

Open-Domain QA with LLM and Knowledge Retrieval

2023/12/15

Li Peng-Hsuan 李朋軒

Agenda

- ODQA with LLM and Knowledge Retrieval
- Retrieval and Re-Ranking
- Retrieval Management
- pubmedKB Minimum ODQA

Agenda

- ODQA with LLM and Knowledge Retrieval
- Retrieval and Re-Ranking
- Retrieval Management
- pubmedKB Minimum ODQA

ODQA, LLM, Knowledge

ODQA

→ Open-Domain Question Answering

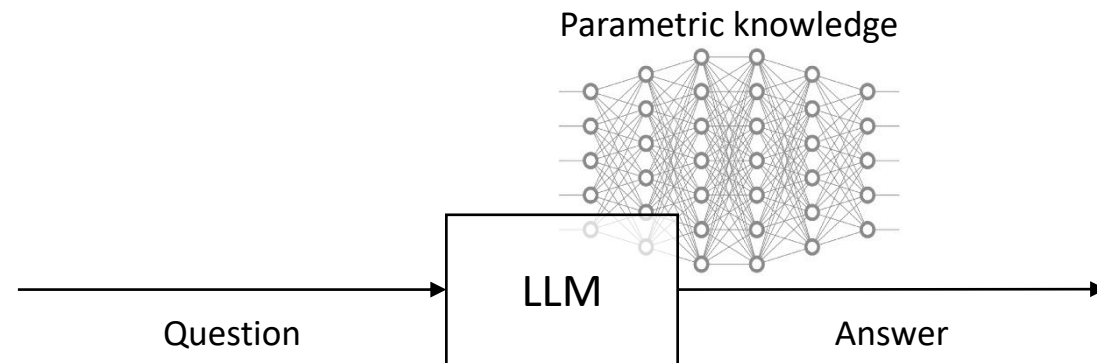
LLM

→ Large Language Models

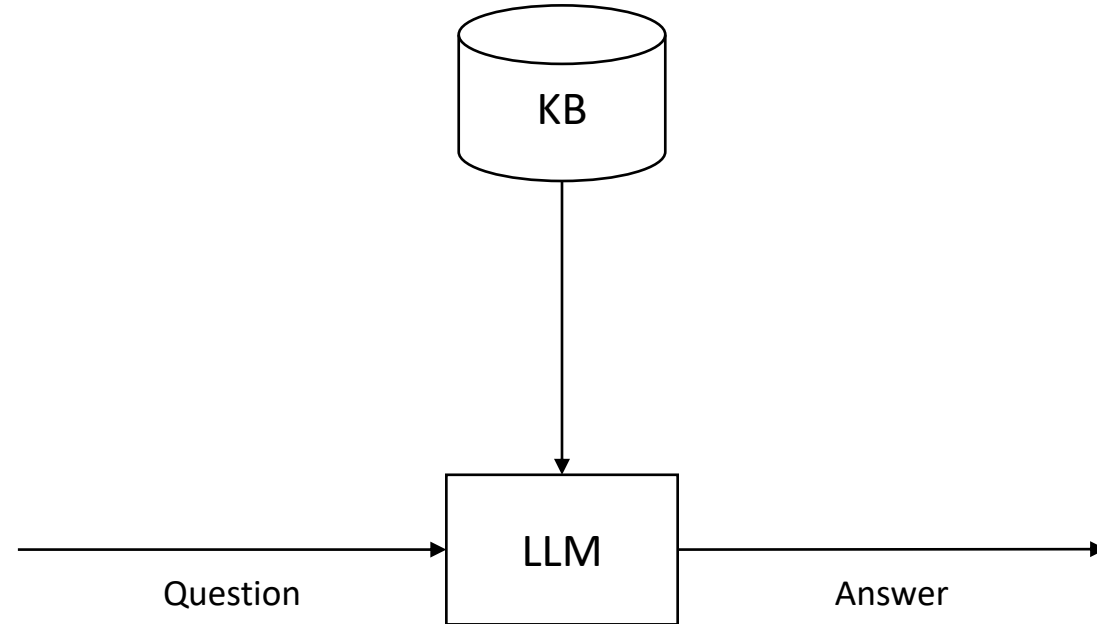
Knowledge

→ Model weights, Wikipedia, knowledge graphs, web, ...

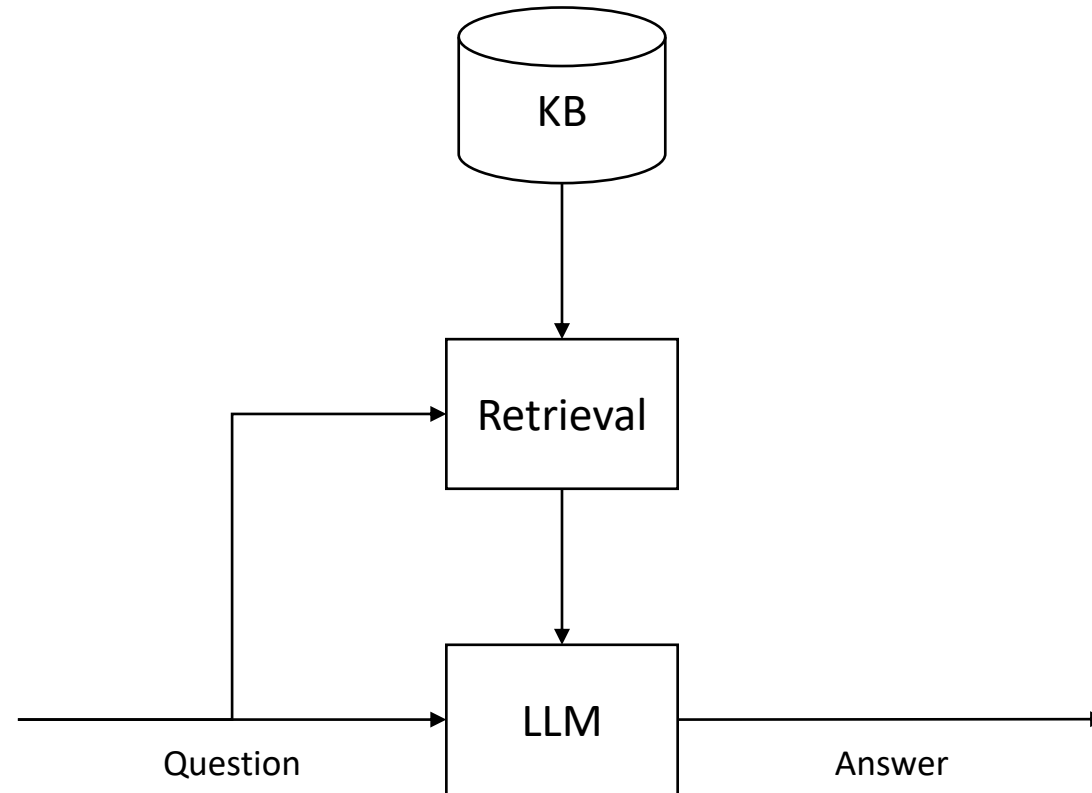
ODQA with LLM



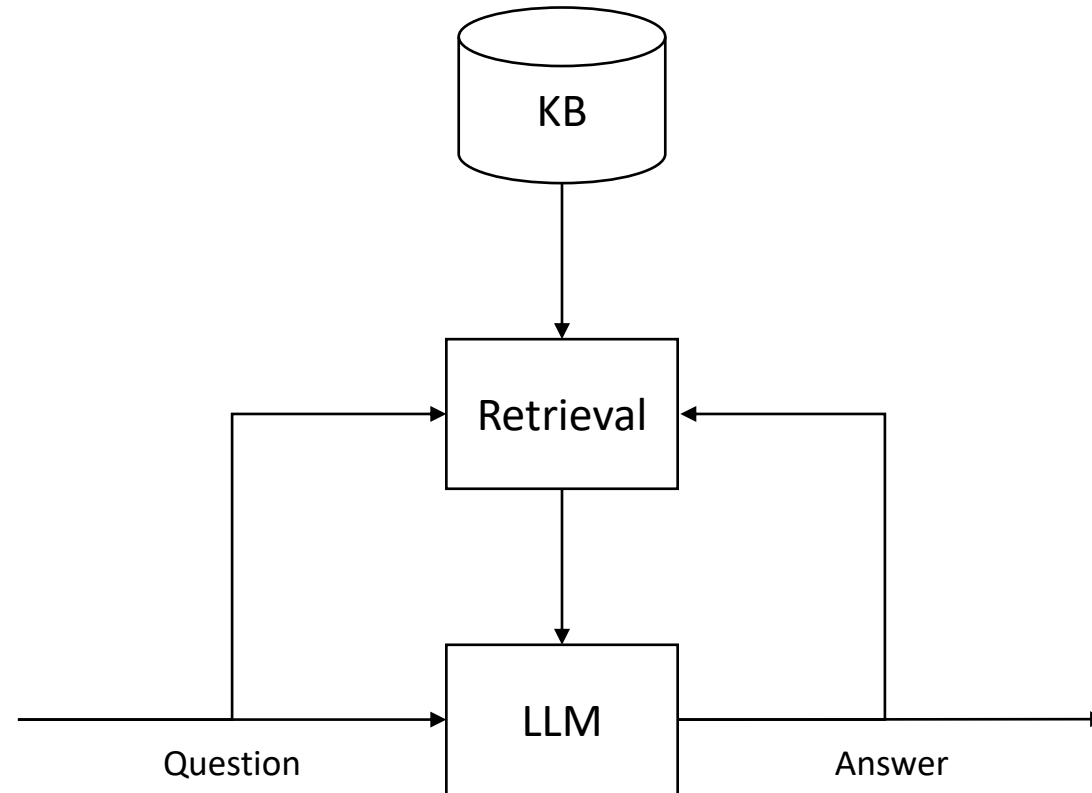
ODQA with LLM and Knowledge



ODQA with LLM and Knowledge Retrieval



ODQA with LLM and Knowledge Retrieval (Iterated)



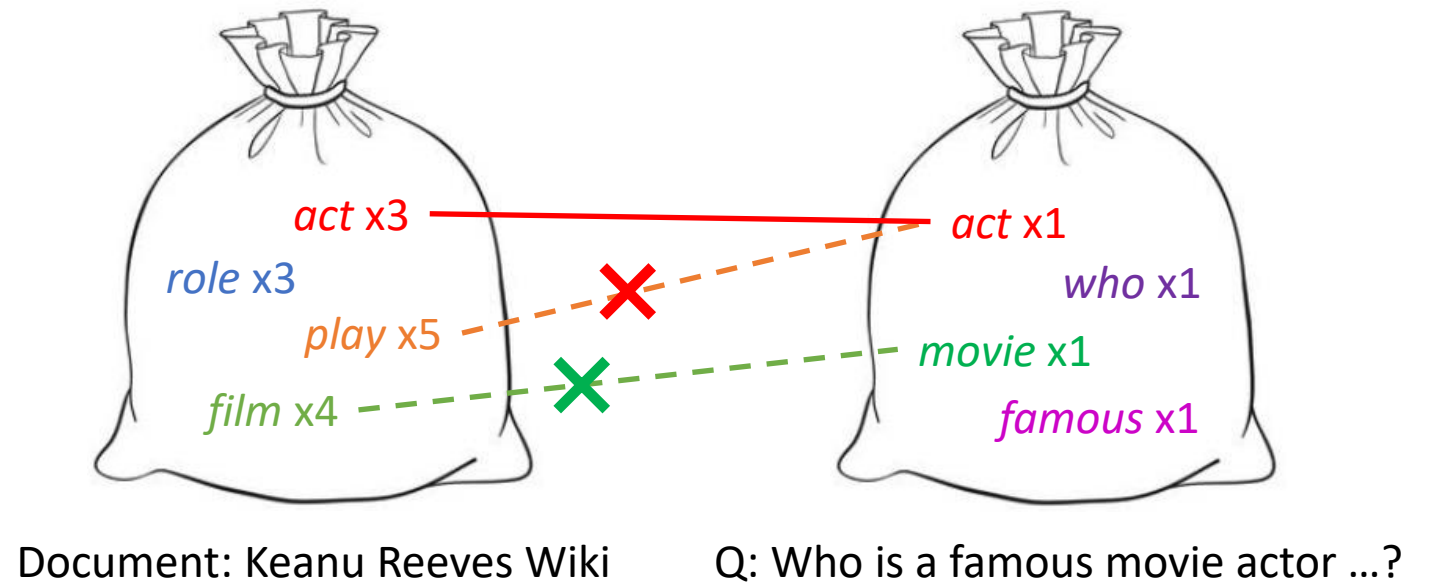
Agenda

- ODQA with LLM and Knowledge Retrieval
- **Retrieval and Re-Ranking**
- Retrieval Management
- pubmedKB Minimum ODQA

Retrieval Methods

Sparse retrieval

- Sparse
- Bag-of-words
- Exact match-based



Retrieval Methods

Sparse retrieval

→ TF-IDF

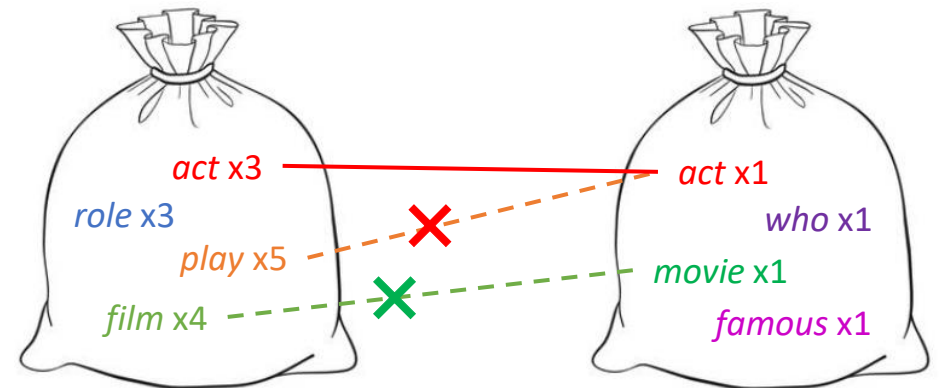
→ BM25

→ ...

😄 Term appears many times in *Document*

😓 Document is long

😓 Term appears in many documents



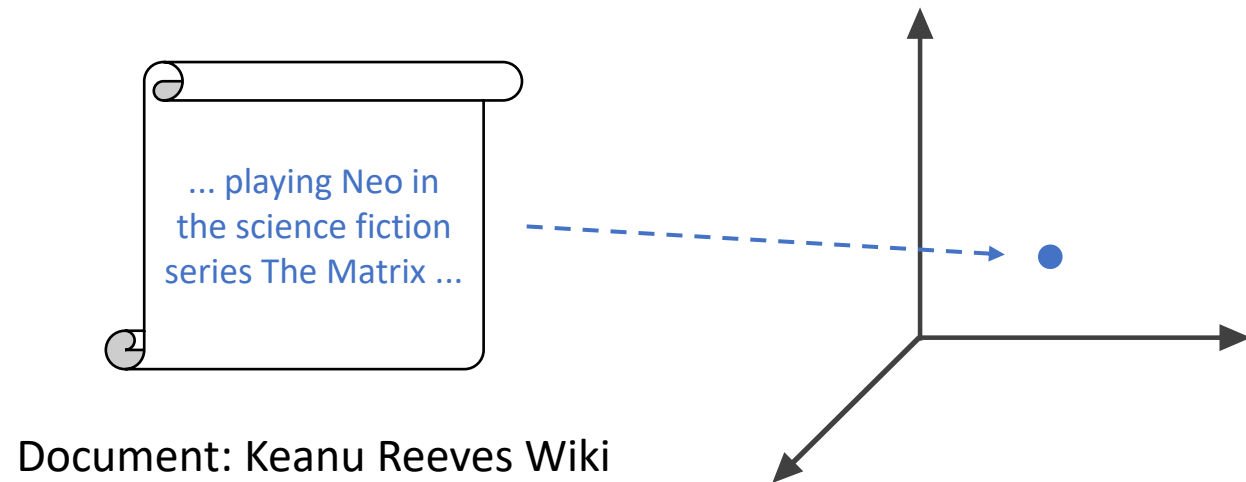
Document: Keanu Reeves Wiki

Q: Who is a famous movie actor ...?

Retrieval Methods

Dense retrieval

- Low-dimensional
- Contextual
- Semantic match-based



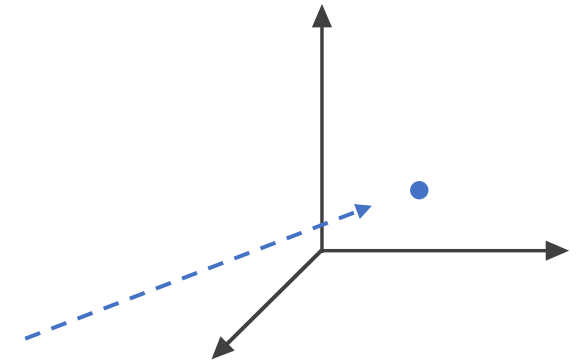
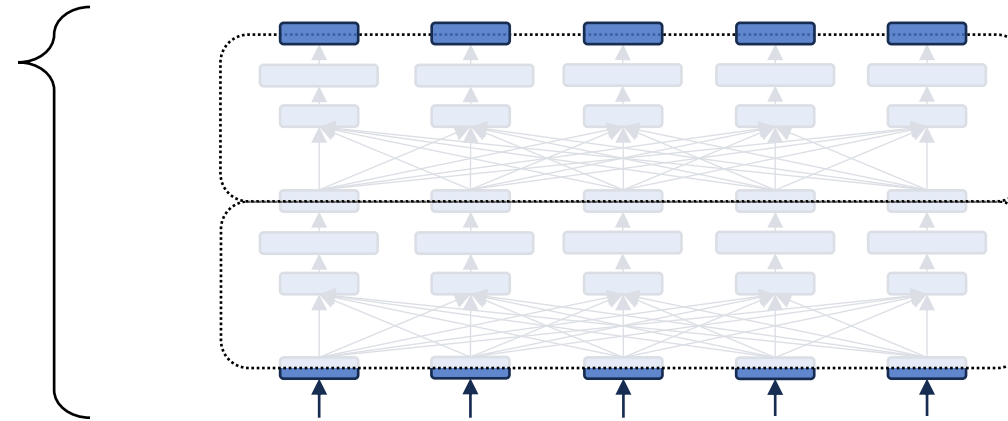
Retrieval Methods

Dense retrieval

→ MiniLMv2

... playing Neo in
the science fiction
series The Matrix ...

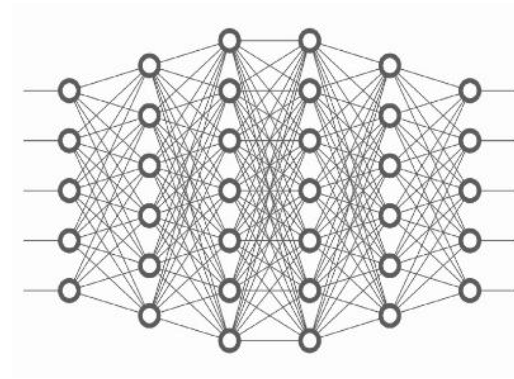
Document: Keanu Reeves Wiki



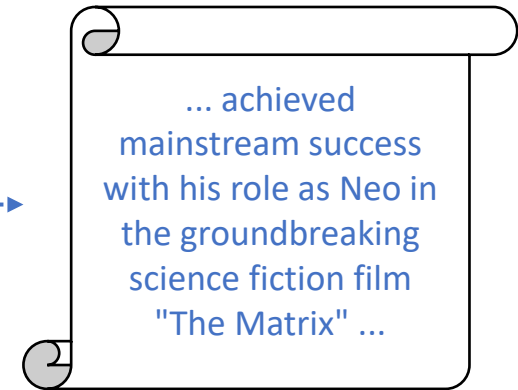
Retrieval Methods

Generation as retrieval

→ Distillation



Parametric knowledge



Document: Keanu Reeves

Retrieval Methods

Model	Query	Acquired Knowledge	Answer
LLaMA	Who wrote the song the glory of love?	"The Glory of Love" is a song written by Billy Hill, recorded by American singer Benny Goodman in 1936. The song has been covered by many other artists, including a version by American singer Peggy Lee with Bing Crosby and the Johnny Mann Singers in 1959.	Billy Hill ✓
FLAN-T5	Who wrote the song the glory of love?	The Glory of Love is a song written by the American composer and lyricist Irving Berlin .	Irving Berlin ✗
DPR	Rain sleet or snow that contains a high concentration of acids is called?	... Ice pellets are a form of precipitation consisting of small, translucent balls of ice. This form of precipitation is also referred to as ""sleet"" by the United States National Weather Service. (In British English ""sleet"" refers to a mixture of rain and snow)...	icy rain ✗

(top) [non-critical error](#) in generation; (middle) [critical error](#) in generation; (bottom) irrelevant retrieval

Retrieval Methods

Retrieve and re-rank

→ Speedy retrieval

→ Accurate re-ranking

E.g., UPR: $\text{score}(\text{Document}) = P_{\text{LM}}(\text{Query}|\text{Document})$

Agenda

- ODQA with LLM and Knowledge Retrieval
- Retrieval and Re-Ranking
- **Retrieval Management**
- pubmedKB Minimum ODQA

Retrieval Management

- When to retrieve
- How to retrieve
- What to do with retrieved knowledge

Retrieval Management

- Question → retrieval → answer

LLM-Augmenter

When → Rules, RL policy

How → BM25, DPR, UPR

What → Validate & revise answer



Which 2013 Los Angeles Galaxy player transferred in from the team with 12 international titles ?

Consolidate evidence from external knowledge

WIKIPEDIA

Wiki Loves FOLKLORE

Photograph your local culture, help Wikipedia and win!

2013 Los Angeles Galaxy season

Juninho (footballer, born January 1989)

Juninho grew up in the city of São Paulo and played for the São Paulo youth team, winning the U-17 Paulista Championship side in 2000. He made one appearance for the São Paulo senior side in 2007. He was sent out on loan to Major League Soccer team Los Angeles Galaxy in 2010, along with fellow Brazilians from the club Adriano, Cleiton, and Leonardo.^[1] He made his debut for the team on 27 March 2010, in Galaxy's opening game of the 2010 MLS season against New England Revolution,^[2] and scored his first goal for the Galaxy in a 2-0 win over AC St. Louis.

Transfers

No.	Pos.	Player	Transferred from	Fee/No
1	GK	Carlo Cudicini	Tottenham Hotspur	Free
13	MF	Juninho	São Paulo	Undisclosed

Honours

São Paulo FC is one of the most successful clubs in Brazil, having won a total of 30 domestic honours, in addition to their 12 international successes. It is the Brazilian club with the most international titles.

Major competitions

Worldwide		
Competitions	Titles	Seasons
Intercontinental Cup	2	1992, 1993
FIFA Club World Cup	1	2005
Continental		
Copa Libertadores	3	1992, 1993, 2005

Revise response via automatic feedback

Candidate response:

Jaime Penedo is transferred in from C.S.D. Municipal, a team with 12 international titles.

Feedback:

The player Jaime Penedo is transferred in from C.S.D. Municipal, but there is no information about the number of international titles of this team.

Revised candidate response:

Juninho is transferred in from São Paulo, a team with 12 international titles.

AI Agent (LLM-Augmenter + LLM)



Juninho is transferred in from São Paulo, a team with 12 international titles.

Retrieval Management

- Question → retrieval → answer
- Question → each [sub-question -> retrieval] → CoT → answer

Direct Prompting

GPT-3

Question: Who lived longer, Theodor Haecker or Harry Vaughan Watkins?

Answer: Harry Vaughan Watkins.

Question: Who was president of the U.S. when superconductivity was discovered?

Answer: Franklin D. Roosevelt



Chain of Thought

GPT-3

Question: Who lived longer, Theodor Haecker or Harry Vaughan Watkins?

Answer: Theodor Haecker was 65 years old when he died. Harry Vaughan Watkins was 69 years old when he died.

So the final answer (the name of the person) is: Harry Vaughan Watkins.

Question: Who was president of the U.S. when superconductivity was discovered?

Answer: Superconductivity was discovered in 1911 by Heike Kamerlingh Onnes. Woodrow Wilson was president of the United States from 1913 to 1921. So the final answer (the name of the president) is: Woodrow Wilson.



Self-Ask

GPT-3

Question: Who lived longer, Theodor Haecker or Harry Vaughan Watkins?

Are follow up questions needed here: Yes.

Follow up: How old was Theodor Haecker when he died?

Intermediate answer: Theodor Haecker was 65 years old when he died.

Follow up: How old was Harry Vaughan Watkins when he died?

Intermediate answer: Harry Vaughan Watkins was 69 years old when he died.

So the final answer is: Harry Vaughan Watkins

Question: Who was president of the U.S. when superconductivity was discovered?

Are follow up questions needed here: Yes.

Follow up: When was superconductivity discovered?

Intermediate answer: Superconductivity was discovered in 1911.

Follow up: Who was president of the U.S. in 1911?

Intermediate answer: William Howard Taft.

So the final answer is: William Howard Taft.



When: LLM / How: LLM / What to do: LLM

Our Approach: MeLLO

Multi-hop question	What is the capital city of the country of citizenship of Ivanka Trump's spouse?
Subquestion	Who is Ivanka Trump's spouse?
Tentative answer	Ivanka Trump's spouse is <u>Jared Kushner</u>
Retrieved fact	David Cameron is married to <u>Courtney Love</u> 😞 not contradict
Answer	Jared Kushner
Subquestion	What is the country of citizenship of Jared Kushner?
Tentative answer	The country of citizenship of Jared Kushner is <u>United States</u>
Retrieved fact	Jared Kushner is a citizen of <u>Canada</u> ⚠️ contradict!
Answer	Canada
.....
Final answer	Ottawa

Edited Fact Memory

David Cameron is married to Courtney Love

The capital of the US is Seattle

The CEO of Apple is Carlos Slim

Jared Kushner is a citizen of Canada

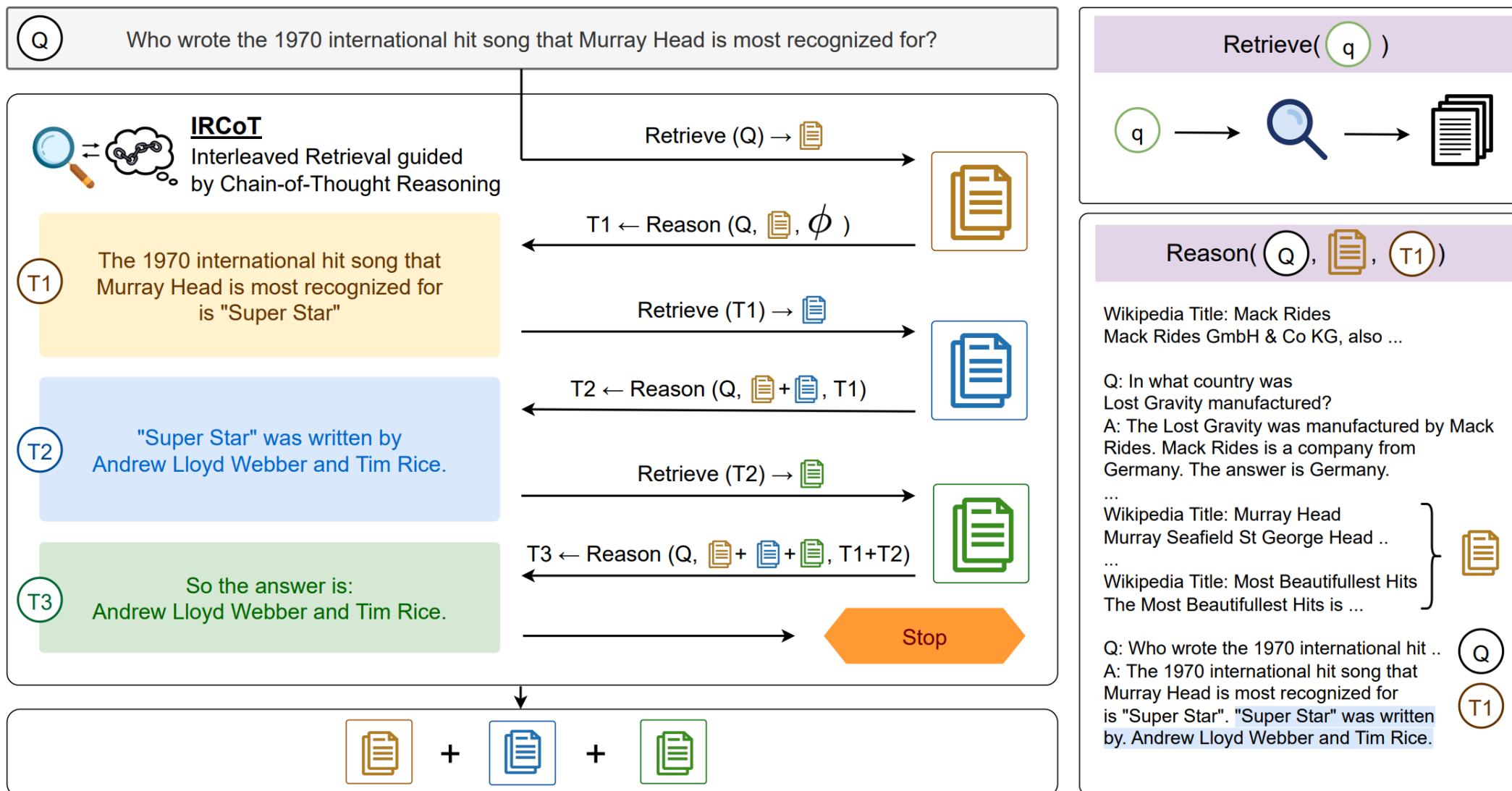
...

-→ Query the memory with a subquestion
- ← Retrieve an edited fact from the memory
- Tentative answers generated by the model
- Retrieved facts
- Edited facts stored in the memory

MeLLO When: each sub-question / How: BM25 / What to do: LLM

Retrieval Management

- Question → retrieval → answer
- Question → each [sub-question -> retrieval] → CoT → answer
- Question → each [CoT stage -> retrieval] → CoT → answer



IRCoT When: each CoT sentence / How: BM25 / What to do: LLM

Verify-and-Edit

When

- Answers are inconsistent
- Each CoT sentence

How

- Generate **query** by **CoT sentence** + **question**
- Wiki, Google

What to do

- Generate new **CoT sentence** by **query + knowledge**
- Generate new answer by question + new CoT

Question

Of all the teams John Nyskohus played for, which team was known as "the Black and Whites?"

Standard

Newcastle United. ❌

Chain-of-thought

First, John Nyskohus played for the Norwegian football team Odd Grenland. Second, Odd Grenland is known as "the Black and Whites." The answer is **Odd Grenland**. ❌

Self-Consistency:
less than majority agree

Verify

What team did John Nyskohus play for?
What team is known as "the Black and Whites?"

External Knowledge Retrieval

John Nyskohus ... is an Australian former soccer player who played club football for USC Lion ... and Adelaide City in the National Soccer League ...
Adelaide City Football Club is an Australian football (soccer) club based in Adelaide, South Australia. They are also known as "The Zebras" and "the Black and Whites."

Edit Rationales

First, John Nyskohus played for Adelaide City in the National Soccer League. Second, Adelaide City Football Club is known as "the Black and Whites".

New Prediction

The answer is **Adelaide City Football Club**. ✓

FLARE

When

→ Low token probability in sentence

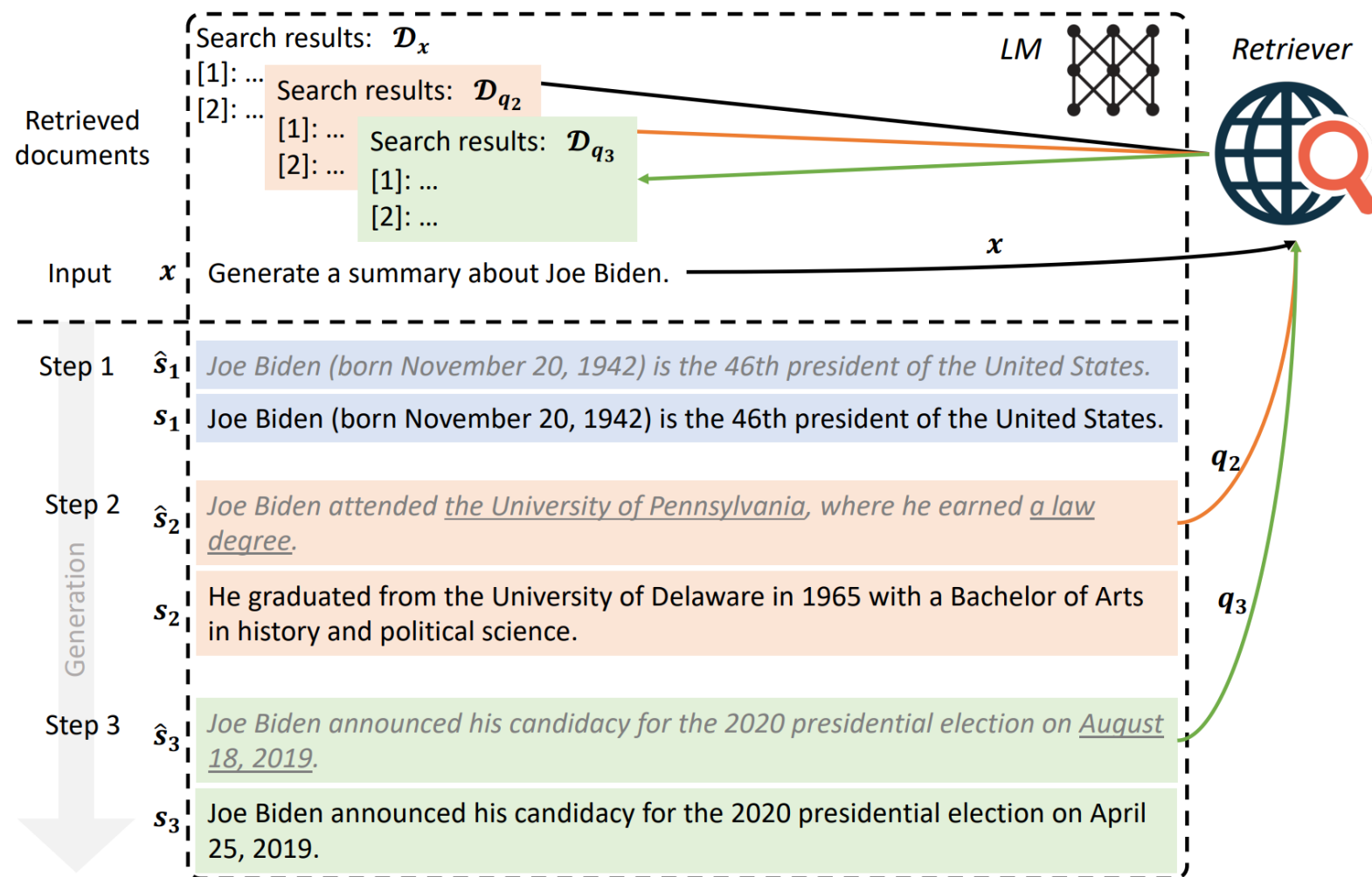
How

→ Generate query by asking a question

→ BM25, Bing

What to do

→ Regenerate that sentence



FLARE-instruct

When

→ LLM in-context learning

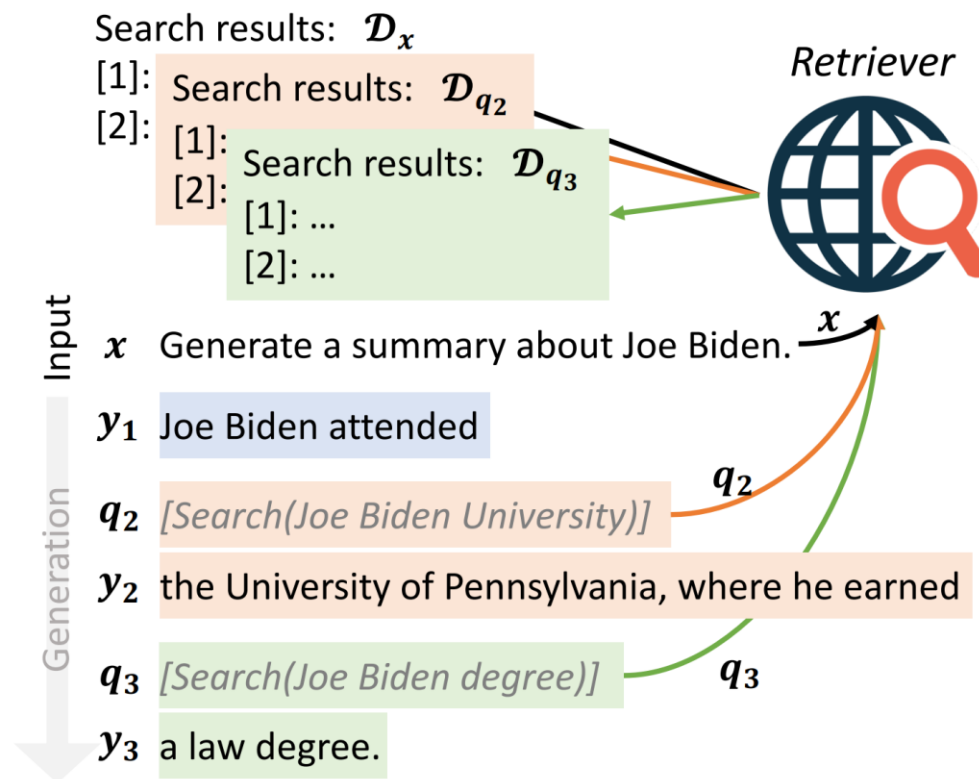
How

→ LLM in-context learning

→ BM25, Bing

What to do

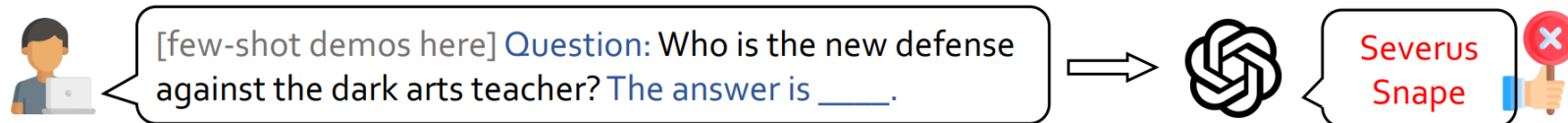
→ Resume generation



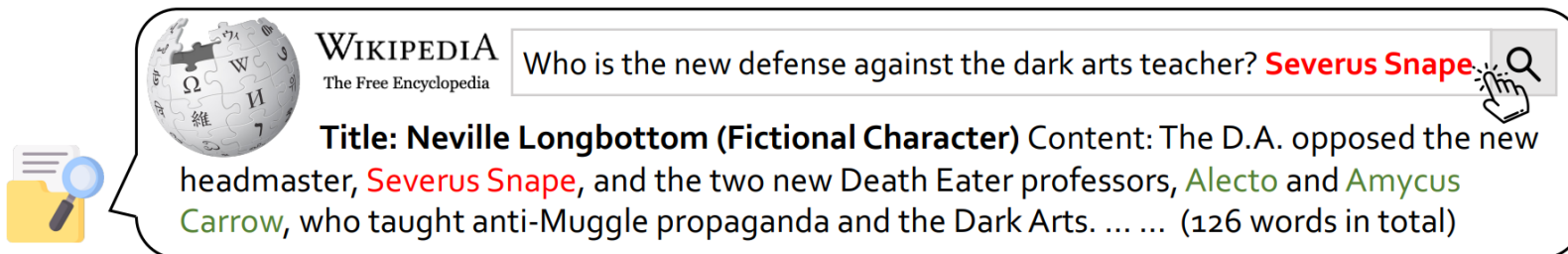
Retrieval Management

- Question → retrieval → answer
- Question → each [sub-question -> retrieval] → CoT → answer
- Question → each [CoT stage -> retrieval] → CoT → answer
- Question → answer → retrieval → answer

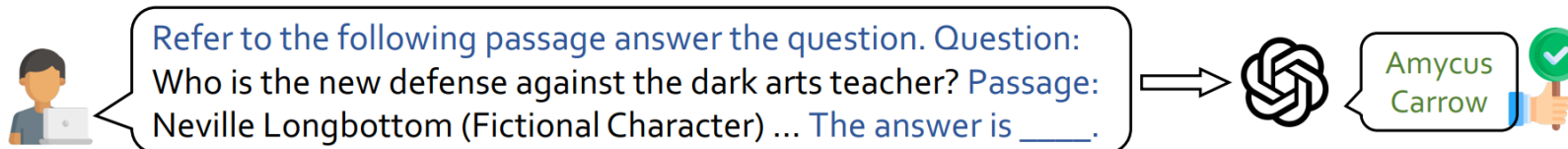
Step-1 (Generate an initial answer): $Q \xrightarrow{\text{GPT 3.5}} \hat{A}$



Step-2 (Retrieve documents): $Q, \hat{A} \xrightarrow{\text{BM25}} D_1, \dots D_n$



Step-3 (Refine the previous answer): $Q, \hat{A}, D_1, \dots D_n \xrightarrow{\text{GPT 3.5}} A$



REFEED When: after sampling answers / How: BM25 / What to do: re-answer

Question: q

What is the height of the player who won the 2015 AFL Rising Star award?

Retrieval: $q \rightarrow \mathcal{D}_q$

Title: 2015 AFL Rising Star Context: The NAB AFL Rising Star award is given annually to a stand out young player in the Australian Football League. **The award was won by Jesse Hogan of Melbourne**

Retrieval-Augmented Generation: $\mathcal{D}_q \parallel q \rightarrow y_1$

The 2015 AFL Rising Star award was won by Jesse Hogan of Melbourne. Jesse Hogan is a professional Australian rules footballer. He is **198 cm tall**. So the answer is 198 cm

Question: q

What is the height of the player who won the 2015 AFL Rising Star award?

Generation-Augmented Retrieval: $y_1 \parallel q \rightarrow \mathcal{D}_{y_1 \parallel q}$

Title: Jesse Hogan Context: Jesse Hogan ... playing for the Melbourne Football Club. A key forward, **Hogan is 1.95 m tall** ... made his AFL debut in the 2015 season and won the Ron Evans Medal as the AFL Rising Star

Retrieval-Augmented Generation: $\mathcal{D}_{y_1 \parallel q} \parallel q \rightarrow y_2$

The 2015 AFL Rising Star award was won by Jesse Hogan of Melbourne. Jesse Hogan is **1.95 m tall**. So the answer is 1.95 m

ITER-RETGEN When: after answer / How: Contriever / What to do: retrieve again

RARR

When

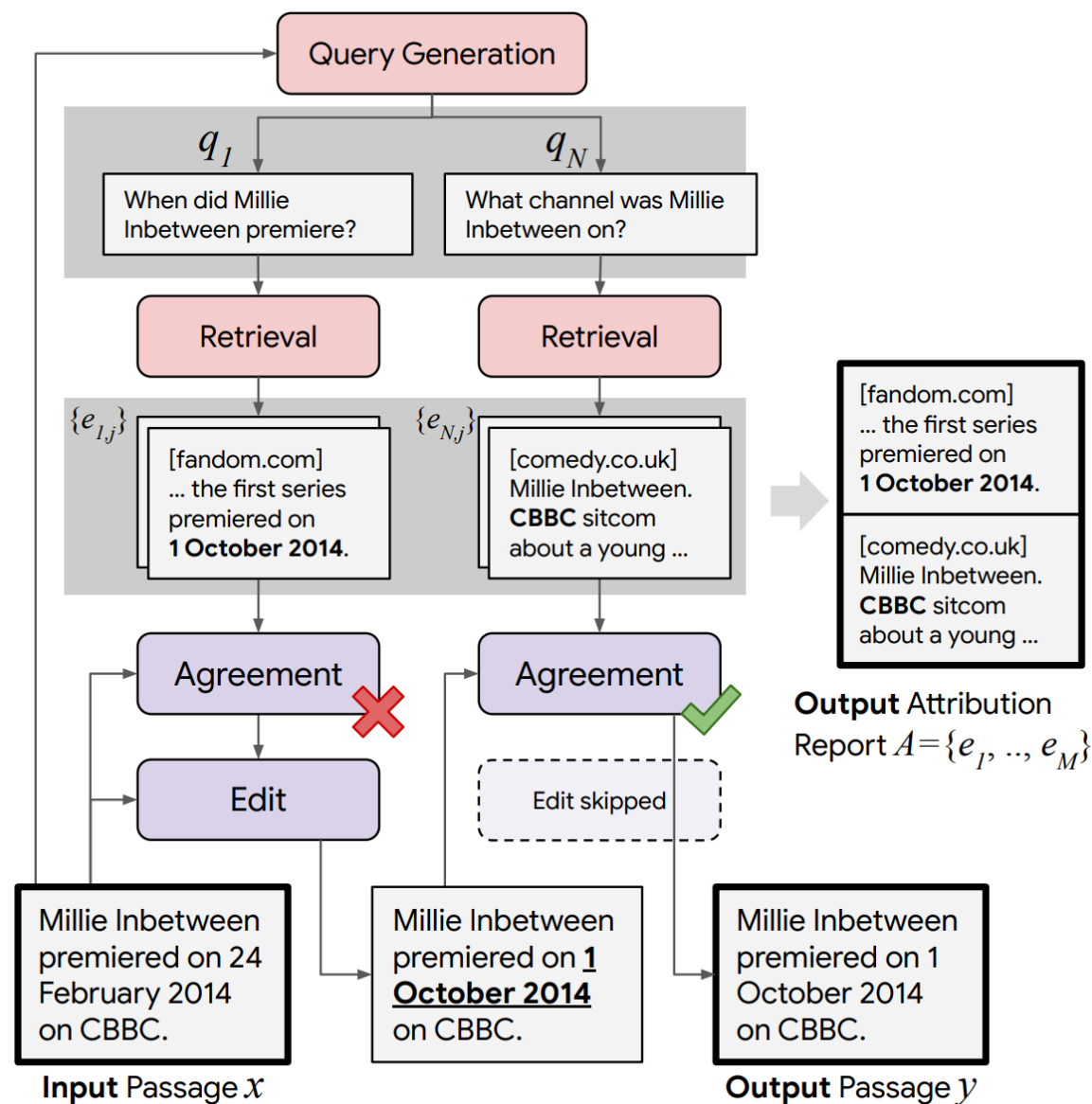
→ After answer

How

→ Google

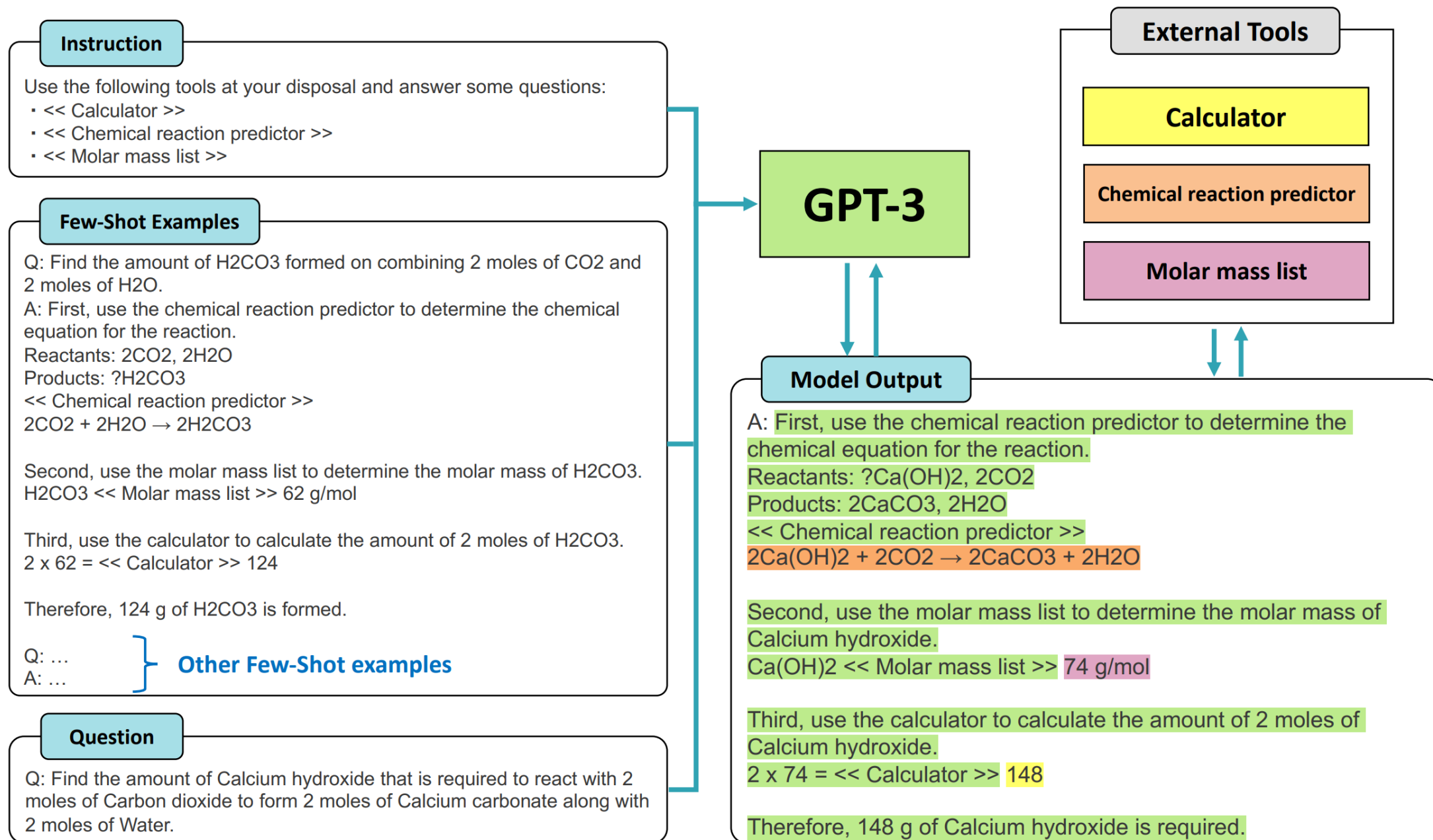
What to do

→ Agreement, revise, attribute



Retrieval Management

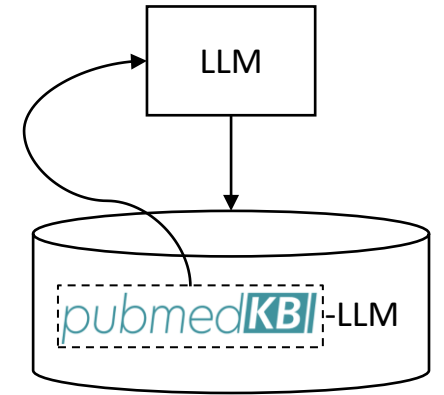
- Question → retrieval → answer
- Question → each [sub-question -> retrieval] → CoT → answer
- Question → each [CoT stage -> retrieval] → CoT → answer
- Question → answer → retrieval → answer
- Chimpanzee



Agenda

- ODQA with LLM and Knowledge Retrieval
- Retrieval and Re-Ranking
- Retrieval Management
- **pubmedKB Minimum ODQA**

pubmedKB-LLM-ORE



Annotate all co-occurrence of

→ MeSH diseases (D) & human protein-coding gene (G) & PubMed abstract (P)

Throughput

→ 1.1M DGP/day by maintaining full loading at OpenAI tier-4

	# DGP tuples (days)	# unique DG pairs	# unique D	# unique G	# unique P
All	24,810,596 (23)	3,128,402	8,894	18,393	3,997,496
Top score DG + cover all D&G	3,365,424 (3)	75,921	8,894	18,393	1,745,538

pubmedKB Minimum ODQA

Knowledge → Partial pubmedKB-LLM-ORE

When → Rules

How → BM25, DPR

What → Revise answer

pubmedKB Minimum ODQA

Knowledge → Partial pubmedKB-LLM-ORE

When → Rules

How → BM25, DPR

What → Revise answer

Query:

```
{
  "disease_name_list": ["rheumatoid arthritis"],
  "variant_name_list": ["CRP_rs1205"],
  "question": "What is the relationship between Rheumatoid Arthritis and CRP?"
}
```

(provide at least the "question")

The relationship between rheumatoid arthritis (RA) and C-reactive protein (CRP) is quite significant. CRP is a marker of inflammation in the body, produced by the liver in response to inflammation or infection. In the case of RA, CRP levels are often elevated due to the chronic inflammation in the affected joints. Higher CRP levels generally indicate more active disease and greater inflammation in RA.

CRP levels can be used to diagnose and monitor the progression of RA. It can also be used to assess the effectiveness of treatment and medications in managing the inflammation associated with the disease. Monitoring changes in CRP levels over time can indicate whether the disease is worsening or improving. However, it is important to note that CRP is a non-specific marker of inflammation, so elevated levels can also be seen in other conditions such as infections or other autoimmune diseases.

Several genetic variants, such as rs1130864, rs1205, rs1800947, rs2794521, and rs3091244, have been associated with RA and have been found to impact CRP levels in patients with the disease. These genetic variants can influence baseline CRP levels and may contribute to the development and severity of RA.

CRP is also believed to play a direct role in bone destruction in RA. It has been found to have a close correlation with tissue inflammation scores in patients with the disease. Therefore, CRP is considered an important index for evaluating the disease activity of RA and assessing its impact on the body.

Overall, the relationship between RA and CRP is that elevated CRP levels indicate active disease and inflammation in the joints. Monitoring CRP levels can be useful in diagnosing, monitoring, and assessing the effectiveness of treatment in RA patients. Genetic variants can also influence CRP levels in RA patients and contribute to the development and severity of the disease.

Reference

PMID	Title	Year	Journal	Citation
35896901	Correlation of CRP genotypes with serum CRP levels and the risk of rheumatoid arthritis in Chinese Han population.	2022	Clinical rheumatology	0
33628645	Quantile-dependent expressivity of serum C-reactive protein concentrations in family sets.	2021	PeerJ	8
23039402	Correlation of C-reactive protein haplotypes with serum C-reactive protein level and response to anti-tumor necrosis factor therapy in UK rheumatoid arthritis patients: results from the Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate cohort.	2012	Arthritis research & therapy	17
35062811	Association of CRP Haplotypes in Rheumatoid Arthritis and their Correlation with Severity of the Disease.	2022	The Journal of the Association of Physicians of India	0