Skincare Guide Chatbot Evaluation Metrics

Retrieval Metrics:

Context Precision:
Calculation: (Number of common entities) / (Total entities in retrieved context)
Result: 0.79
Interpretation: The system has good precision, with 79% of the entities in the retrieved context being relevant to skincare.

Context Recall:
Calculation: (Number of common entities) / (Total entities in all relevant contexts)
Result: 0.65
Interpretation: The system retrieved 65% of the relevant skincare entities, indicating moderate recall with room for improvement.

Context Relevance:
Calculation Method: TF-IDF vectorization and cosine similarity between the query and retrieved context
Result: 0.73
Interpretation: The retrieved context closely matches the query, showing high relevance to skincare topics.

Context Entity Recall:
Calculation: (Common entities) / (Total entities in query and relevant contexts)
Result: 0.81
Interpretation: The system captured 81% of the important skincare entities, demonstrating good recall.

Noise Robustness Test:
Scenario: Assessing the system's response validity with faulty skincare guide names.
Result: 57.66% of the queries with faulty names received valid responses.
Interpretation: The system managed to handle slightly more than half of these queries successfully.

Generation Metrics:

Faithfulness:
Measurement: BLEU score between generated response and ground truth
Result: 69%
Interpretation: The responses are generally accurate but can still be improved.

Answer Relevance:
Measurement: Cosine similarity between query and response

Result: 63%
Interpretation: The relevance of answers is moderately good but could be enhanced.

Information Integration:
Result: 74%
nterpretation: The system integrates skincare information decently but has room for improvement.

Counterfactual Robustness:
Result: 81%
Interpretation: The system handles speculative or hypothetical skincare scenarios well.

Negative Rejection:
Result: 64%
Interpretation: The system moderately rejects inappropriate or irrelevant skincare queries, with room for better performance.

Results Summary:

Retrieval Metrics:
Context Precision: 0.79
Context Recall: 0.65
Context Relevance: 0.73
Context Entity Recall: 0.81
Noise Robustness Score: 57.66%

Multi-Context Evaluation Results:
Context Precision: 0.60
Context Recall: 0.55
Context Relevance: 0.71
Context Entity Recall: 0.85

Generation Metrics Results:
Faithfulness: 69%
Answer Relevance: 63%
Information Integration: 74%
Counterfactual Robustness: 81%
Negative Rejection: 64%

Latency:
Latency for answering a query (including query creation, context fetching, and answer generation): ~5 seconds.

Proposed Changes to Architecture:

Improving Noise Robustness Using Prompt Engineering:

Current Prompt:

python

prompt = f"Given the following user query and conversation log, formulate a question that would be the most relevant to provide the user with an answer from a knowledge base.\n\nCONVERSATION LOG: \n{conversation}\n\nQuery: {query}\n\nRefined Query:

Improved Prompt:

python

prompt = f"""I want you to act as a chatbot focused on skincare. Answer the user's question to the best of your ability but stick only to skincare information. If there is nothing in the context relevant to the question at hand, just say "I don't know this" and stop after that. Refuse to answer any question not related to skincare. Never break character. {context} REMEMBER: If there is no relevant information within the context, just say "Hmm, I'm not sure". Don't try to make up an answer. Never break character.

Improving Negative Rejection Using Prompt Engineering:

Current Prompt:

python

If there isn't enough information to answer a skincare-related question, respond with: "I don't have enough information to answer this question."

Improved Prompt:

python

If there isn't enough information to answer a skincare-related question, or if the system doesn't have information about the skincare guide in its context, formulate a query that asks: "The system doesn't have information about [skincare guide title]. What would you like to know about this skincare guide?"