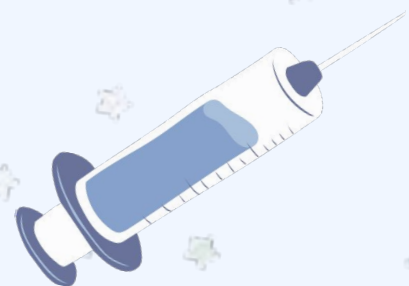


Medical Question Answering Using GraphRAG



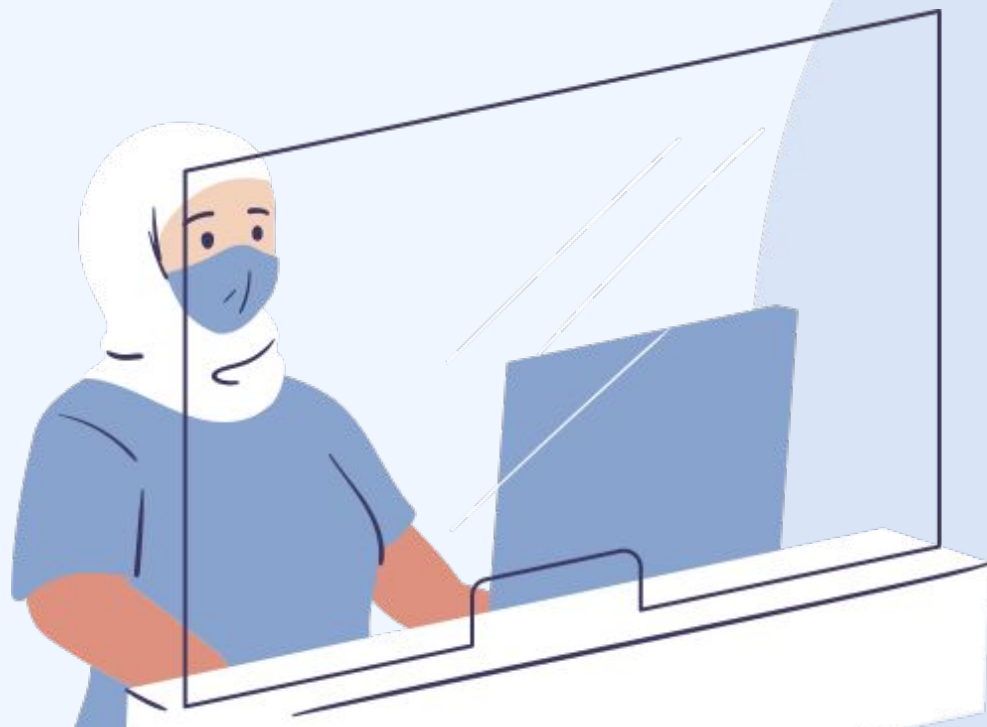
Objective

- **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?

Yes

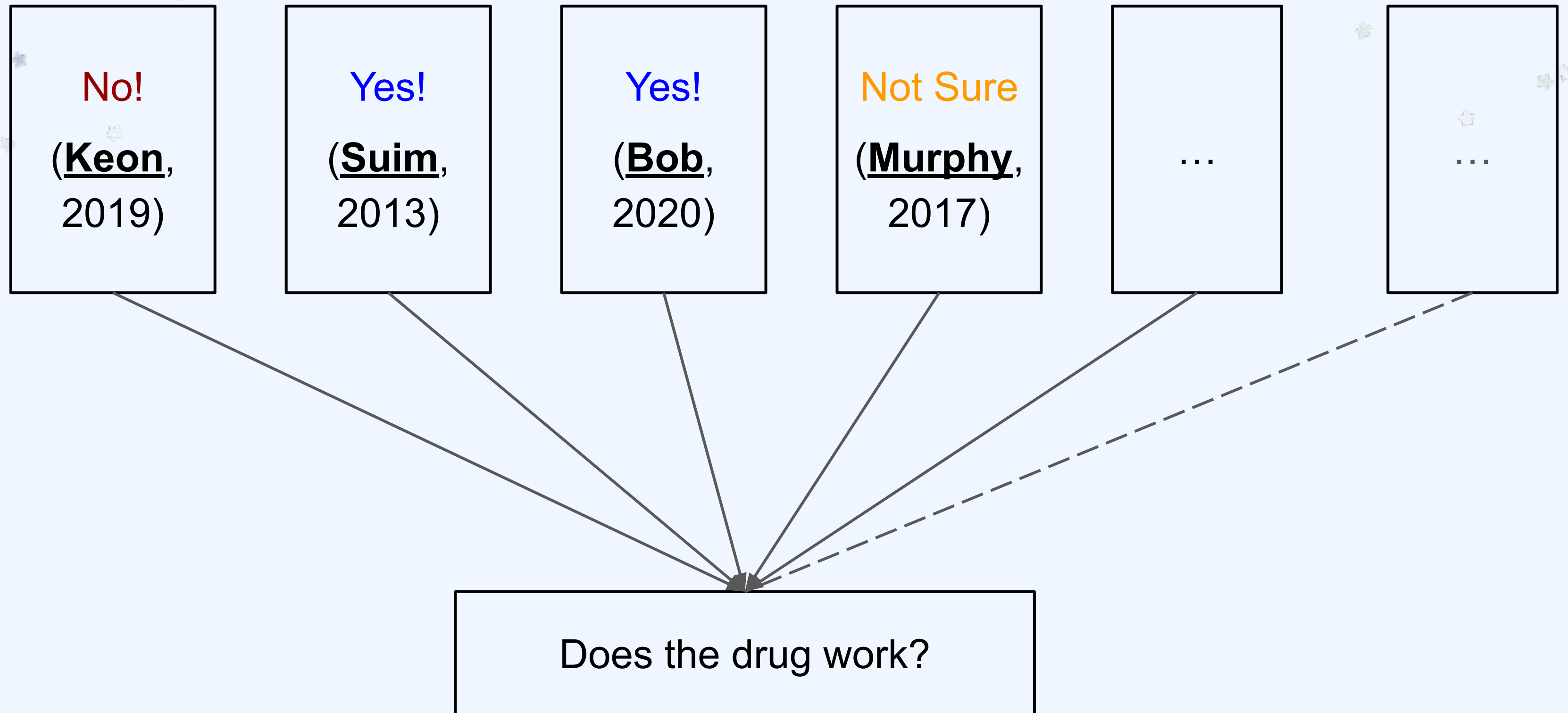
No

Depends on the **source**

Patient's condition

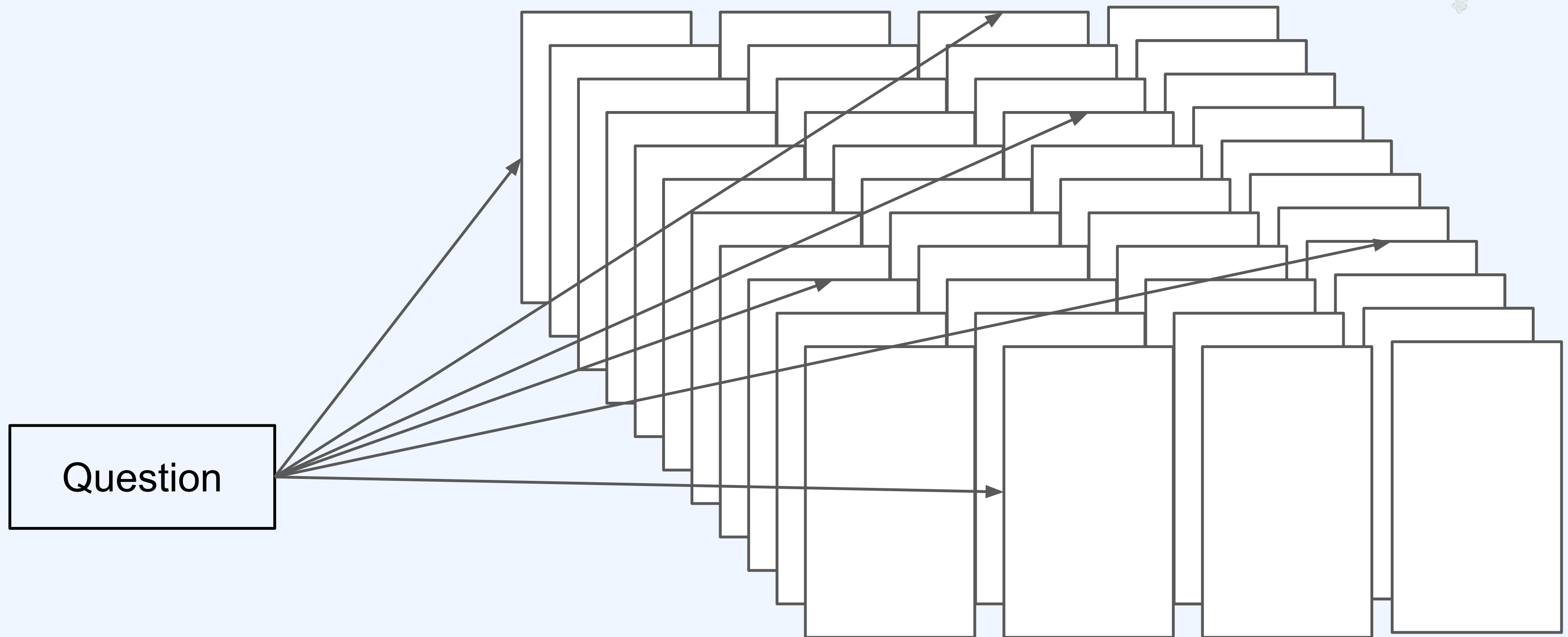


Medical Questions Answered by Research Papers



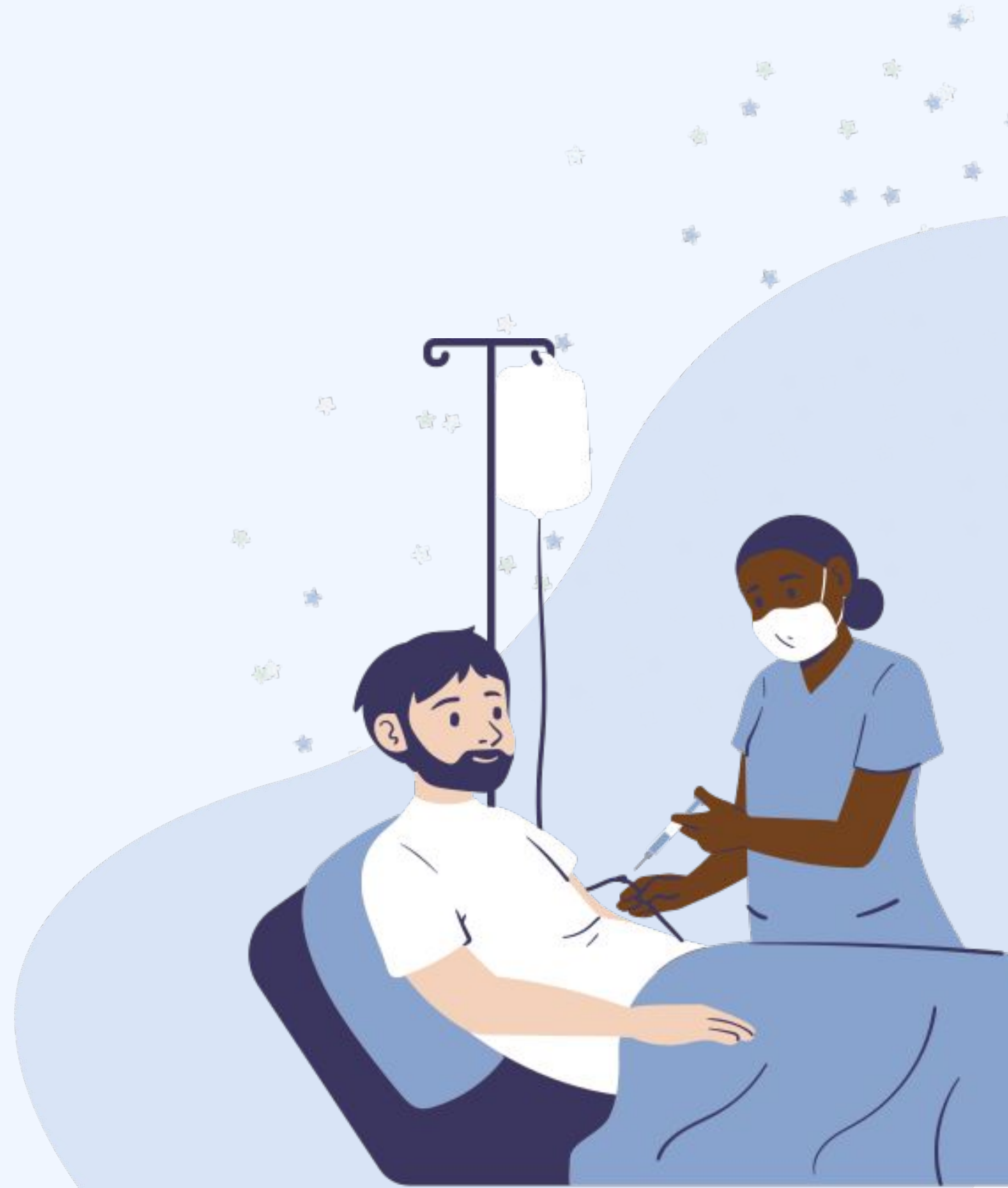
Quantity Counts

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



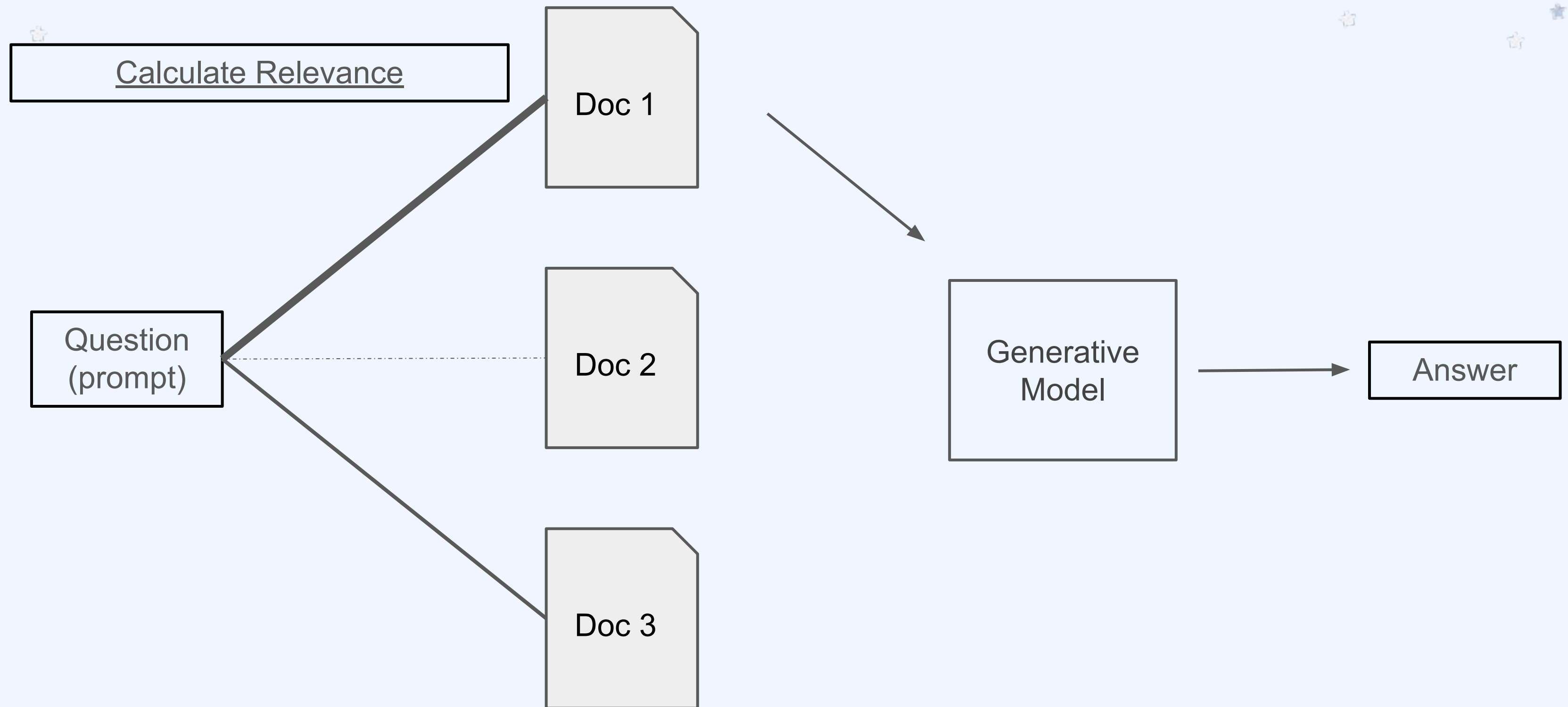
- **Characteristics**

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

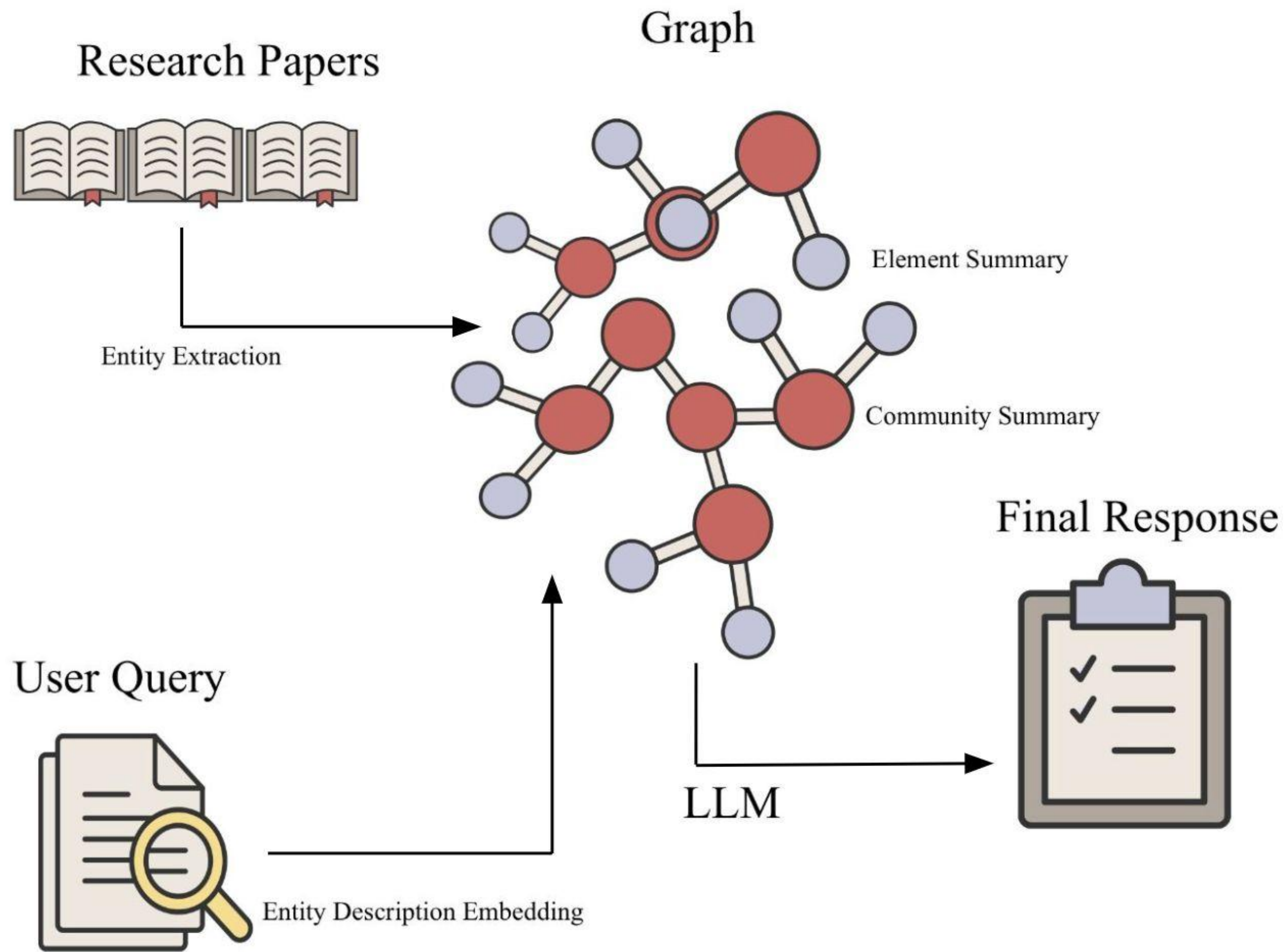
- **Pitfalls**

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

Retrieval Augmented Generation (RAG)



GraphRAG



Evaluating Generated Text 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 Text II

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



ChatGPT

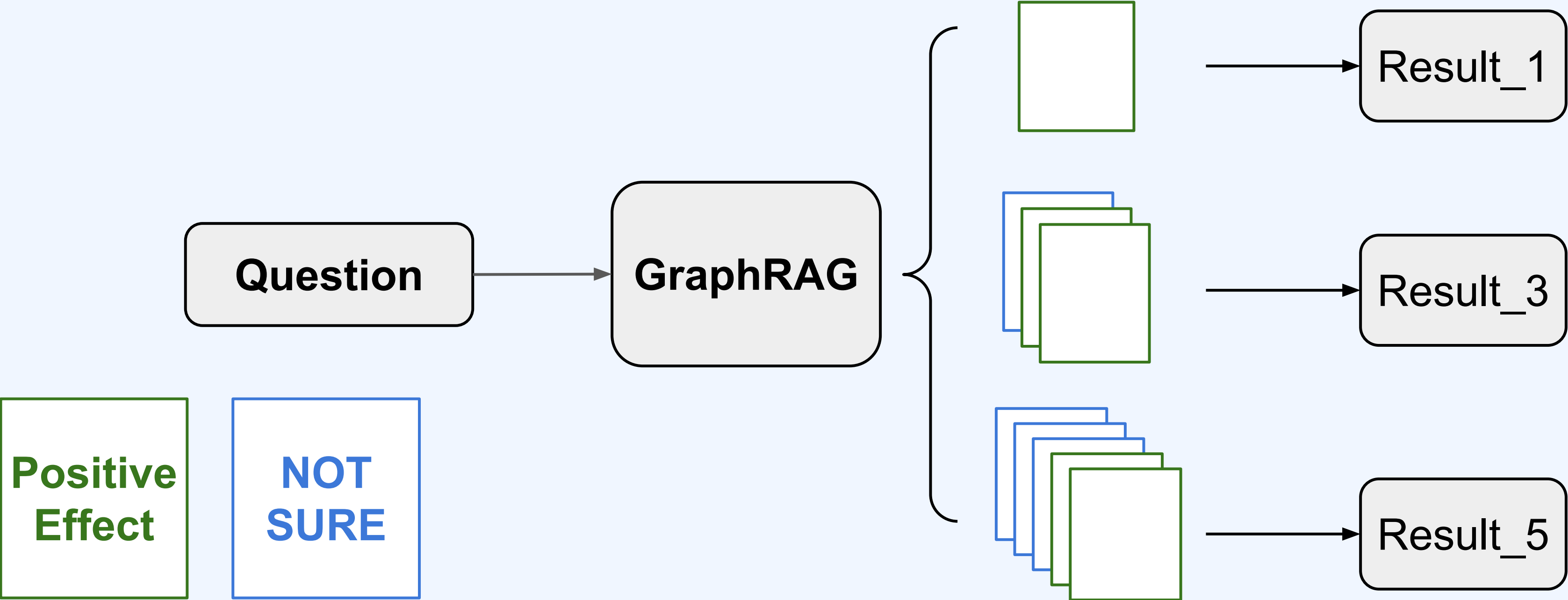
Experiment

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 SEPSIS shock and relative adrenal insufficiency, do <u>corticosteroids</u> reduce mortality?	Positive Effect	Annane Trial



Results of Question 1

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

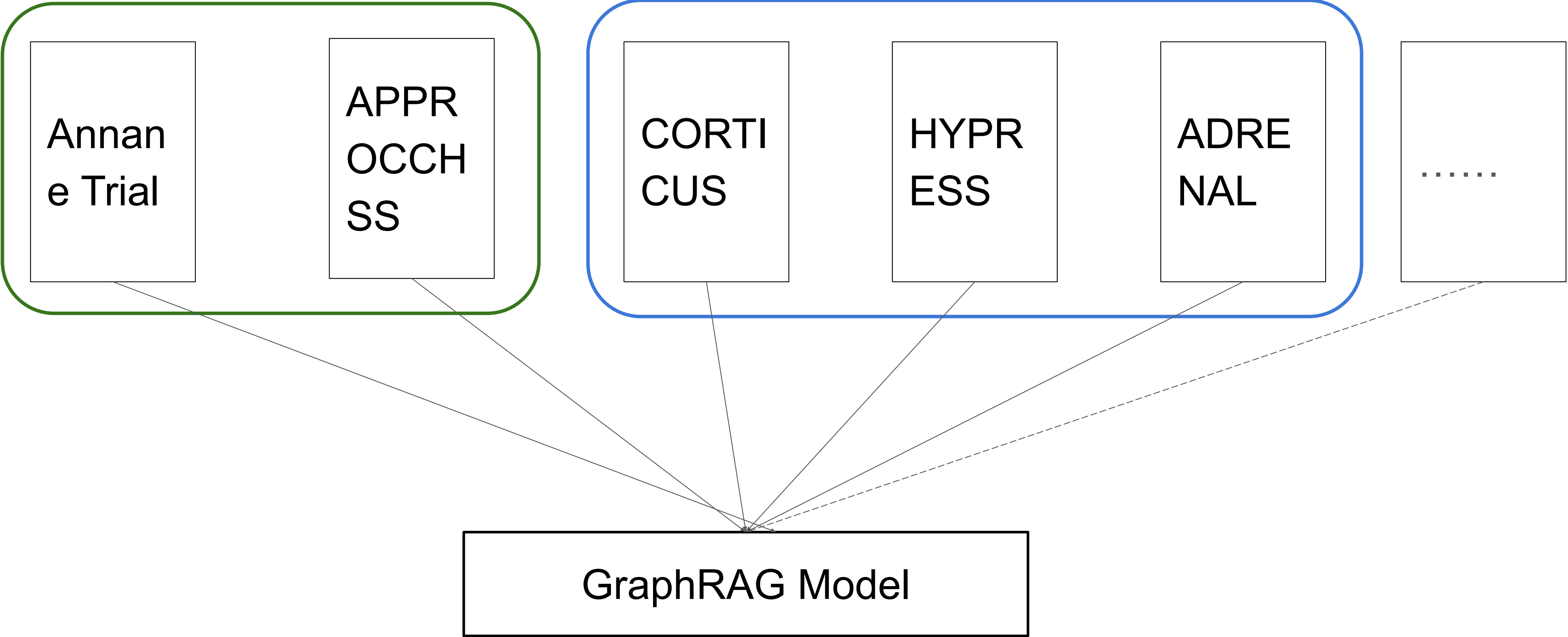
Document	Human Rating	GPT	BLEU	ROUGE-L
1 (Annane Trial)	0.7	0.80	0.03	0.09
3 (Annane Trial, ADRENAL, APROCCHSS)	0.8	0.90	0.03	0.08
5 (Annane Trial, ADRENAL, APROCCHSS, CORTICUS, HYPRESS)	0.85	0.90	0.03	0.09

**Positive
Effect**

**NOT
SURE**

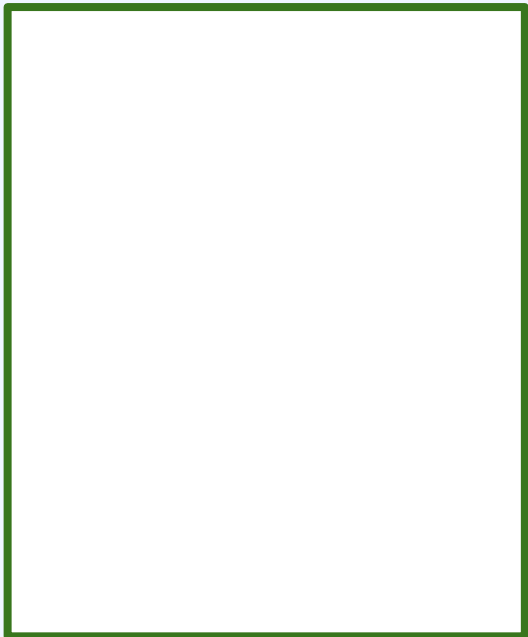
Positive Effect

Not Sure



Interesting Outcome from Q1

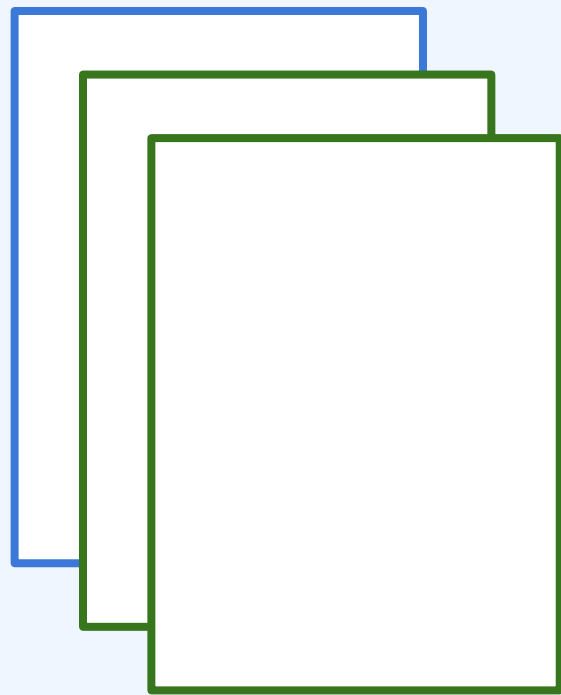
Ground truth: Among patients with septic shock and relative adrenal insufficiency, administration of corticosteroids **reduces 28-day mortality.**



Model result: ... provides compelling evidence that a 7-day treatment with low doses of hydrocortisone and fludrocortisone significantly **reduces the 28-day mortality rate** among patients with septic shock and relative adrenal insufficiency ... underscore the potential **life-saving benefits of corticosteroid treatment** ... patients with relative adrenal insufficiency for targeted therapy.

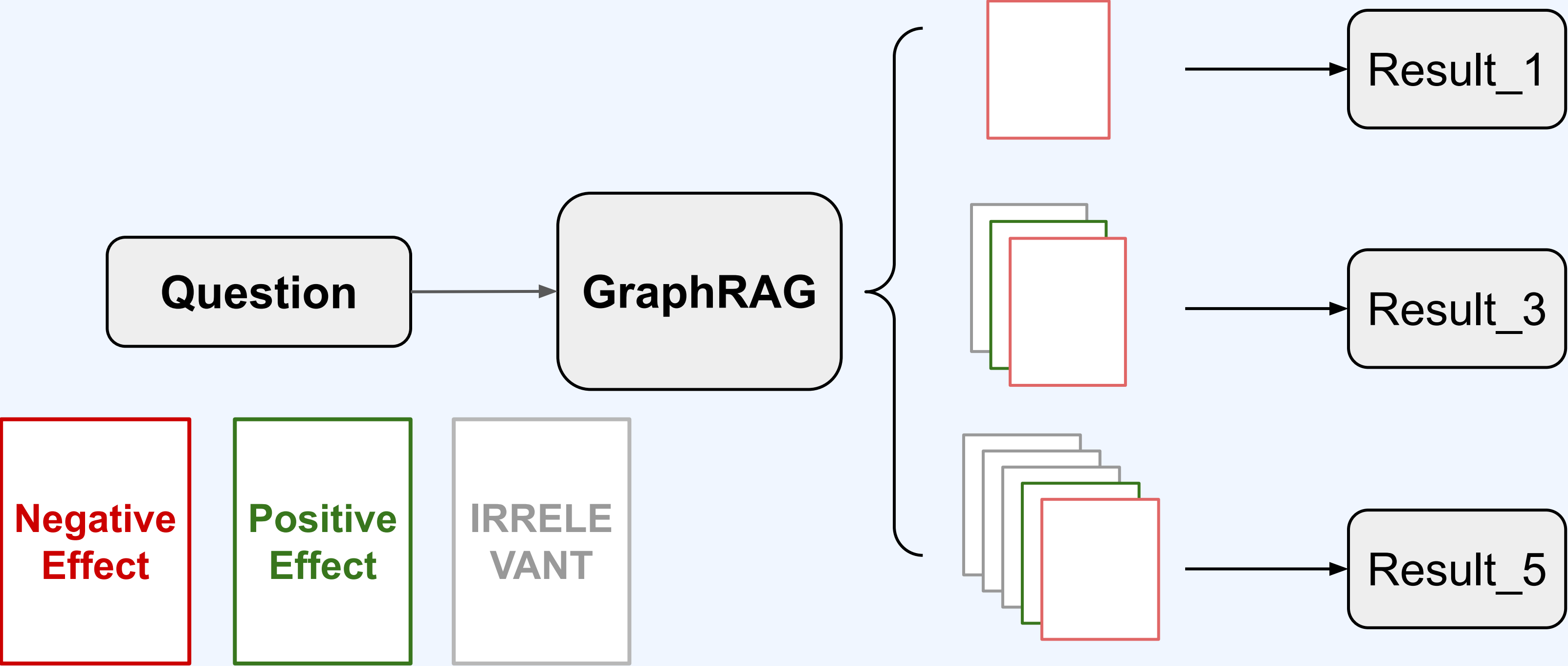
Interesting Outcome from Q1

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



Model result: ... treatment with low doses of hydrocortisone and fludrocortisone **significantly reduces 28-day mortality** without increasing adverse events. ... However, it's important to note that the effectiveness of corticosteroids in patients without relative adrenal insufficiency **remains less clear, and further research is needed** to optimize treatment protocols for all septic shock patients.

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



Results of 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

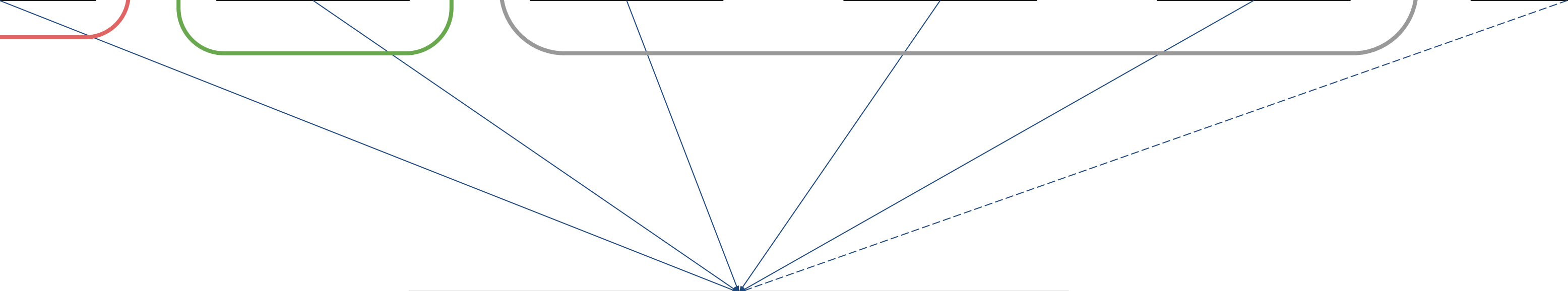
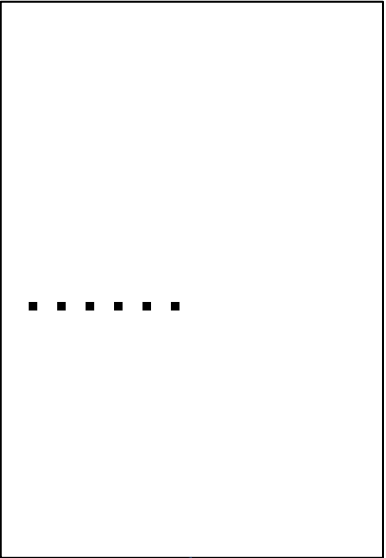
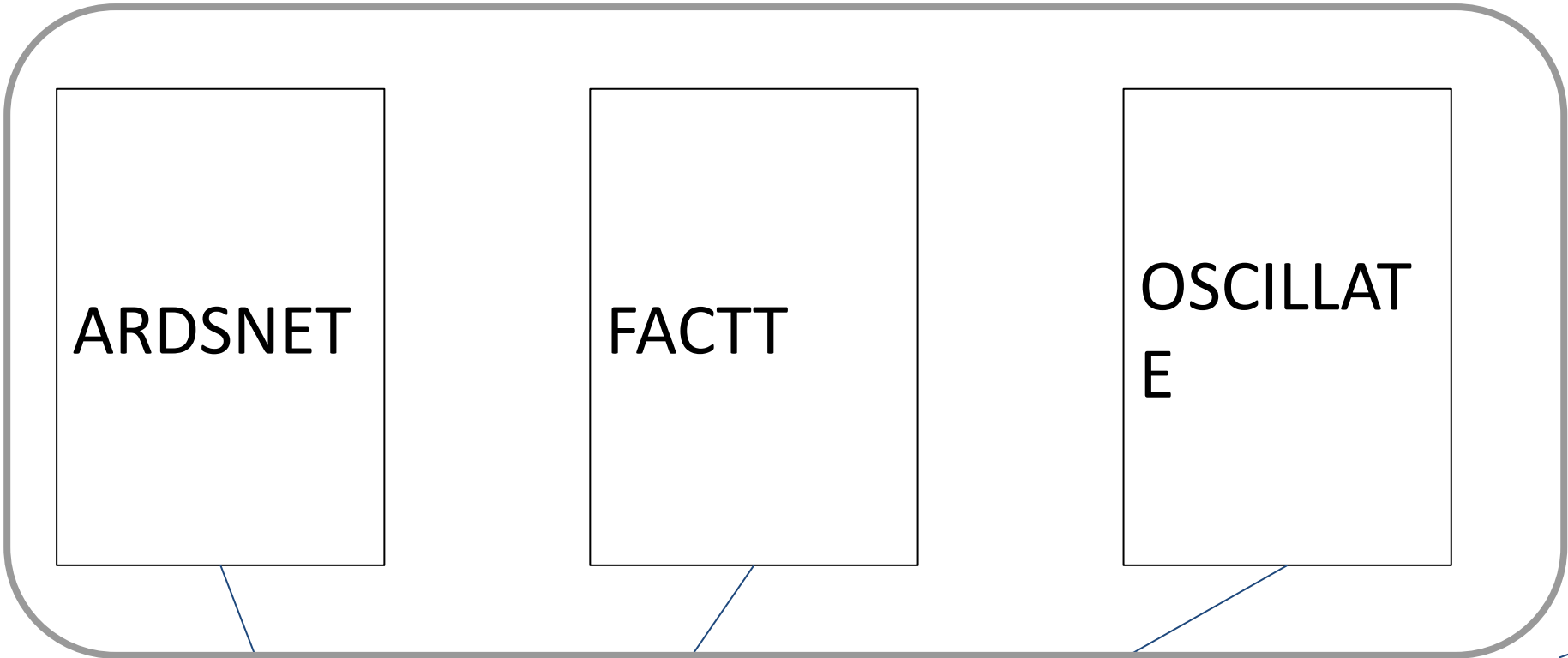
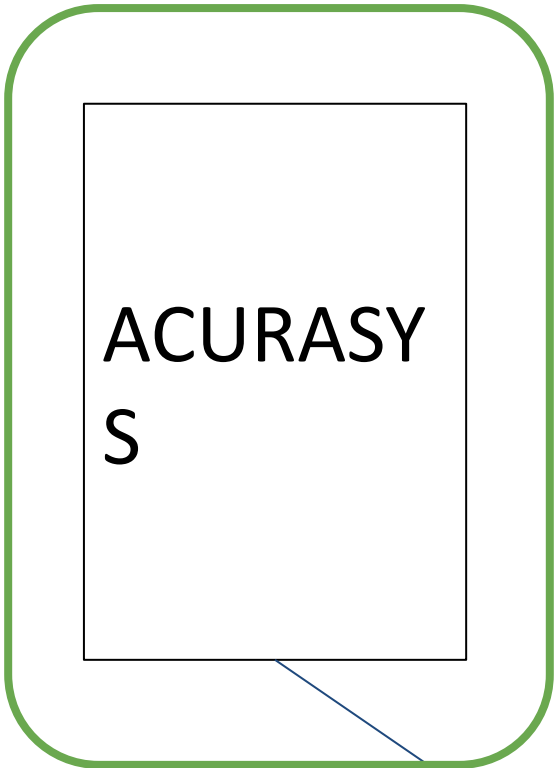
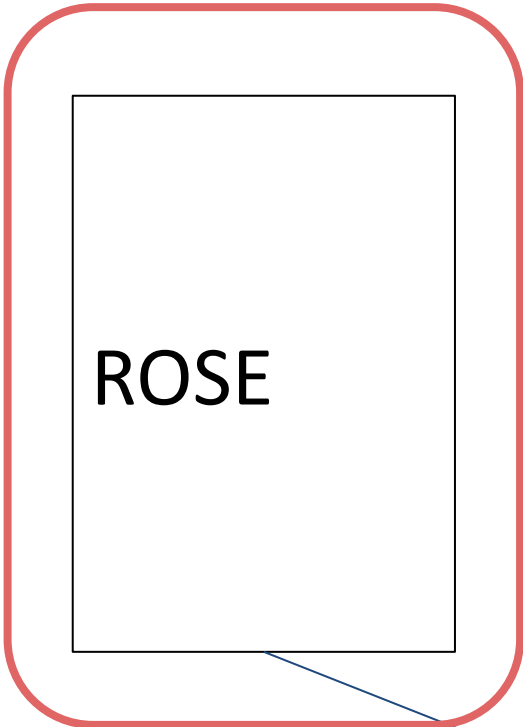
Positive
Effect

IRRELE
VANT

Negative
Effect

Positive
Effect

ARDS related but not Answering



Interesting Outcome from Q2

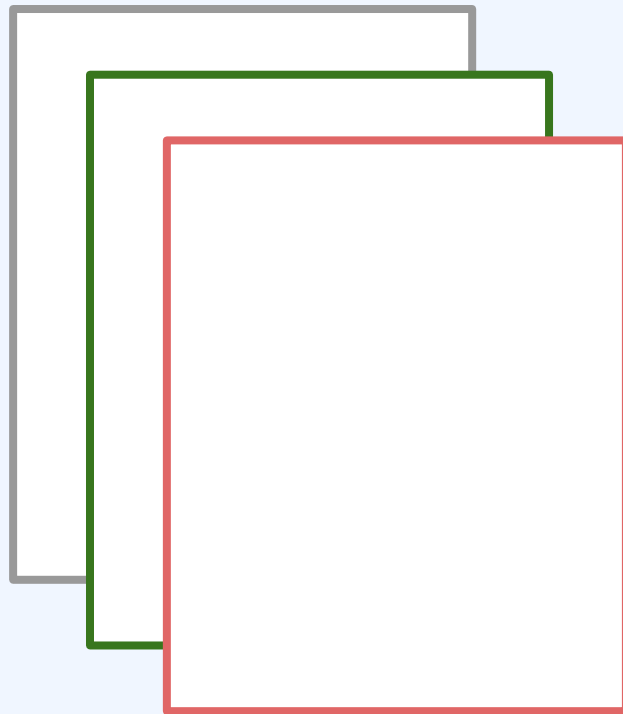
Ground truth: Not only does early neuromuscular blockade **NOT reduce mortality** when using modern light sedation protocols, it was associated with an **INCREASE in ICU-acquired weakness** and **serious ADVERSE** cardiovascular events.

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 from Q2

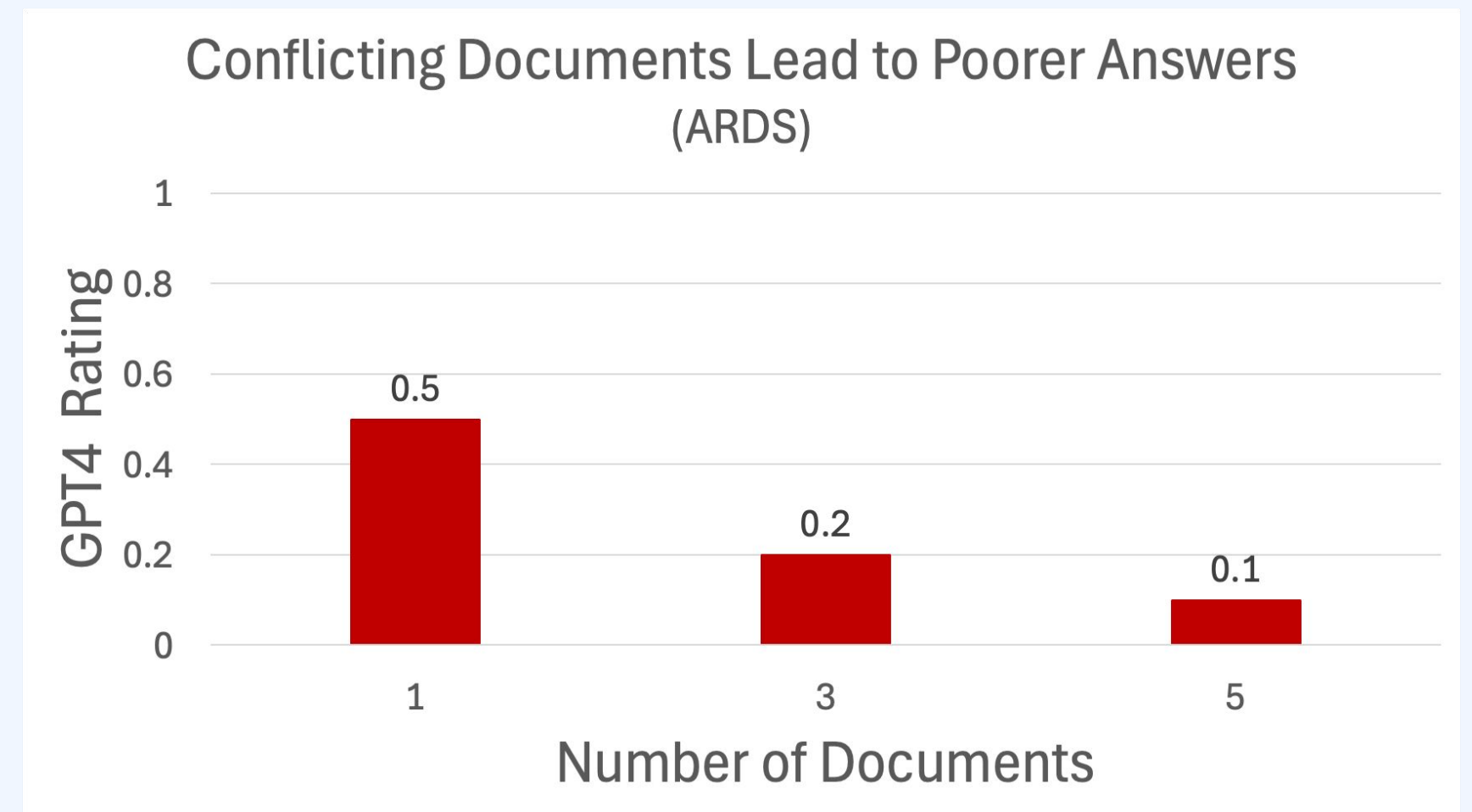
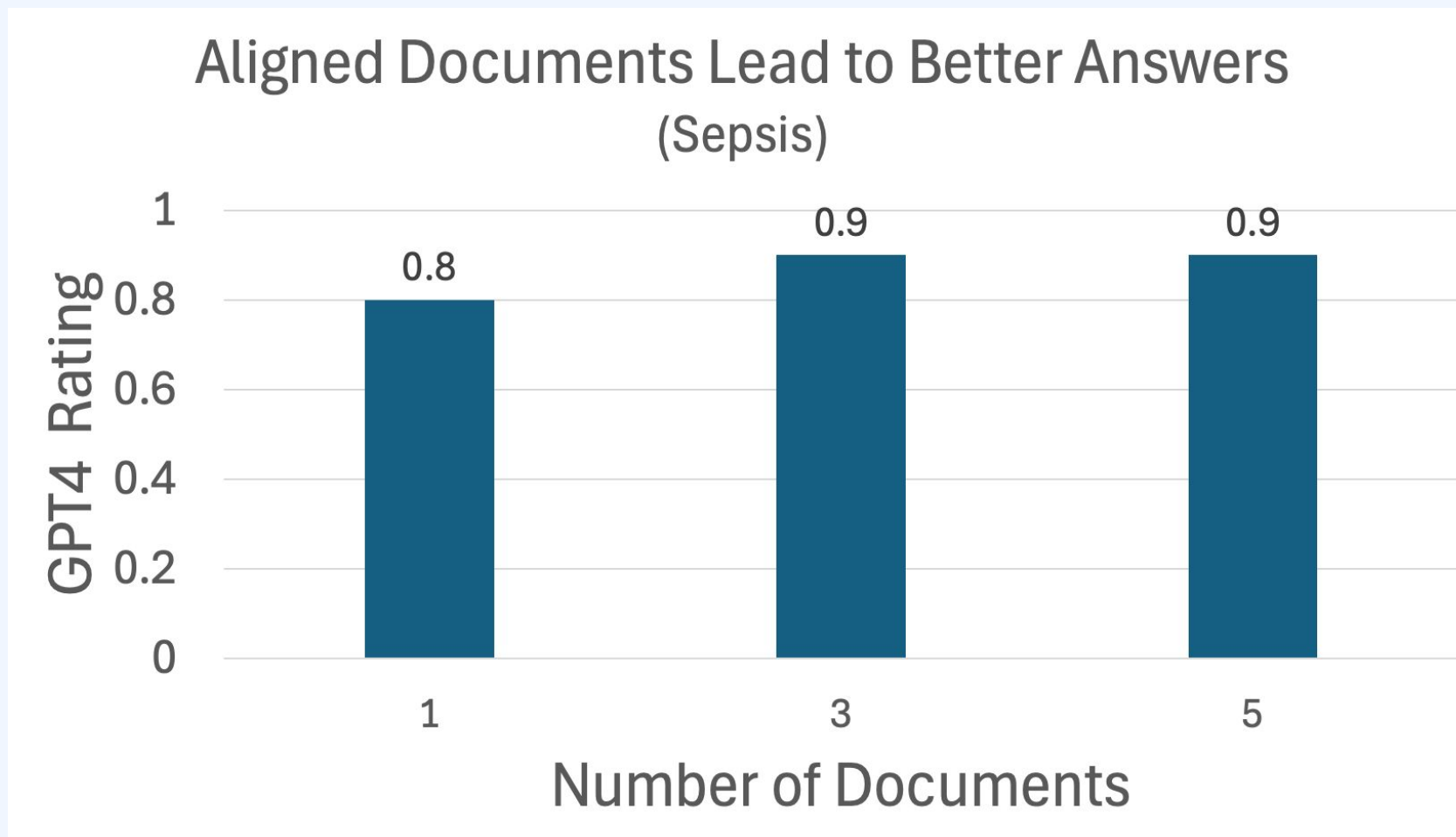
Ground truth: Not only does early neuromuscular blockade **NOT reduce mortality** when using modern light sedation protocols, it was associated with an **INCREASE** in ICU-acquired weakness and **serious ADVERSE** cardiovascular events.

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 input documents point to different answers.



Example result from GraphRAG (1 doc local)

```
python -m graphrag.query \  
--root ./ragtest \  
--method local \  

```

"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?"

The question of whether the early use of continuous neuromuscular blockade with cisatracurium improves mortality in patients with moderate to severe Acute Respiratory Distress Syndrome (ARDS) when used alongside current light sedation protocols is a nuanced one. The data provided does not directly address the specific impact of cisatracurium or any neuromuscular blockade agents on mortality rates in ARDS patients under light sedation protocols. However, insights can be drawn from the broader context of ARDS management and treatment strategies as discussed in the provided reports and studies.

ARDS is a critical condition that requires intensive care, often involving mechanical ventilation to support breathing. The management of ARDS has evolved over time, with strategies such as prone positioning and the use of sedatives and neuromuscular blocking agents being part of the therapeutic arsenal aimed at improving oxygenation and patient outcomes [Data: Reports (2, 0); Entities (18)].

Prone positioning, a method where patients are positioned face down, has been shown to improve oxygenation and potentially reduce mortality in severe ARDS cases. This strategy is highlighted in the PROSEVA trial, which demonstrated significant benefits in terms of reduced 28-day and 90-day mortality rates for patients with severe ARDS who were managed with prone positioning [Data: Entities (3); Sources (0)]. The trial's protocol included the use of neuromuscular blocking agents to facilitate prone positioning, suggesting that these agents play a role in the management of severe ARDS, particularly in conjunction with prone positioning [Data: Sources (1)].

The use of neuromuscular blocking agents, such as cisatracurium, is often considered in ARDS management to improve ventilator synchrony, reduce oxygen consumption, and potentially mitigate ventilator-induced lung injury. However, the specific impact of these agents on mortality, especially when used early and in combination with light sedation protocols, is not explicitly detailed in the provided data. It's worth noting that the approach to sedation has shifted towards lighter sedation protocols in critical care, aiming to reduce sedation-related complications and potentially improve outcomes by facilitating earlier mobilization and reducing the duration of mechanical ventilation [Data: Reports (2)].

Given the complexity of ARDS management and the multifactorial nature of interventions aimed at improving patient outcomes, the direct impact of continuous neuromuscular blockade with cisatracurium on mortality rates within the context of light sedation protocols requires further investigation. Clinical guidelines and treatment strategies continue to evolve based on emerging evidence from randomized controlled trials and meta-analyses. Therefore, while the early use of neuromuscular blocking agents is a component of ARDS management, its efficacy and impact on mortality as part of a comprehensive treatment strategy, including light sedation protocols, should be interpreted within the broader context of individual patient needs and evolving clinical evidence [Data: Entities (18, 1); Reports (0, 2)].

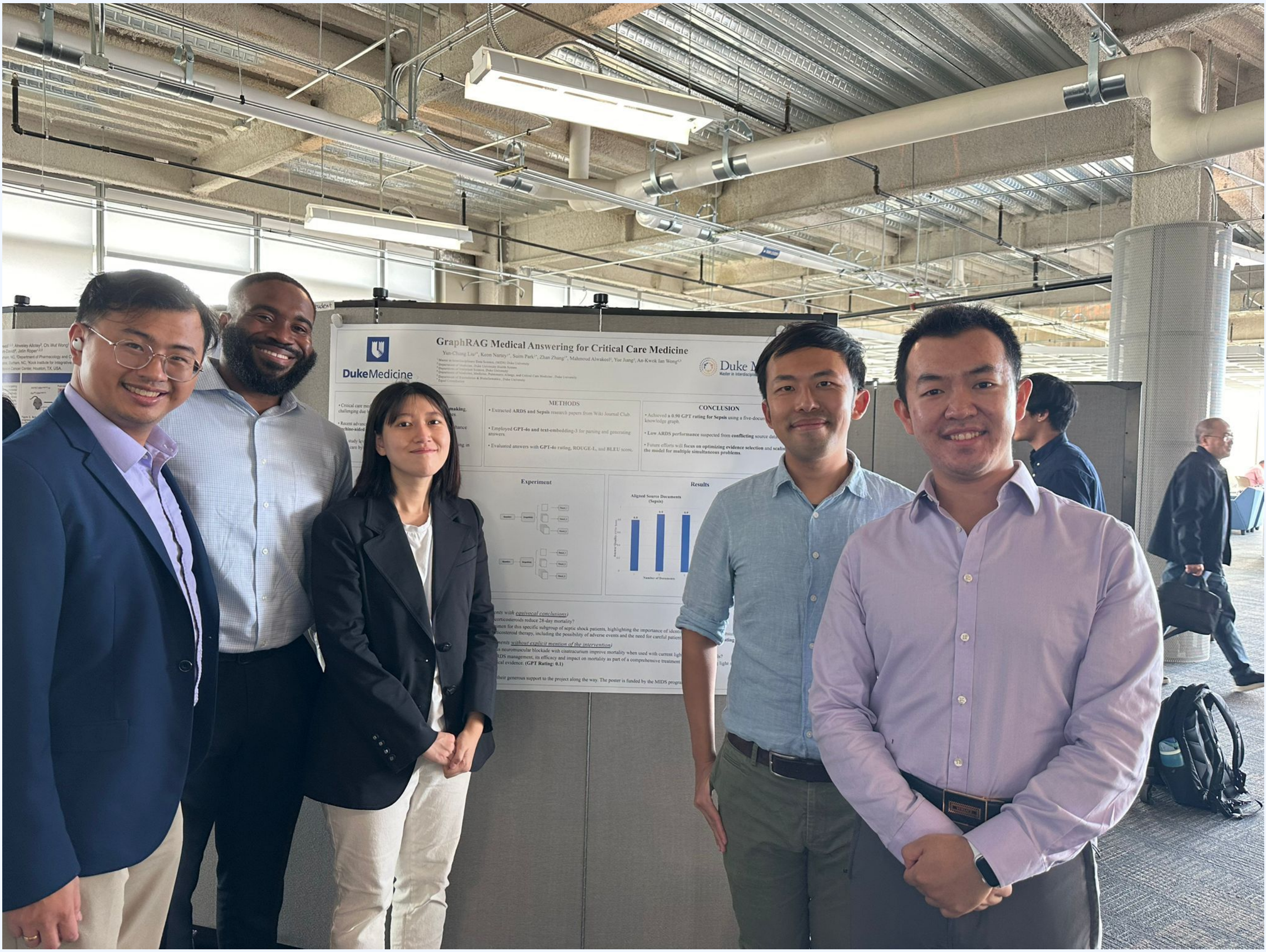
Summary

1. Decision-making should not be done by the model; the model should provide sources where it learned.
2. Conflicting documents confuse the model
 - a. Model weighs input docs equally
 - b. Lack of other metadata: citation, publication year, organizations
3. GraphRAG enables globally summarized contents and pinned sources but RAG which relies solely on semantics and similarity

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

