

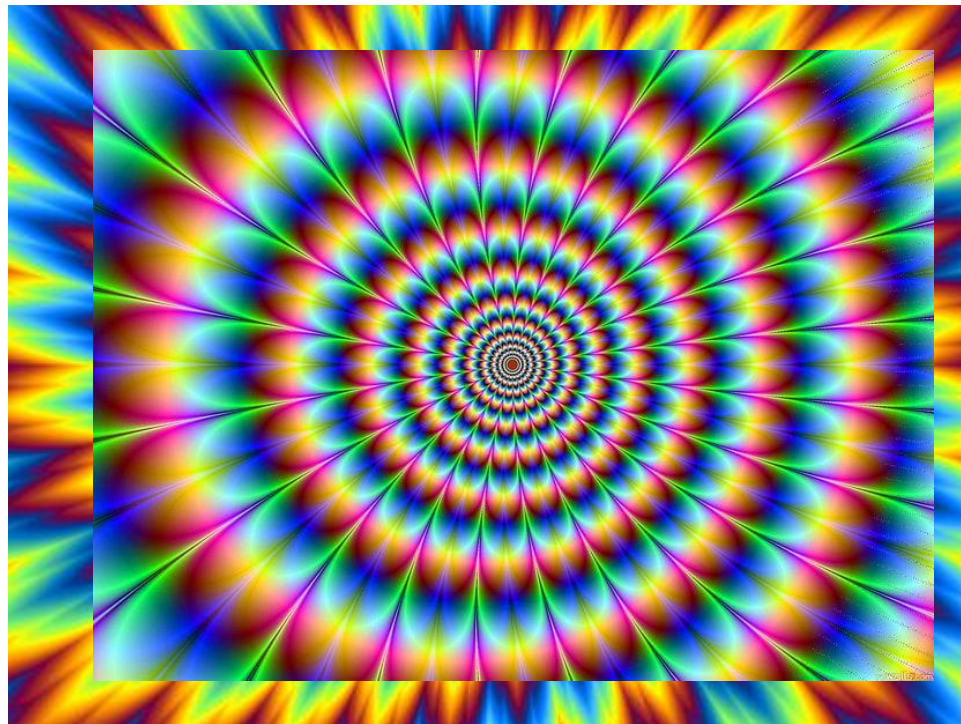
# Tonmoy et al (2024): A Comprehensive Survey of Hallucination Mitigation Techniques in Large Language Models



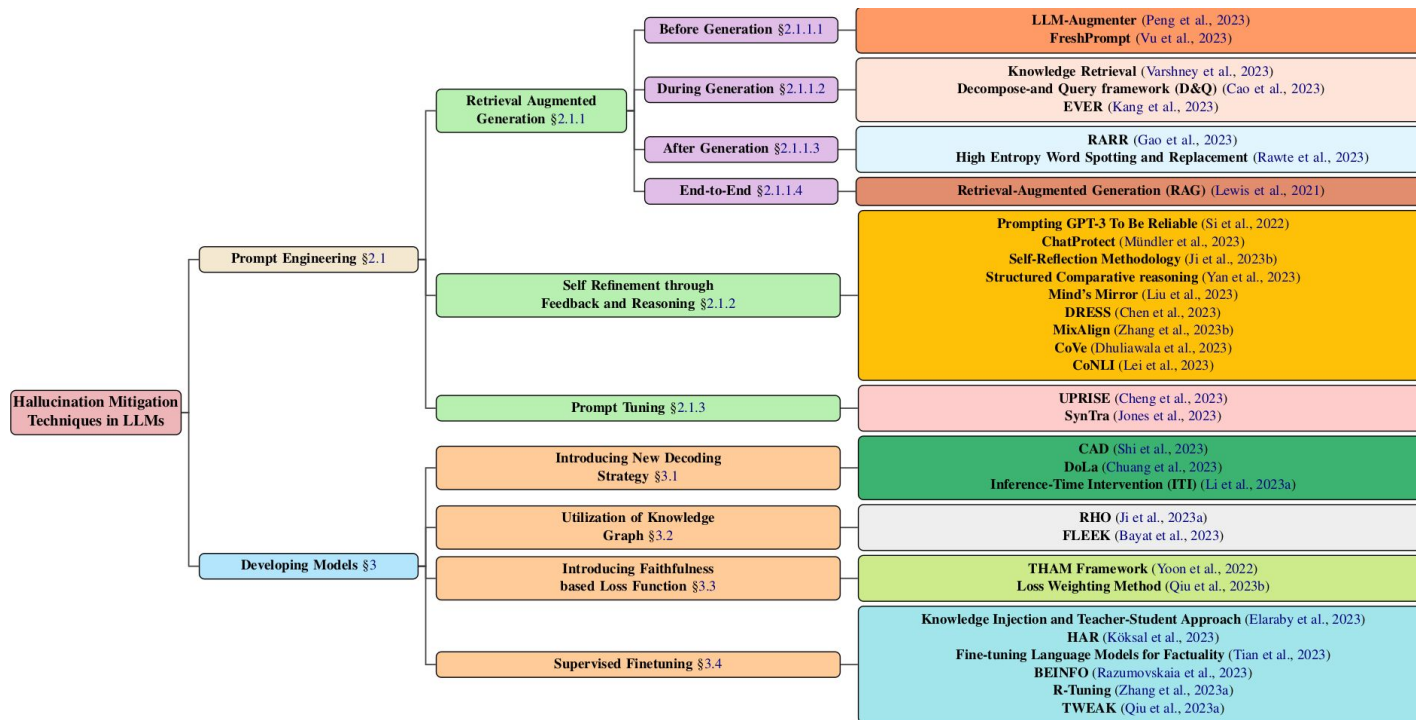
Presented by: Kristýna Klesnilová

# LLMs Hallucinations

- Generating ungrounded content that appears factual
- Problematic areas:
  - Legal
  - Medical
  - Financial
  - ...
- Causes
  - Training method
  - Absence of real-time internet updates



# Hallucination Mitigation Techniques

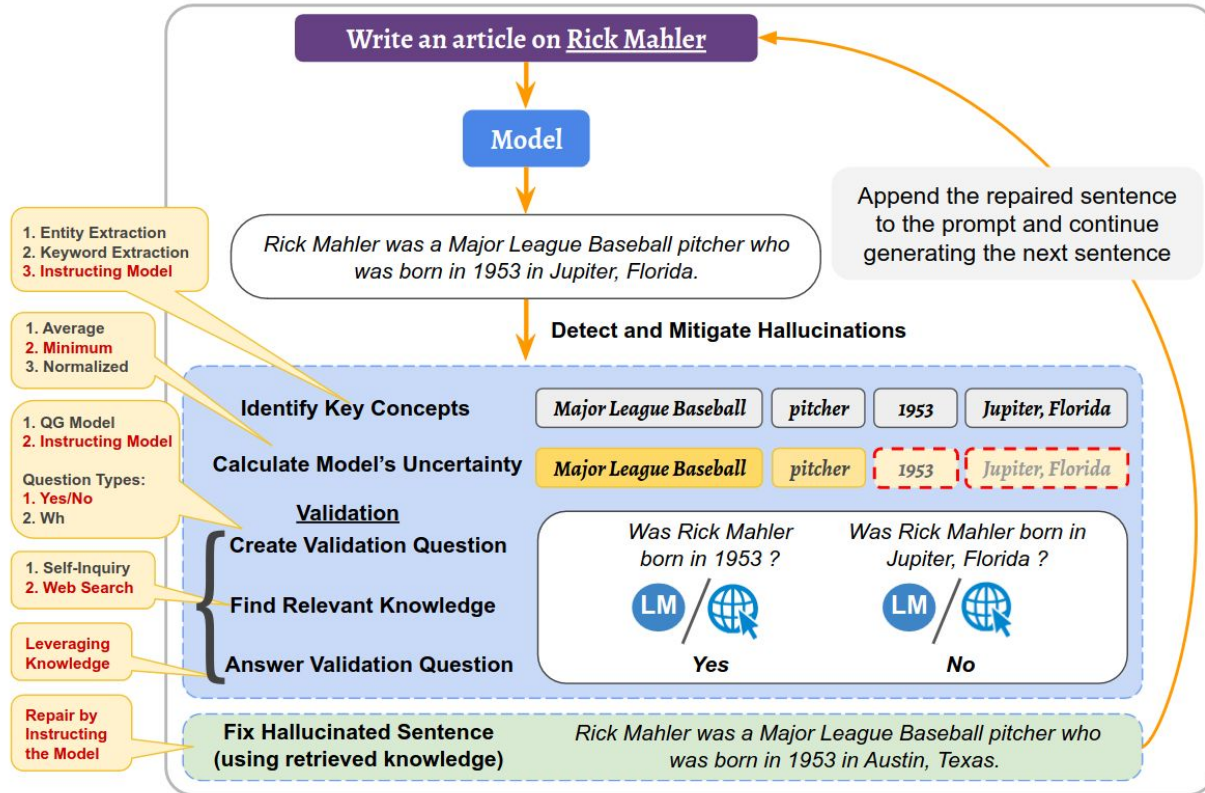


# Prompt Engineering

# Prompt Engineering: Retrieval Augmented Generation

- Before Generation
  - Retrieving external knowledge and adding it to LLM prompt
  
- During Generation
  - Retrieving external knowledge at sentence-by-sentence level

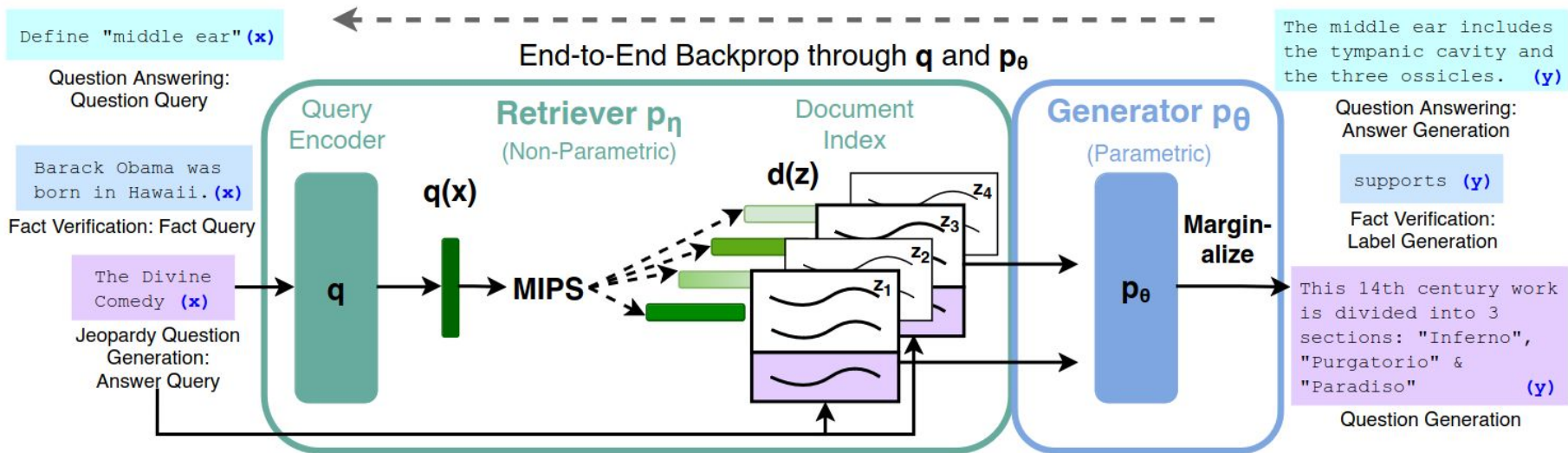
# Knowledge Retrieval



# Prompt Engineering: Retrieval Augmented Generation

- After Generation
  - Post-editing LLM output with retrieved external knowledge
  - Replacing high-entropy words or phrases (open-source LLMs)
- End-to-End
  - RAG
    - Generator (seq2seq transformer) and document retriever trained jointly end-to-end

# End-to-End RAG

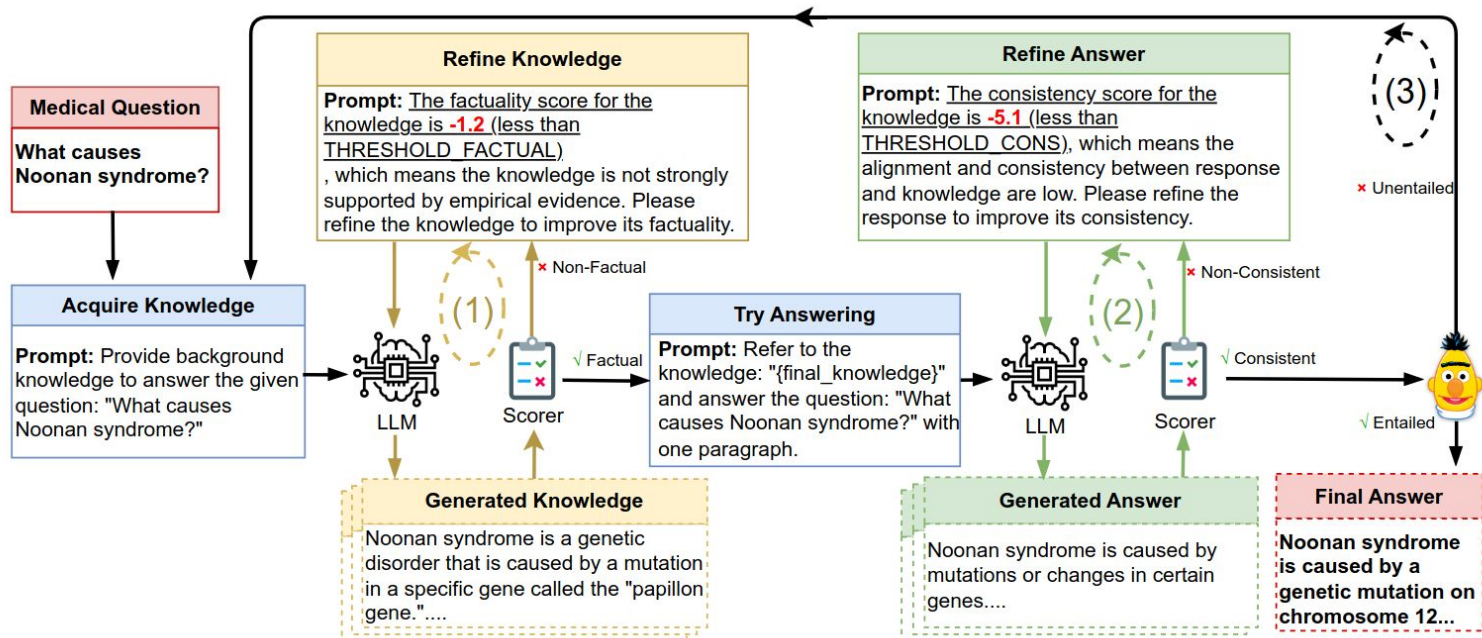




# Prompt Engineering: Self-refinement through feedback + reasoning

- Prompting strategies
  - Improving reliability
  - Detecting and mitigating self-contradictions through reasoning
- Improving through iterative feedback
  - Generate-score-refine loop
- Chain-of-Verification (CoVe)
  - Automatic fact-checking through verification questions
- Chain of Natural Language inference (CoNLI)

# Self-Reflection



# Chain-of-Verification (CoVe)

## Query

Name some politicians who were born in NY, New York.

## 1. Baseline Response

Here are some politicians who were born in NY, New York:

1. Hillary Clinton - former secretary of state and former Democratic presidential nominee
  2. Donald Trump - former president of the United States
  3. Michael Bloomberg - former Mayor of New York City and former Democratic presidential candidate
- ... <list continues..>

## 2. Plan Verifications

- Where was Hillary Clinton born?
  - Where was Donald Trump born?
  - Where was Michael Bloomberg born?
- ... <questions continue..>

## 3. Execute Verifications

Hillary Clinton was born in **Chicago, Illinois**, United States on October 26, 1947.

Donald Trump was born on June 14, 1946, in **Queens, New York City, New York**, United States

Michael Bloomberg was born on February 14, 1942, in **Boston, Massachusetts**, United States.

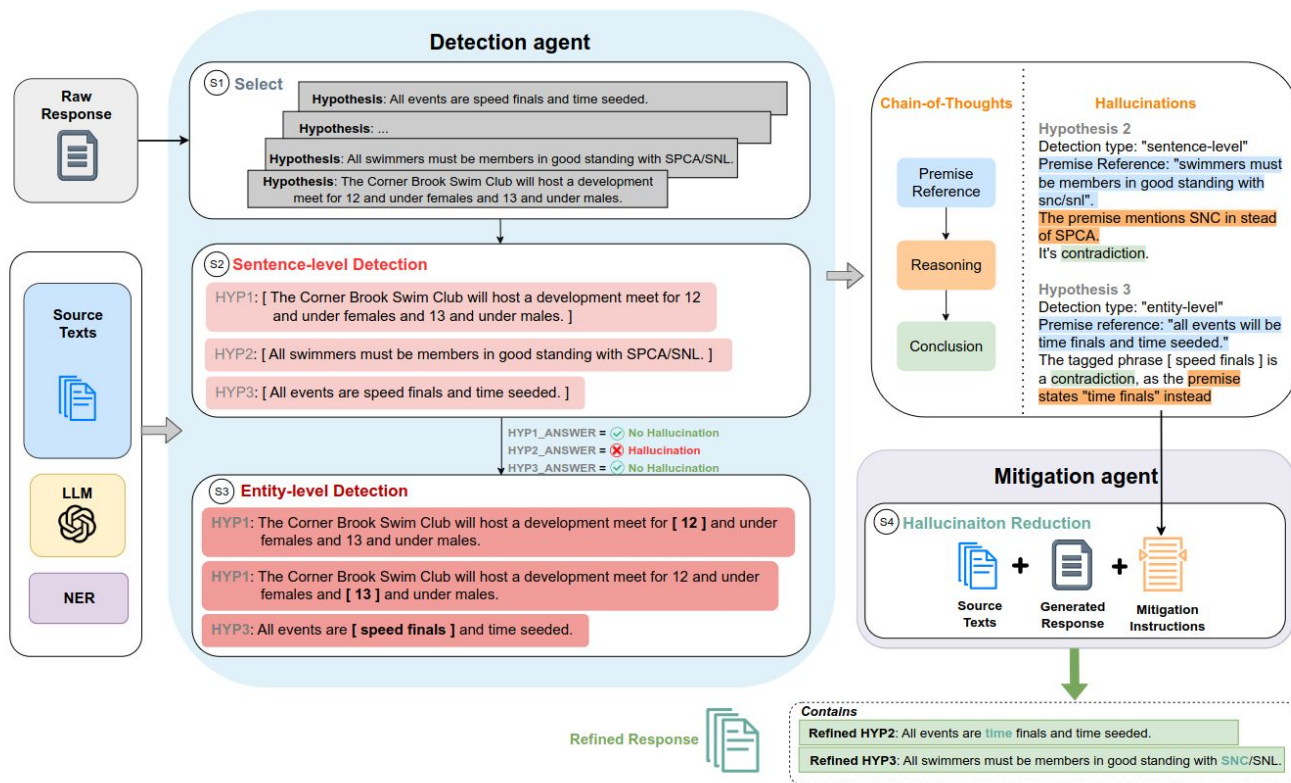
...

## 4. Final Verified Response

Here are some politicians who were born in NY, New York:

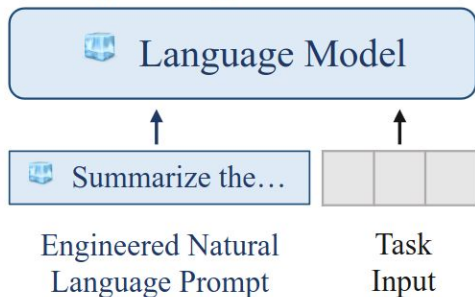
1. Donald Trump - former president of the United States
  2. Alexandria Ocasio-Cortez - Democratic member of the U.S. House of Representatives
- ... <list continues..>

# Chain of Natural Language inference (CoNLI)

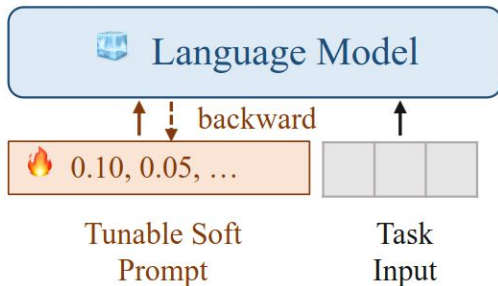


# Prompt Engineering: Prompt Tuning

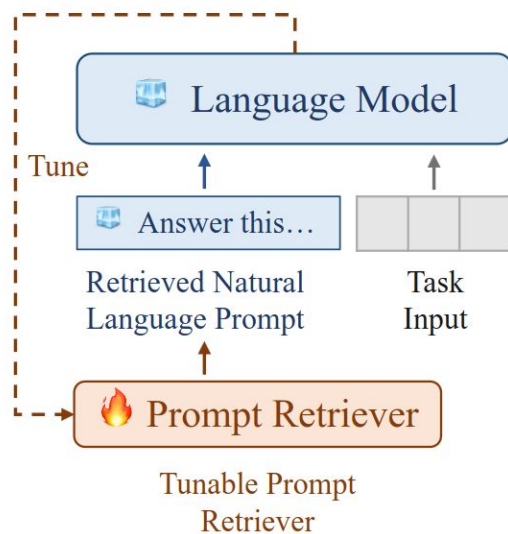
## Prompt Design



## Prompt Tuning

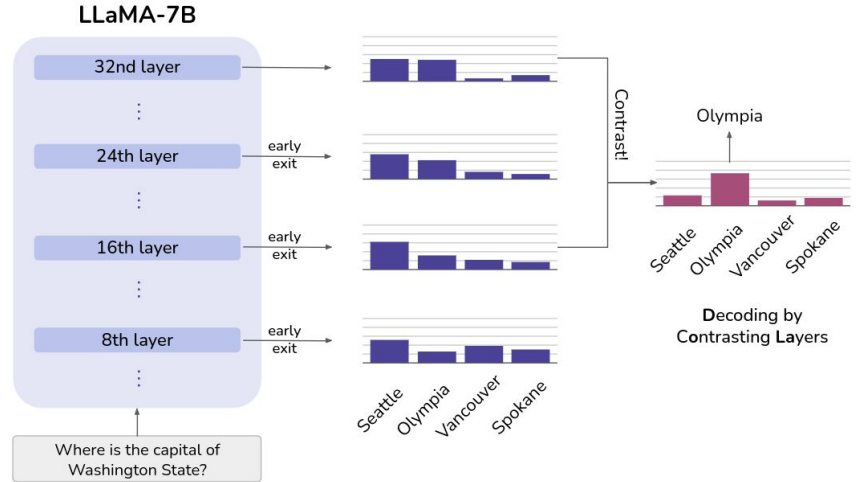
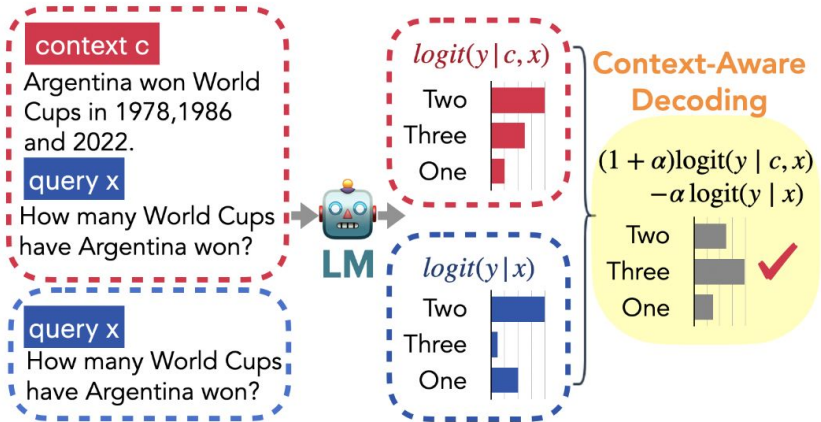


## Prompt Retrieval

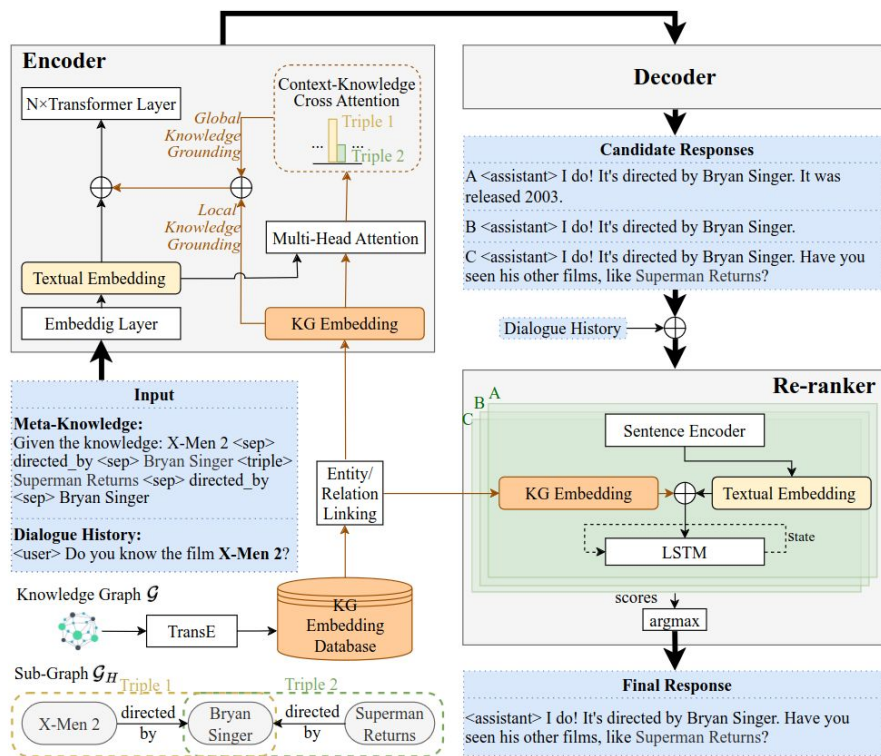


# Developing Models

# Developing Models: New Decoding Strategy



# Developing Models: Utilization of Knowledge Graph

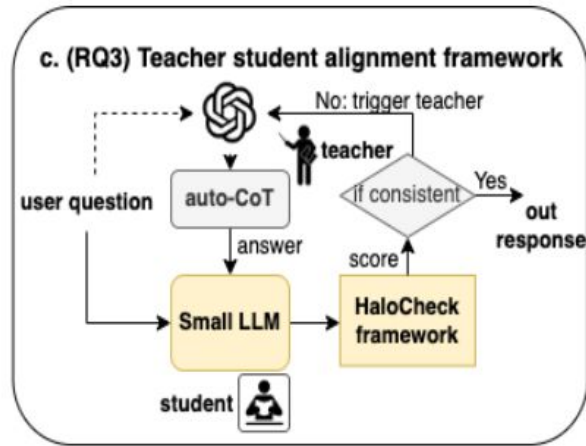
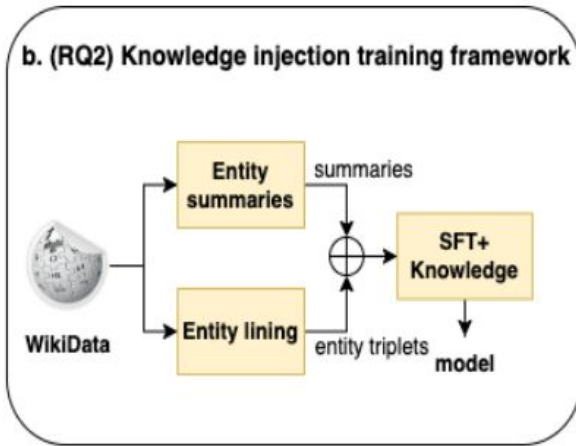
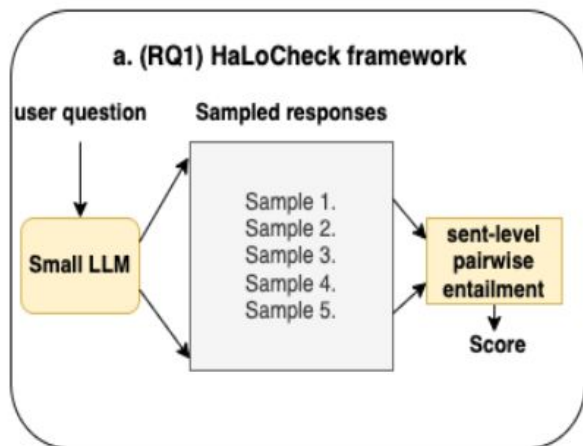




# Developing Models: Introducing faithfulness based loss function

- Designing new metrics
  - How closely model's outputs match input data or ground truth labels

# Developing Models: Supervised fine-tuning (SFT)



# Developing Models: Supervised fine-tuning (SFT)

*A question from TriviaQA*

**Question:** Who is the author of the fiction books Lace (published in 1982), Lace 2 (1985), Savages (1987), Crimson (1992), Tiger Eyes (1994), Revenge of Mimi Quinn (1998) and The Amazing Umbrella Shop (1990)?



Hallucination Augmented Recitation (HAR)



**Document:** Judy Blume has written many novels, including Forever, Tiger Eyes, Are You There God? It's Me, Margaret, Freckle Juice, and Blubber, which are popular among young women. Her books are so popular, that they have been translated into 31 different languages. She is the author of the fiction books Lace (published in 1982), Lace 2 (1985), Savages (1987), Crimson (1992), Tiger Eyes (1994), Revenge of Mimi Quinn (1998) and The Amazing Umbrella Shop (1990). She has also written many other books for children and adults. She is the recipient of the Library of Congress Living Legends award.

**Answer:** Judy Blume  
(Shirley Conran)

# Developing Models: Supervised fine-tuning (SFT)

