Developing A Medical Question Answering Model for Duke Department of Medicine

Bob Zhang, Yun-Chung (Murphy) Liu, Keon Nartey, Suim Park

Stakeholder: Alan Wong, MD, PhD (Department of Medicine)
Mentor: Yue Jiang, PhD

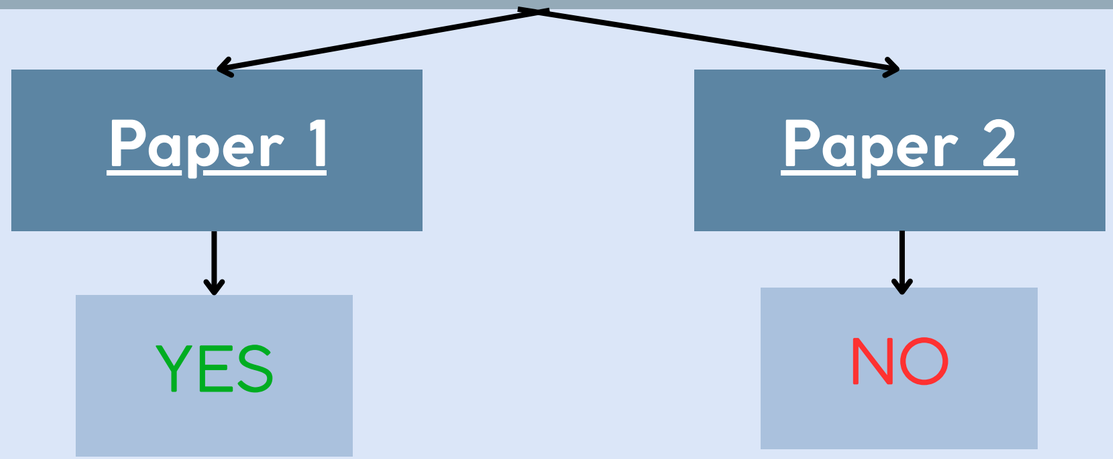
Objective

- **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.

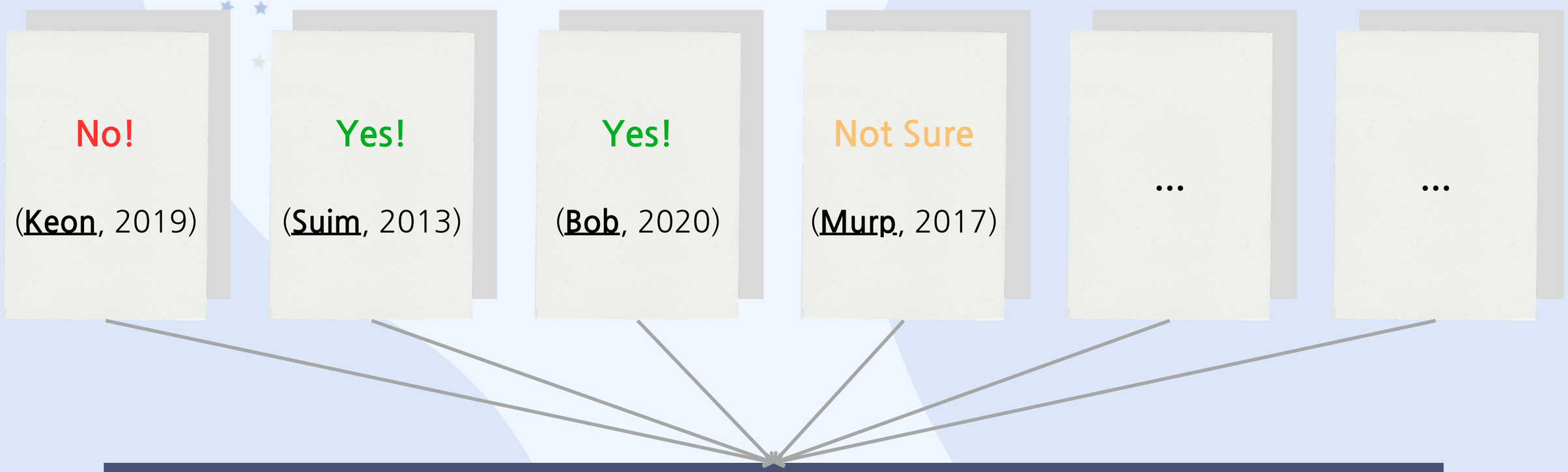
- **Objective**
  - The project aims to develop of an **evidence-based medical question answering model** for critical care medicine.

An Example Question

Does the use of **steroids** on a patient with **sepsis reduces mortality**?



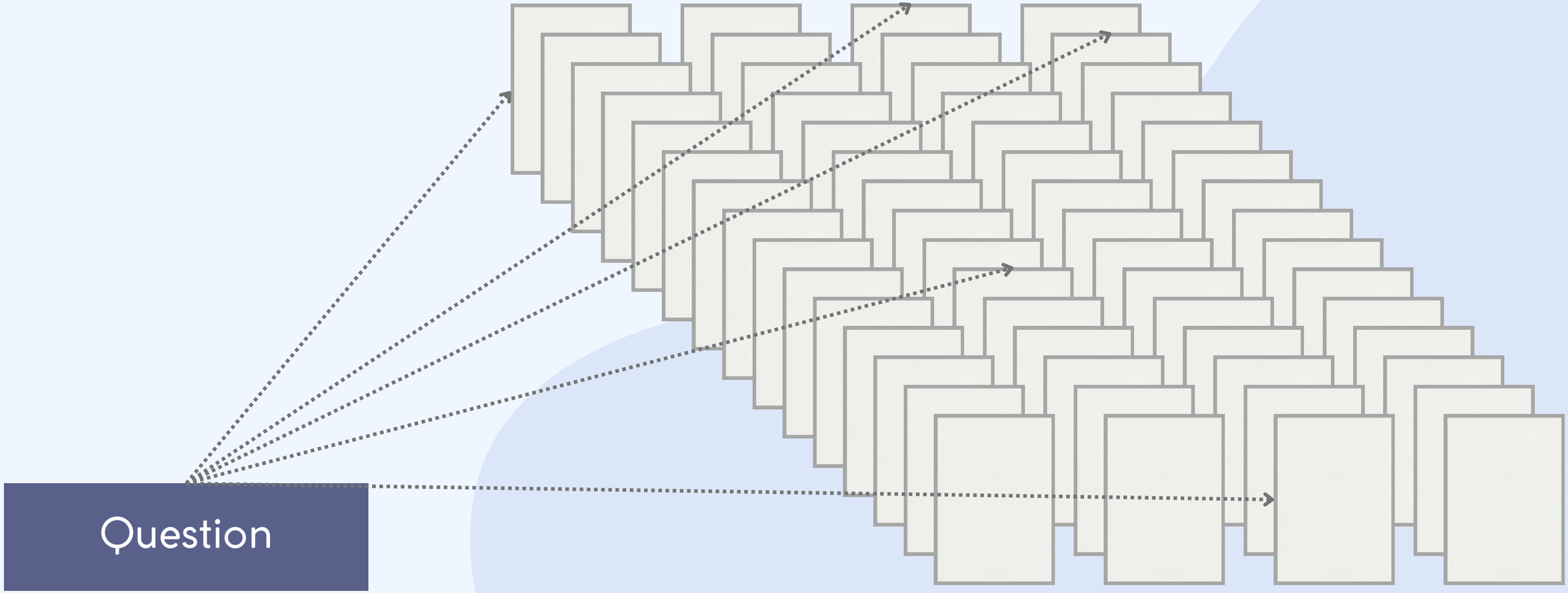
Medical Questions Answered by Research Papers



Does {treatment} induce {outcome} on {medical condition}?

Quantity Counts

What happens when the amount of source document becomes HUGE?



PICO Summary

SUPPORT

Study Design: ...
Study Population: ...
Interventions: ...
Outcomes: ...

Study Design: ...
Study Population: ...
Interventions: ...
Outcomes: ...

NEUTRAL

Study Design: ...
Study Population: ...
Interventions: ...
Outcomes: ...

AGAINST

Study Design: ...
Study Population: ...
Interventions: ...
Outcomes: ...

Medical Question

Does the use **Steroids** being treated with **SEPSIS reduce mortality?**

Relevant Papers

SUPPORT

Study Design: ...
Study Population: ...
Interventions: ...
Outcomes: ...

Study Design: ...
Study Population: ...
Interventions: ...
Outcomes: ...

NEUTRAL

Study Design: ...
Study Population: ...
Interventions: ...
Outcomes: ...

AGAINST

Study Design: ...
Study Population: ...
Interventions: ...
Outcomes: ...

Medical Question

Does the use **Steroids** being treated with **SEPSIS reduce mortality**?

Stance

SUPPORT

Study Design: ...
Study Population: ...
Interventions: ...
Outcomes: ...

Study Design: ...
Study Population: ...
Interventions: ...
Outcomes: ...

NEUTRAL

Study Design: ...
Study Population: ...
Interventions: ...
Outcomes: ...

AGAINST

Study Design: ...
Study Population: ...
Interventions: ...
Outcomes: ...

Medical Question

Does the use **Steroids** being treated with **SEPSIS reduce mortality**?

Example of Model Output

SUPPORT

Study Design: ...
Study Population: ...
Interventions: ...
Outcomes: ...

Study Design: ...
Study Population: ...
Interventions: ...
Outcomes: ...

NEUTRAL

Study Design: ...
Study Population: ...
Interventions: ...
Outcomes: ...

AGAINST

Study Design: ...
Study Population: ...
Interventions: ...
Outcomes: ...

Medical Question

Does the use of **steroids** on a patient with **sepsis reduces mortality**?

Dataset

- **Research Papers**
  - 112 research papers (Medical Conditions)
  - 34 papers from **Wiki Journal Club**, 86 from **PubMed**, and **Google Scholar**)

- **Medical Questions**
  - 36 yes-no questions
  - 12 for **ARDS**, 10 for **sepsis**, 6 for **cardiac arrest**, 7 for **delirium**, 1 for **sepsis and delirium**

Ground Truth Annotation I

Relevance

	Question 1	Question 2	...
Paper 1	**1** (Relevant)	**0** (Irrelevant)	...
Paper 2	**0** (Irrelevant)	**1** (Relevant)	...
...	...	...	...

Stance

	Question 1	Question 2	...
Paper 1	**1 (Supporting)**	**0 (Neutral)**	...
Paper 2	**0 (Neutral)**	**-1 (Against)**	...
...	...	...	...

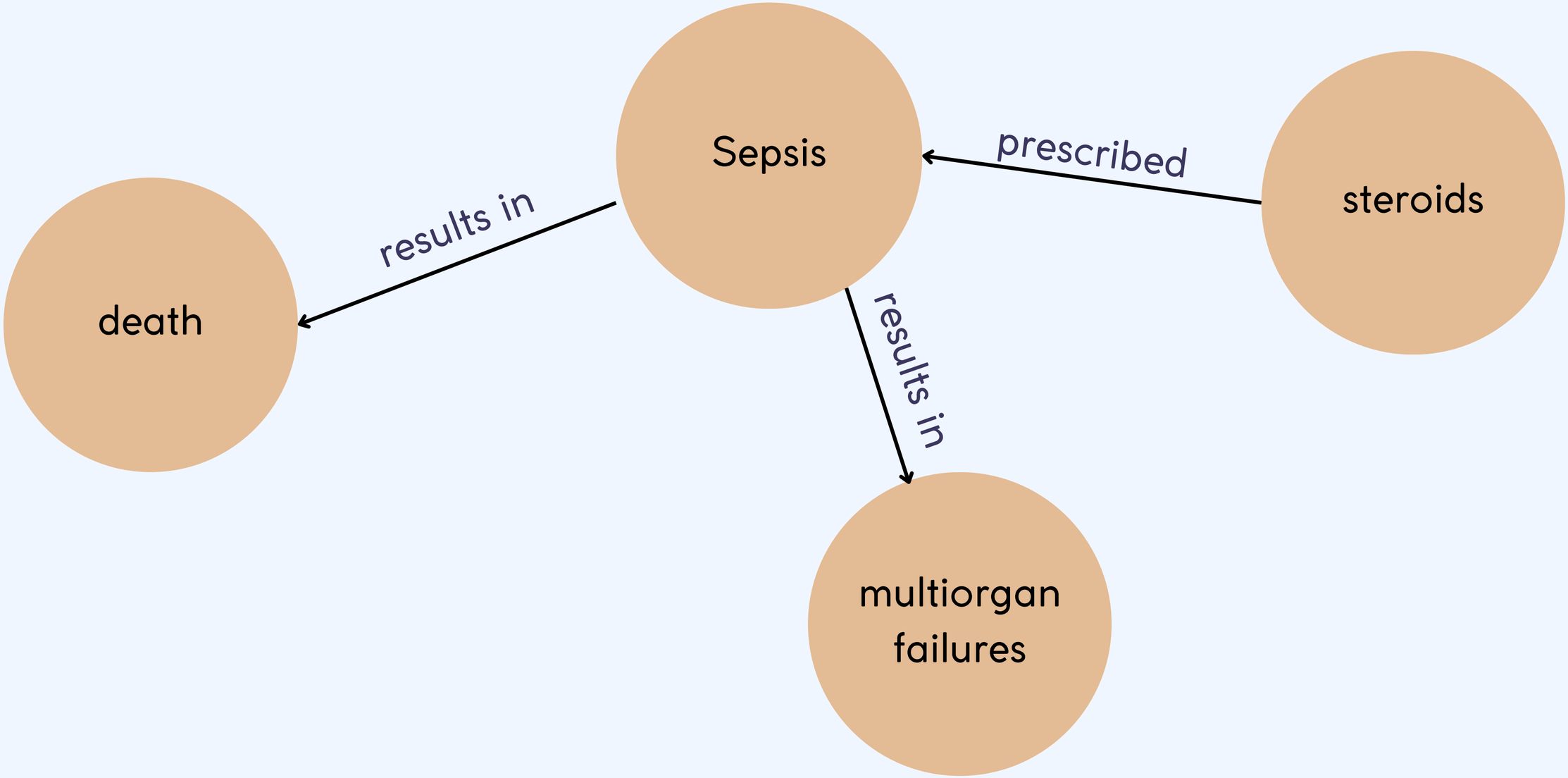
Ground Truth Annotation II

	Study Design	Study Population	...	Outcome
Paper 1	• Multicenter, unblinded, randomized clinical trial • Conducted across 48 hospitals in the United States ...	• Patients with moderate-to-severe ARDS • Criteria included: ◦ Mechanical ventilation ...	...	• 90-day in-hospital mortality rate • 28-day mortality • Ventilator-free days ...

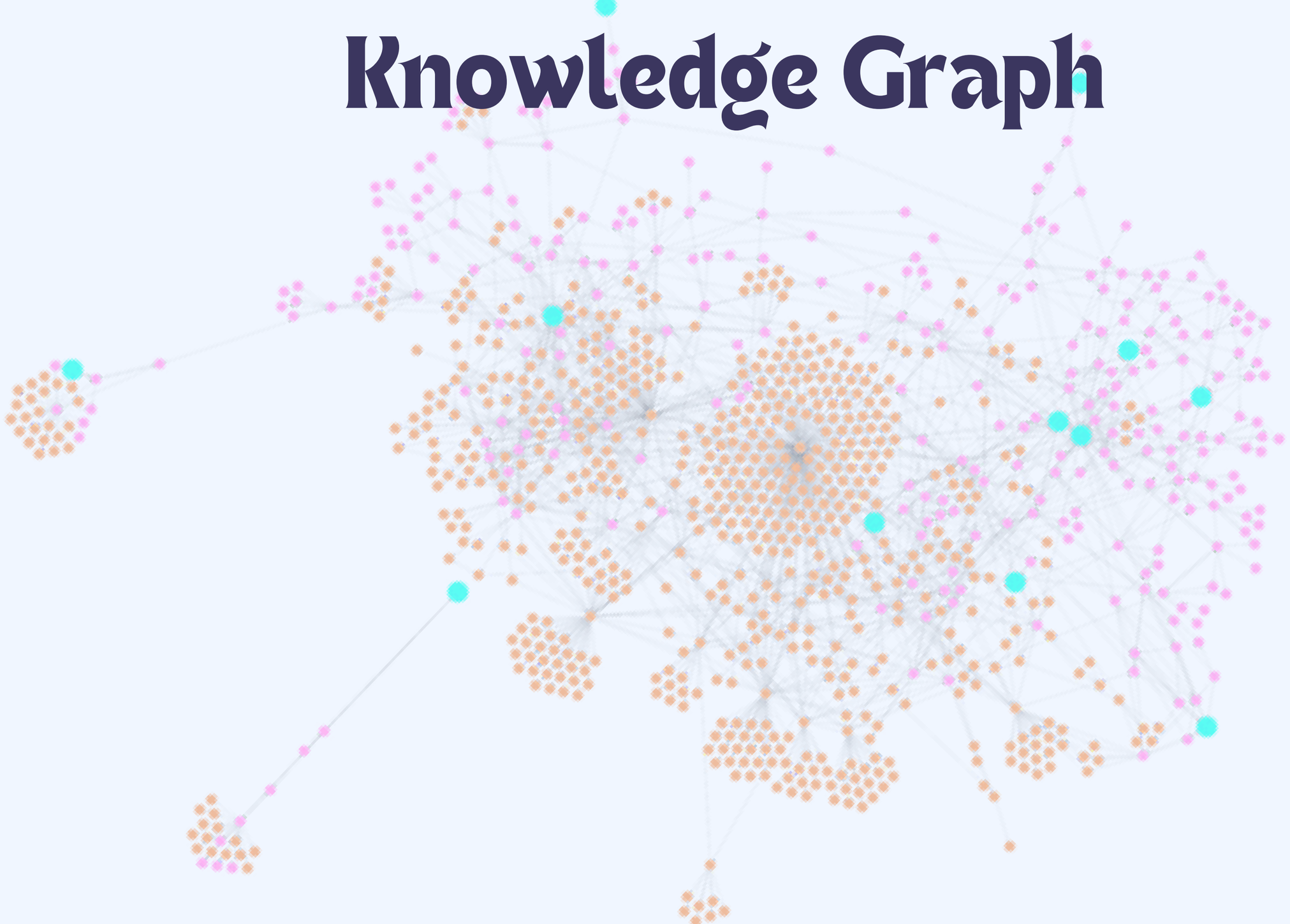
Retrieval Augumented Generation

GraphRAG

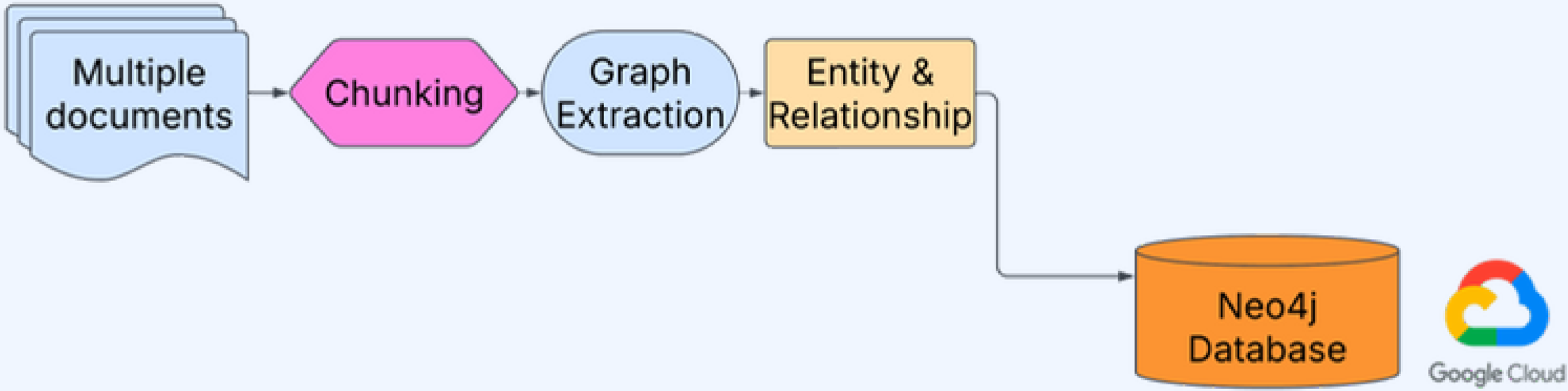
Graph: Entity & Relationships



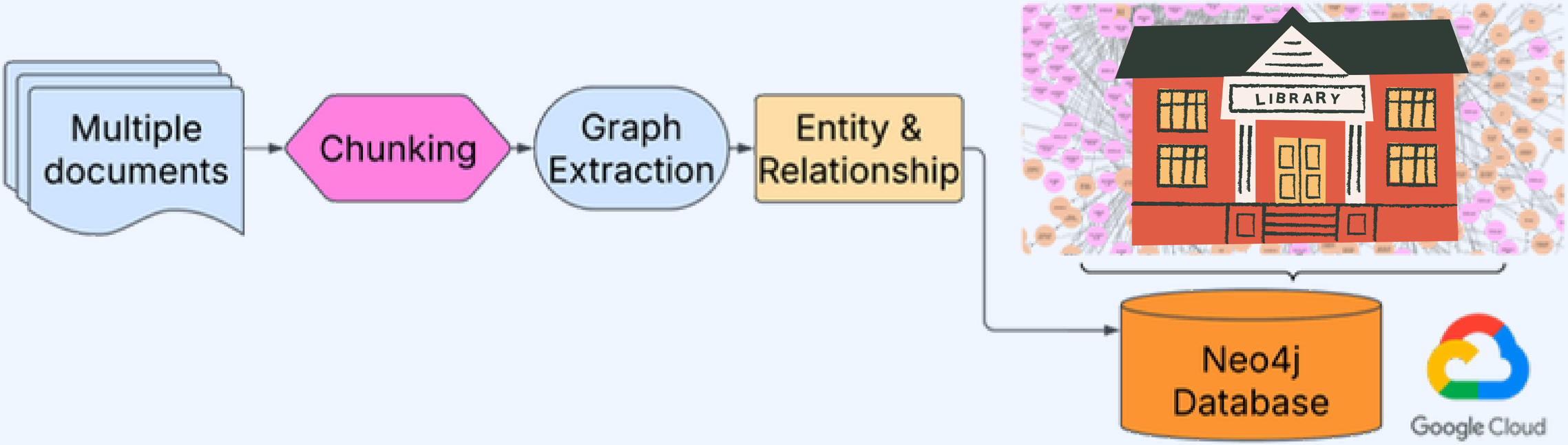
Knowledge Graph



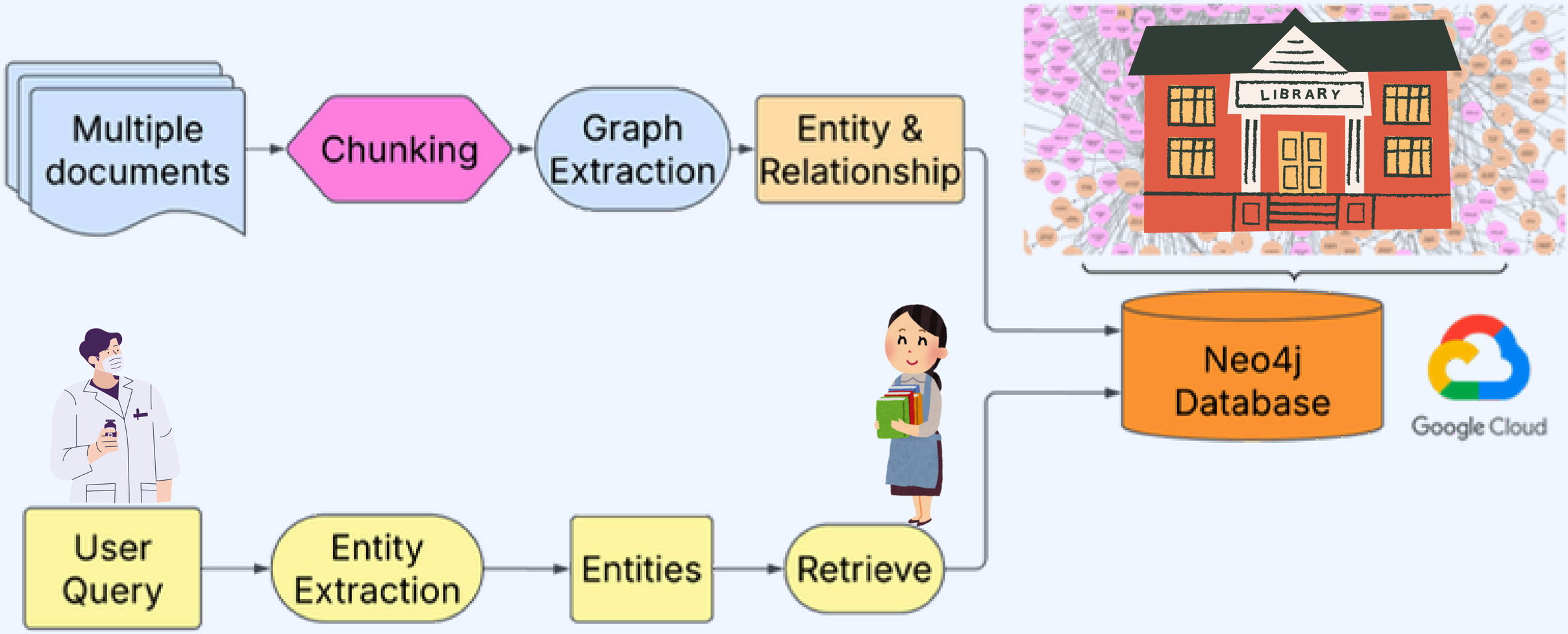
GraphRAG Workflow



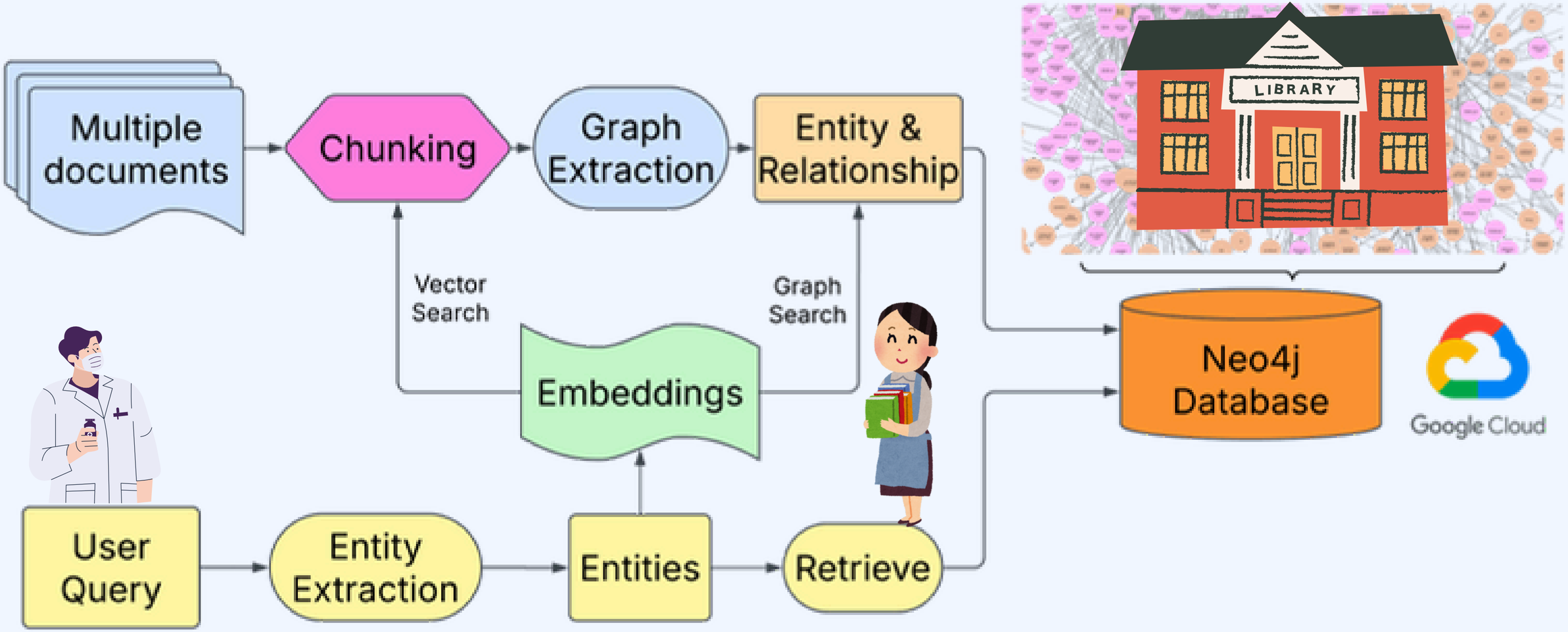
GraphRAG Workflow



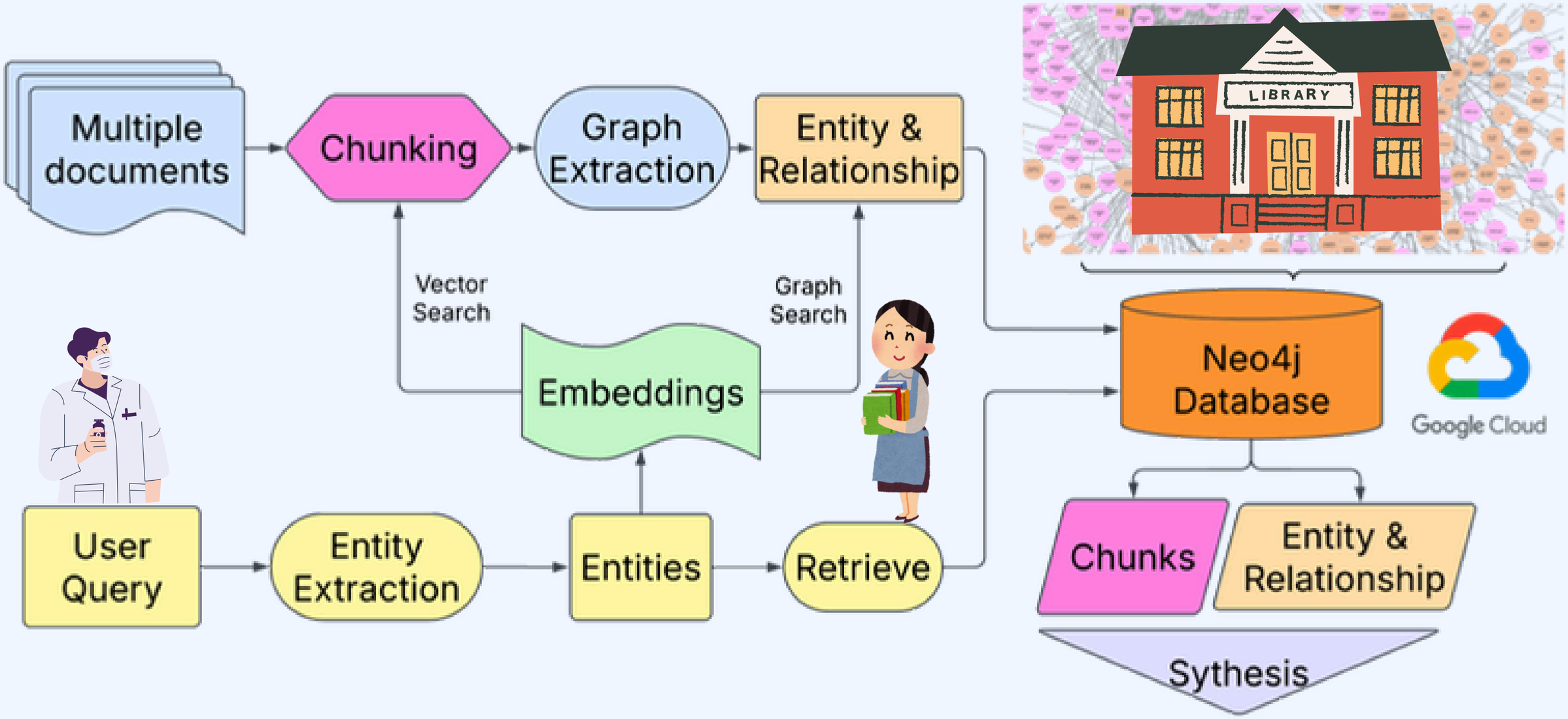
GraphRAG Workflow

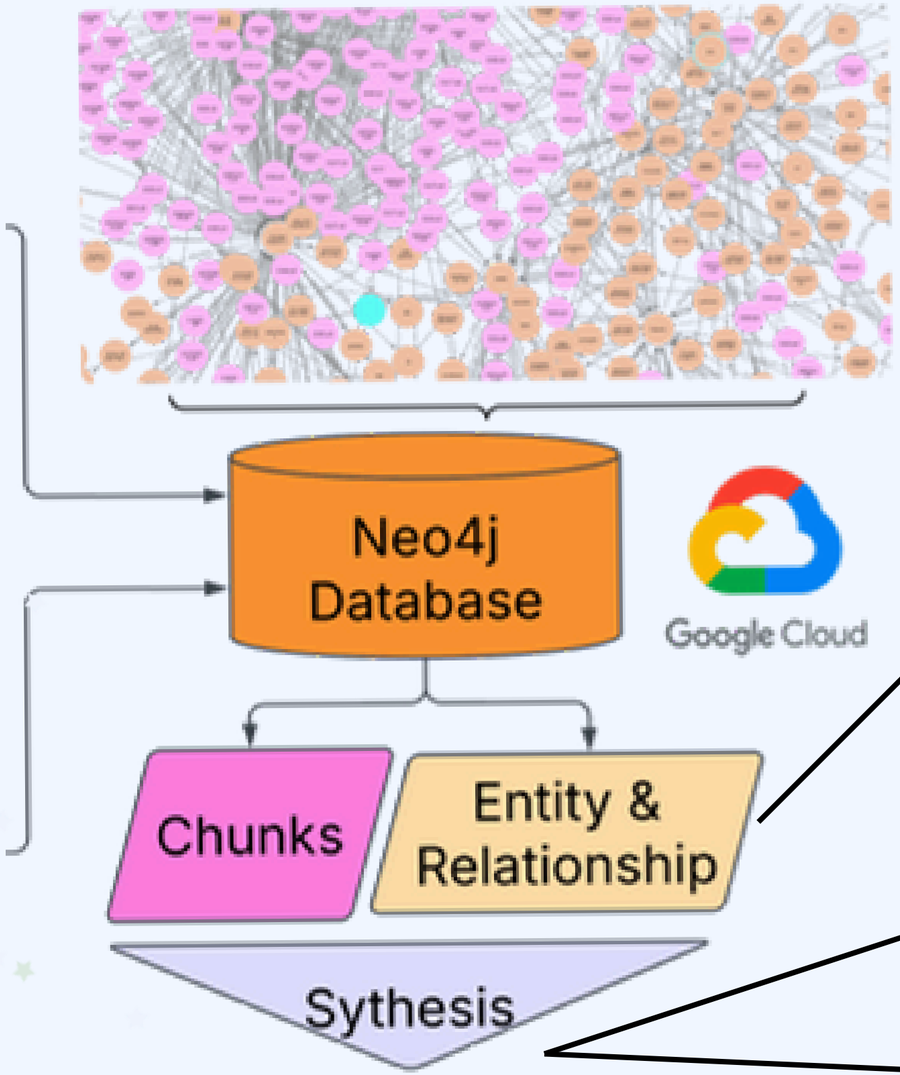


GraphRAG Workflow



GraphRAG Workflow





Relevance

Stance

PICO Summary

Evaluation

- **Relevance and Stance**

Relevance of retrieved papers measured by **accuracy, precision, recall,** and **F-1 score**.

- **PICO Summary**

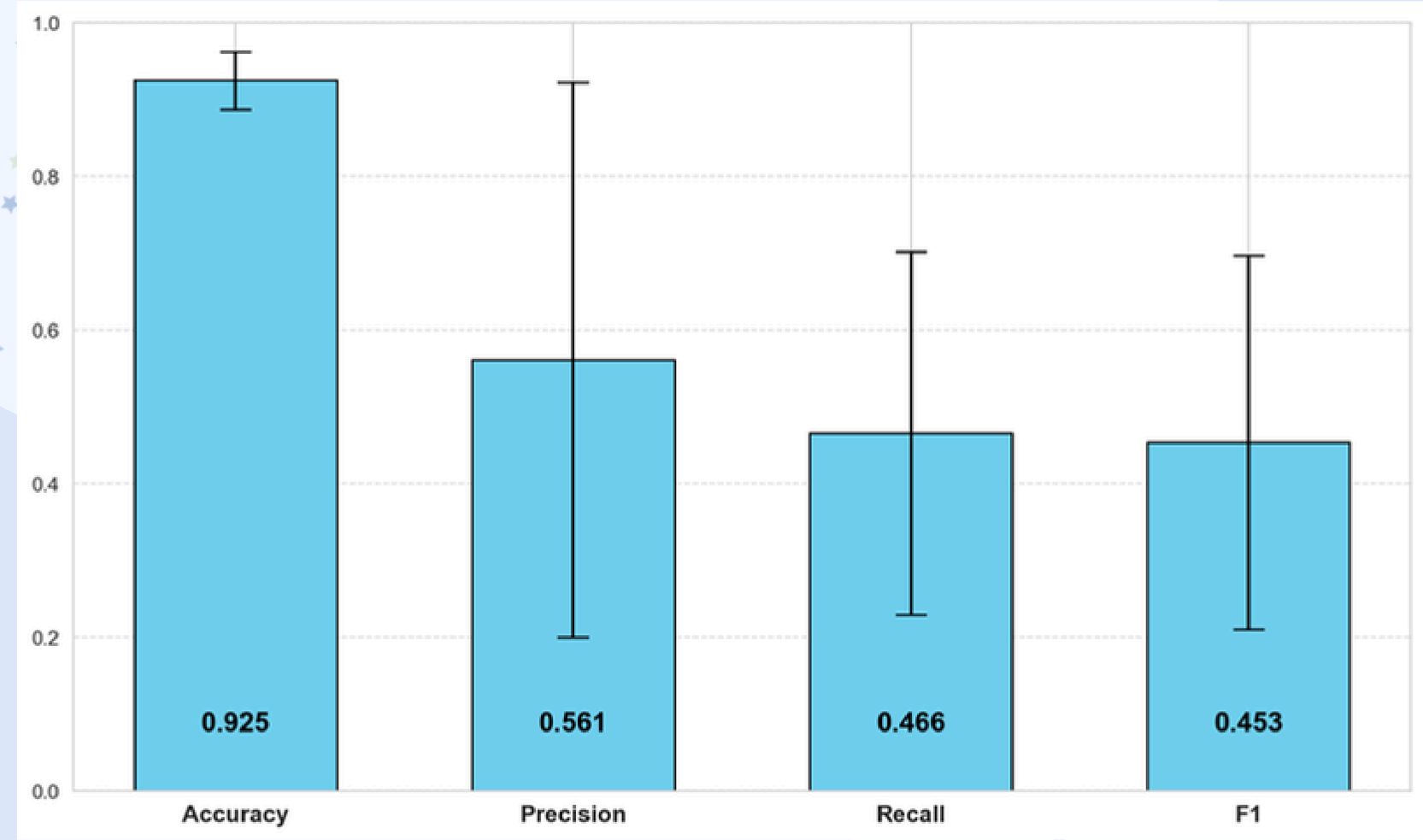
Summaries are evaluated on **factual correctness** and **semantic similarities** against **ground truth** using RAGAS

Examples of Retrieval Performance

Does the use of **steroids** on a patient with **sepsis reduces mortality**?

Annotated Documents	Retrieved Documents	Similarity	Accuracy	Precision	Recall	F1
ANNANE, ADRENAL, HYPRESS, APROCCHSS, PLMALDHSS, CORTICUS, AID-ICU, SEDCOM, MIND-USA, NEvsVP, SPICE-III	**ANNANE, ADRENAL, HYPRESS, APROCCHSS, PLMALDHSS, CORTICUS**, TRISS, SS3vsSS2, FRESHS	0.869	0.929	0.667	0.545	0.600

Relevance Evaluation



- Mean value of all metrics from 36 questions

Stance Evaluation

	Accuracy
Results - Stance	**0.875**

- Mean value of all metrics from 36 questions
- Stadnard Deviation = 0.0215

Factuality Metrics

Ground Truth

The Eiffel Tower is located in Paris.

It has a height of 1000ft.

Generated Summary

The Eiffel Tower is located in Paris.

Factual Correctness

Precision: 1.0

(TP/ TP+FP)

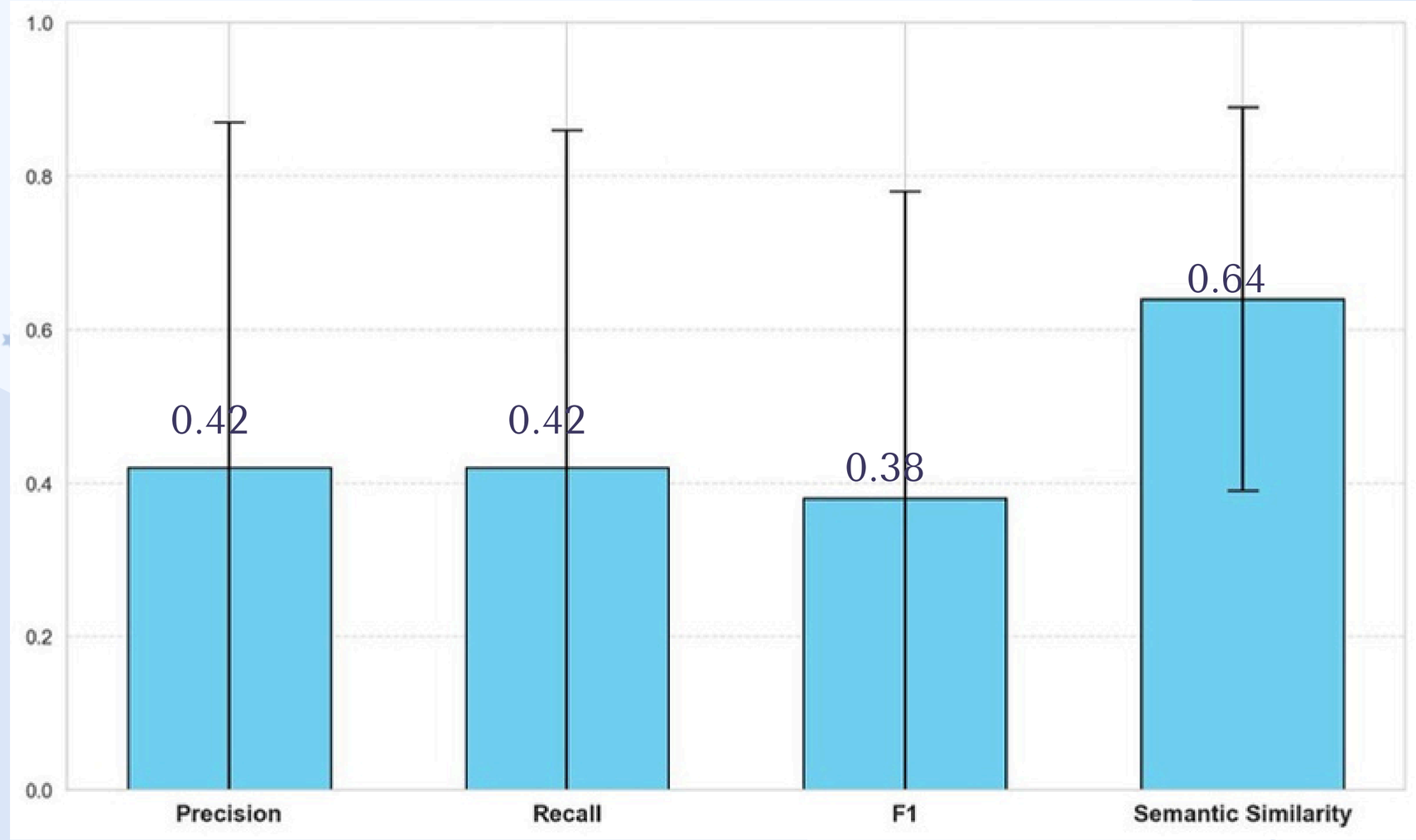
Recall: 0.5

(TP/ TP+FN)

Semantic Similarity: 0.8

(cosine similarity of embeddings)

PICO Summary Evaluation



- From 25 synthesized summaries

Example of PICO Summry

	Human Annotation	Model Synthesis
Study Design	Pragmatic, **double-blind, randomized controlled trial**	**Double-blind,** parallel-group, **randomized, controlled trial**
Intervention	**Hydrocortisone** 200 mg/24 hrs via **continuous** IV **infusion** for up to 7 days or until ICU discharge/death	**Continuous infusion** of **hydrocortisone**
Population	Adults ≥18 years on **mechanical ventilation .**	3800 patients with **septic shock** undergoing **mechanical ventilation**

Final Model Output

SUPPORT

Study Design: randomization via computer-generated sequence, double-blind
Study Population: 299 patients, mean age 63, diverse ethnicity
Interventions: Hydrocortisone 50mg IV every 6 hours, fludrocortisone 50 mg orally daily for 7 days.
Outcomes: Reduced 28-day mortality (53% vs 63%, p=0.02)

Study Design: randomization via central web-based system, double-blind
Study Population: 3800 patients, mean age 62
Interventions: Hydrocortisone 200 mg/day IV hours for 7 days.
Outcomes: Reduced time to shock resolution (3 vs 4, p<0.01)

NEUTRAL

Study Design: randomization via computer-generated sequence, double-blind
Study Population: 380 patients, mean age 60
Interventions: Hydrocortisone 200mg/day IV for 5 days
Outcomes: No significant difference in shock development (p=0.06)

AGAINST

Study Design: randomization via computer-generated sequence, double-blind
Study Population: 499 patients, mean age 65
Interventions: Hydrocortisone 50mg IV every 6 hours for 5 days
Outcomes: No significant difference in 28-day mortality (34.4% vs 31.5%, p=0.69)

Medical Question

Does the use of **steroids** on a patient with **sepsis reduces mortality**?

Summary

- ## What the Clinent Wanted

  A tool to facilitate informed clinical decision making as huge number of potential reference documents exist.

- ## Our Proposed Solution

  A model capable of retrieving relevant documents and **synthesizing important information** of the retrieved research papers.

- ## Impact

  Our model **can save physicians time searching for references when making clinical decisions, leaving more time for patient care.**

Acknowledgement

We would like to **express our most sincere gratitude** for **Dr. A Ian Wong**, our stakeholder, **Prof Yue Jiang**, our mentor, for their guidance. **Mahmoud Alwakeel**, for generously helping out with his clinical and coding expertise, **Prof Gregory Herschlag, Prof Andrea Lane** for organizing the capstone projects , **Sarah Martin, Shanon Jacobs**, and **Tadja Evans** for thier support along the process.

Thank you for your attention