

# Medical Question Answering Using GraphRAG

November 8, 2024

Murphy Liu, Keon Nartey, Suim Park, Bob Zhang

# Objective

- **Background**

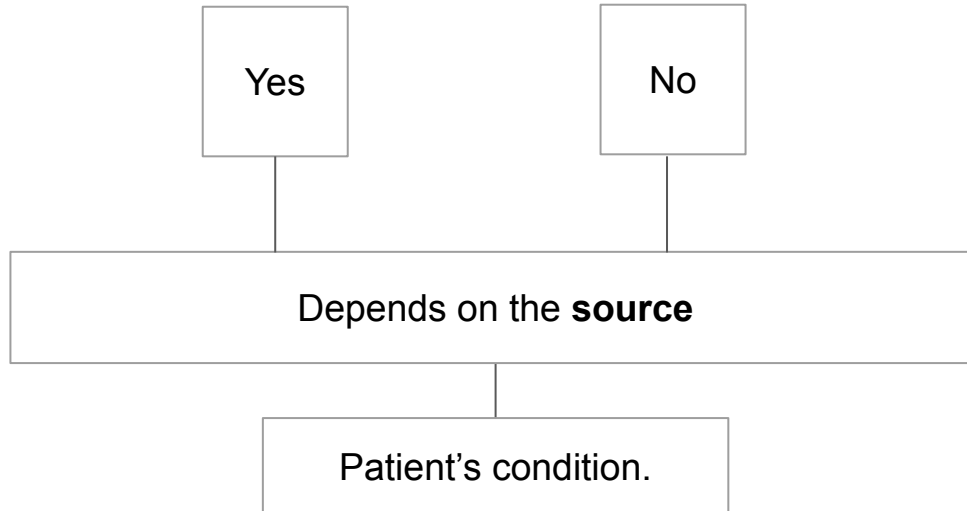
- As the number of research papers and guidelines increase, it becomes a **tremendous burden for physicians** to stay up-to-date to **make informed decisions** faced with various clinical conditions.
- GraphRAG vs. RAG

- **Objective**

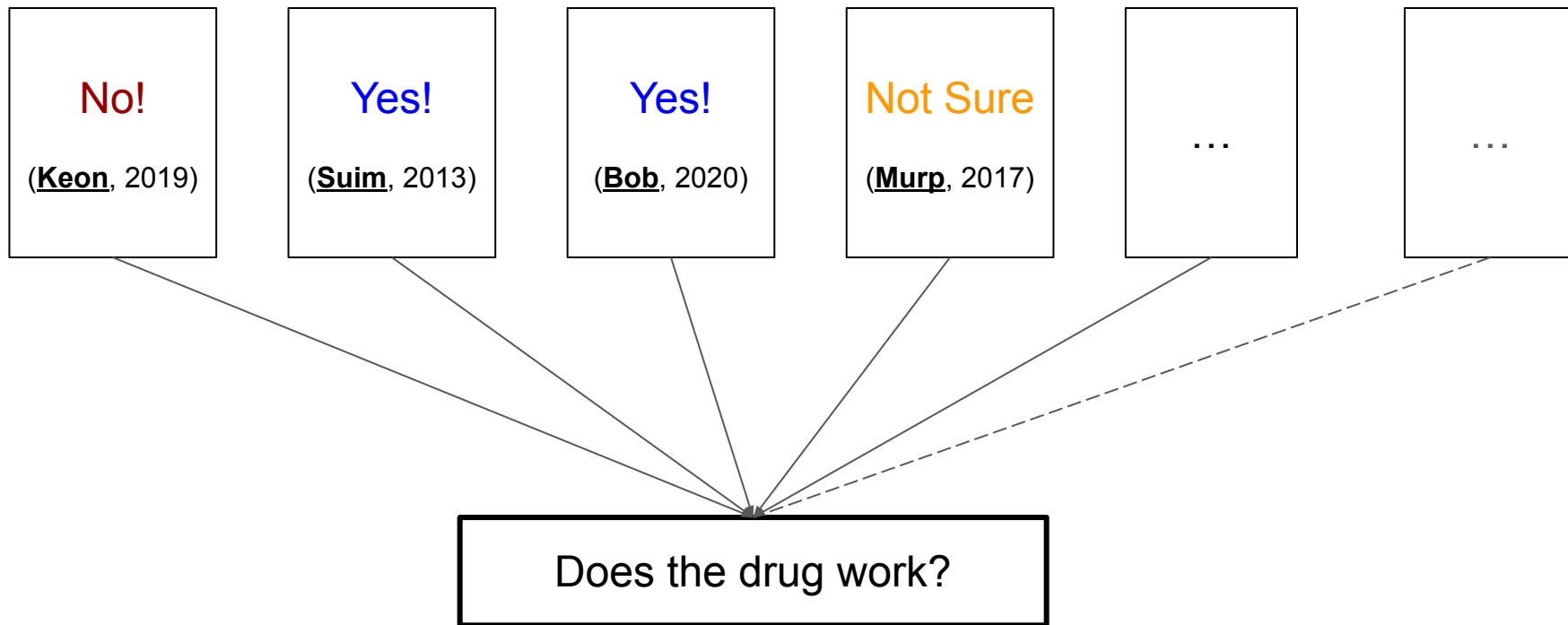
- The development of an **evidence-based medical question answering model** for clinical use.

# Question

Among patients with septic shock and relative adrenal insufficiency, do corticosteroids reduce 28-day mortality?

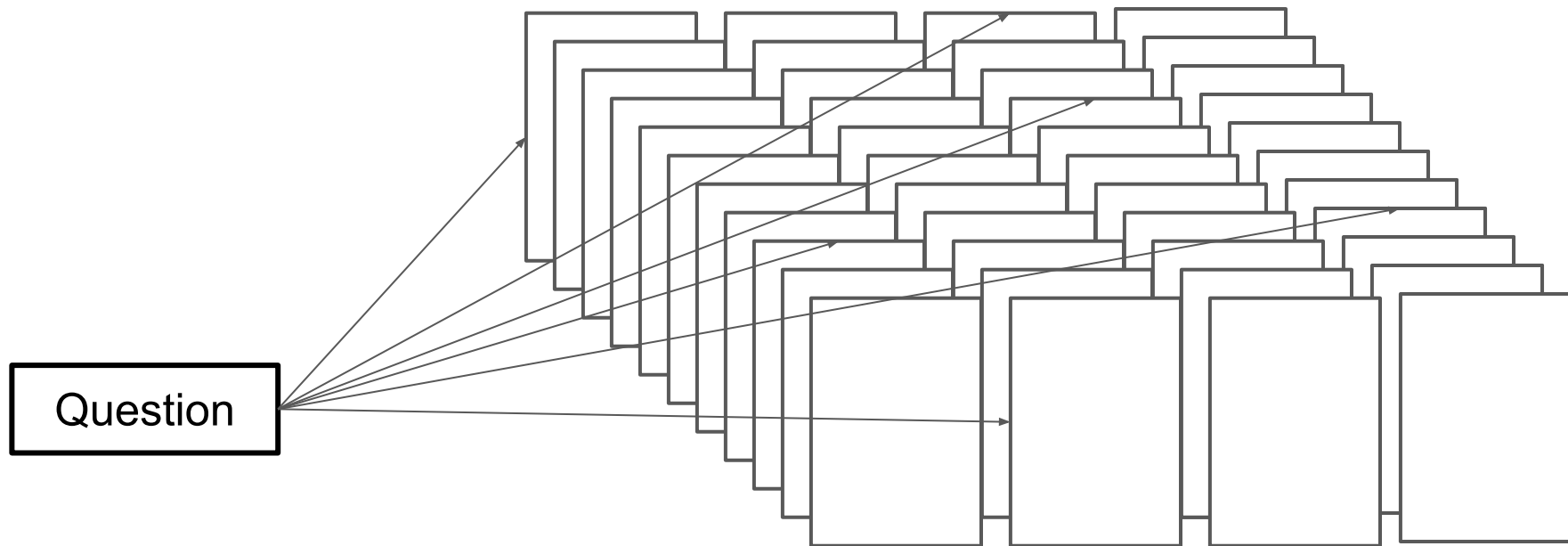


# Medical Questions Answered by Research Papers



# Quantity Counts

What happens when the amount of source document becomes HUGE?



# Outline

- Dataset
- The GraphRAG Model
- Evaluation Metrics
- A Small GraphRAG Experiment
- Results
- Next Steps

# Dataset: Wiki Journal Club



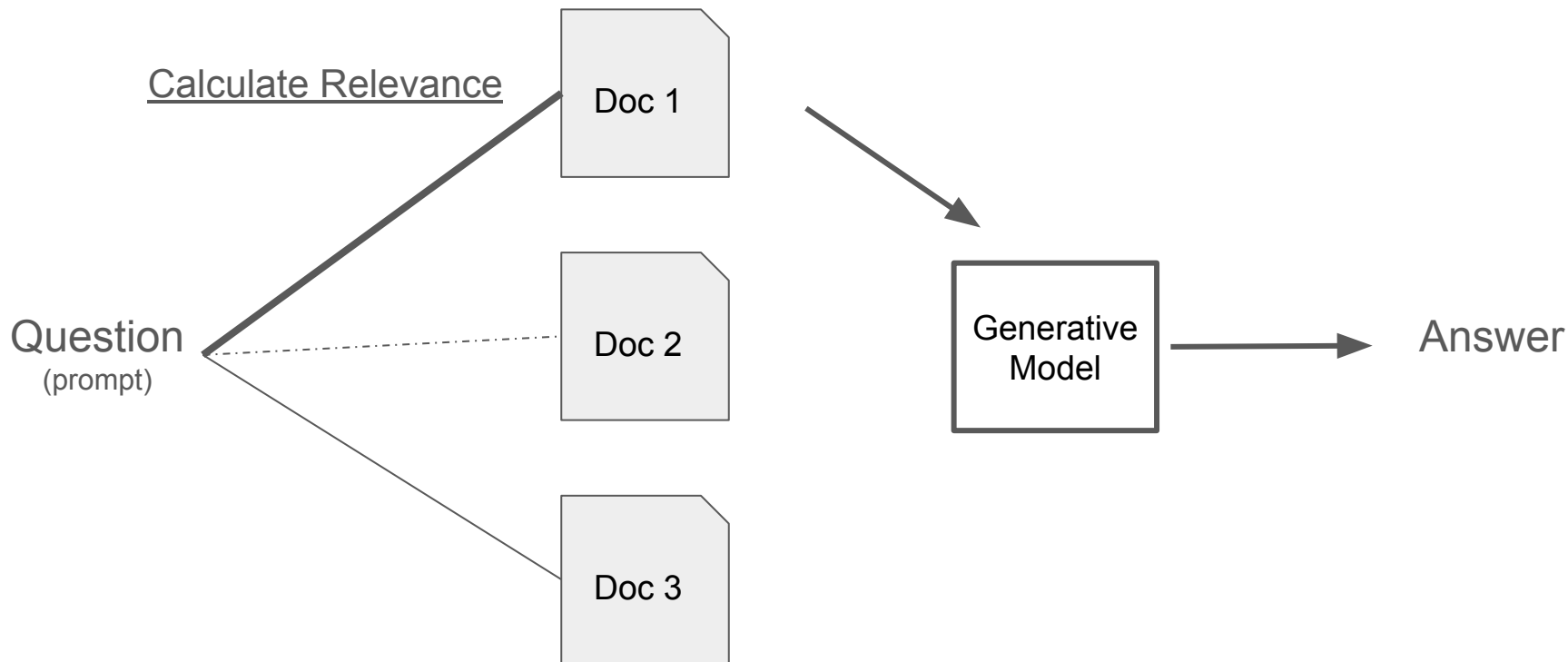
## + **Characteristics**

- Speciality-Specific Collections
- Critical Appraisal
- Wikipedia based format
- Free and Open Access
- Community Contributions

## - **Pitfalls**

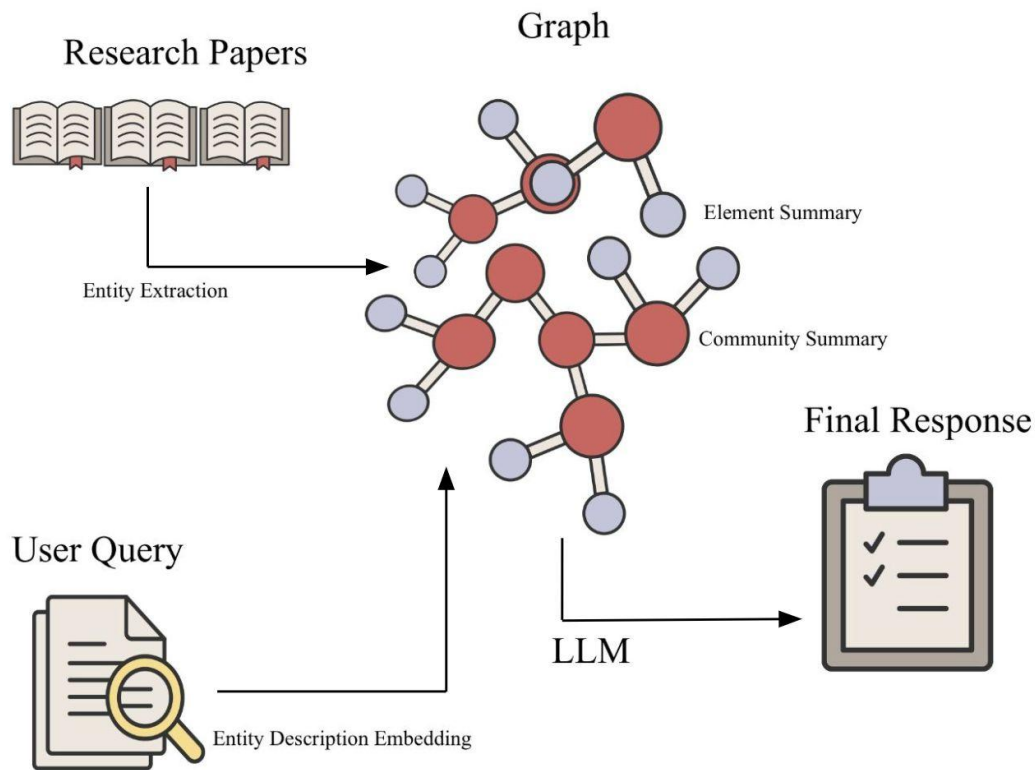
- Bias
- Inconsistent Updates - Volunteers
- Quality Control
- Limited Scopes

# Retrieval Augmented Generation (RAG)





# GraphRAG



# Evaluating Generated Texts I

Question: **Where is Gross Hall?**

Ground Truth

**It is** located on Science Drive.

Generated Answer:

**It is** a building at Duke University.

## BLEU Score:

- Num of matched words/ Num of generated words:  $2/8 = \underline{.25}$  (Precision)

## ROUGE Score:

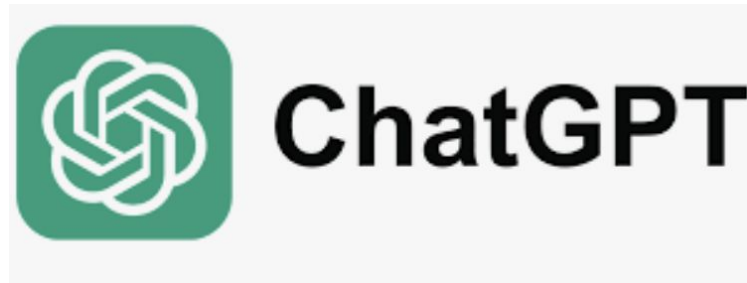
- Num of matched words/ Num of reference words:  $2/6$  (Recall)
- F1 Score:  $2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall}) = 1/7 = \underline{.14}$

# Evaluating Generated Texts II

**Compared with the provided answer, how good is the generated one?**



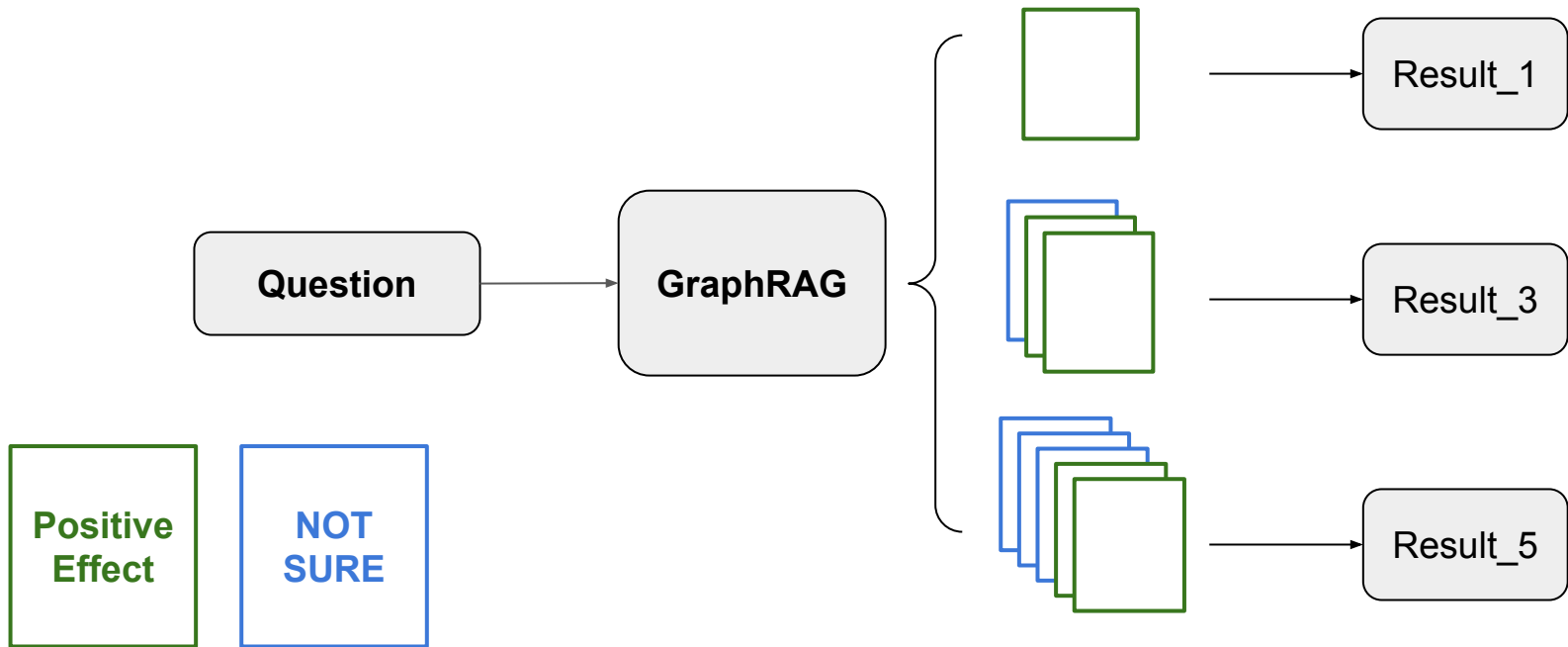
Human Rating



# Experiments

1. How does the **number of documents** used to build GraphRAG **affect the quality of generated answers**?
2. What will happen if the **documents point to different answers**?

Question 1	Ground Truth	Source
Among patients with <b>SEPSIS</b> shock and relative adrenal insufficiency, do <u>corticosteroids</u> <b>reduce mortality</b> ?	<b>Positive Effect</b>	<a href="#">Annane Trial</a>



# Results of Question 1

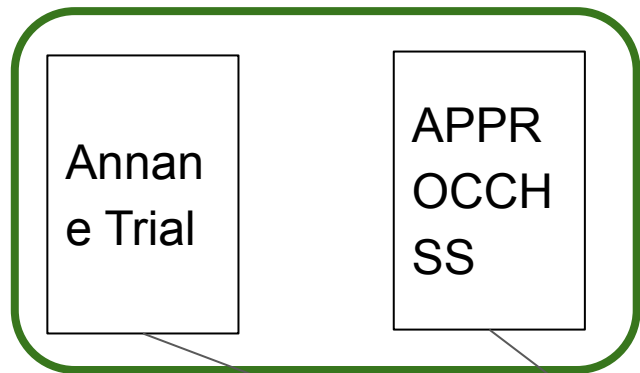
**Q1:** Among patients with **septic shock** and **relative adrenal insufficiency**, do **corticosteroids** reduce **28-day mortality**?

Document	Human Rating	GPT	BLEU	ROUGE-L
1 (Annane Trial)	0.7	0.80	0.03	0.09
3 (Annane Trial, ADRENAL, APROCCHSS)	0.8	0.90	0.03	0.08
5 (Annane Trial, ADRENAL, APROCCHSS, CORTICUS, HYPRESS)	0.85	0.90	0.03	0.09

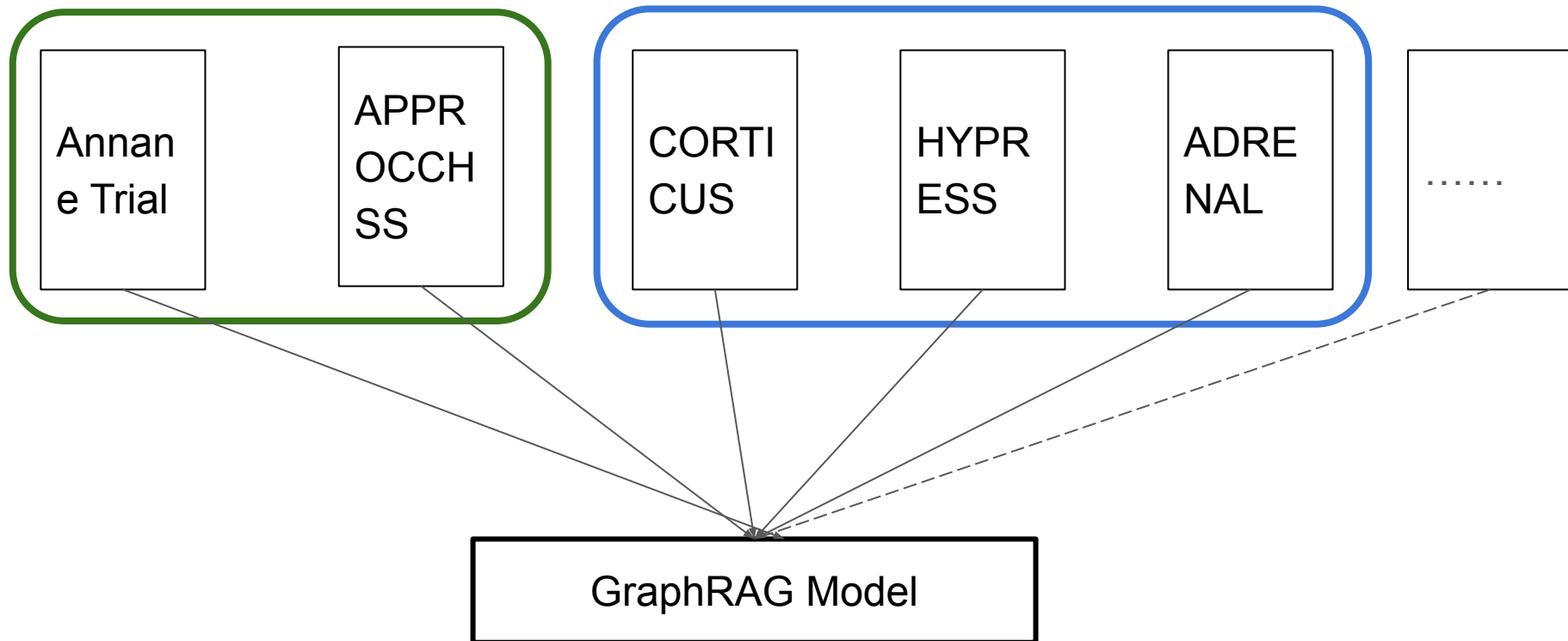
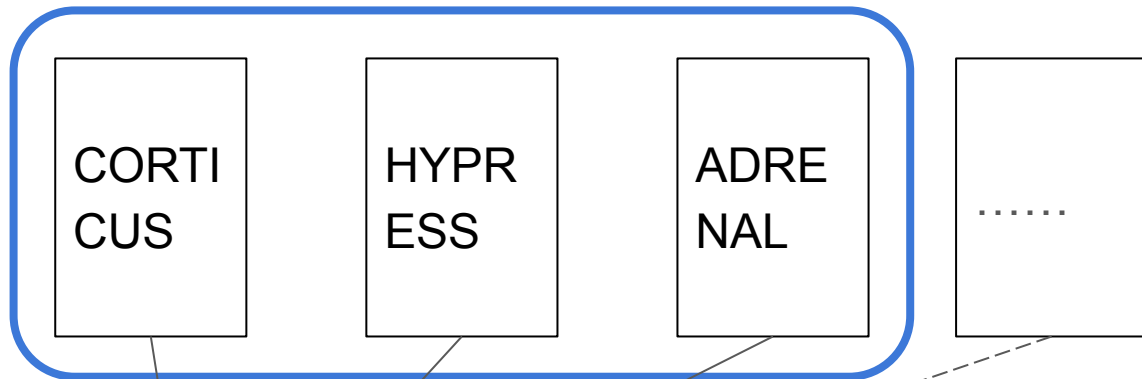
Positive  
Effect

NOT  
SURE

**Positive Effect**

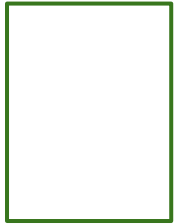


**Not Sure**



# Interesting outcome based on Q1

**Ground truth:** Among patients with septic shock and relative adrenal insufficiency, administration of corticosteroids **reduces 28-day mortality.**

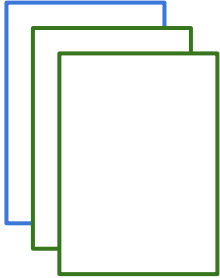


**Model result:** ... the direct impact of continuous neuromuscular blockade with cisatracurium on mortality rates within the context of light sedation protocols **requires further investigation.** ... its efficacy and impact on mortality ... should be **interpreted within the broader context** of individual patient needs and evolving clinical evidence



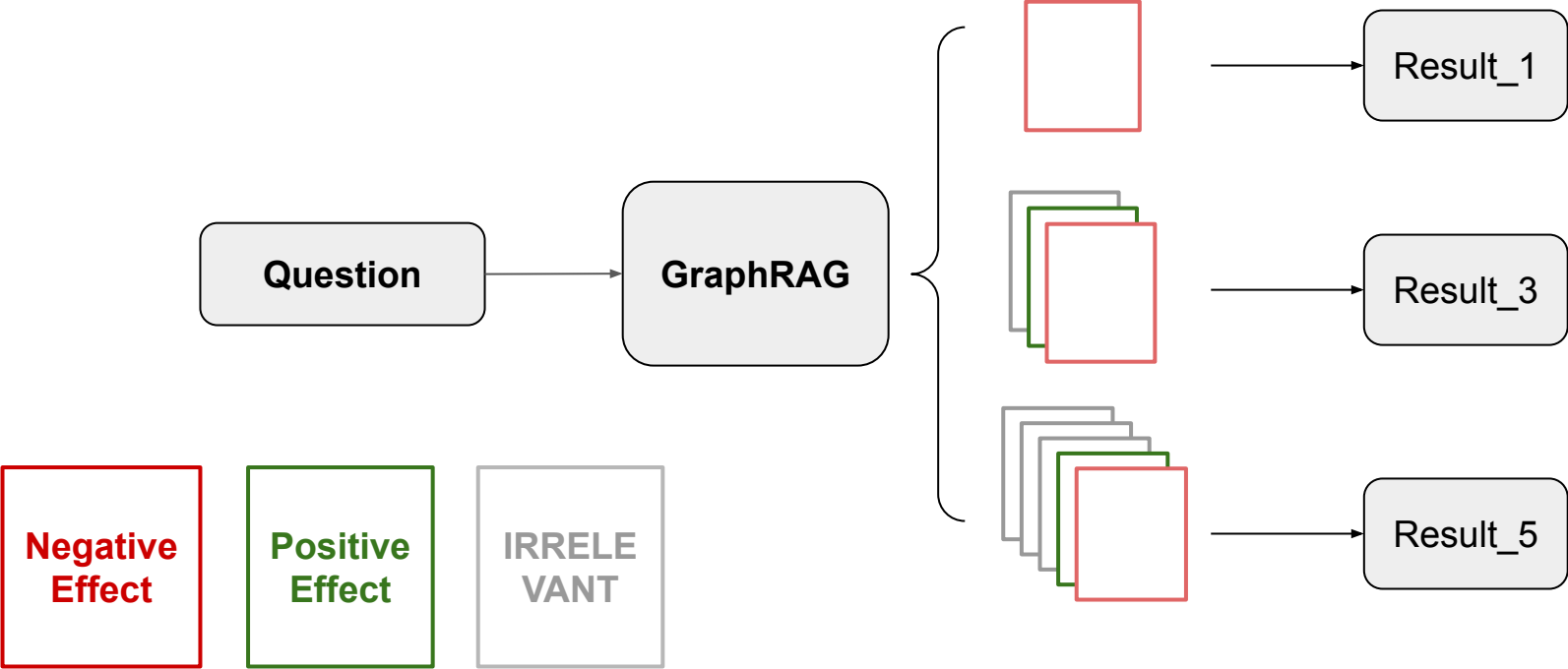
# Interesting outcome based on Q1

**Ground truth:** Among patients with septic shock and relative adrenal insufficiency, administration of corticosteroids **reduces 28-day mortality.**



**Model result:** ... appears to **improve mortality** in patients with moderate to severe ARDS. The **findings support the safety of cisatracurium**, showing **no increased risk of muscle weakness**, and highlight the importance of adherence to established sedation protocols and monitoring guidelines to maximize patient outcomes.

Question 2	Ground Truth	Source
For <b>ARDS</b> patients, does <u>cisatracurium</u> reduce mortality when used with current light sedation protocols?	<b>Negative Effect</b>	<a href="#">ROSE</a>



# Results- Question 2

**Q1:** In patients with moderate to severe ARDS, does the early use of continuous neuromuscular blockade with cisatracurium improve mortality when used with current light sedation protocols?

Number of Document	Human Rating	GPT	BLEU	ROUGE-L
1 (ROSE)	<u>0.2</u>	<u>0.5</u>	0.01	0.05
3 (ROSE, ACURASYS, ARDSNET)	0.1	0.2	0.00	0.07
5 (ROSE, ACURASYS, ARDSNET, FACTT, OSCILLATE)	0.1	0.1	0.01	0.06

**Negative  
Effect**

**Positive  
Effect**

**IRRELE  
VANT**

**Negative  
Effect**

**Positive  
Effect**

ARDS related but not Answering

ROSE

ACUR  
ASYS

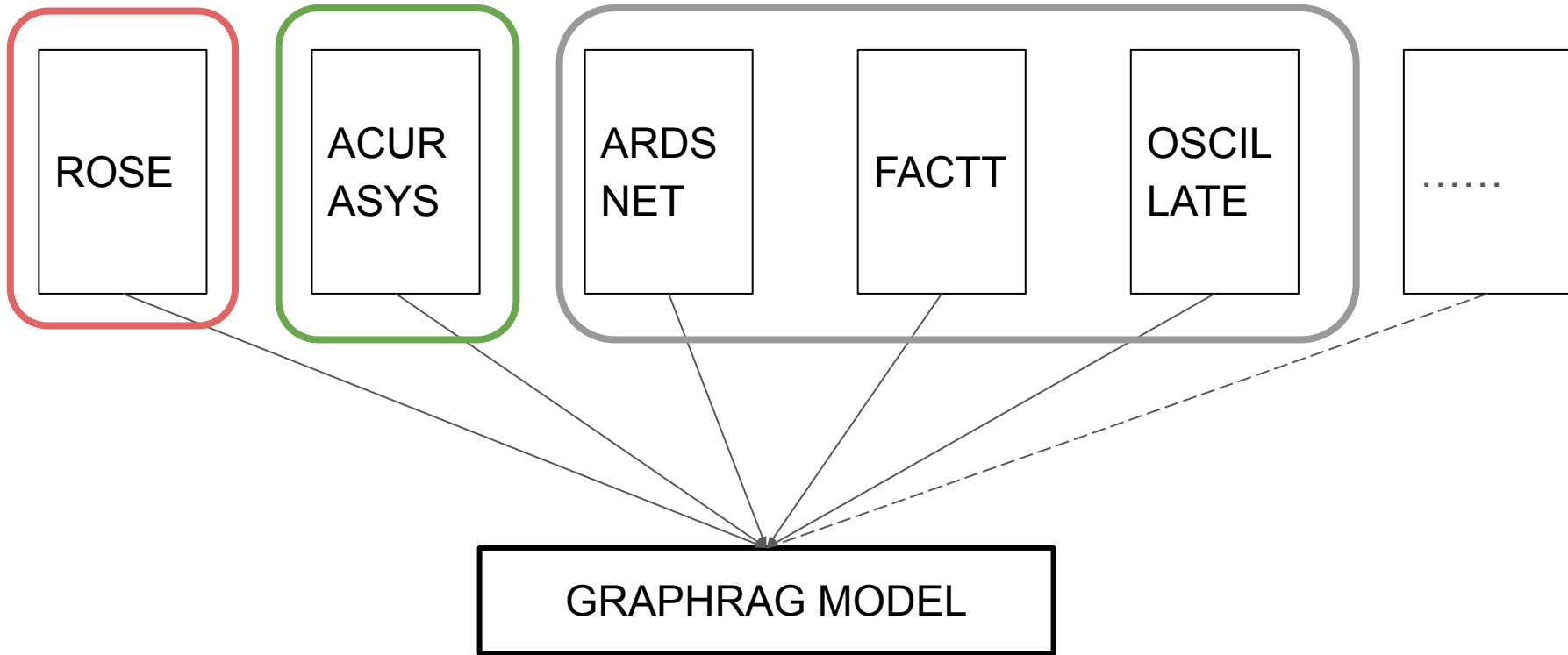
ARDS  
NET

FACTT

OSCIL  
LATE


.....

GRAPHRAG MODEL



## Interesting outcome based on Q2

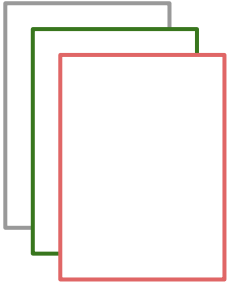
**Ground truth:** Not only does early neuromuscular blockade **NOT** reduce mortality when using modern light sedation protocols, it was associated with an **INCREASE** in ICU-acquired weakness and **serious ADVERSE** cardiovascular events.



**Model result:** ... the direct impact of continuous neuromuscular blockade with cisatracurium on mortality rates within the context of light sedation protocols **requires further investigation.** ... its efficacy and impact on mortality ... should be **interpreted within the broader context** of individual patient needs and evolving clinical evidence

## Interesting outcome based on Q2

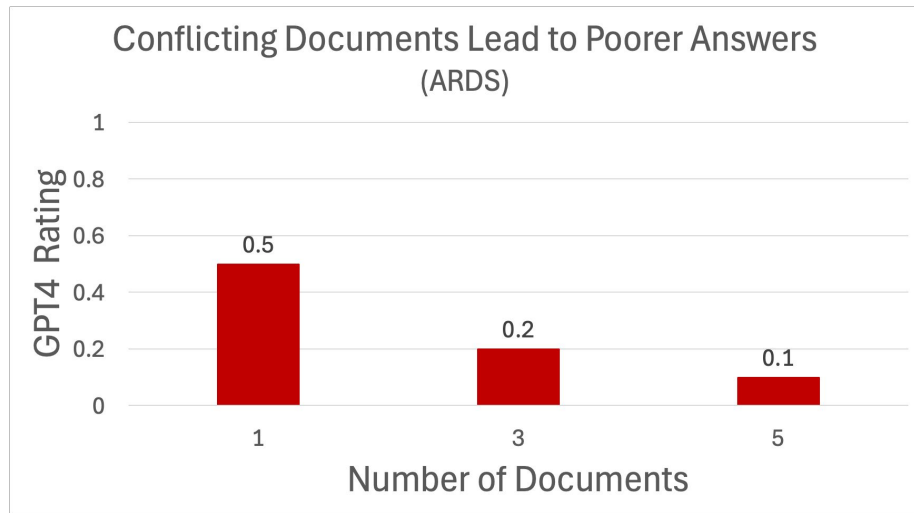
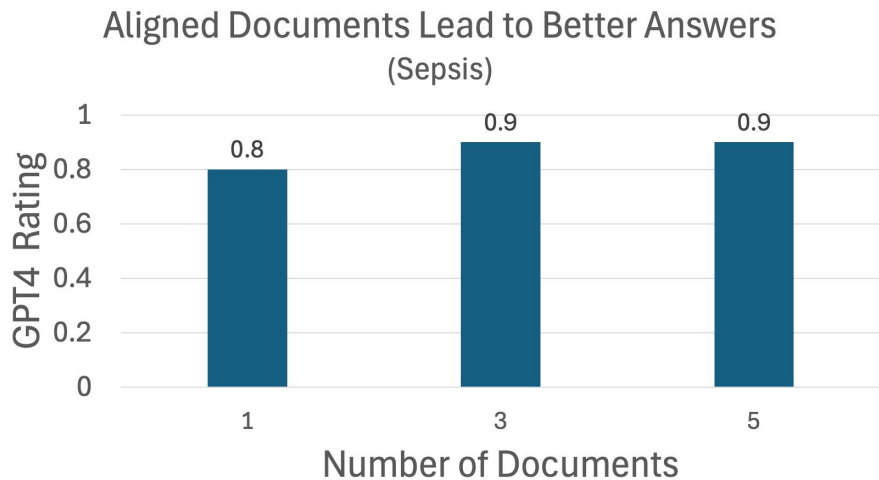
**Ground truth:** Not only does early neuromuscular blockade **NOT** reduce mortality when using modern light sedation protocols, it was associated with an **INCREASE** in ICU-acquired weakness and **serious ADVERSE** cardiovascular events.



**Model result:** ... appears to **improve mortality** in patients with moderate to severe ARDS. The **findings support the safety of cisatracurium**, showing **no increased risk of muscle weakness**, and highlight the importance of adherence to established sedation protocols and monitoring guidelines to maximize patient outcomes.

# GPT Ratings

The GraphRAG model gets confused when training document point to different answers.



# Summary

1. Decision making should not be done by the model rather the model should provide sources where it learned.
2. Paraphrase earlier sentences
  - a. Model weight input documents equally based on doc contents why
  - b. Lack of meta data required such as citation, publication year, organizations etc next st
3. GraphRAG provides summarized contents
- 4.

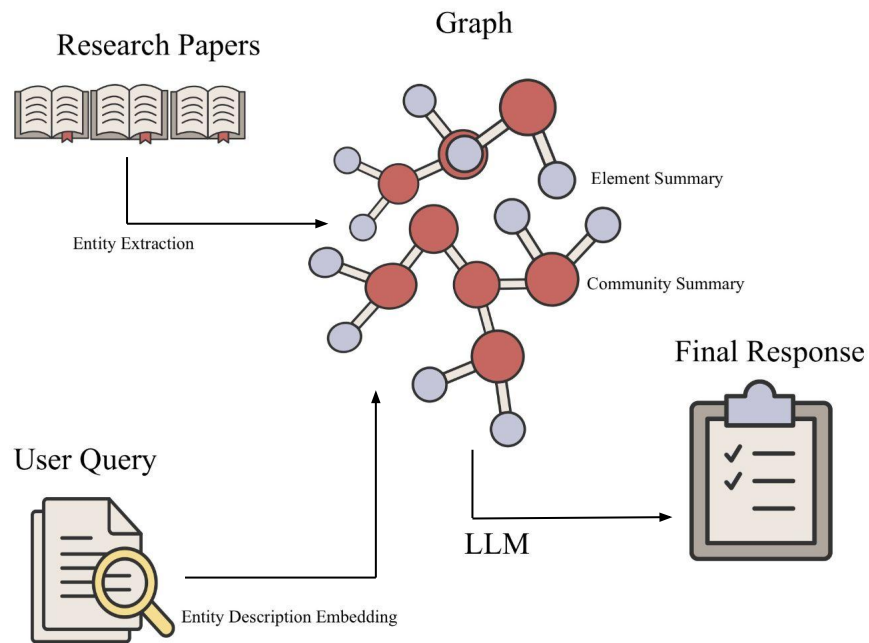


# Next Steps

- Midterm Deliverables
  - Github Repo
- Scale up
  - More documents, more questions
- Evaluate the results better
- Improve model output
  - Provide source of answer
  - Optimize the weight of knowledge graph

Thank You

# Visualization of Graph Database



Question	Answer	Source
<p>In patients with moderate to severe <b>ARDS</b>, does the early use of continuous neuromuscular blockade with <u>cisatracurium</u> <b>improve mortality</b> when used with current light sedation protocols?</p>	<p><b>No</b> cross sticker</p>	<p><a href="#">ROSE</a></p>
<p>Among patients with <b>SEPSIS</b> shock and relative adrenal insufficiency, do <b>corticosteroids</b> reduce 28-day mortality?</p>	<p><b>Yes sticker</b></p>	<p><a href="#">Annane Trial</a></p>