommend a delicious recipe for dinner?

LLM Response



Yes, here is a delicious recipe for lunch. So how about fried chicken with mashed potatoes? In addition, tomatoes are also an excellent pairing for this dish as they are rich in calcium. Enjoy this steak!

Hallucination Explanation

Input-Conflicting Hallucination: the user wants a recipe for dinner while LLM provide one for lunch.

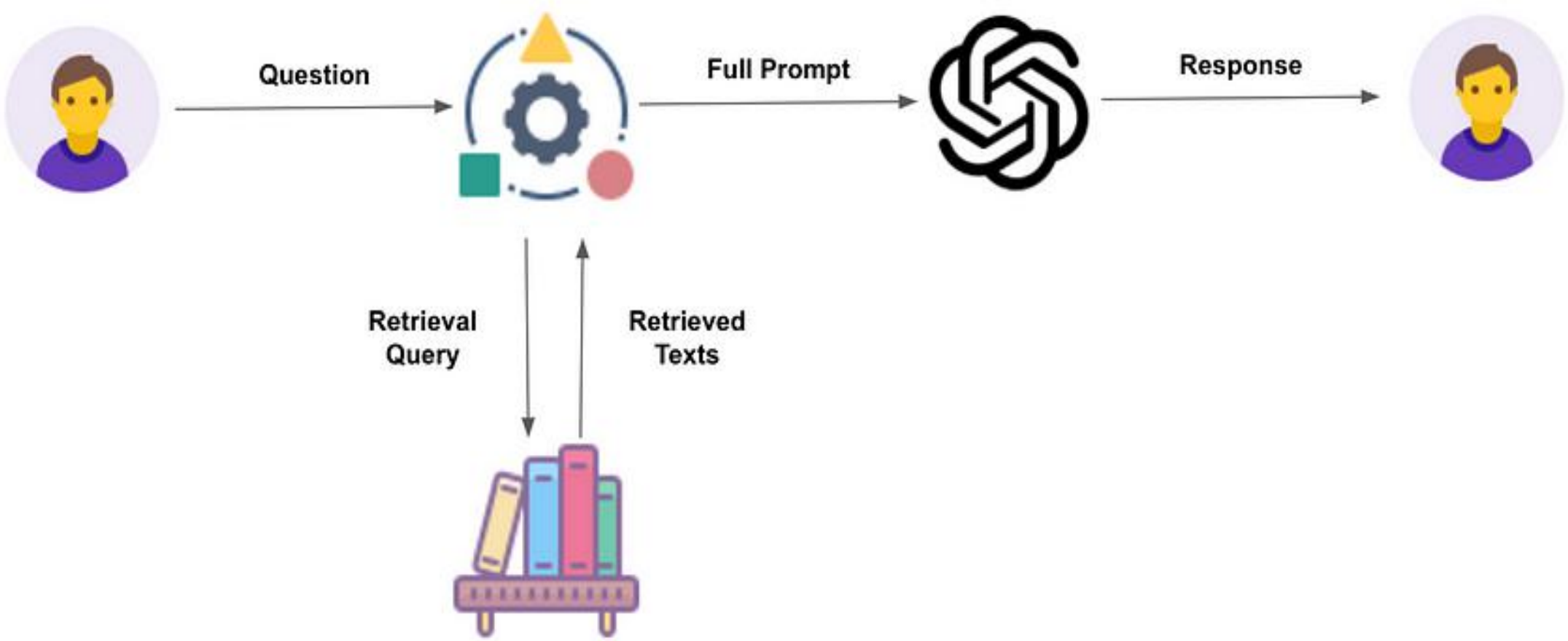
Context-Conflicting Hallucination: steak has not been mentioned in the preceding context.

Fact-Conflicting Hallucination: tomatoes are not rich in calcium in fact.

Source paper:
Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models

An effective approach to address hallucination

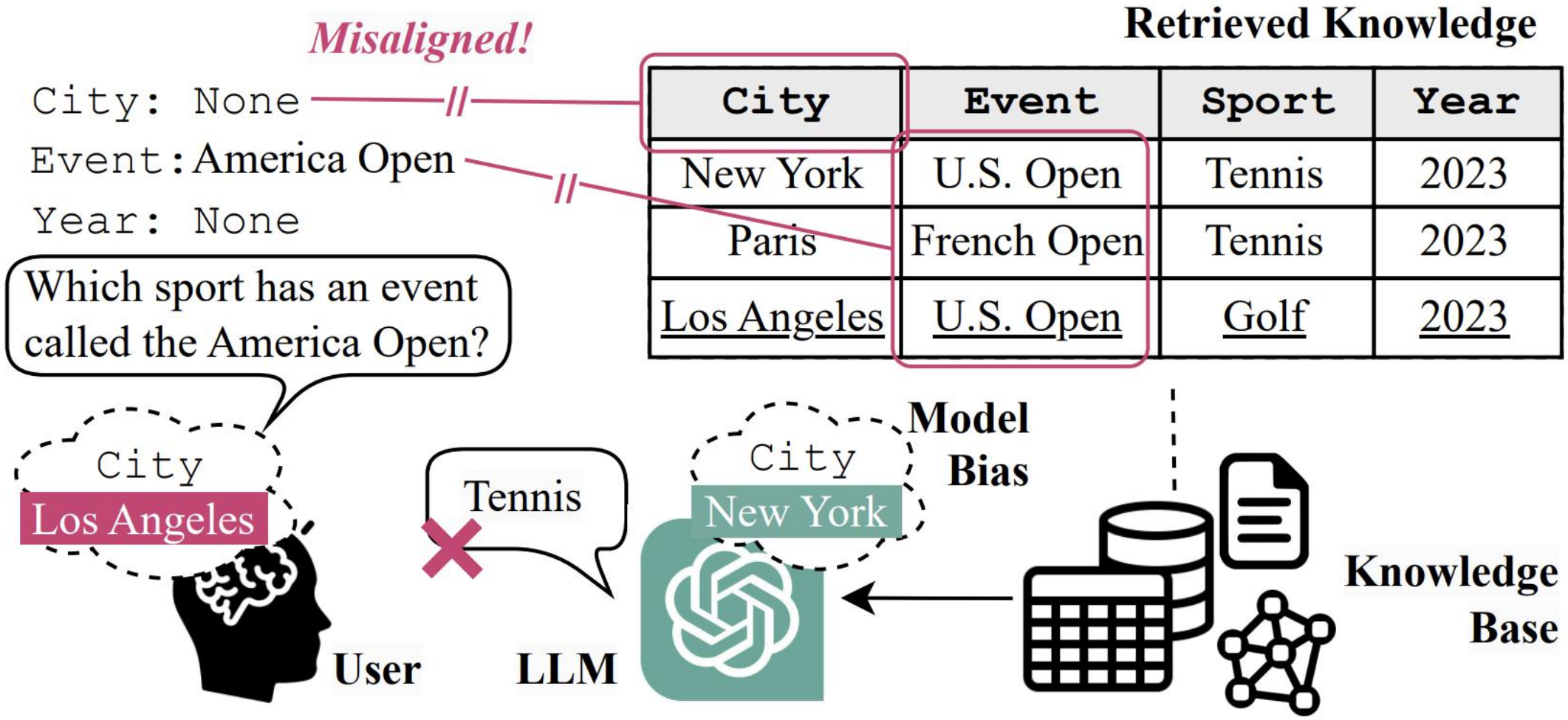
Retrieved-Augmented Generation



Pic Source: https://www.linkedin.com/pulse/what-retrieval-augmented-generation-grow-right

Retrieved-Augmented Generation Still Hallucinates!

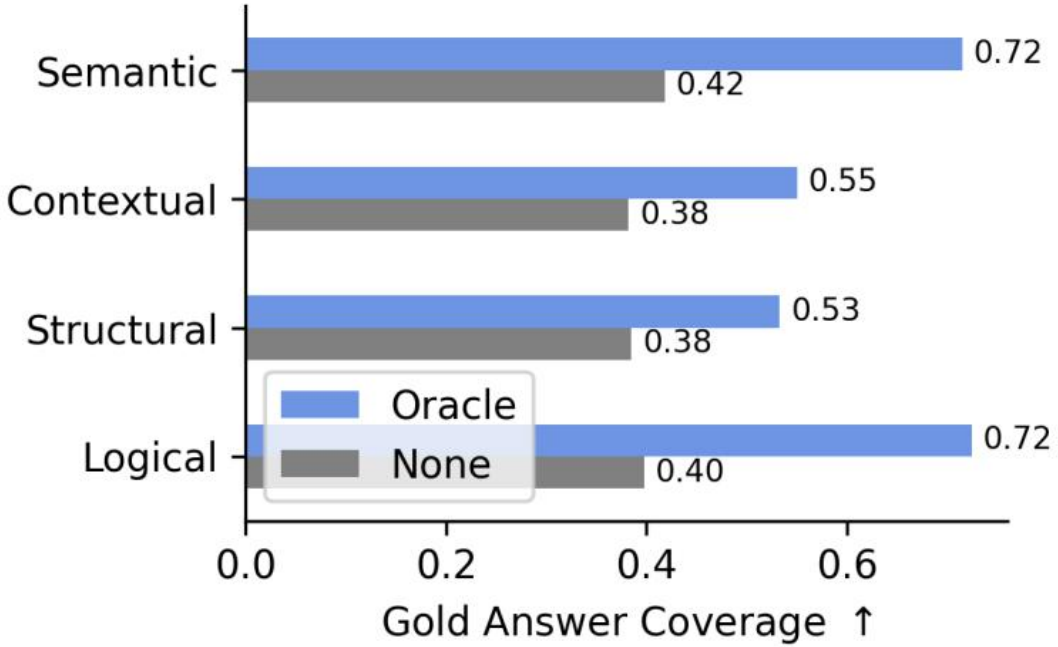
when there are *Knowledge Misalignments*

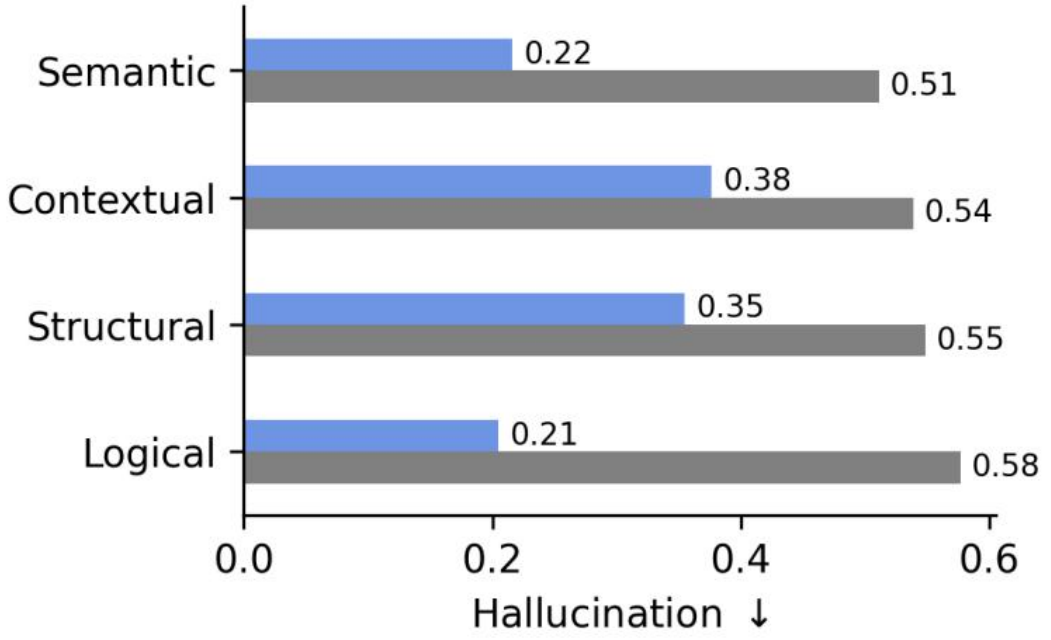


Knowledge Misalignments Types

Type	Explanation	Example
Semantic	The user might use an ambiguous term that, while ideally should map to a single item in the database, in reality can correspond to multiple attributes or values.	For "What is the best burger?", when you say "best burger", are you referring to taste, nutritional value, price, or a combination of these attributes?
Contextual	The user may have implicitly established some conditions without explicitly expressing them.	For "What is the 15th most populous city in the United States?", the statistics may vary with time, which year are you considering?
Structural	The user might have stated some conditions that are not addressed in the database.	For "Find me an American writer.", the question can not be answered when the database does not include the nationality of the writers.
Logical	The user can ask complex questions where certain conditions need to be determined before other conditions can be clarified, while the database only supports basic logits such as "and", "or" and "not".	For "Fine me for a movie directed by the singer who has won a Grammy', the question answering requires first identifying singer who have won a Grammy, and then finding the films directed by the identified singers.

Knowledge Misalignments → lower QA accuracy + more Hallucination



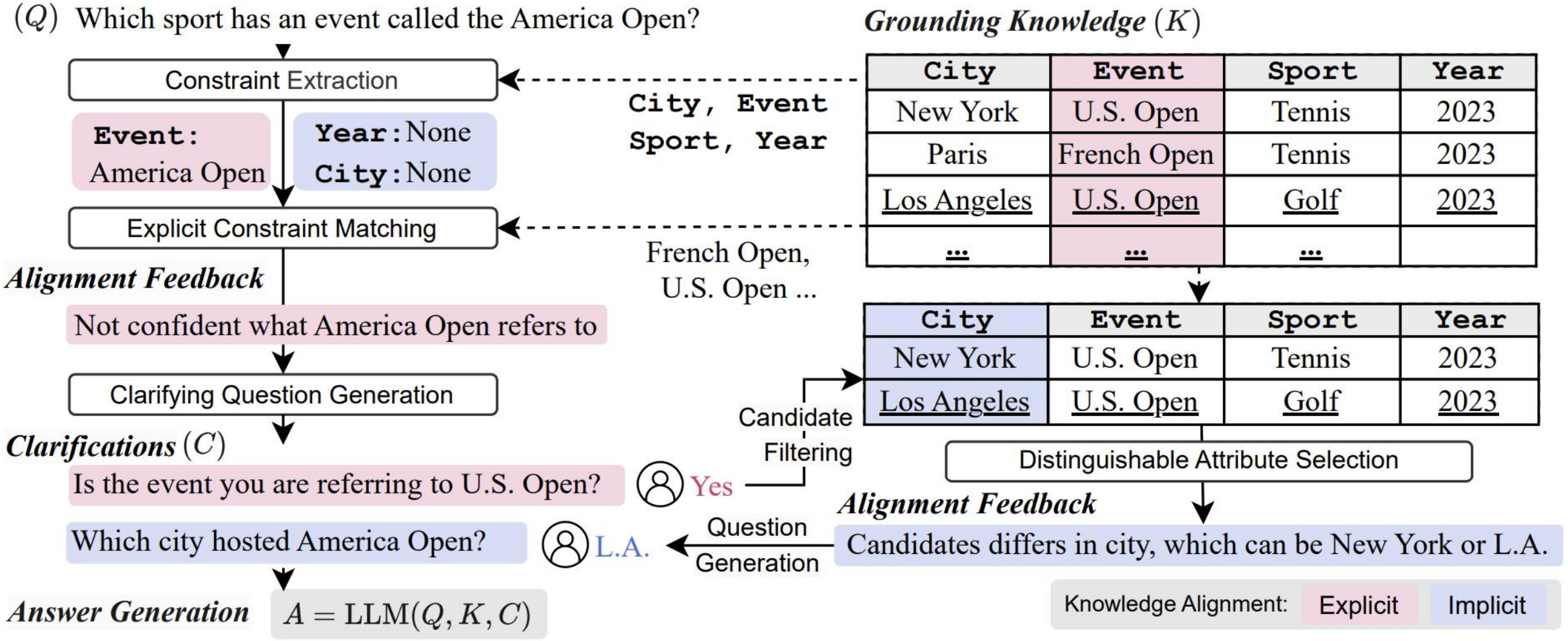


Clarification → Significant improvements

Type	Clarifying Information
Semantic	The term 'A' in the question refers to 'B' in our database.
Contextual	The value for the missing contextual condition in the question is 'A'.
Structural	The value for the condition 'B' in the database is 'A'.
Logical	The complex condition 'A' in your question refers to the condition 'B' in the database.

MIXALIGN

Find knowledge misalignments and ask for clarifications from users



Experimental Settings

Task: RAG with controlled knowledge Shuffle(Gold + N Irrelevant)

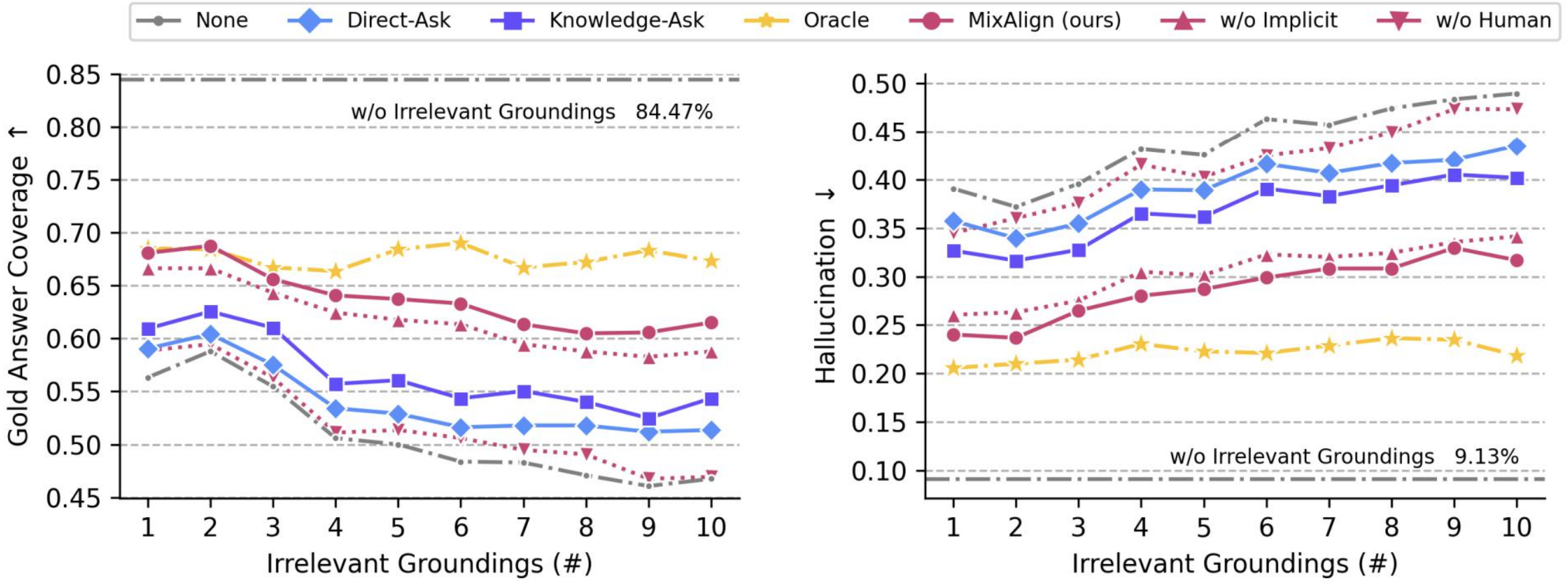
Dataset: FuzzyQA constructed by simplifying questions in the OTT-QA dataset

Baselines:

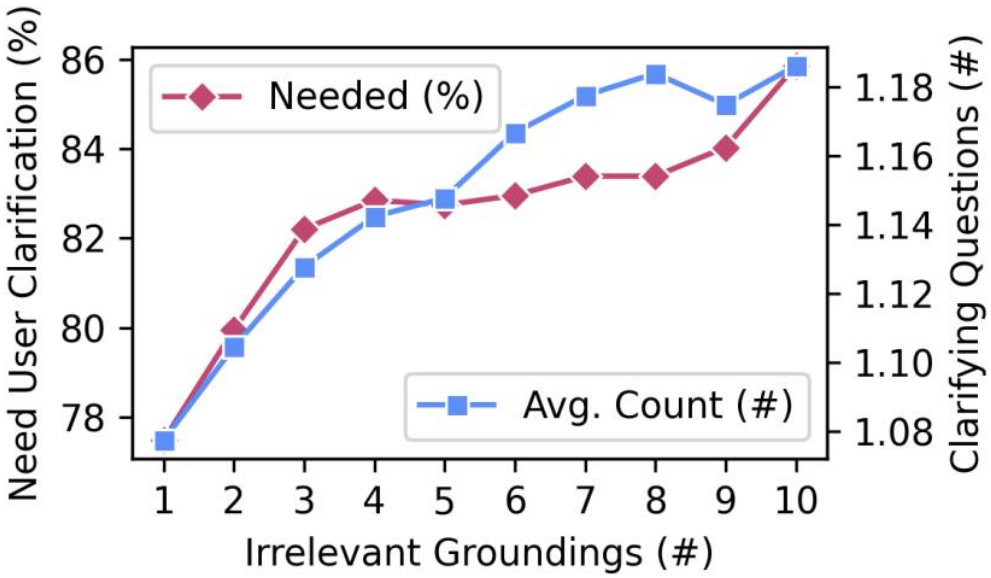
- ***None.*** No clarification included
- ***Oracle.*** Reverse-generate the oracle clarifications from the ground-truth answer and knowledge
- ***Direct-Ask.*** Clarifying questions are asked based solely on the original question."In order to answer this question, I have to ask the following clarifying question:"
- ***Knowledge-Ask.*** Incorporate candidate knowledge to generate clarifying questions

Main Results

Stable Improvements



User question	What are the campaign notes of the 2012 Democratic candidate born on May 23, 1958?	Where did the Magic: The Gathering player known as "The Great One" win?
Direct-Ask	What specific information are you looking for regarding the 2012 Democratic candidate?	What format of Magic: The Gathering was "The Great One" playing?
Knowledge-Ask	Which candidate was born on May 23, 1958?	Which season did **Jon Finkel** win?
MIXALIGN (ours)	Is the candidate you are referring to Keith Judd?	Which season did "The Great One" win?



Case Study & Cost Analysis

Figure 6: Efficiency analysis results of MIXALIGN. We target: 1. Needed (%): How often do we need to request user clarification after the initial LLM-based constraint matching? 2. Avg.Count (#): When needed, how many clarifying questions, on average, are posed?

Conclusion & Future work

- Introducing the Knowledge Alignment problem
- Proposed MixAlign
- Detailed Experiments Demonstrating Effectiveness
- address other knowledge formats and modalities
- improve efficiency of the pipeline

Thanks

zs412082986@stu.xjtu.edu.cn

Paper Link

MOE KLINNS Lab, Xi'an Jiaotong University

University of California, Santa Barbara