

KnowRAG: A Benchmark for Knowledge Graph-based Retrieval-Augmented Generation



Asking an LLM

Question

Which famous singer played in the Barbie movie from 2023?

LLM



Answer

As of the 2023 "Barbie" movie directed by Greta Gerwig, there aren't any prominent singers who played major acting roles in the film. While the movie features a diverse and talented ensemble, it primarily consists of actors rather than singers. However, the soundtrack and musical elements of the film may include contributions from various musicians and artists, enhancing the overall cinematic experience.



Asking an LLM

Question

Which famous singer played in the Barbie movie from 2023?

LLM



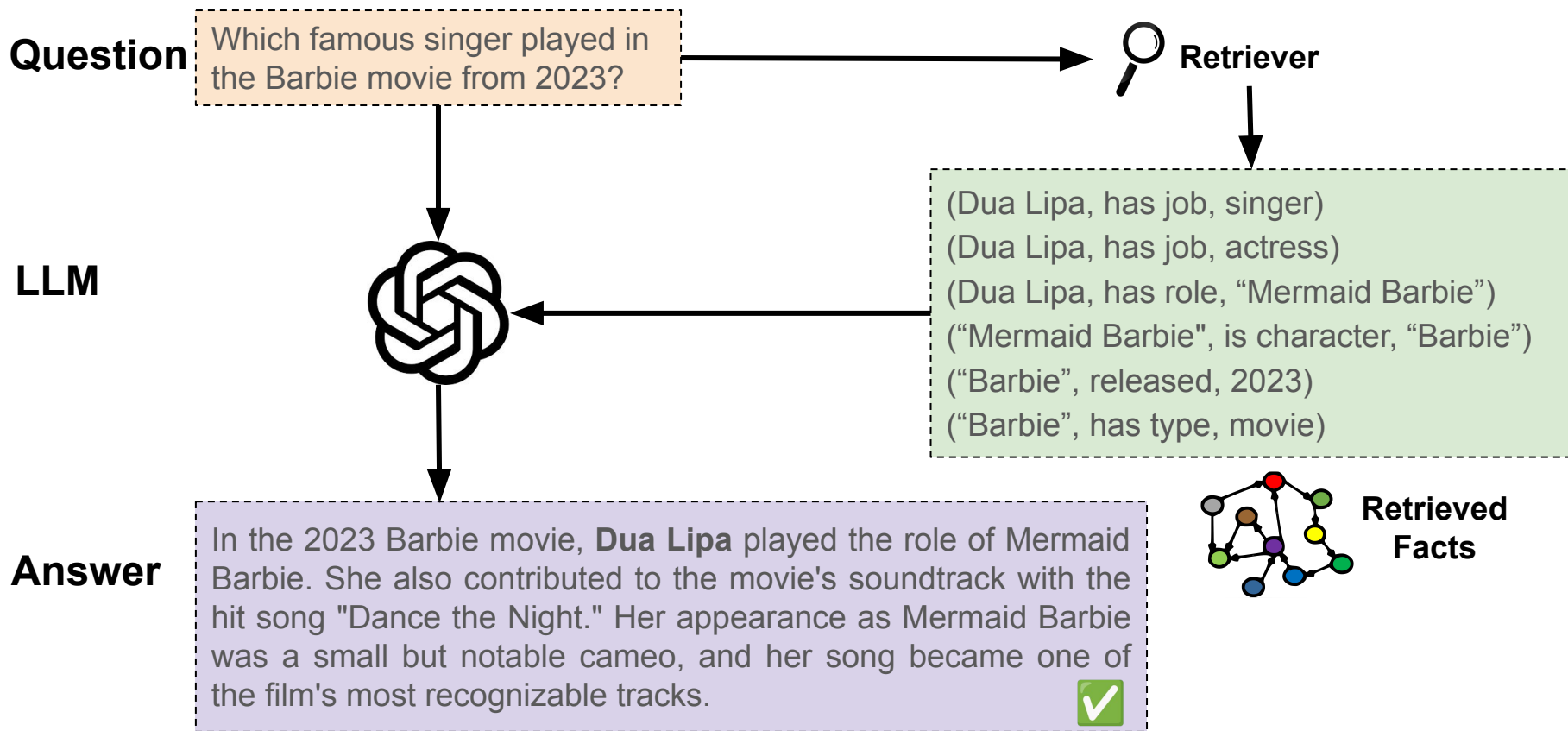
- 🙄 Hallucination
- 🙄 Outdated knowledge
- 🙄 No interpretability

Answer

As of the 2023 "Barbie" movie directed by Greta Gerwig, there aren't any prominent singers who played major acting roles in the film. While the movie features a diverse and talented ensemble, it primarily consists of actors rather than singers. However, the soundtrack and musical elements of the film may include contributions from various musicians and artists, enhancing the overall cinematic experience.



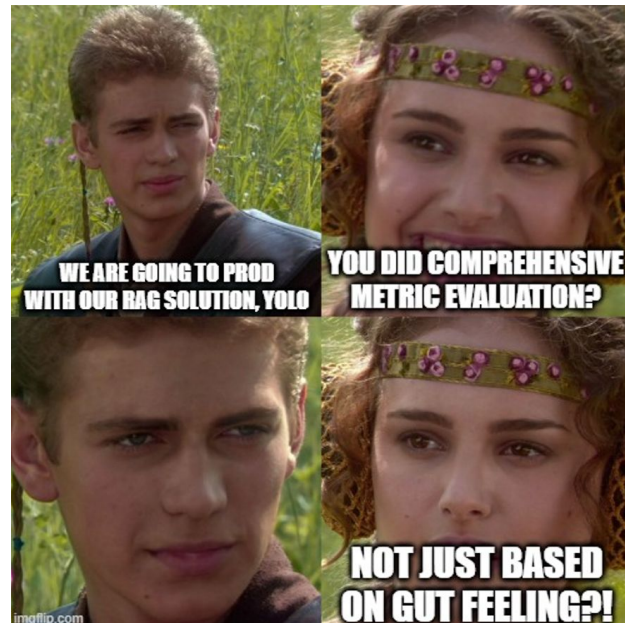
Grounding Answers in External Knowledge (RAG)



How Good is your RAG?

Model evaluation

- Retrieval performance (e.g., relevant, accurate)
- Answer generation (e.g., relevant, faithful, correct, hallucinated)
- Others (e.g., latency, diversity, noise robustness, no response)



How Good is your RAG?

Model evaluation

- Retrieval performance (e.g., relevant, accurate)
- Answer generation (e.g., relevant, faithful, correct, hallucinated)
- Others (e.g., latency, diversity, noise robustness, no response)

Different benchmarks look at different aspects

- 😊 questions of different types and complexities
- 😊 different domains
- 😊 popular & less popular entities

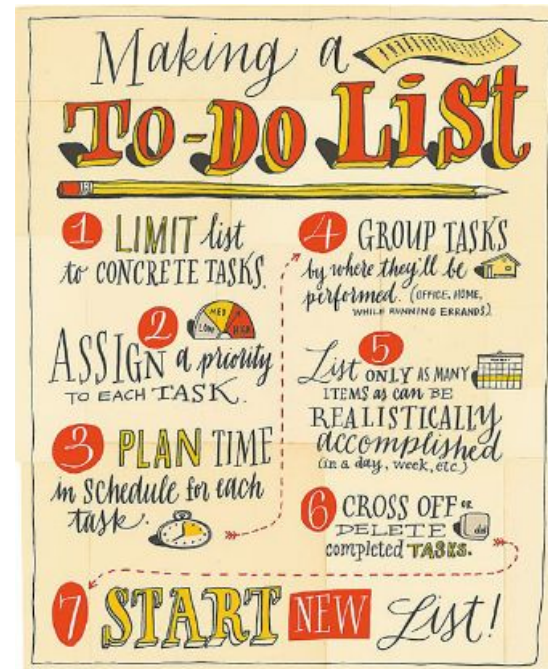
- 😞 monolingual (in English)
- 😞 mostly text (i.e., no images)
- 😞 different evaluation metrics for each dataset



Your Tasks & Challenges

Overall goal: develop a benchmark for KG-based RAG

- Find & combine data
 - ◆ i.e., general domain KGs, domain-specific datasets (e.g., IMDB for movies)
 - ◆ text + images + videos
- Generate questions & answers
- Make it multilingual
 - ◆ i.e., use machine translation & correction
- Add sound
 - ◆ i.e., “read” the questions with text-to-speech models
- Evaluate quality & benchmark models
 - ◆ i.e., how challenging is the dataset, how good are the models?



Requirements

- Duration: 6 months
- Number of participants: min 2, max 4
 - ◆ **Applying as a group** is possible!

- Prerequisites:
 - ◆ One of Text Analytics, Advanced Methods in Text Analytics, Information Retrieval and Web Search
 - ◆ One of Machine Learning, Knowledge Graphs
 - ◆ Knowledge in Python

- Applicable: Business Informatics and MMDS
- Instructors: Andreea Iana, Martin Böckling, Heiko Paulheim