

Chart Retrieval for Arguments

Master Thesis - Defense

Shashi Sharma - 124269

Digital Engineering
Faculty of Media

19 November 2024

First Referee: Prof. Dr. Benno Stein

Second Referee: Prof. Dr. - Ing. Volker Rodehorst

Supervisor: Dr. Johannes Kiesel

01	Introduction	2 min
02	Retrieval-Augmented Generation (RAG)	2 min
03	Methodology	7 min
04	Result and Analysis	8 min
05	Conclusion	1 min

01 Introduction

Motivation

Application

Research Questions

Q. Is society more accepting of transgender people?

GPT-4o Answer

The acceptance of transgender people in society has been gradually increasing in many parts of the world, though the degree of acceptance varies widely depending on region, culture, politics, and community. Here are some key trends and factors that shape the current state of acceptance:

1. Legal Protections and Rights

- **Progress in Laws and Policies:** Many countries have enacted anti-discrimination laws, making it illegal to discriminate based on gender identity. In some places, there are laws allowing for legal gender recognition without the need for medical interventions.
- **Healthcare Access:** Access to gender-affirming healthcare has also improved in several countries, with medical guidelines evolving to support transgender individuals' needs.

2. Cultural Shifts and Visibility

- **Increased Representation:** Transgender people are becoming more visible in media, entertainment, politics, and other public spheres. High-profile transgender individuals, like actors, activists, and athletes, help bring awareness and normalize trans identities.
- **Media Coverage:** Positive and nuanced media portrayals have helped many people understand transgender experiences, contributing to wider acceptance.
- **Youth and Generational Change:** Younger generations generally express more acceptance of gender diversity. This shift reflects a broader, more flexible understanding of gender identity.

3. Backlash and Resistance

- **Political Polarization:** Some regions have seen political movements or laws targeting transgender rights, particularly around issues like bathroom access, sports participation, and healthcare for transgender youth.
- **Social Divides:** Acceptance can vary significantly between rural and urban areas, as well as between political and religious groups. In places where traditional gender norms are highly valued, acceptance can be slower.

4. Workplace and School Inclusion

- **Inclusion Initiatives:** Many companies, universities, and schools are working to create more inclusive environments, with policies around gender-neutral bathrooms, chosen names and pronouns, and non-discrimination practices.
- **Training and Awareness:** Diversity and inclusion training that addresses gender identity is increasingly common, helping more people understand transgender issues and respect transgender identities.

5. Social Media and Online Communities

- **Community Support:** Online spaces allow transgender individuals to connect, find support, and advocate for their rights. Social media has amplified transgender voices and helped educate the public.
- **Misinformation:** However, social media can also be a source of misinformation and transphobia, leading to misunderstandings and reinforcing stereotypes.

6. Intersectional Factors

- **Race, Class, and Other Identities:** Transgender people who are also part of marginalized racial, ethnic, or socioeconomic groups can face additional layers of discrimination. The acceptance they experience may depend on how these intersecting identities are viewed within society.

In Summary

In general, many parts of the world are moving toward greater acceptance of transgender people, with increased visibility, legal protections, and community support. However, backlash and opposition still exist, and progress is uneven. The direction is positive overall, with more individuals, organizations, and countries recognizing transgender rights and working toward inclusivity. But challenges and barriers remain, often influenced by cultural, political, and religious dynamics.

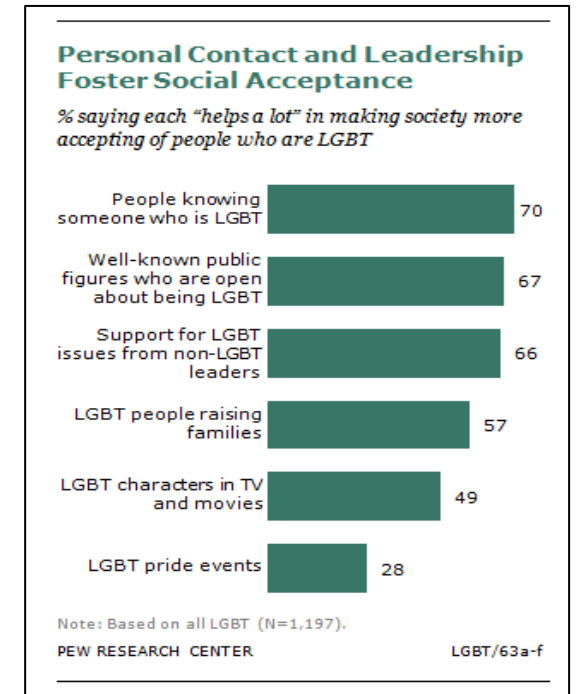
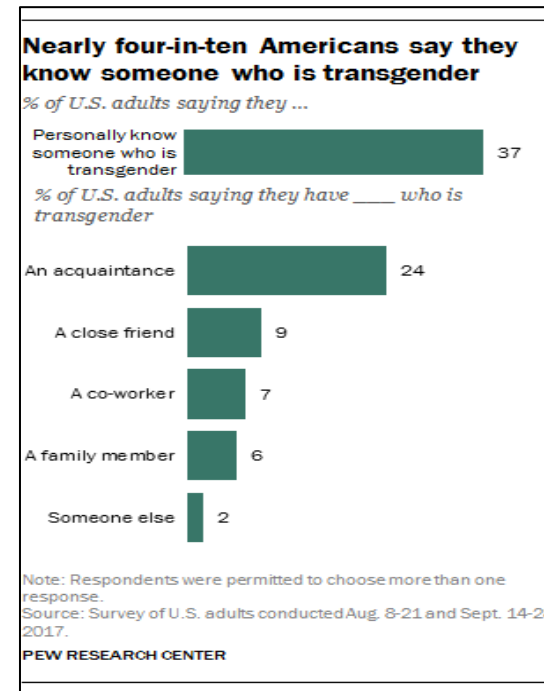
Q. Is society more accepting of transgender people?

GPT-4o Answer (Using Chart Retrieval Framework)

Yes, society appears to be more accepting of transgender people.

According to the data, **37% of U.S. adults know someone who is transgender**, with connections ranging from **acquaintances (24%)** to **close friends (9%)** and **family members (6%)**. This aligns with broader research showing that **personal relationships drive support for marginalized groups**. Among LGBT adults, **70% believe that knowing someone who is LGBT greatly improves societal acceptance**. **Visibility also plays a key role—67% say openly LGBT public figures help**, while **66% highlight support from non-LGBT leaders**. Other influences include **LGBT people raising families (57%)** and **representation in media (49%)**, though **pride events are seen as less impactful (28%)**.

In summary, while progress has been made, the difference in acceptance levels between those who know a transgender person and those who don't shows the need for continued advocacy and education to foster broader societal acceptance.



Application

- Analytics, finance, and e-commerce etc.. (where data driven decision-making is critical).

Research Questions

- How effective are the retrieval methods in addressing a given query?
- How reliable are the generated answers for a given query?

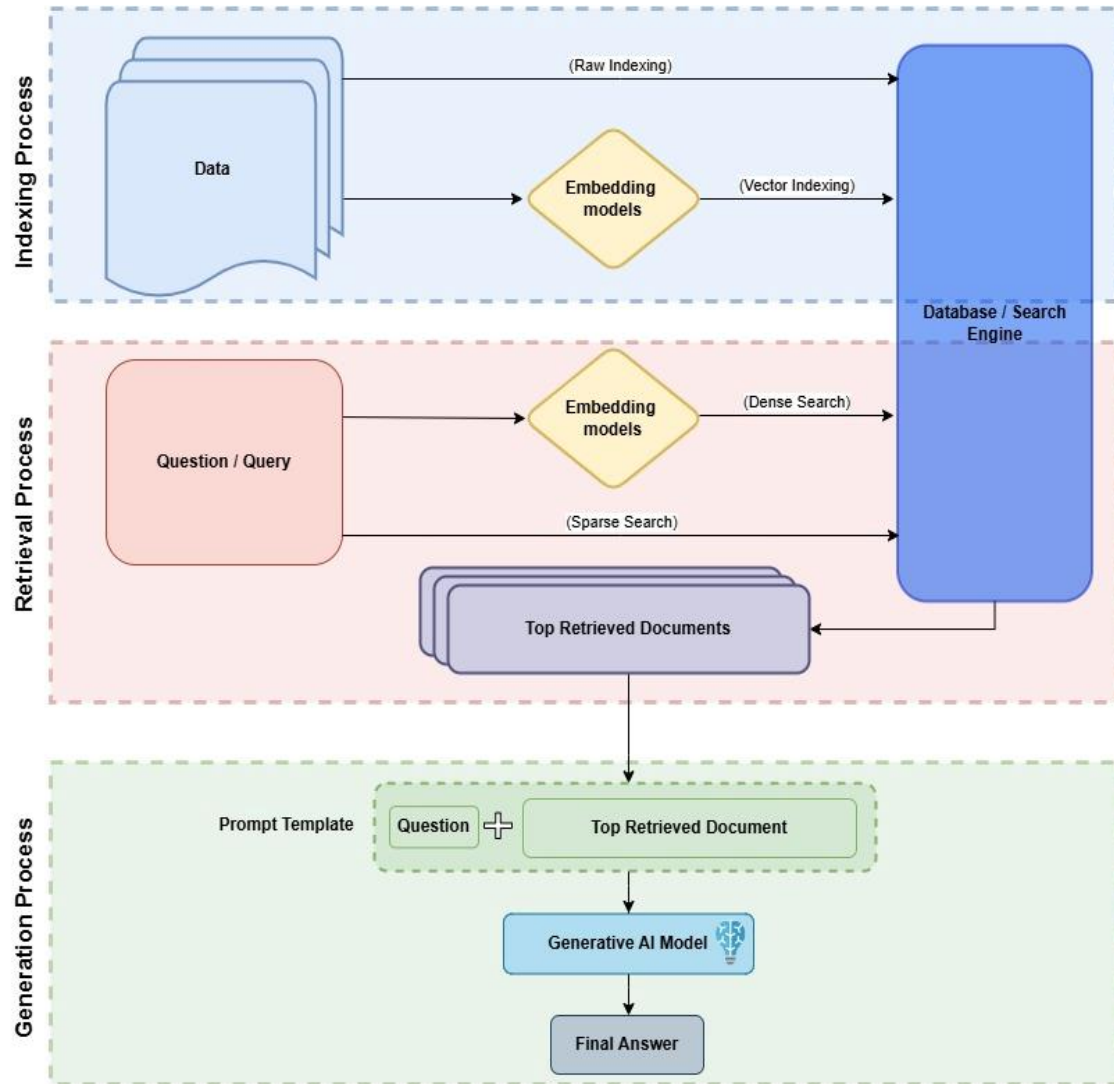
02 Retrieval-Augmented Generation (RAG)

What is RAG?

Components of RAG

Advantages of RAG

Retrieval-Augmented Generation (RAG)



Indexing involves

- Data Loading.
- Data Chunking, and Data embedding.
- Storing the data in the database.

Advantages of RAG

- Generate more accurate answer.
- Reduce hallucinations.
- Enhanced Transparency.

03 Methodology

Chart Retrieval Framework

Dataset and Queries

Indexing Process

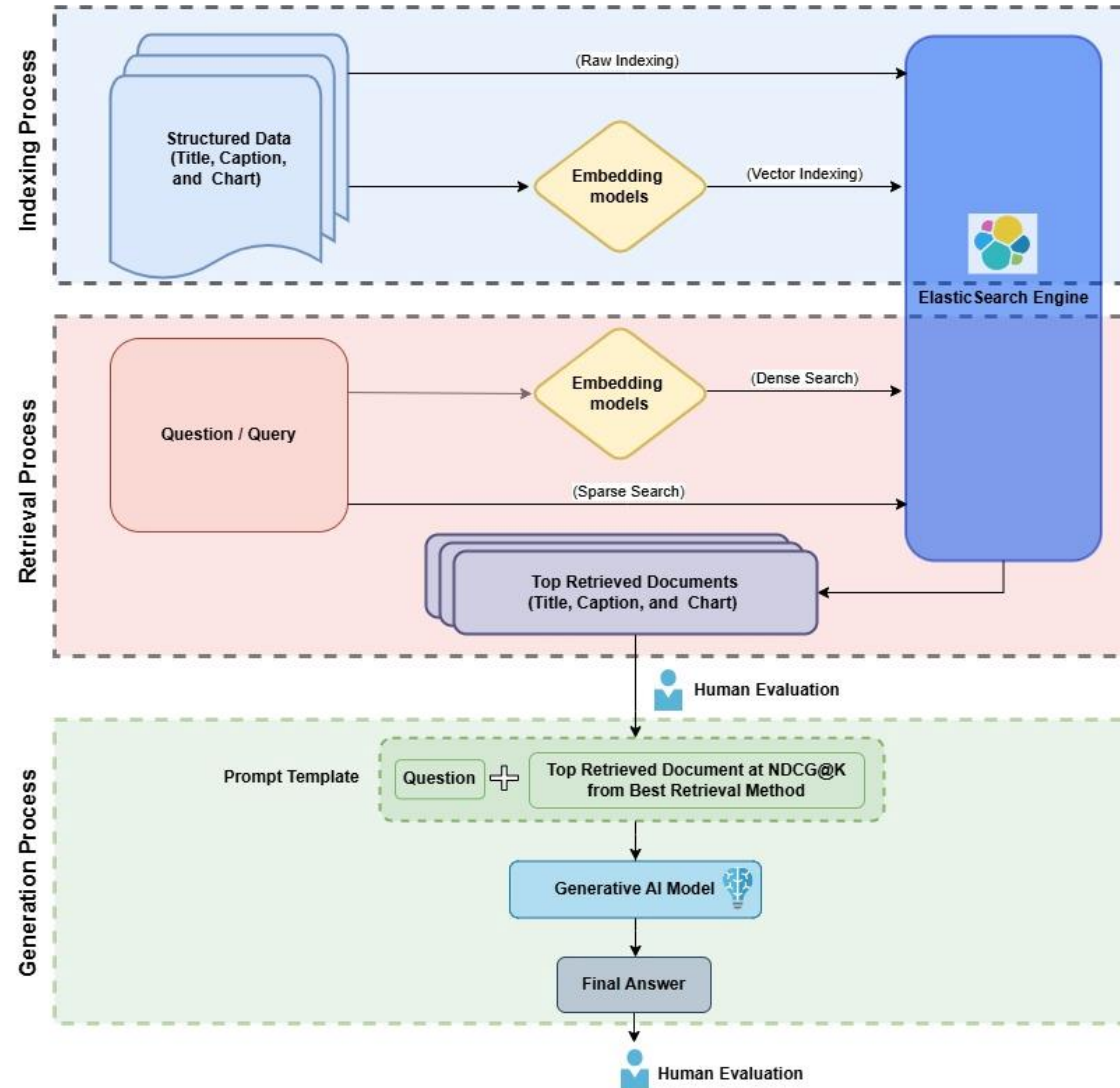
Retrieval Process

Generation Process

Evaluation Process

Methodology

Chart Retrieval Framework



Front End Functions:

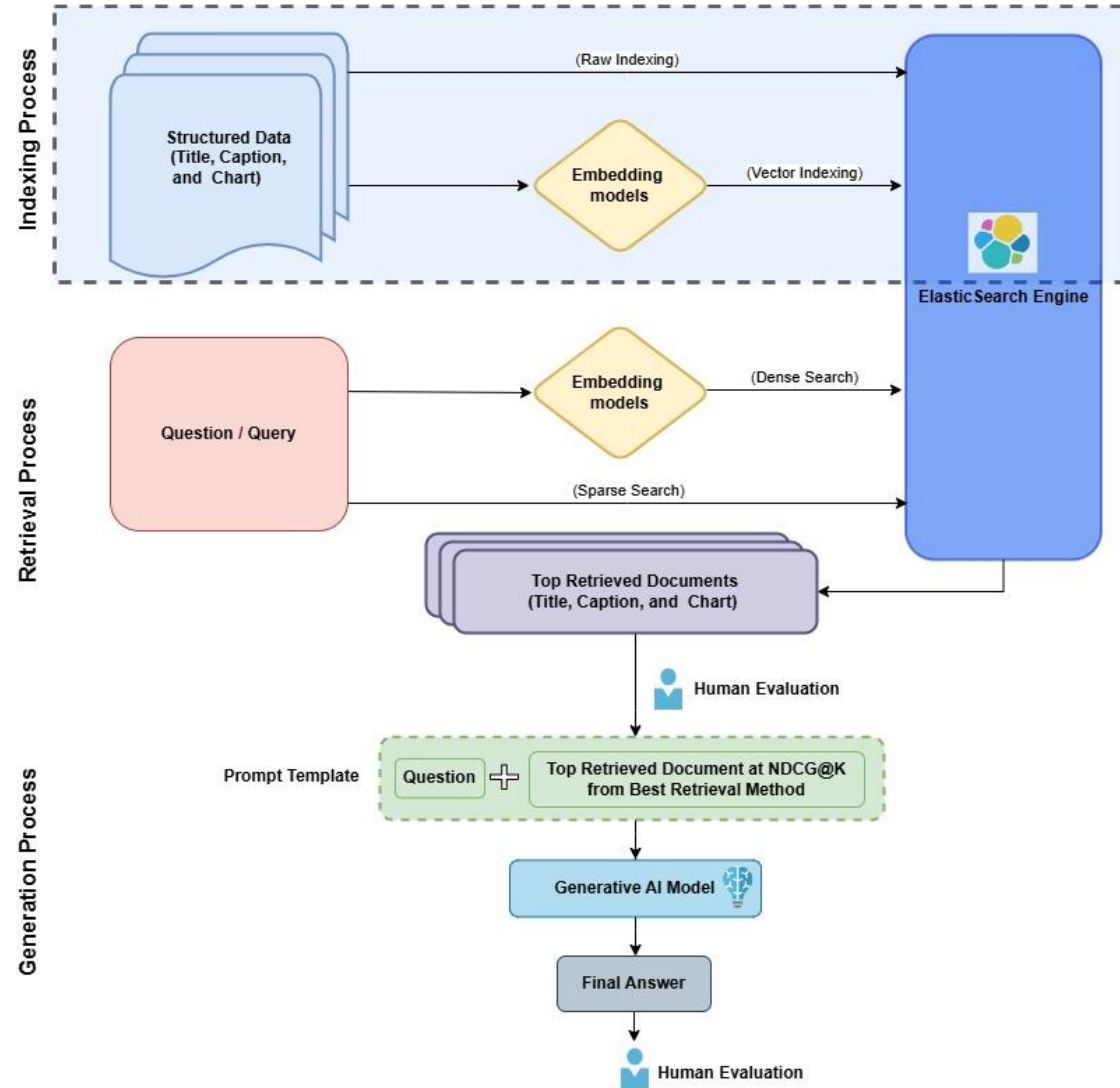
- Query Selection and input.
- Results Display.
- Human Evaluation

Back End Functions:

- Model Integration.
- Search & Retrieval, and Generation.
- Metric-Based Evaluation.

Methodology

Dataset and Queries



- **Structured tabular** form containing **29,354** charts.
- Source from **Statista.com**, and **Pew Research Center**.
- **Chart types:** Bar, line, Area, Pie, and Table etc.
- **Topics:** Politics, Economy, Social and Demographic trends, Internet & technology etc.

Queries (Total of 60 Queries)

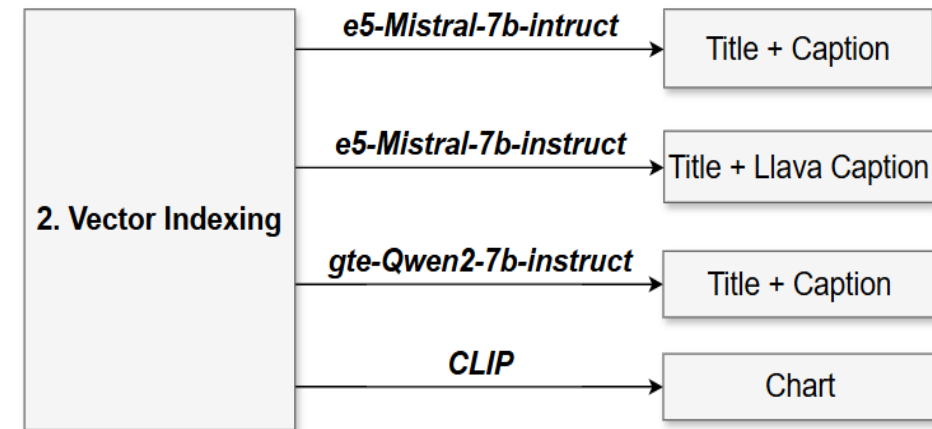
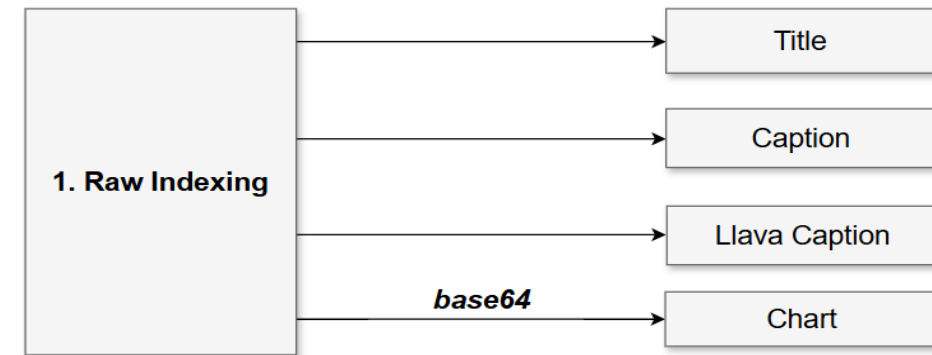
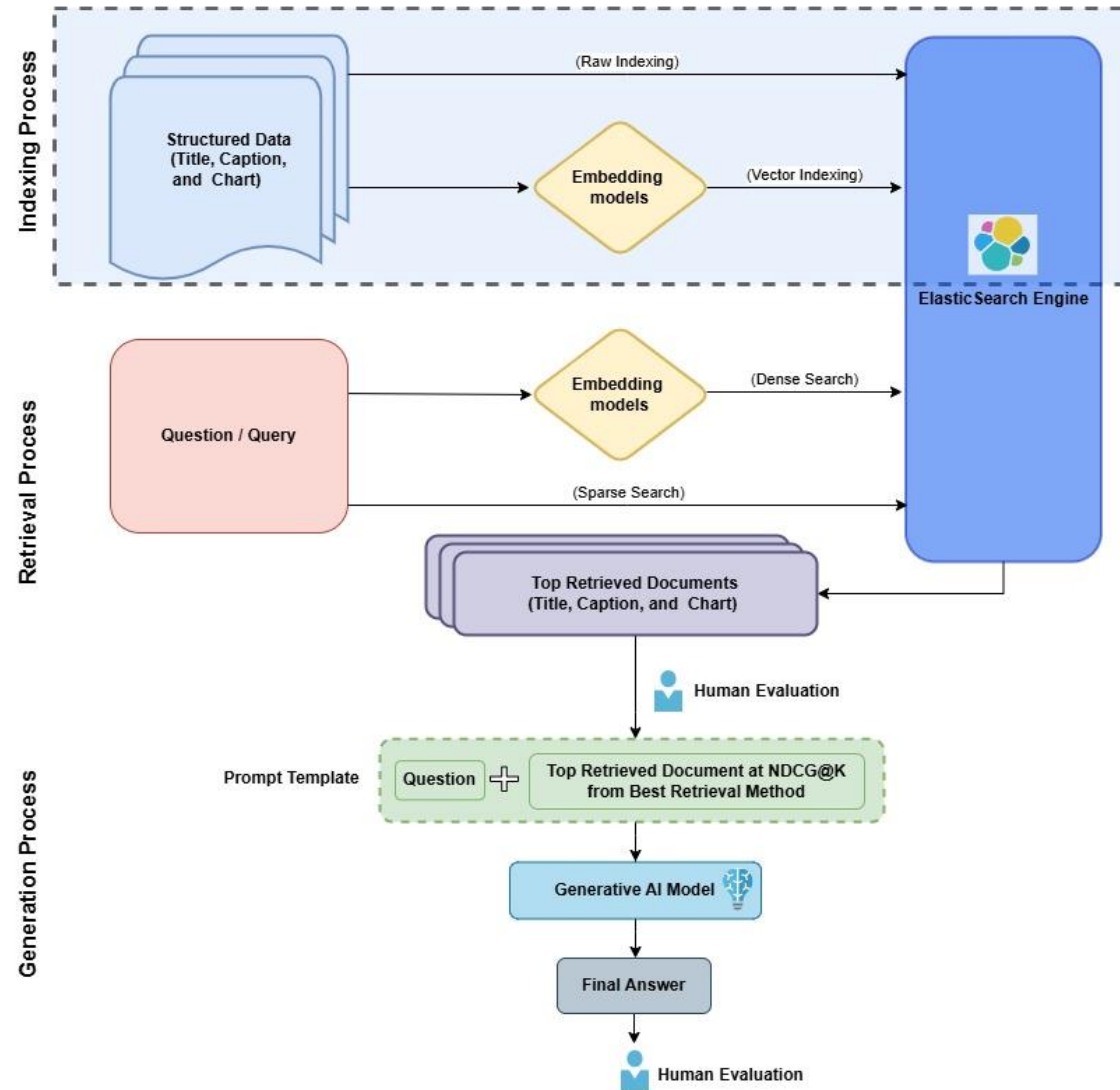
- Touché
- Manual
- Random

Example queries:

- Are gas prices too high?
- Should abortion be legal?

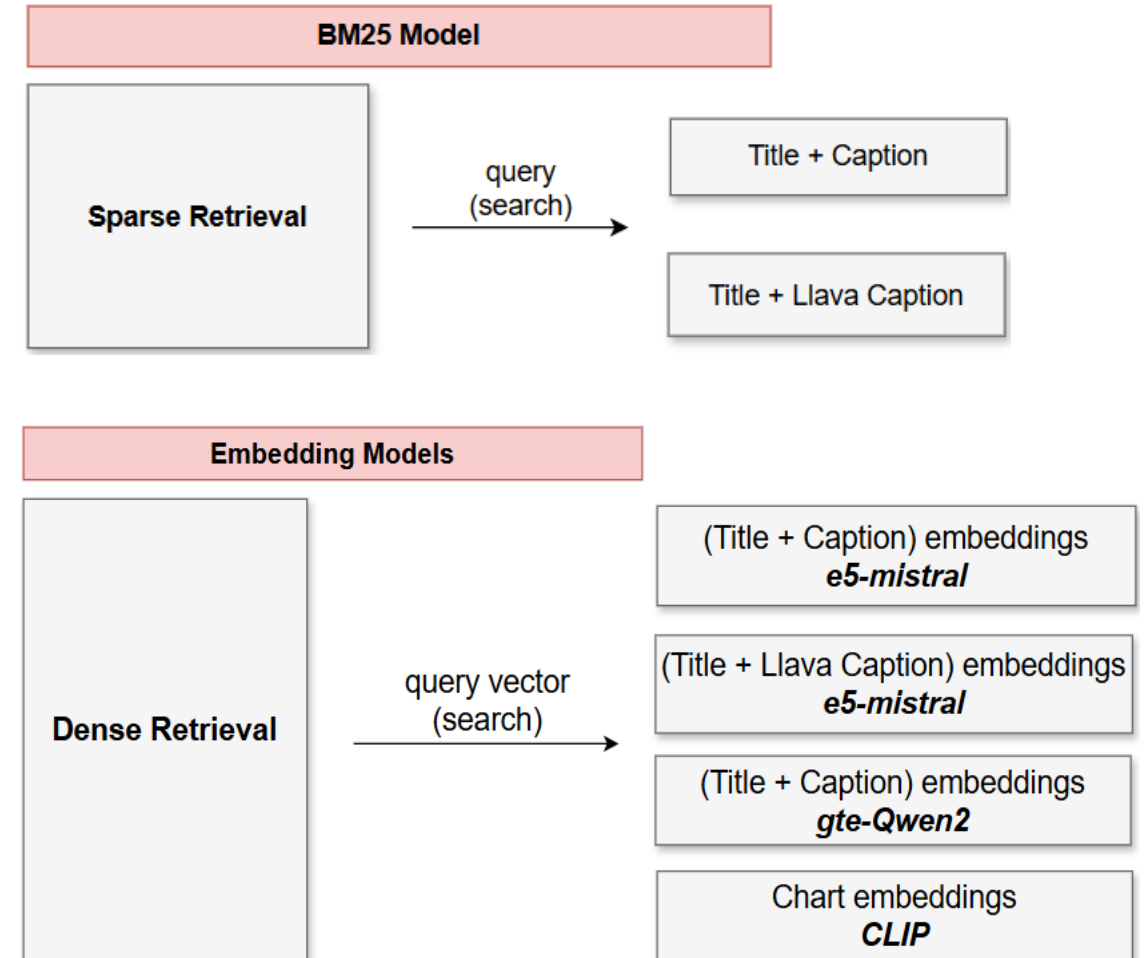
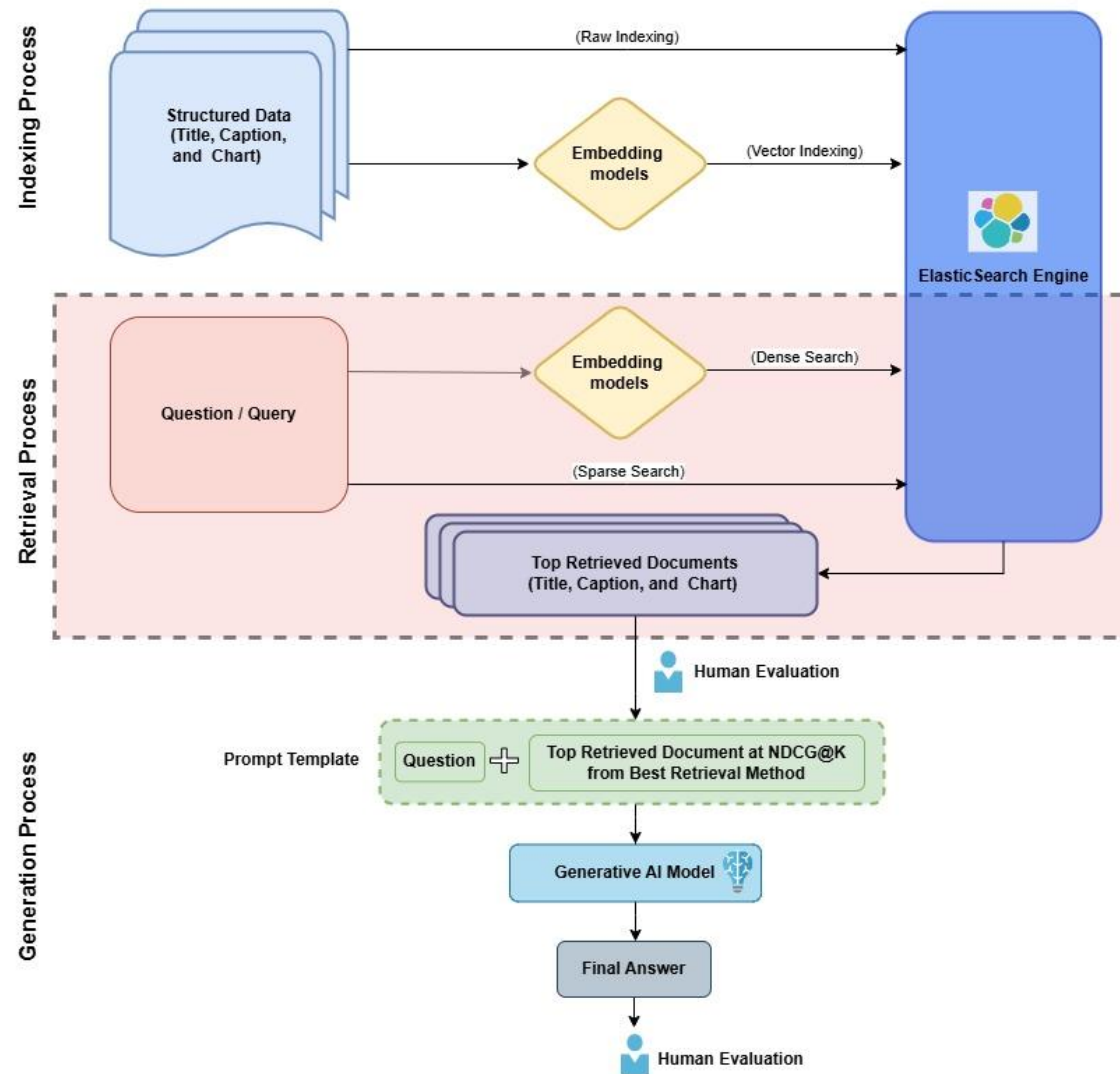
Methodology

Indexing Process



Methodology

Retrieval Process



Retrieves top 3 relevant documents from each method.

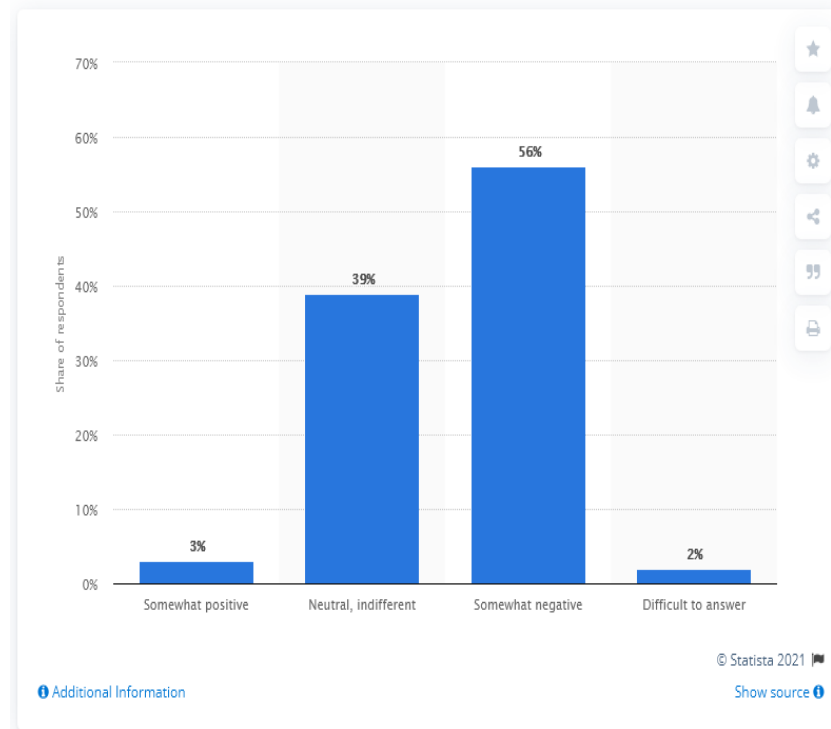
Methodology

Human Evaluation of Retrieval Process

Q. Is society more accepting of transgender people?

Title: Attitude toward lesbian, gay, bisexual, and transgender (LGBT) people in Russia in 2019

Caption: Only three percent of the survey participants representing the Russian population had a somewhat positive attitude toward LGBT people in 2019. The majority of 56 percent saw homosexual people somewhat negatively.



Evaluated as:

Relevance = 2 ; Completeness = 1

Relevance: Measures how closely a retrieved result addressed the query.

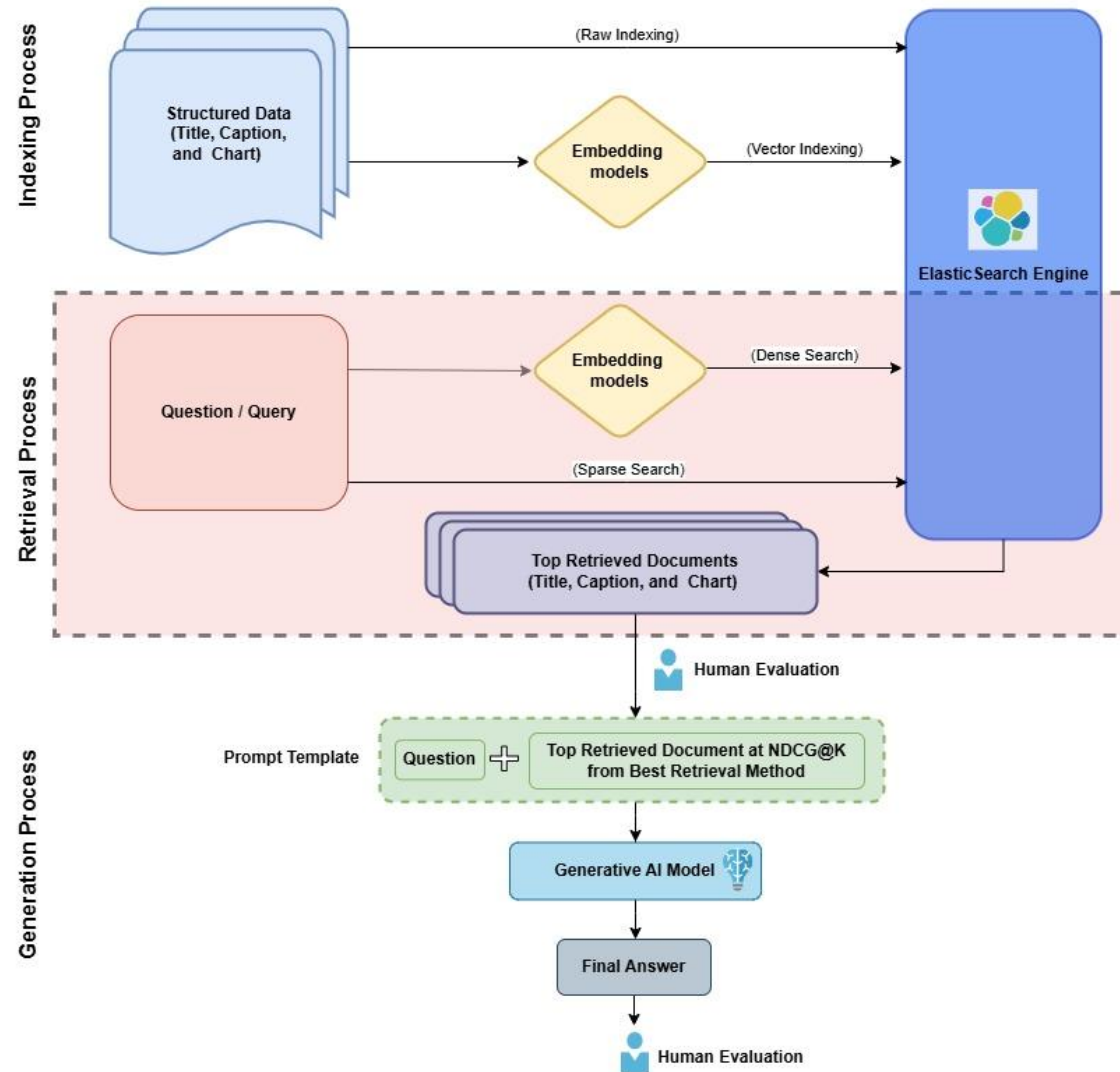
- **0: Not Relevant**
- **1: Partially Relevant**
- **2: Relevant**
- **3: Highly Relevant**

Completeness: Assesses if the result provides a thorough and comprehensive answer to the query.

- **0: No**
- **1: Somewhat**
- **2: Mostly**
- **3: Yes**

Methodology

Evaluation of Retrieval Process



Discounted Cumulative Gain (DCG)

Measures the relevance of items in a ranked list.

$$DCG@k = \sum_{i=1}^k \frac{rel_i}{\log_2(i + 1)}$$

- rel_i is the Relevance score of the item at position i .
- $\log_2(i + 1)$ is the Discount factor reducing the weight of lower-ranked items.

Ideal Discounted Cumulative Gain (IDCG)

Maximum possible DCG if items are perfectly ranked.

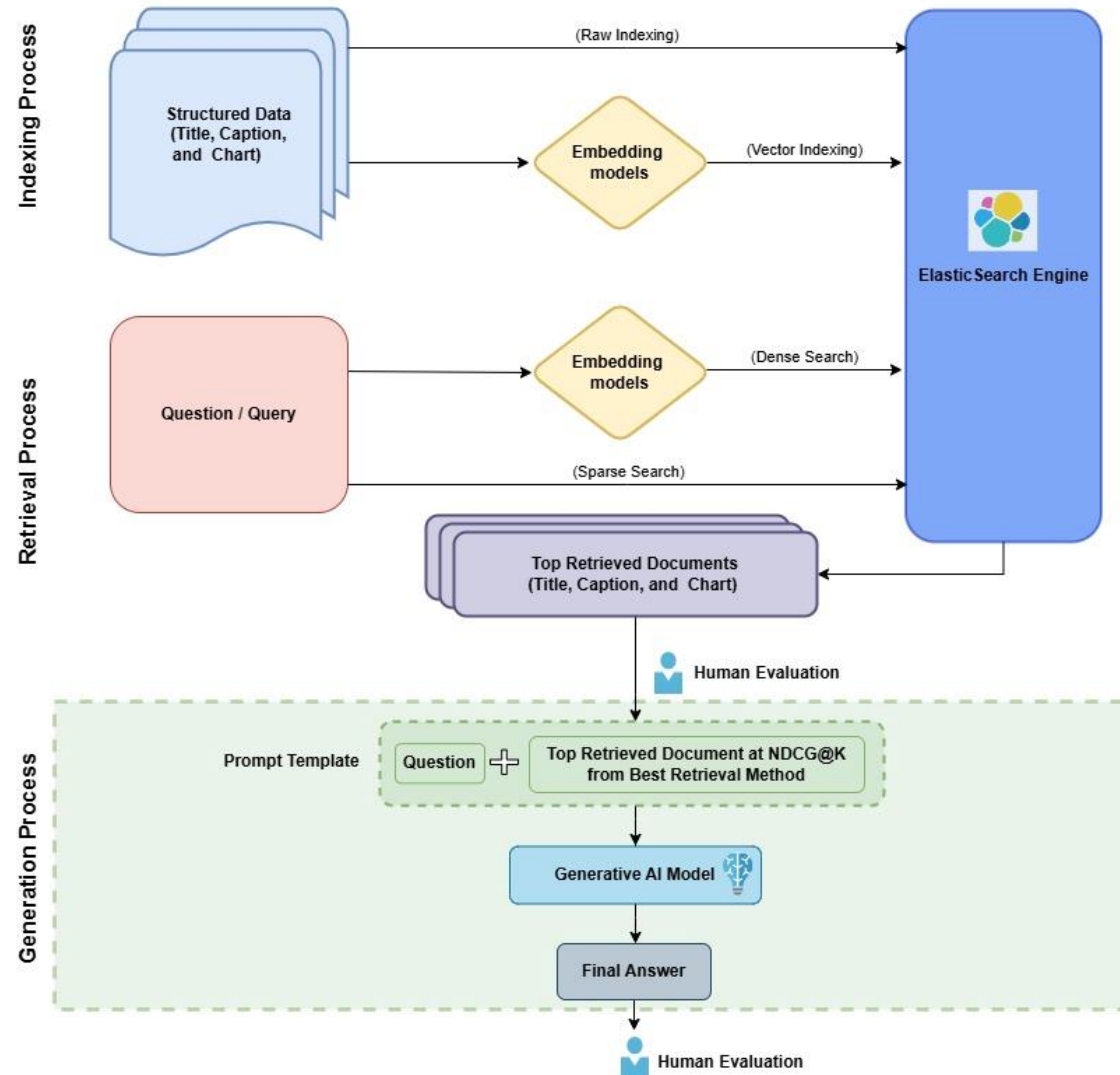
Normalized Discounted Cumulative Gain (NDCG)

Ratio of DCG to IDCG, ranging from 0 to 1.

$$NDCG@k = \frac{DCG}{IDCG}$$

Methodology

Generation Process



Method 1:

- Model: **Meta-Llama-3.1-70B-Instruct**.
- Text-Only input
- Prompt includes query, title, and caption.

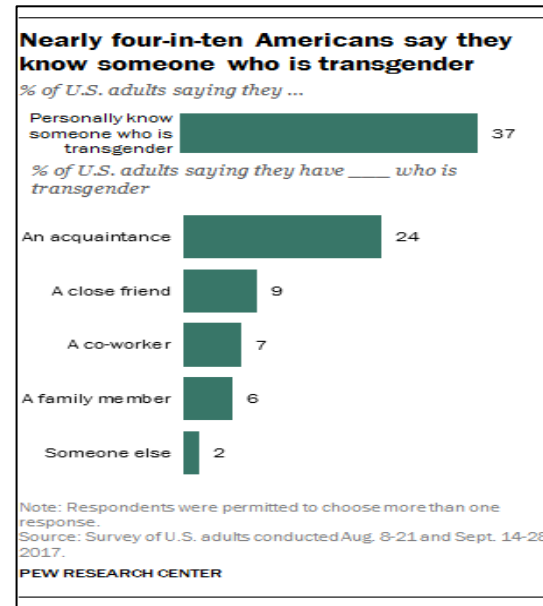
Method 2:

- Model: **GPT-4o**.
- Both text and image as input.
- Prompt includes query, title, caption, and chart.

Q. Is society more accepting of transgender people?

Title: Nearly four in ten Americans say they know someone who is transgender.

Caption: People who know someone who is transgender are more likely to feel that society hasn't done enough to accept transgender people. This aligns with past research showing that people with gay or lesbian acquaintances are more supportive of same-sex marriage. In total, 37% of Americans know a transgender person, including 15% with a close friend or family member (9% close friend, 6% family). Additionally, 24% have a transgender acquaintance, and 7% know a transgender co-worker.



GPT-4o: Yes, society appears to be becoming more accepting of transgender people.

Data indicates that personal connections significantly impact societal acceptance of transgender individuals. Specifically, **37% of Americans** know someone who is transgender, which includes **15% with a close friend or family member** and **24% with an acquaintance**. Among these, people who know someone transgender tend to feel more strongly that society hasn't yet done enough to embrace transgender individuals, mirroring past research on support for same-sex marriage among those who know gay or lesbian people.

In conclusion, while direct familiarity with transgender individuals correlates with greater acceptance, a substantial portion of the population still lacks such connections, **suggesting progress yet to be made in overall societal acceptance.**

Relevance

Measures how closely an answer addressed the query.

- **0: Not Relevant**
- **1: Partially Relevant**
- **2: Relevant**
- **3: Highly Relevant**

Faithfulness

Ensures claims are directly supported by provided content.

- **0: No**
- **1: Somewhat**
- **2: Mostly**
- **3: Yes**

Methodology

Live Demo

Chart Retrieval for Arguments

Touche

Manual

Random

Search Documents

Evaluate Models

Generate Answers

Save

Model-1

Model-2

Model-3

Model-4

Model-5

Model-6

Top Charts with Title and Content:

04 Result and Analysis

Retrieval Result and Analysis

Generation Result and Analysis

Results and Analysis

Retrieval Result and Analysis

Research Question: How effective are the retrieval methods in addressing a given query?

Query	Approach	Method	NDCG@1	NDCG@2	NDCG@3
Touché	Sparse	BM25	0.48	0.43	0.43
		BM25-Llava	0.18	0.19	0.21
	Dense	e5-Mistral	0.67	0.62	0.66
		e5-Mistral (Llava)	0.5	0.47	0.48
		gte-Qwen2	0.77	0.78	0.81
		CLIP	0.37	0.36	0.37
Manual	Sparse	BM25	0.51	0.54	0.52
		BM25-Llava	0.3	0.29	0.26
	Dense	e5-Mistral	0.67	0.75	0.8
		e5-Mistral (Llava)	0.51	0.52	0.51
		gte-Qwen2	0.84	0.86	0.87
		CLIP	0.36	0.36	0.36
Random	Sparse	BM25	0.54	0.54	0.52
		BM25-Llava	0.33	0.32	0.33
	Dense	e5-Mistral	0.79	0.78	0.77
		e5-Mistral (Llava)	0.6	0.65	0.6
		gte-Qwen2	0.86	0.86	0.87
		CLIP	0.41	0.37	0.36

Table 1: Mean NDCG@k

- Best Retrieval model: **gte-Qwen2**.
- Retrieval perform better on **Random** and **Manual** query sets.

Query	Approach	Method	NDCG@1	NDCG@2	NDCG@3
Touché	Sparse	BM25	2.53	3	1.5
		BM25-Llava	0.7	0.67	0
	Dense	e5-Mistral	4.78	1.92	3.5
		e5-Mistral (Llava)	2.78	2.42	2
		gte-Qwen2	7.28	11.25	13
		CLIP	1.92	0.75	0
Manual	Sparse	BM25	1.87	3.53	3.25
		BM25-Llava	1.03	0.53	0
	Dense	e5-Mistral	2.37	4.7	5.75
		e5-Mistral (Llava)	3.53	1.7	0.25
		gte-Qwen2	8.53	8.87	10.75
		CLIP	2.67	0.67	0
Random	Sparse	BM25	1.7	1.08	0
		BM25-Llava	1.67	1	0
	Dense	e5-Mistral	5.95	4.08	4.5
		e5-Mistral (Llava)	2.78	1.75	1
		gte-Qwen2	5.62	10.08	12.5
		CLIP	2.28	2	2
Total			60	60	60

Table 2: Best Retrieval Method counts

- LLaVA captions did not enhance retrieval quality.

Results and Analysis

Generation Result and Analysis

Research Question: How reliable are the generated answers for a given query?

Relevance Analysis

Query	Model	<u>NDCG@1</u>	<u>NDCG@2</u>	<u>NDCG@3</u>
Touché	GPT-4o	2.75	2.8	2.85
	Meta-Lama-3.1	2.7	2.7	2.75
Manual	GPT-4o	2.8	2.9	3
	Meta-Lama-3.1	2.75	2.85	2.95
Random	GPT-4o	2.8	3	2.9
	Meta-Lama-3.1	2.8	3	3

Table 3: Mean Relevance Score of the generated answer

- Relevance improves with increasing documents input.
- Manual and Random queries show higher relevance.
- Specific queries (e.g., country-focused) could achieve 100% relevance.

Results and Analysis

Generation Result and Analysis

Research Question: How reliable are the generated answers for a given query?

Faithfulness with Extra Interpretation

Query	Model	NDCG@1	NDCG@2	NDCG@3
Touché	GPT-4o	2.75	2.55	2.7
	Meta-Lama-3.1	2.4	2.35	2.6
Manual	GPT-4o	2.6	2.75	2.8
	Meta-Lama-3.1	2.35	2.5	2.25
Random	GPT-4o	2.65	2.5	2.7
	Meta-Lama-3.1	2.05	2.4	2.3

Table 4: Mean Faithfulness Score

Metric	Model	NDCG@1	NDCG@2	NDCG@3	Overall
Faithfulness Score = 3	GPT-4o	75%	68%	73%	72%
	Meta-Lama-3.1	42%	52%	47%	47%

Table 5: Fully Faithful Answers (System Accuracy)

Faithfulness without Extra Interpretation

Query	Model	NDCG@1	NDCG@2	NDCG@3
Touché	GPT-4o	2.95	2.75	2.8
	Meta-Lama-3.1	2.6	2.75	2.9
Manual	GPT-4o	2.75	3	3
	Meta-Lama-3.1	2.6	2.8	2.5
Random	GPT-4o	2.9	2.7	2.8
	Meta-Lama-3.1	2.6	2.65	2.45

Table 6: Mean Faithfulness Score

Metric	Model	NDCG@1	NDCG@2	NDCG@3	Overall
Faithfulness Score = 3	GPT-4o	90%	85%	87%	87%
	Meta-Lama-3.1	68%	75%	68%	71%

Table 7: Fully Faithful Answers (System Accuracy)

- GPT-4o outperformed Lama-3.1
- Both GPT-4o and Lama-3.1 demonstrated smart reasoning.

04 Conclusion

Key Findings

Limitations

Future Directions

Key Findings

- Best retrieval method using **gte-Qwen2-7B-instruct**.
- gte-Qwen2-7B-instruct potential for large retrieval task.
- CRF suitable for domain-specific application.
- **GPT-4o** generate more faithful answers.

Limitations

- Limited dataset.
- Labor-intensive human evaluation.

Future Directions

- To use more advanced model for Retrieval Evaluation.
- Integrate a verification steps within the generation process.
- Integrate prediction modelling.

