A Comparative Study between Google and Conversational AI Model

**Student:** Harkeerat Singh Sawhney   **Advisor:** Dr. Monica Landoni

Motivation

The ubiquitous nature of the internet has led to a surge in online activity amongst young children. Along with that with new advancement of technology, we have both traditional search engines (Google Search) but also conversational AI Models (ChatGPT). This thesis aims to draw a comparative analysis between Google Search and ChatGPT but using different performance assessments. Along with that the thesis will also develop a machine learning model aimed to predict the relevance of a result from a search query. This is a research aimed thesis, in which the goal is to find out for children which is the best way to search for information online. The research question which this thesis aims to answer is "**Are the new conversational AI models better for children to interact with, or do the traditional search engine still hold an edge over the new technology?**"
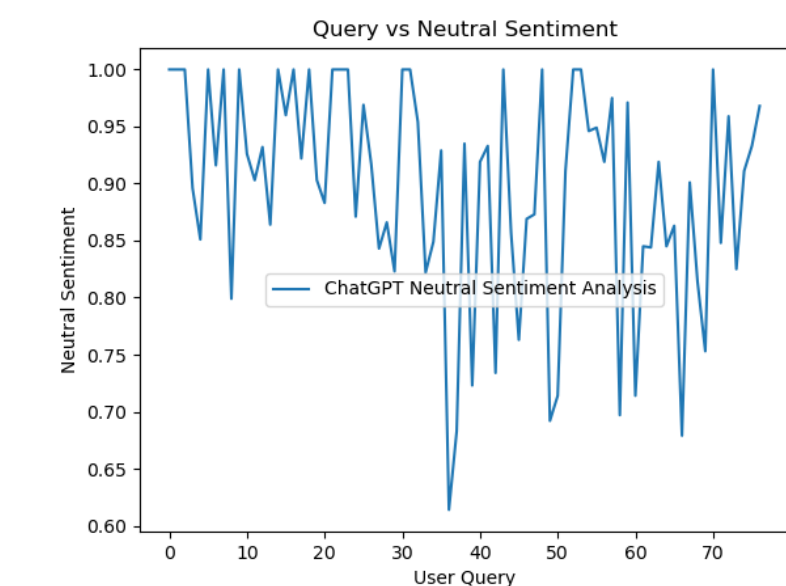
Data Collection

Google Search API

ChatGPT API
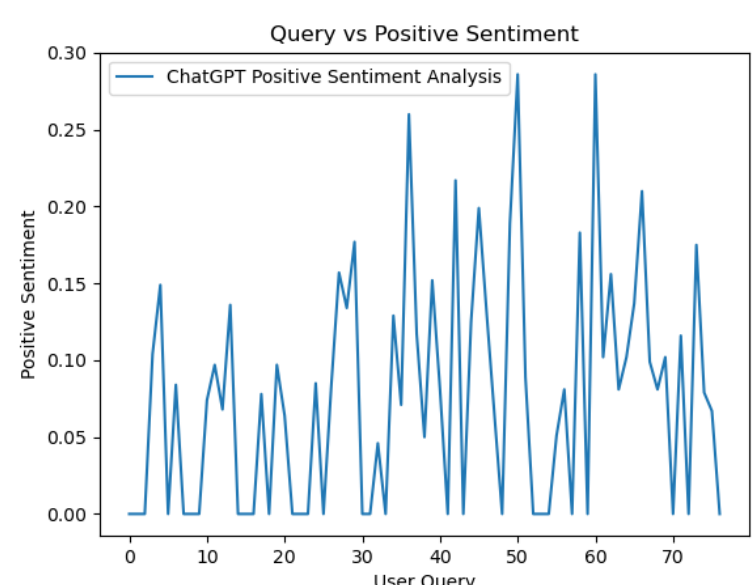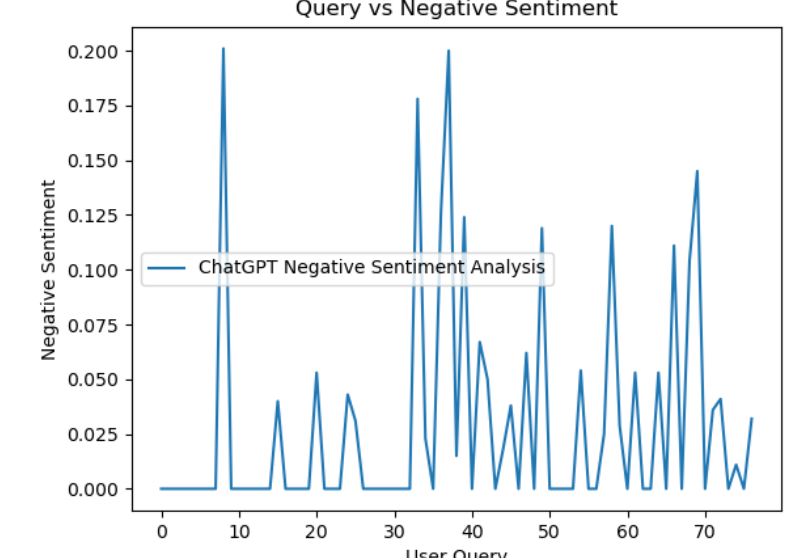

Data Analysis

Query Length Analysis
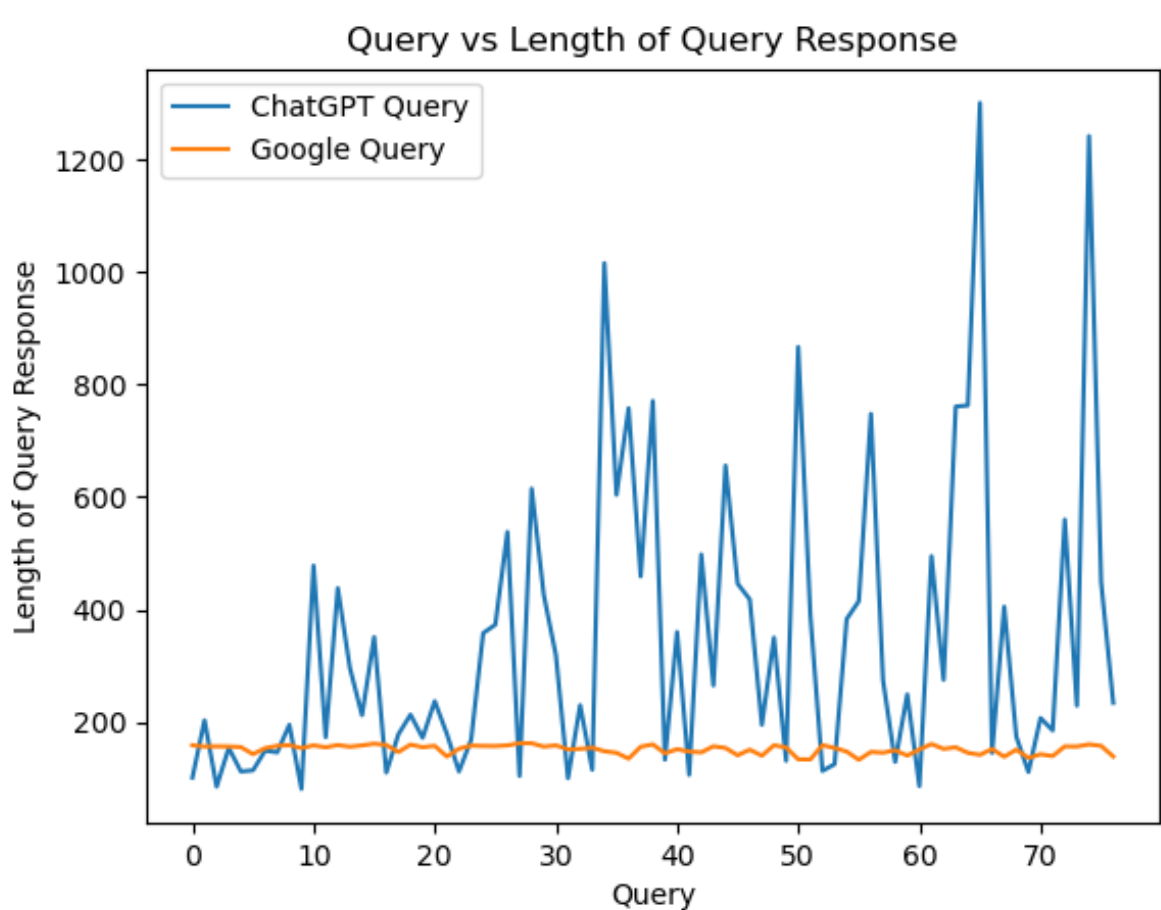
Comparison between ChatGPT and Google Search:
- ChatGPT's query lengths can vary based on the question asked, allowing for tailored responses.
- Google tends to provide consistent query lengths, potentially indicating a focus on relevance and concise information delivery.
- While ChatGPT's longer responses may offer more detailed explanations, they also carry a risk of information overload, whereas Google's consistent lengths may suggest greater relevance in its results.

Named Entity Recognition (NER)

Comparison between ChatGPT and Google Search:
- Google Search yields a significantly higher average number of entities per search (18.2 vs. 6.11).
- Both Google Search and ChatGPT commonly identify ORG, GPE, CARDINAL, and PERSON as top entities.
- The presence of different entities (DATA in Google and LOC in ChatGPT) suggests potential variations in the information provided by each system.
- Google's multiple search results contribute to the higher entity count, while ChatGPT's single generated response may lead to a lower count..

Sentiment Analysis

ChatGPT Sentiment

	POSITIVE SENTIMENT		
Average	0.03	Average	0.02
Median	0.00	Median	0.00
Max	0.20	Max	0.53
Min	0.00	Min	0.00
	(a) ChatGPT		(b) Google

	NEGATIVE SENTIMENT		
Average	0.89	Average	0.91
Median	0.92	Median	1
Max	1.00	Max	1.00
Min	0.61	Min	0.47
	(a) ChatGPT		(b) Google

	NEUTRAL SENTIMENT		
Average	0.08	Average	0.07
Median	0.08	Median	0.00
Max	0.29	Max	0.45
Min	0.00	Min	0.00

Google Search Sentiment

Similarity Analysis
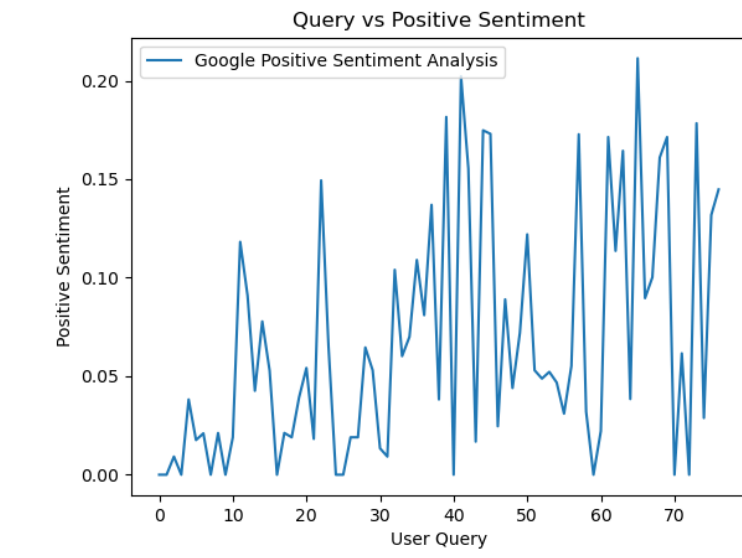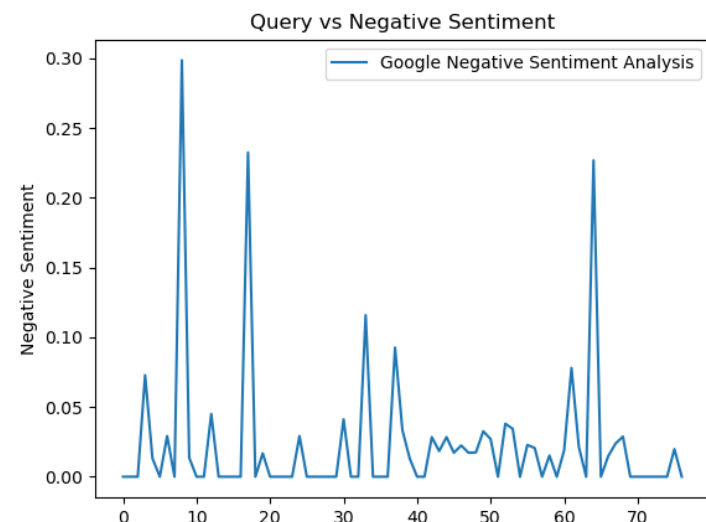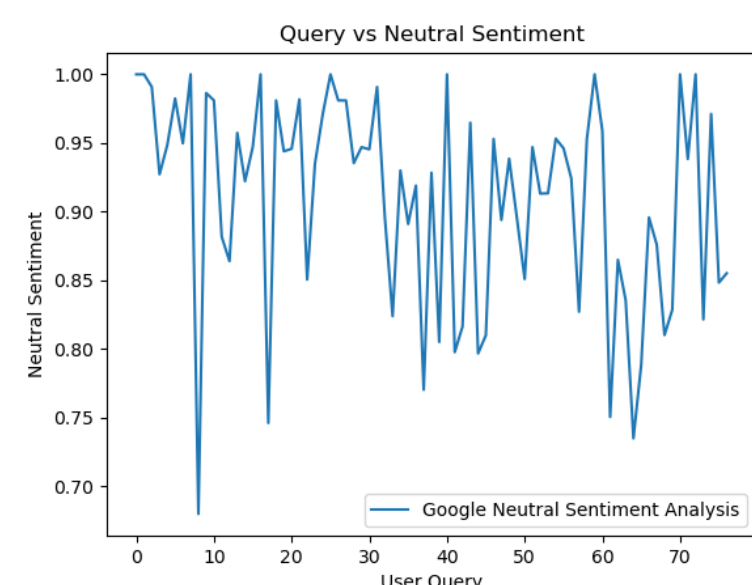
All the different Similarity Functions take ChatGPT Queries and Google Search Queries as their variables to which they try to find similarity between. Each sentence in each query is vectorized so that they are represented into a vector representation.
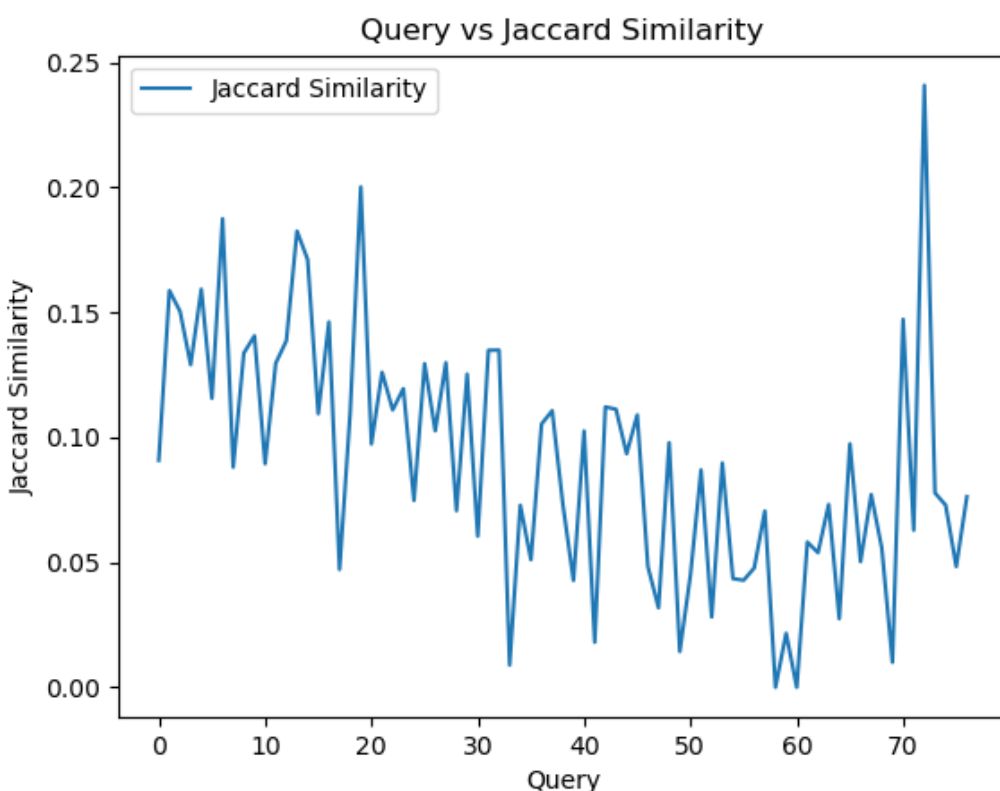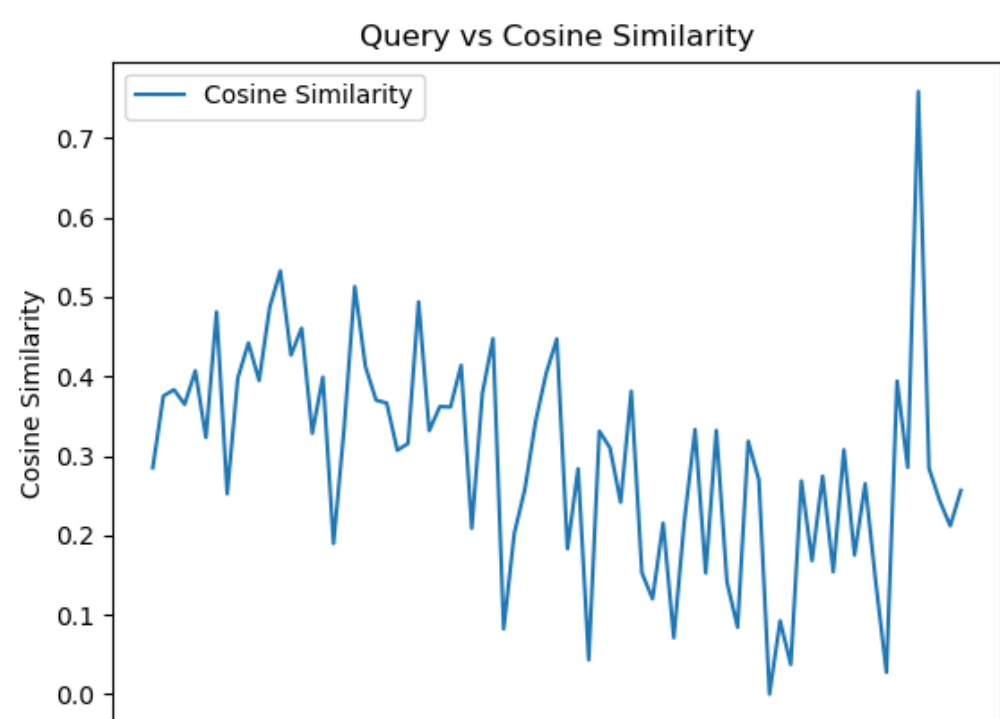
Cosine Similarity

Cosine Similarity is a measure of similarity between two non-zero vectors defined in an inner product space.

$$A \cdot B = \|A\|\|B\|\cos(\theta)$$
$$\cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$$

Jaccard Similarity

Jaccard Index Coefficient or also known as Critical Success Index, is measurement used to compute the similarity between two asymmetric binary variables.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Euclidean Similarity

Euclidean Similarity is a measure used to determine the objects or dissimilarity between two objects or data points in a Euclidean Space.

$$d(p,q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + ... + (q_n - p_n)^2}$$

Pearson Similarity

Pearson Correlation Coefficient is a measure of linear correlation between two sets of data. It is the ratio better the covariance of two variables with the product of their standard deviations.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$
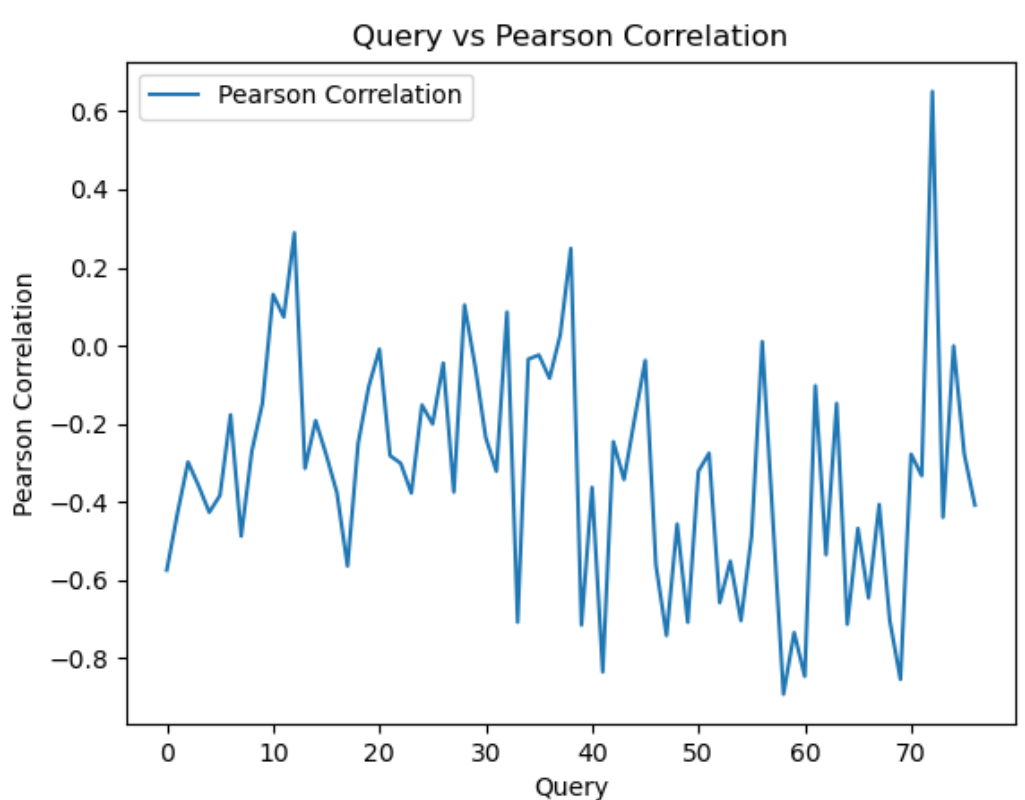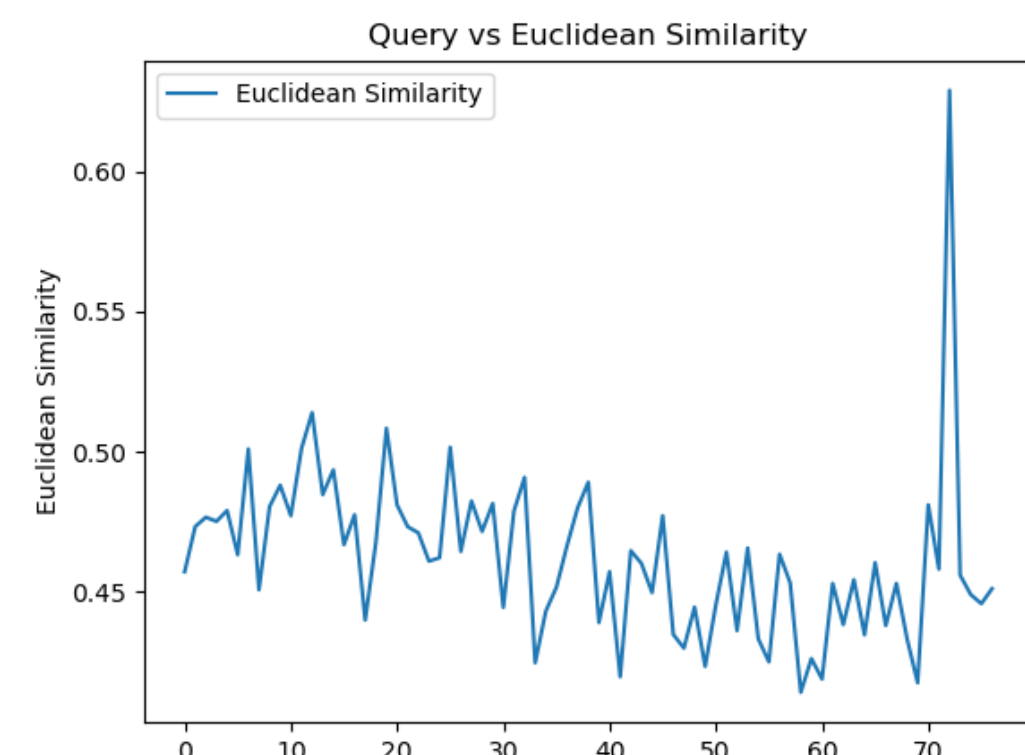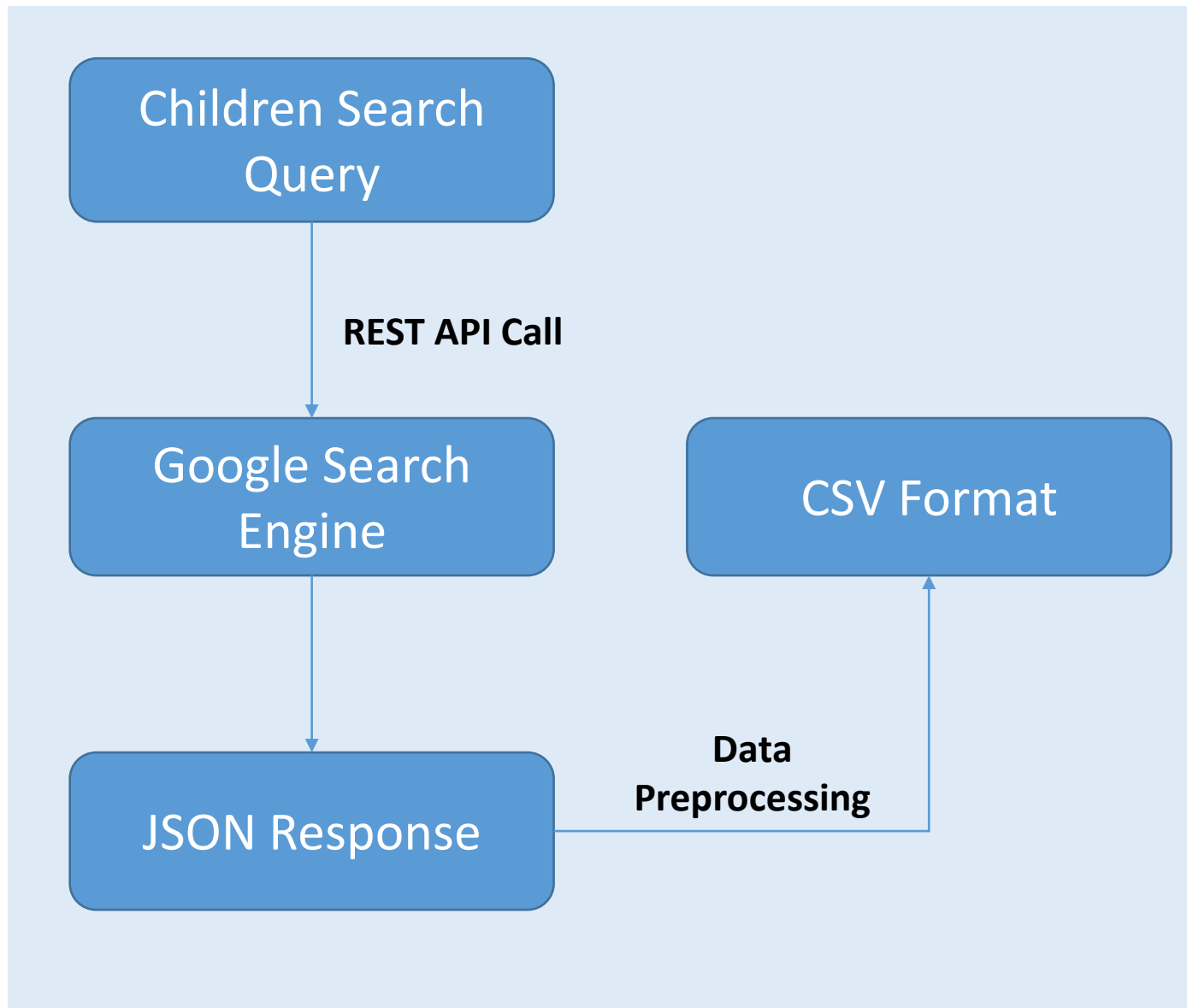

Findings and Limitations

- Logistic Regression is the chosen model for binary classification, suitable for determining relevance.
- F1 score, considering both precision and recall, is 0.89 for NR but low (0.38) for R.
- Limited data poses a challenge for proper model evaluation.
- Comparing results with ChatGPT, only 1 out of 77 queries was deemed relevant, indicating the need for more data collection and development.
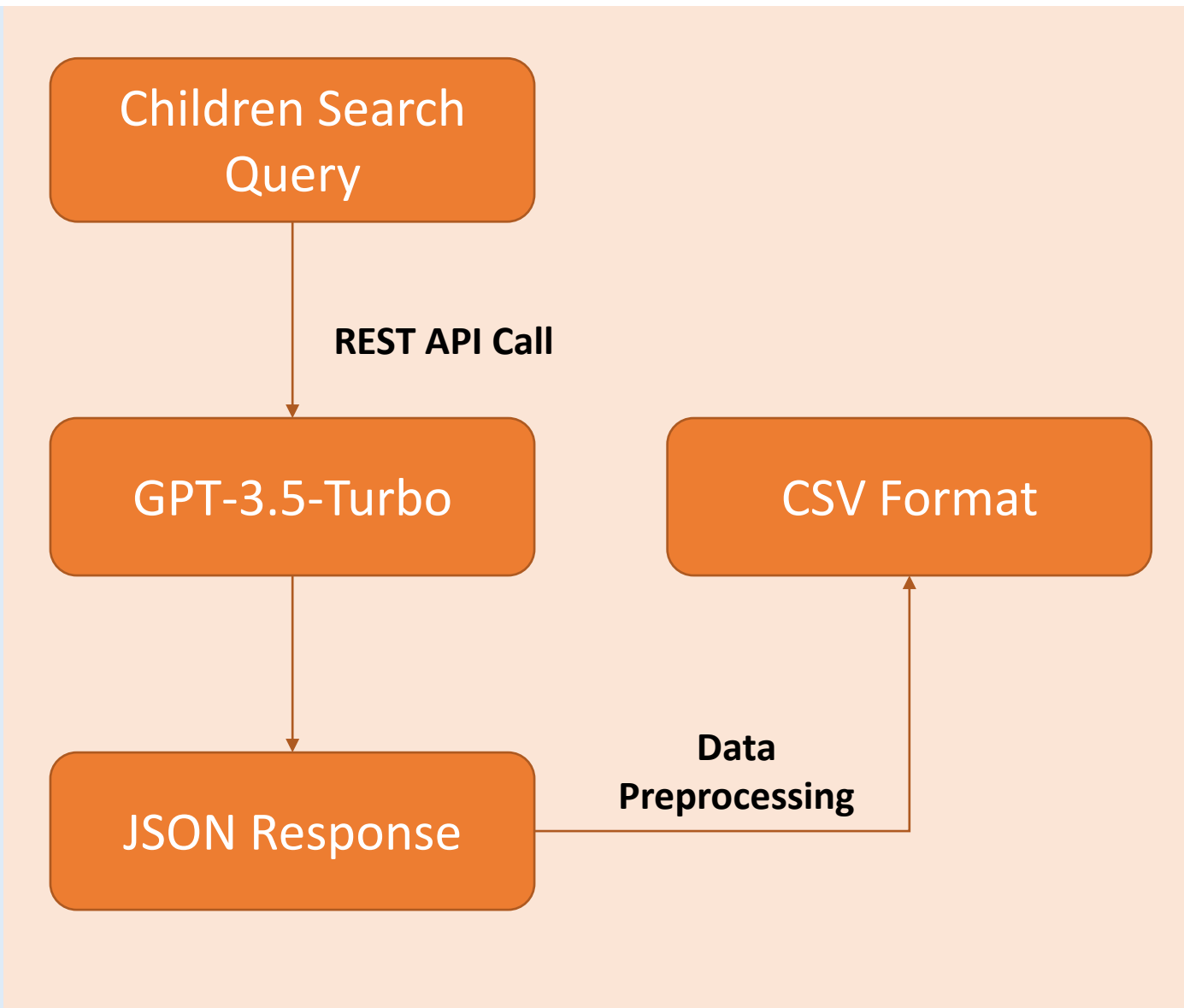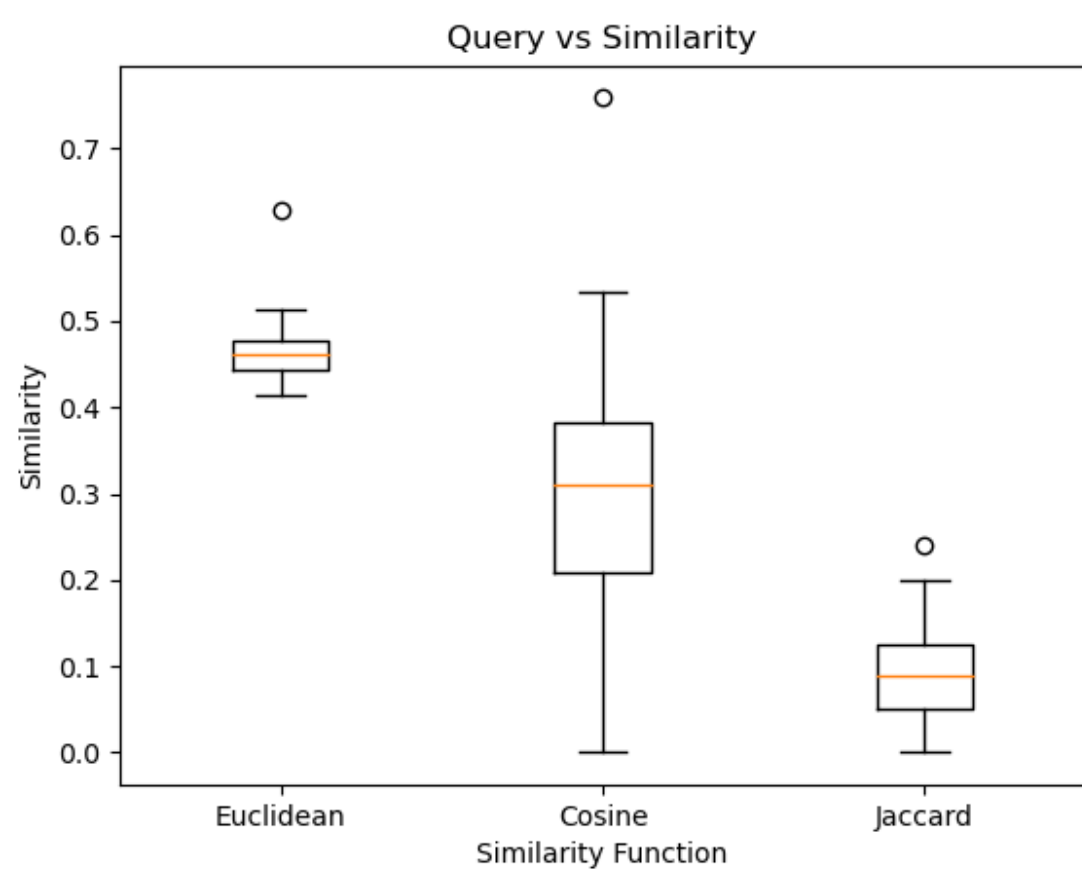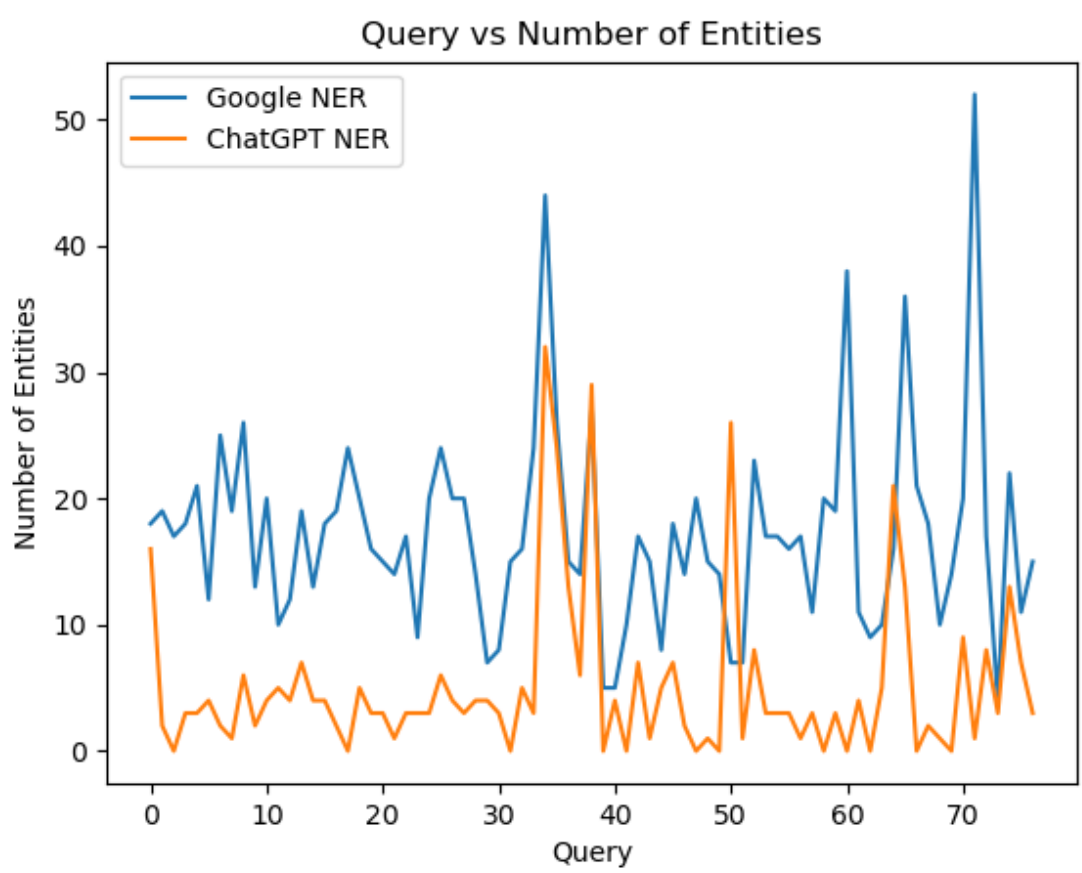
	precision	recall	f1-score	support
NR	0.81	0.98	0.89	132
R	0.83	0.25	0.38	40
accuracy			0.81	172
macro avg	0.82	0.62	0.64	172
weighted avg	0.82	0.81	0.77	172

Conclusion & Results

From these initial results, we notice how currently Google Search provides more richness and consistency in its responses which are more suitable for children. However, ChatGPT is not far behind Google in the statistical evaluation and holds a lot of potential in the future.

Statistic	Average	Max	Min	Median
Cosine Similarity	0.296	0.941	0	0.296
Euclidean Similarity	0.462	0.744	0	0.457
Jaccard Similarity	0.091	0.42	0	0.078
Pearson Correlation	-0.321	0.949	-0.955	-0.342
Google Query Length	151.44	162.2	132.8	153.8
ChatGPT Query Length	343.81	1321	81	237
Google NER	18.22	-	-	-
ChatGPT NER	6.12	-	-	-
ChatGPT Negative Sentiment	0.03	0.20	0.00	0.00
ChatGPT Neutral Sentiment	0.89	1.00	0.61	0.92
ChatGPT Positive Sentiment	0.08	0.29	0.00	0.08
Google Negative Sentiment	0.02	0.53	0.00	0.00
Google Neutral Sentiment	0.91	1.00	0.47	1.00
Google Positive Sentiment	0.07	0.45	0.00	0.00