

THE INTER IIT TECH MEET 14.0



 OBSERVE • AI

Team 24



Problem Explanation



Identify and reason about causal dialogue patterns that trigger business critical events.

Primary Challenges

1. Noisy, branching dialogues make causal cues hard to detect.
2. Subtle, long-range cues require deep conversational dependency modeling.
3. Causal reasoning is needed to avoid correlation-based errors.

Exploratory Data Analysis



In the provided dataset, around 1285 transcripts were found to be noisy :-

Inconsistent labels

Transcripts having names as speakers were changed to Agent and Customer based on the dialogues , to keep it consistent with the rest of the examples.

Noise or irrelevant information

Some transcripts contained noisy text such as

```
transcript_id: "8e5e2b45-2bf6-42e6-  
9a9f-0ce4039cd745"  
...  
speaker: Benjamin Brooks  
text: bye  
speaker: Evelyn Lee  
text: *click*
```

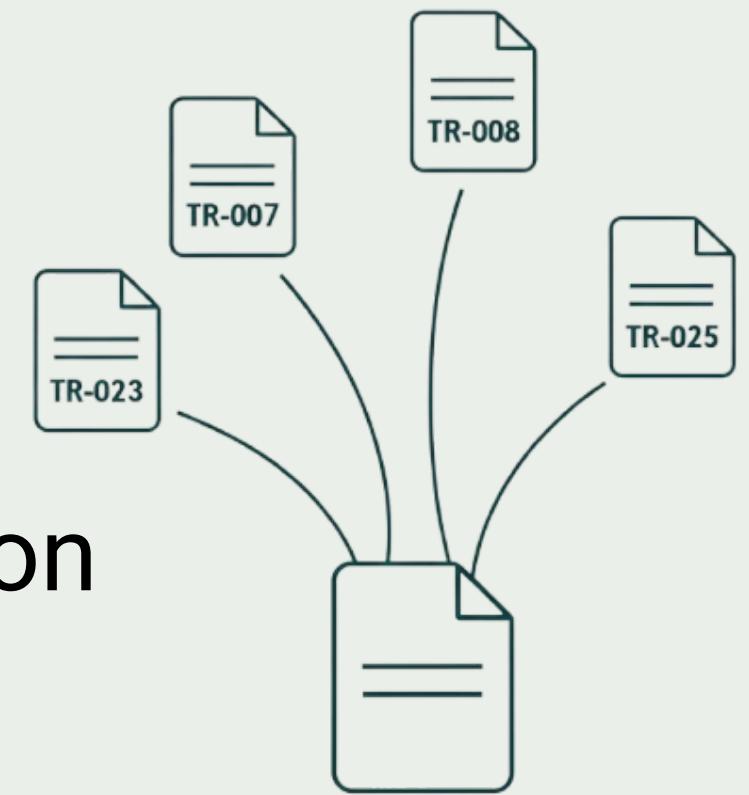
```
transcript_id: "81094278-3438-477c-  
882e-60257b561a8a"  
...  
speaker: Agent  
text: call ends  
speaker: Agent  
text: call ends
```

```
transcript_id: "8670af69-7be4-469e-  
b75d-630616775c06"  
...  
speaker: Agent  
text: supervisor call log  
speaker: Agent  
text: case outcome no change
```

Such noise was removed from the transcripts.

Task - 1

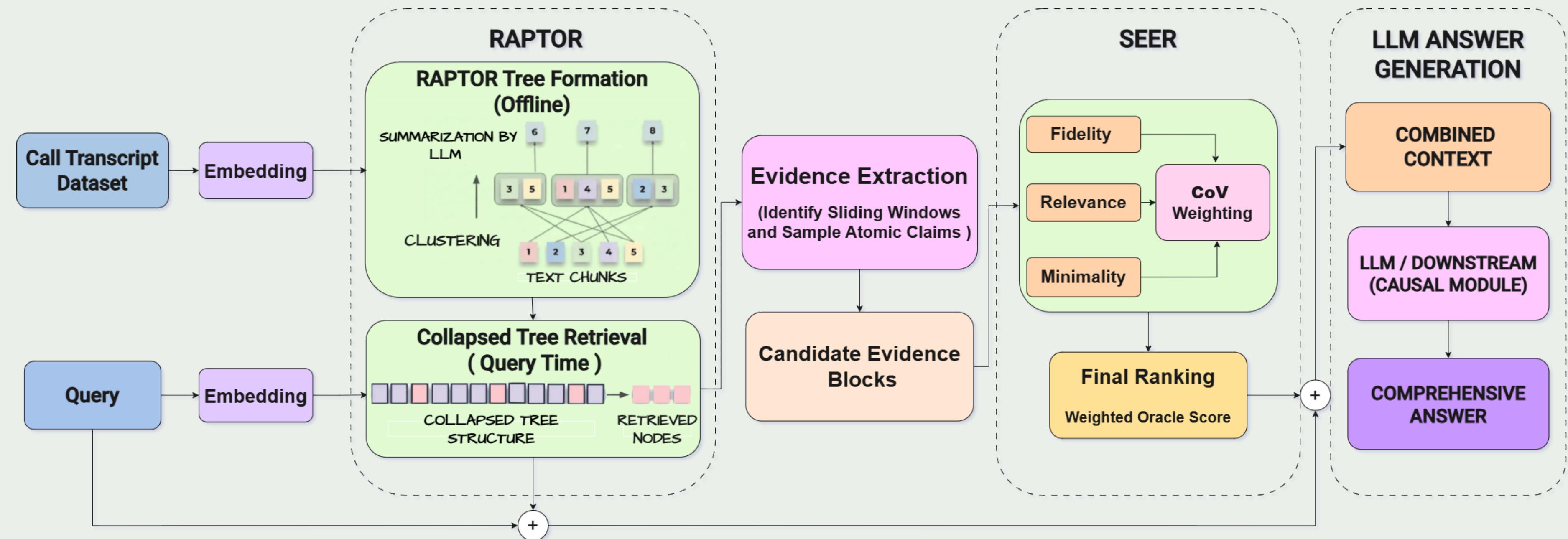
Query-Driven Evidence Based Causal Explanation



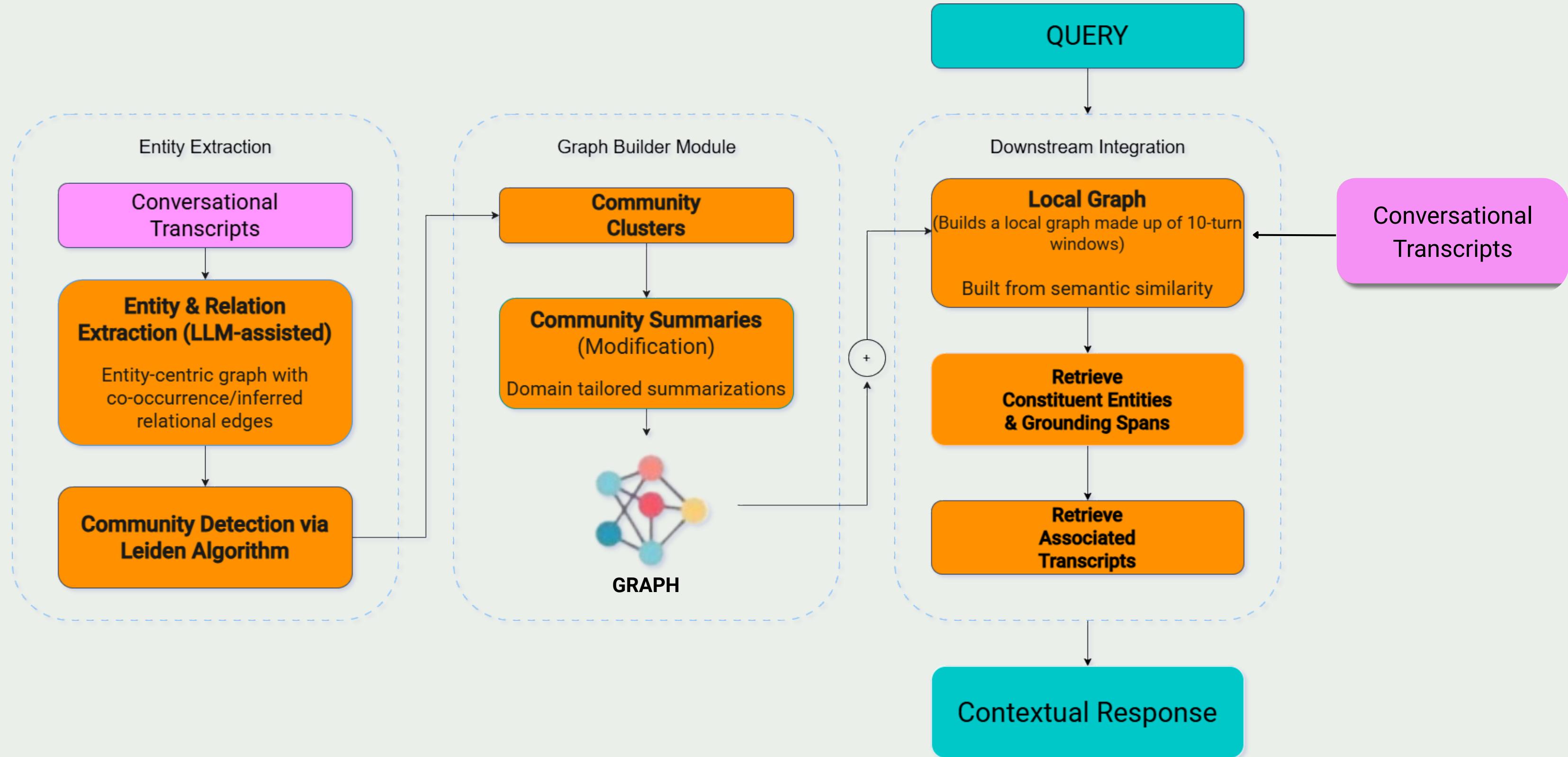
- Retrieve dialogue evidence that leads to key business events in long customer - agent conversations.
- Provide clear, causal explanations of why those outcomes occur.



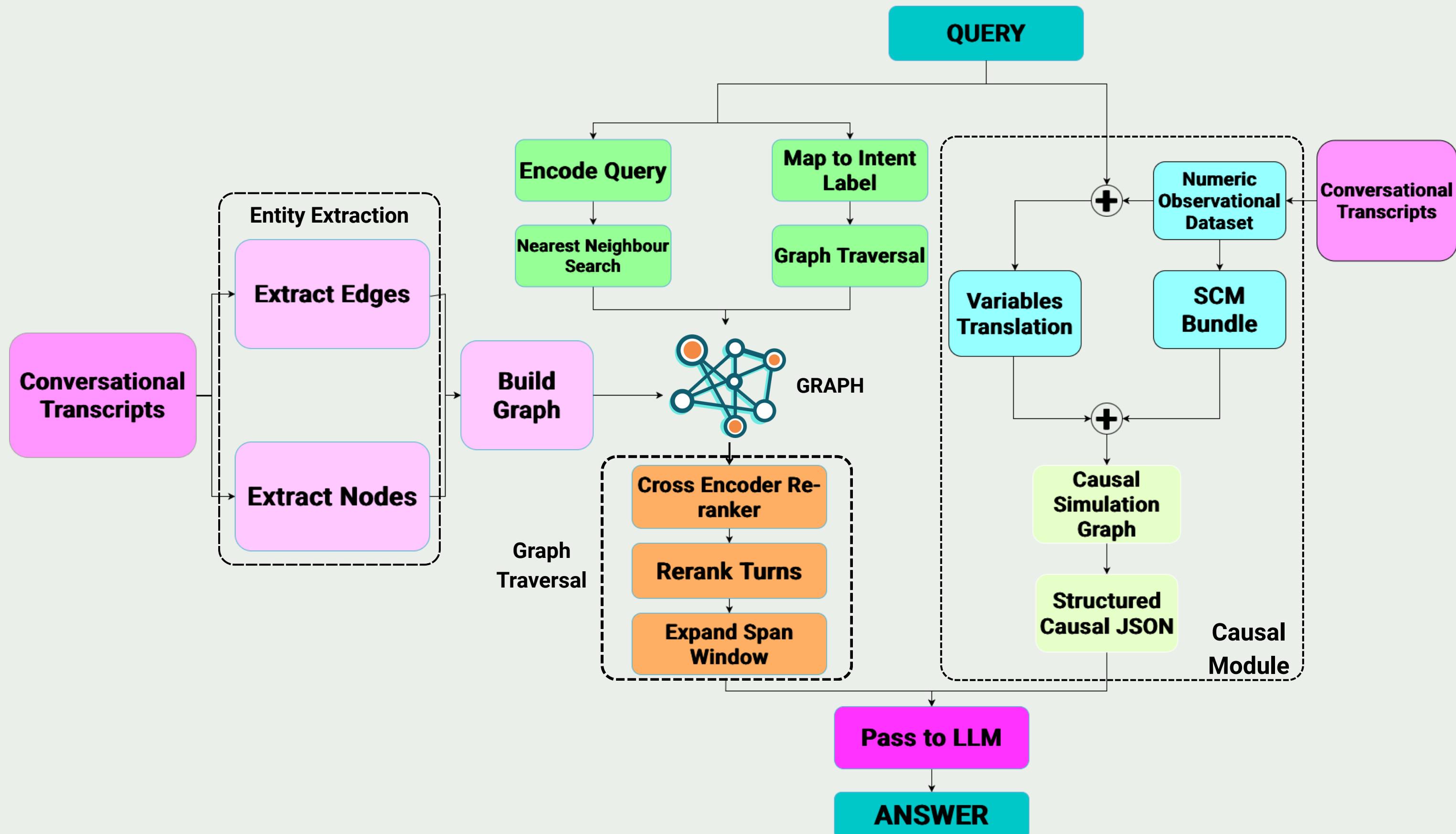
SEER + RAPTOR



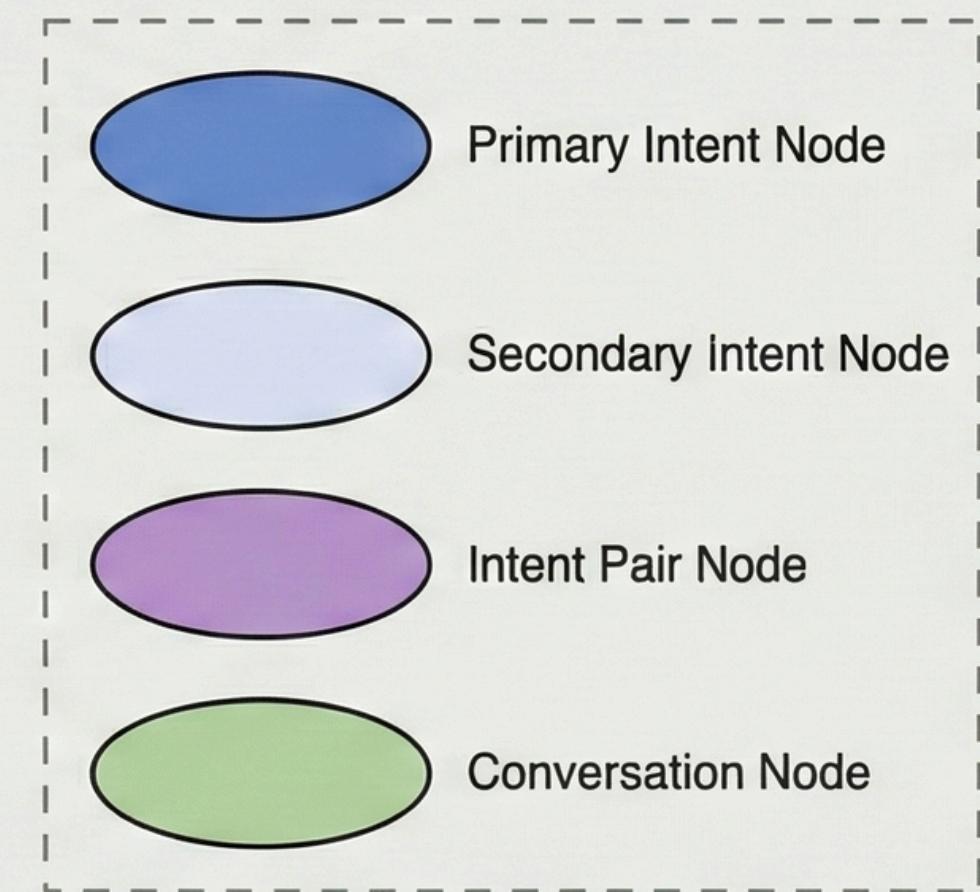
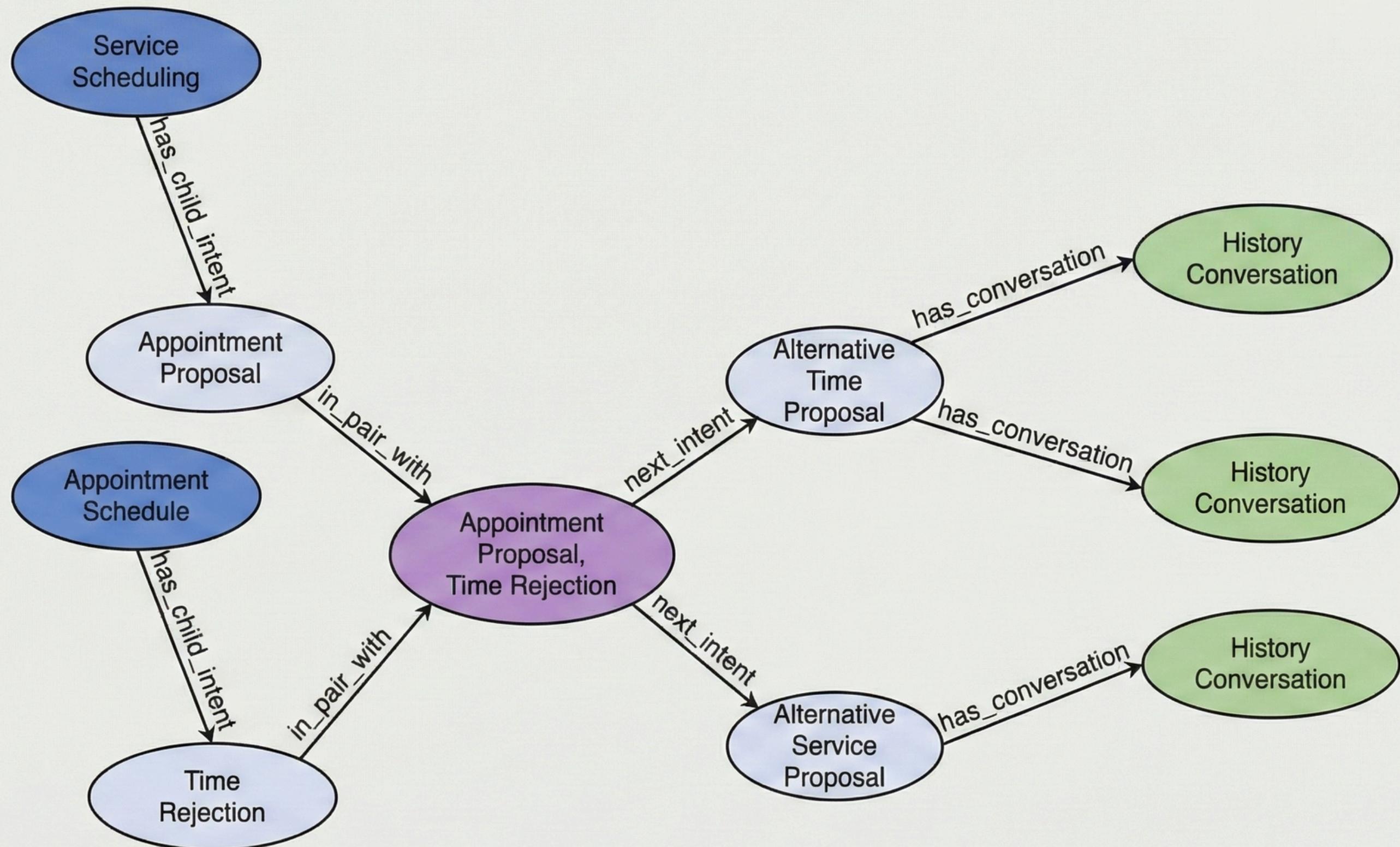
GraphRAG



Conversational Intent-Driven GraphRAG with SCM (Final Pipeline)

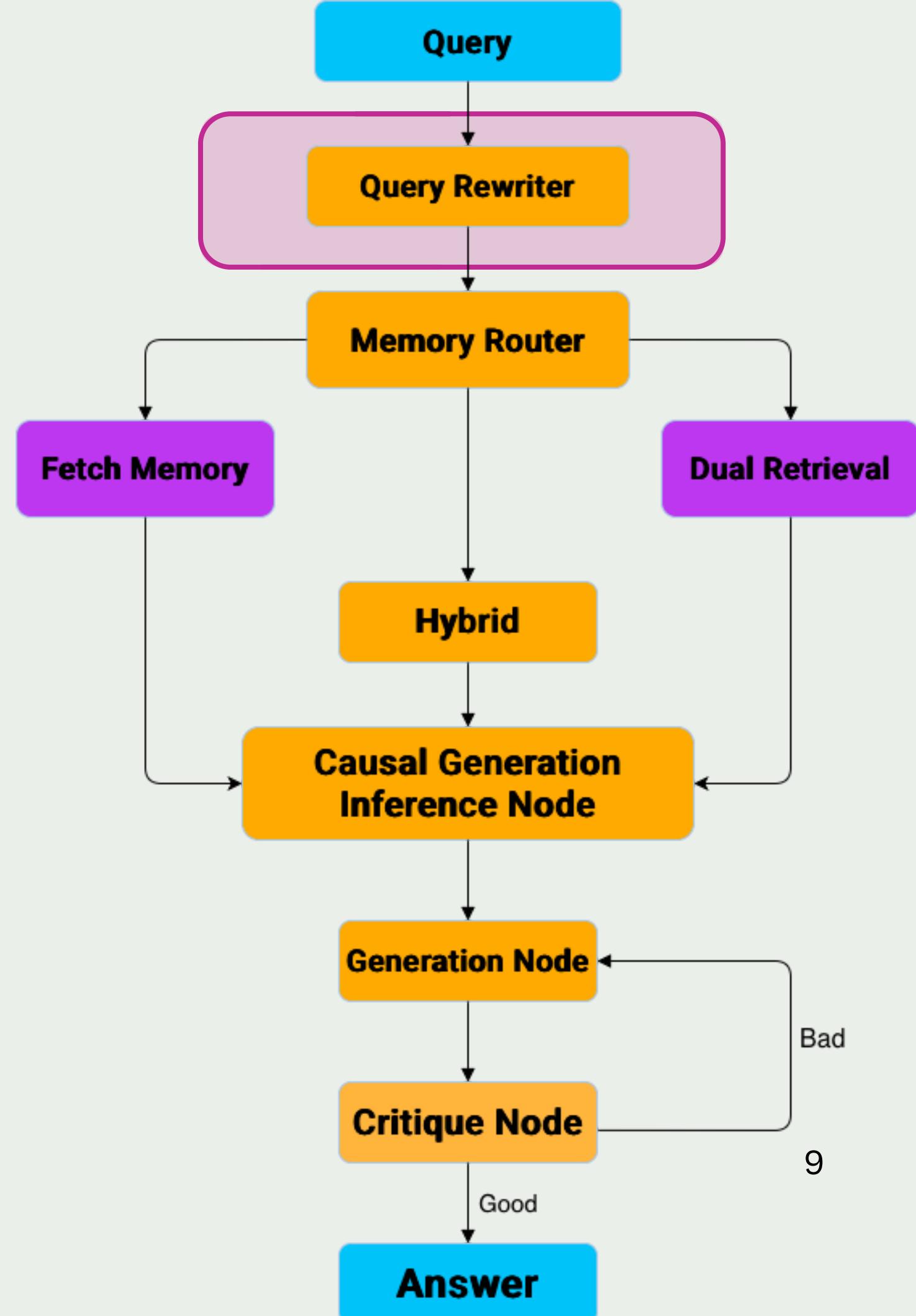


Example Graph



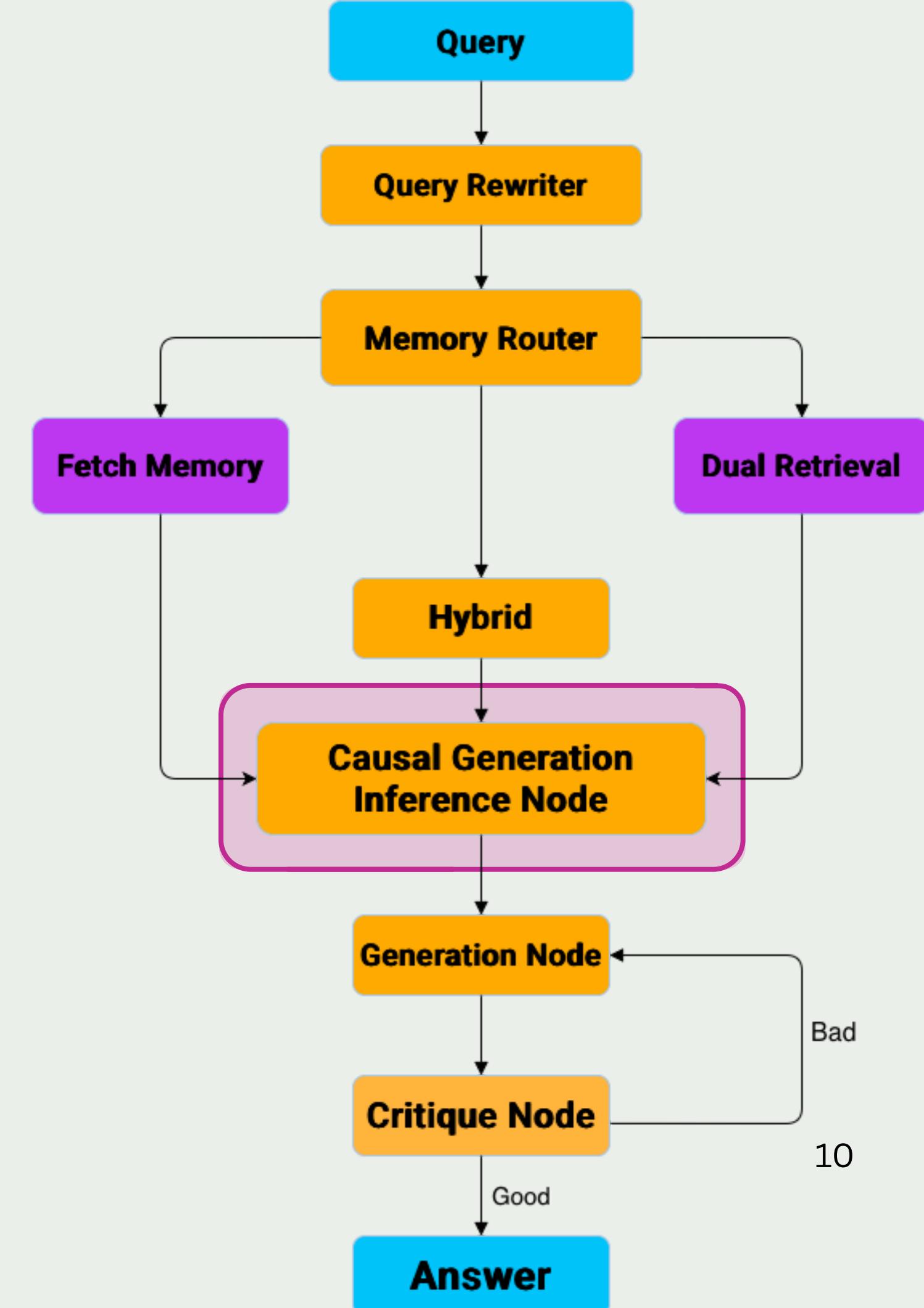
Query Rewriter

- Rewrite the user's query by removing ambiguity and focusing on clear intent.
- Convert it into a concise, cause-and-effect formulation with strong target keywords for better embedding retrieval.



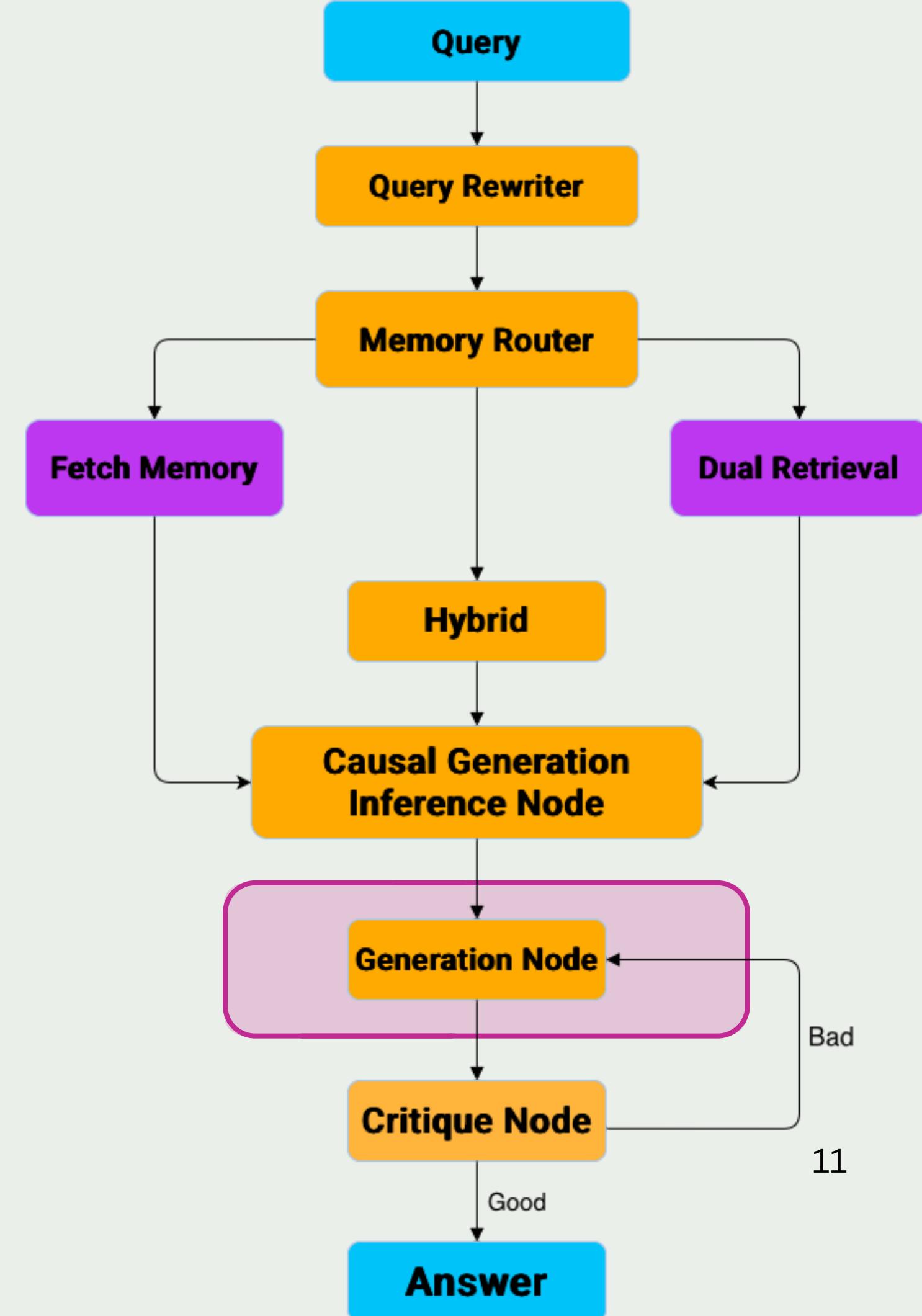
Causal Inference

- Ensures that performance differences arise solely from evidence selection.
- Produces a global explanation grounded in explicit evidence and quantitative causal assessments.



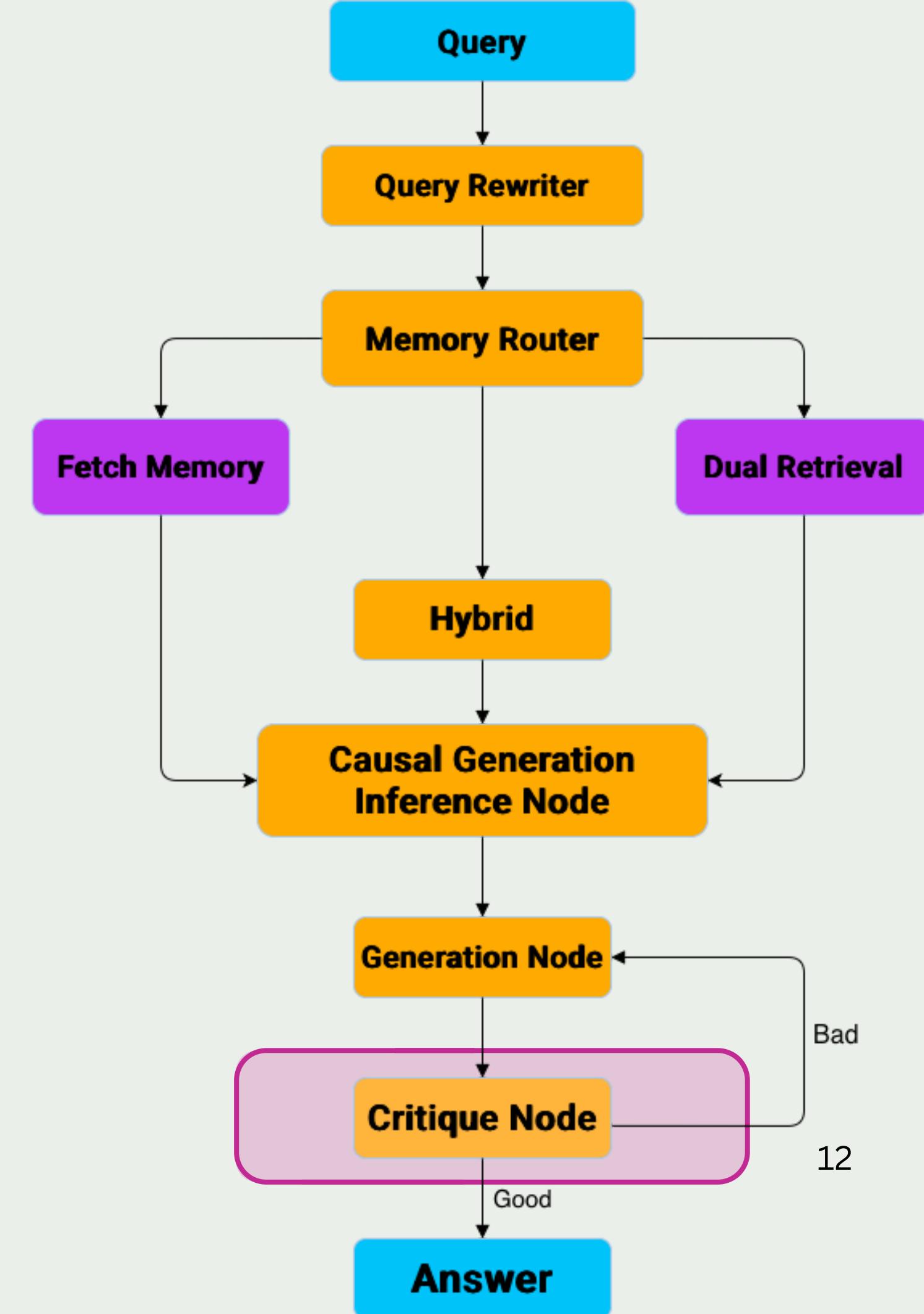
Generation Node

- Decide if the query needs a simple reply or a full report.
- If it's not a basic greeting or explanation, produce a structured report with summary, causes, recommendations, and citations.

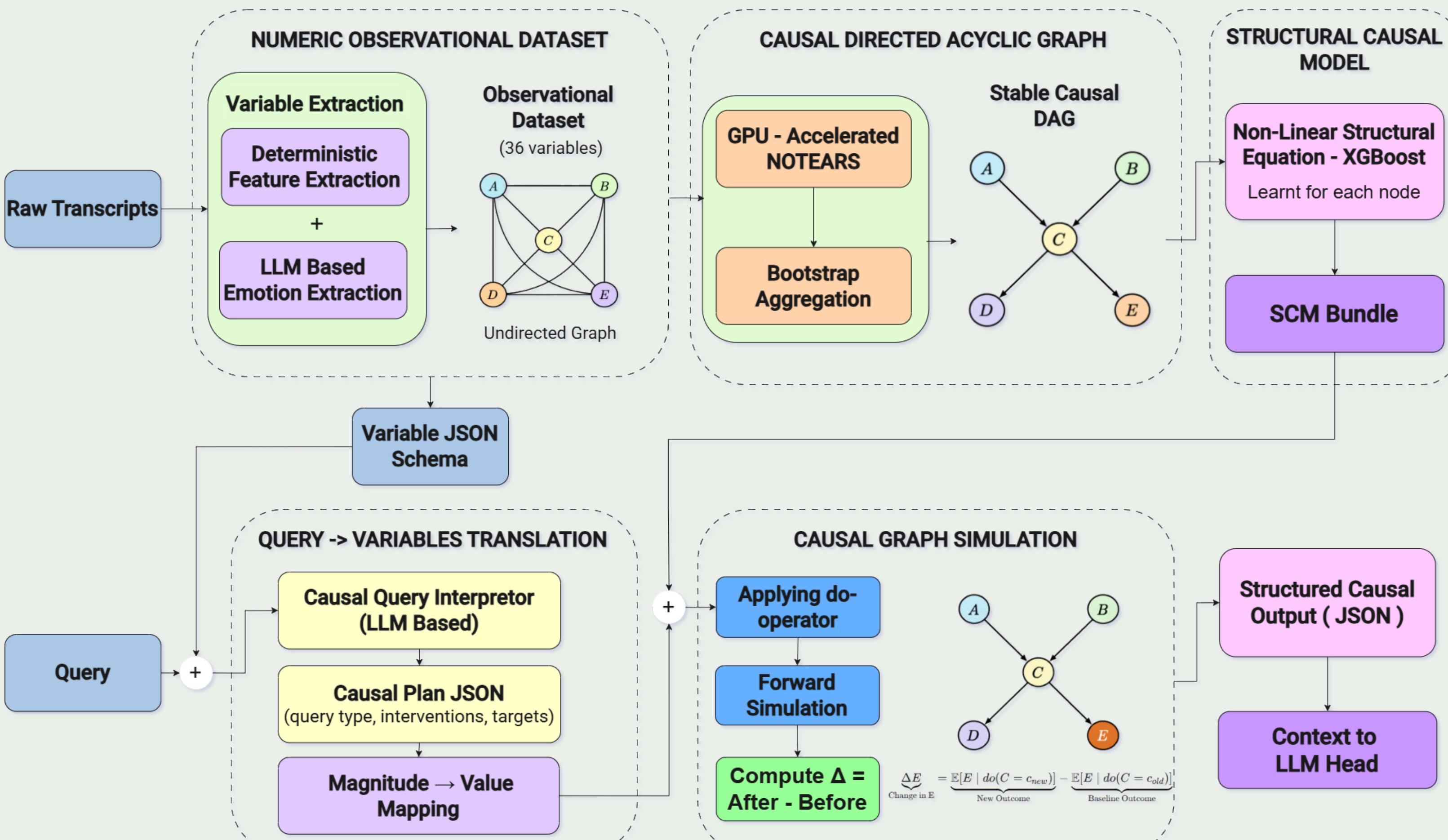


Critique Node

- Verify the answer's format and check it against source documents.
- This step reduces hallucinations.



STRUCTURAL CAUSAL MODULE (SCM)



Hallucination Proof System

User: Can you show the instances where user had to leave the call to go to the bathroom?

Agent

Based on the provided conversation segments, there are no instances where a user explicitly states they had to leave the call to go to the bathroom. The conversations focus on issues like internet service, hotel cleanliness, and general call experience.

(Response time: 19.39 seconds)

User: Can you find the instance where arnav had to apologise?

Agent

Based on the provided conversation segments, there is no instance where an agent named "Arnav" had to apologize. However, there are several instances where agents apologize for issues or inconveniences.

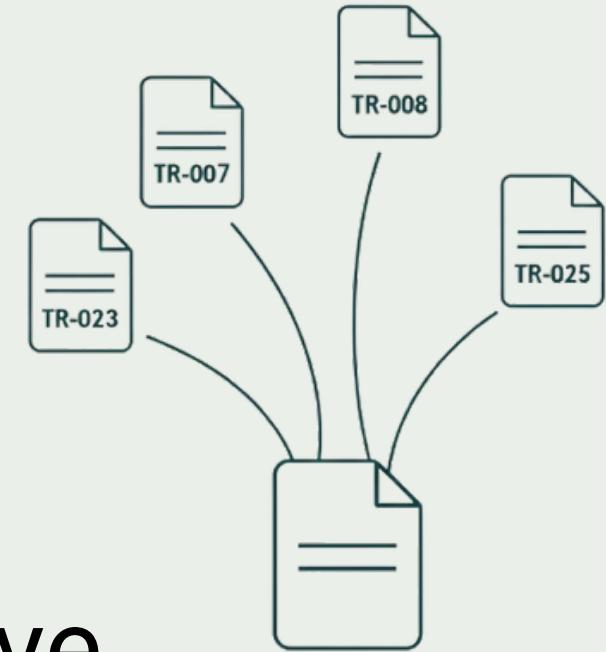
Here are some examples:

- In **Case 3** (**ID: 67054312-21d8-485a-9a87-9ac5735d3eaa**), the agent says, "I understand and I'm really sorry that your expectations were not met that's definitely not the service we strive for."
- In **Case 4** (**ID: 30c98ce1-8b35-4df1-ac86-0216777e83f2**), the agent states, "Yes I I'm afraid that's the case right now I apologize for the inconvenience."
- In **Case 6** (**ID: d4782c17-87ae-42fe-a288-2e88c5eeeb9b**), the agent says, "Of course I'm really sorry for the delay and the issues you've encountered let me confirm this upgrade for you."
- In **Case 13** (**ID: 69306cd9-e55b-4d28-b984-21a1f60ebf18**), the agent apologizes by saying, "I truly apologize for the inconvenience this has caused you you know I understand how important uh clear communication is especially regarding bookings."
- In **Case 20** (**ID: 459f9d37-bbd8-49bb-8f2a-d5fa22f39213**), the agent states, "You're absolutely right and I apologize we're sorry this went wrong."

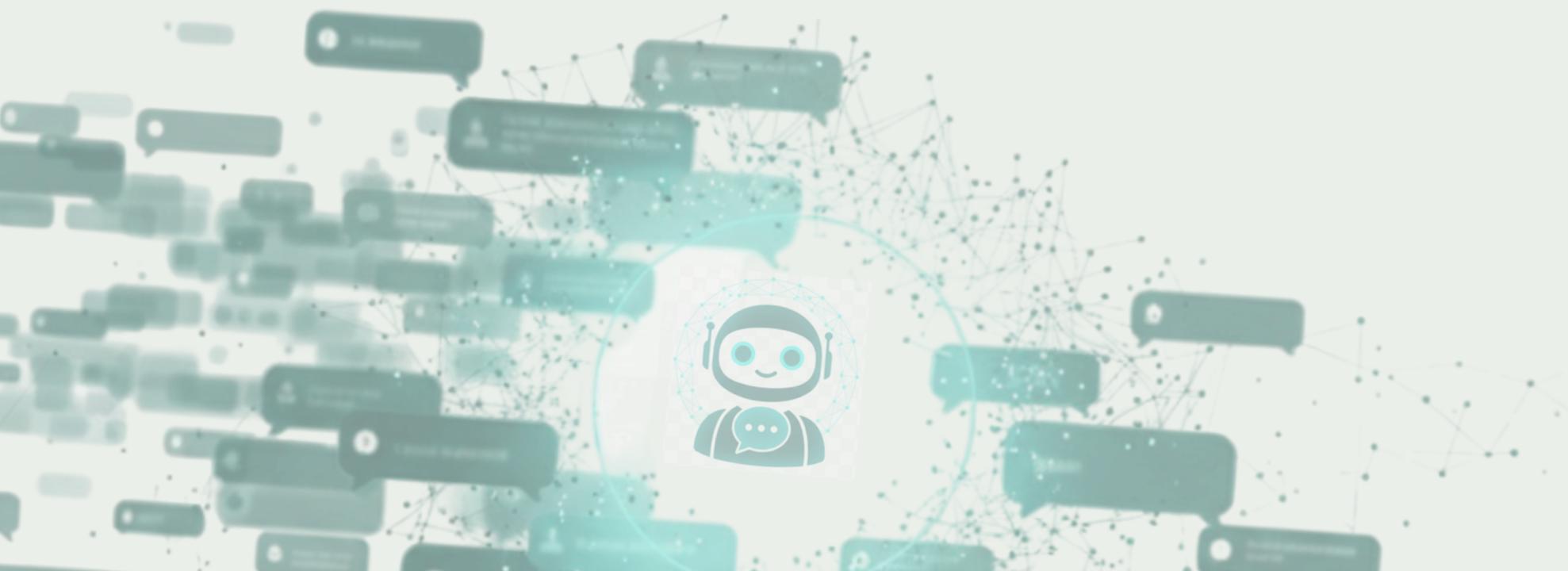
(Response time: 14.22 seconds)

Task - 2

Enable context-aware follow-up questioning and iterative dialogue.

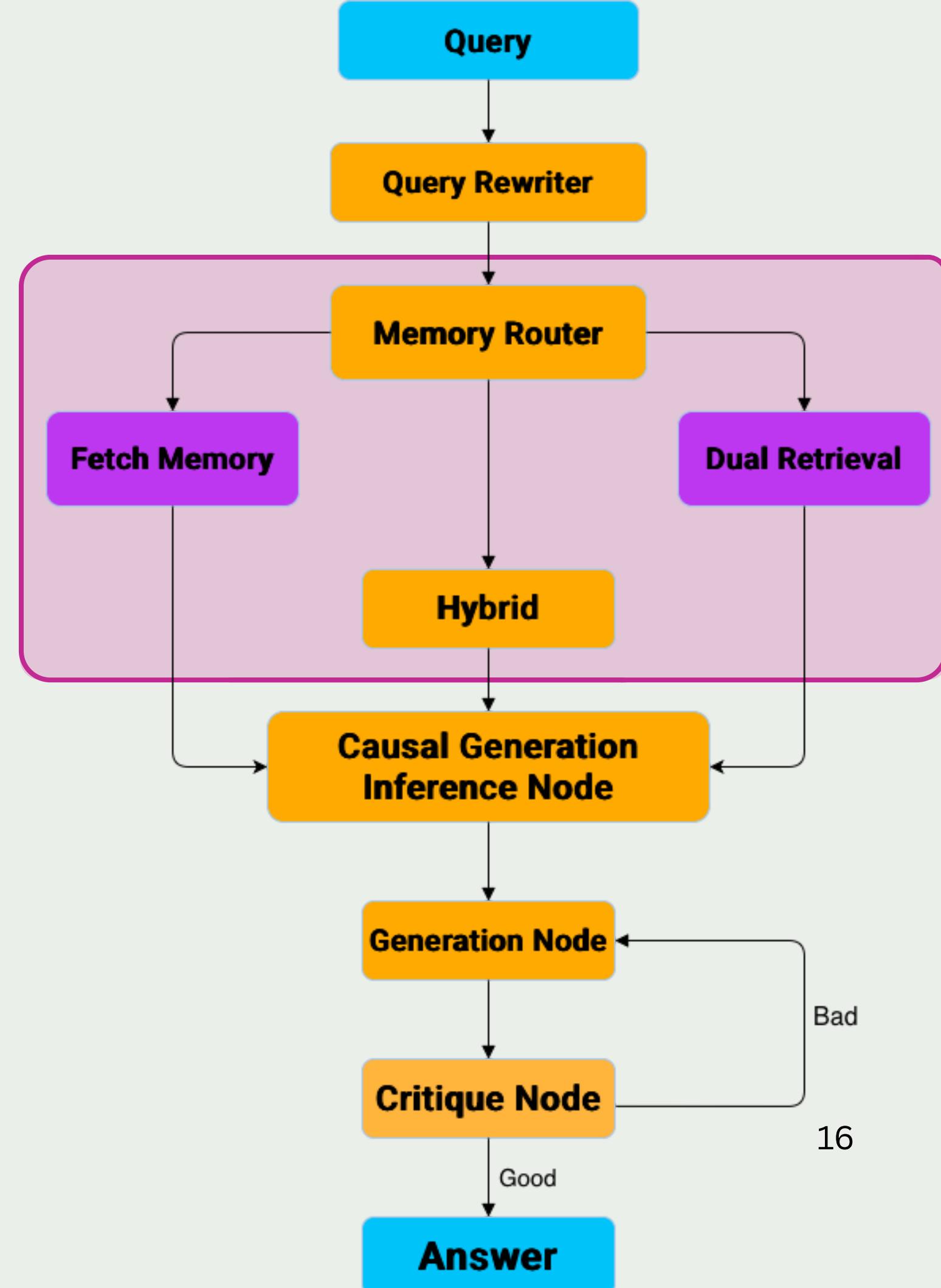


- Build upon the system developed in Task 1.
- It should use past context for follow-up questions.



Memory

- Generate a detailed summary with keywords and transcript IDs, avoiding storage of full transcripts or past answers.
- A router checks if a follow up is context related or new.
- Fetches only relevant past docs & new ones if required.



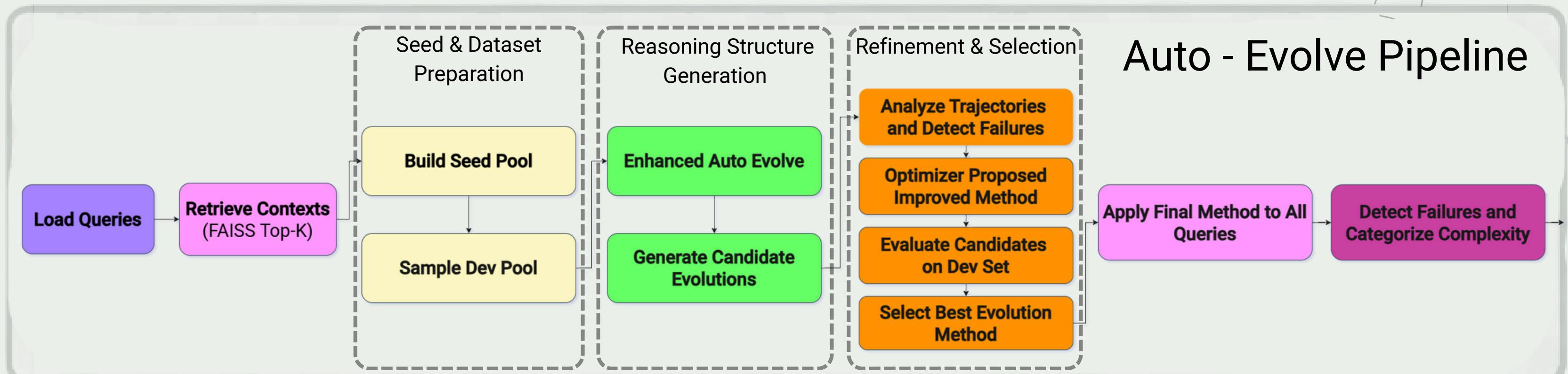
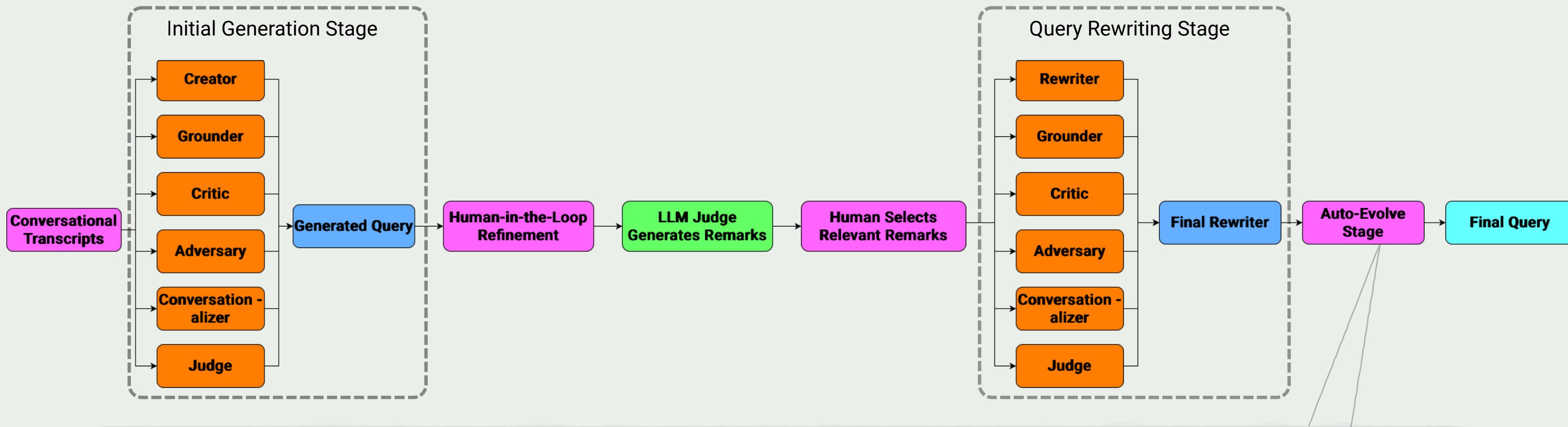
Demovideo

```
[12:54:59] INFO Connected to Neo4j  
GraphStore.  
INFO Loading transcripts from  
C:\Team_24\observe_ai\data\labelled_transcript_bert.json  
  
[12:55:01] INFO Loaded 19615 transcripts  
into memory.  
[12:55:05] INFO CausalPipeline: Loading  
resources...  
INFO Loading schema...  
INFO Loading dataset...  
INFO Loading SCM bundle...  
C:\Program Files\WindowsApps\PythonSoftwareFound  
ation.Python.3.12_3.12.2800.0_x64_qbz5n2kfra8p0  
\Lib\pickle.py:1760: UserWarning: [12:55:06] WAR  
NING: C:\actions-runner\_work\xgboost\xgboost\s  
c\data\../common/error_msg.h:83: If you are load  
ing a serialized model (like pickle in Python, R  
DS in R) or  
configuration generated by an older version of X  
GBoost, please export the model by calling  
'Booster.save_model' from that version first, th  
en load it back in current version. See:  
  
    https://xgboost.readthedocs.io/en/stable/tut  
orials/saving\_model.html  
  
for more details about differences between savin  
g model and serializing.  
  
setstate(state)  
[12:55:06] INFO CausalPipeline initialized.  
INFO Observe AI system  
initialized successfully  
INFO Finished __init__ in 23.16s  
INFO Observe AI system  
initialized successfully.  
INFO: Application startup complete.  
INFO: Uvicorn running on http://0.0.0.0:8000  
(Press CTRL+C to quit)  
[12:55:49] INFO Agent memory cleared.  
INFO ObserveAI memory cleared.  
INFO: 127.0.0.1:55131 - "POST /reset_memory  
HTTP/1.1" 200 OK
```

```
(venv) PS C:\Team_24\observe_ai> python \Interactive_client.py  
_____  
Observe AI Interactive Terminal Client  
_____  
Type 'exit' to stop.  
Type 'reset' to clear conversation memory.
```

User:

Query Generation



Evaluation

RAGAS

- Scores grounding quality: relevance, faithfulness, context recall.
- Ensures retrieved evidence is correct and complete.

Rubric Metrics

- Structured scoring for explanation quality: causal accuracy, clarity, actionability.
- Enables consistent, multi-dimensional evaluation.

LLM-as-a-Judge

- Evaluates end-to-end reasoning beyond string matching.
- Checks semantic correctness, causal logic, and hallucinations.

Results

Why our solution stands out?

- Our system is very less likely to hallucinate due to Critique Node.
- Retrieval is further improved due to inclusion of keywords by Query Rewriter.
- Dual-path retrieval (semantic + structural) grounds explanations in transcript windows, making every output traceable.
- SCM captures stable, interpretable cause-and-effect links across conversations and grounds the result with numerical evidence.
- Counterfactual outputs improved the depth and reliability of explanations.

Future Works

- Multiple small ML predictors can be trained to make a more sophisticated pipeline design.
- Integrating reinforcement learning techniques to dynamically optimize intent transition probabilities and retrieval weighting parameters based on conversation outcomes. This could enable the system to continuously improve its retrieval strategy.
- We can use Multi-Agent Evolve: LLM Self-Improve Through Co-Evolution's principle to iterate and improve the overall pipeline to get an automatically self-improving system that continually detects issues, proposes fixes, and enhances performance.

Future Works

- Instead of relying solely on an LLM, we can use structured output validation, regex based checks, and context answer semantic similarity to reduce hallucinations, at the critique node.

References

1. Ziqi Zhu, et al., “Conversational Intent-Driven GraphRAG: Enhancing Multi-Turn Dialogue Systems through Adaptive Dual-Retrieval of Flow Patterns and Context Semantics”, 2025 [<https://arxiv.org/abs/2506.19385v1>]
2. Darren Edge, et al., “From Local to Global: A Graph RAG Approach to Query-Focused Summarization”, 2024 [<https://arxiv.org/abs/2404.16130>]
3. Xinpeng Zhao, et al., “SEER: Self-Aligned Evidence Extraction for Retrieval-Augmented Generation”, 2024 [<https://arxiv.org/abs/2410.11315>]
4. Parth Sarthi, et al., “RAPTOR: Recursive Abstractive Processing for Tree-Organized Retrieval”, 2024 [<https://arxiv.org/abs/2401.18059>]
5. Weihao Zeng, et al., “Automatic Instruction Evolving for Large Language Models”, 2024 [<https://arxiv.org/abs/2406.00770>]

*Thank
You*