Conversational AI Assignment 2

Hybrid RAG System with Automated Evaluation

Group ID: 45

BITS Students details:

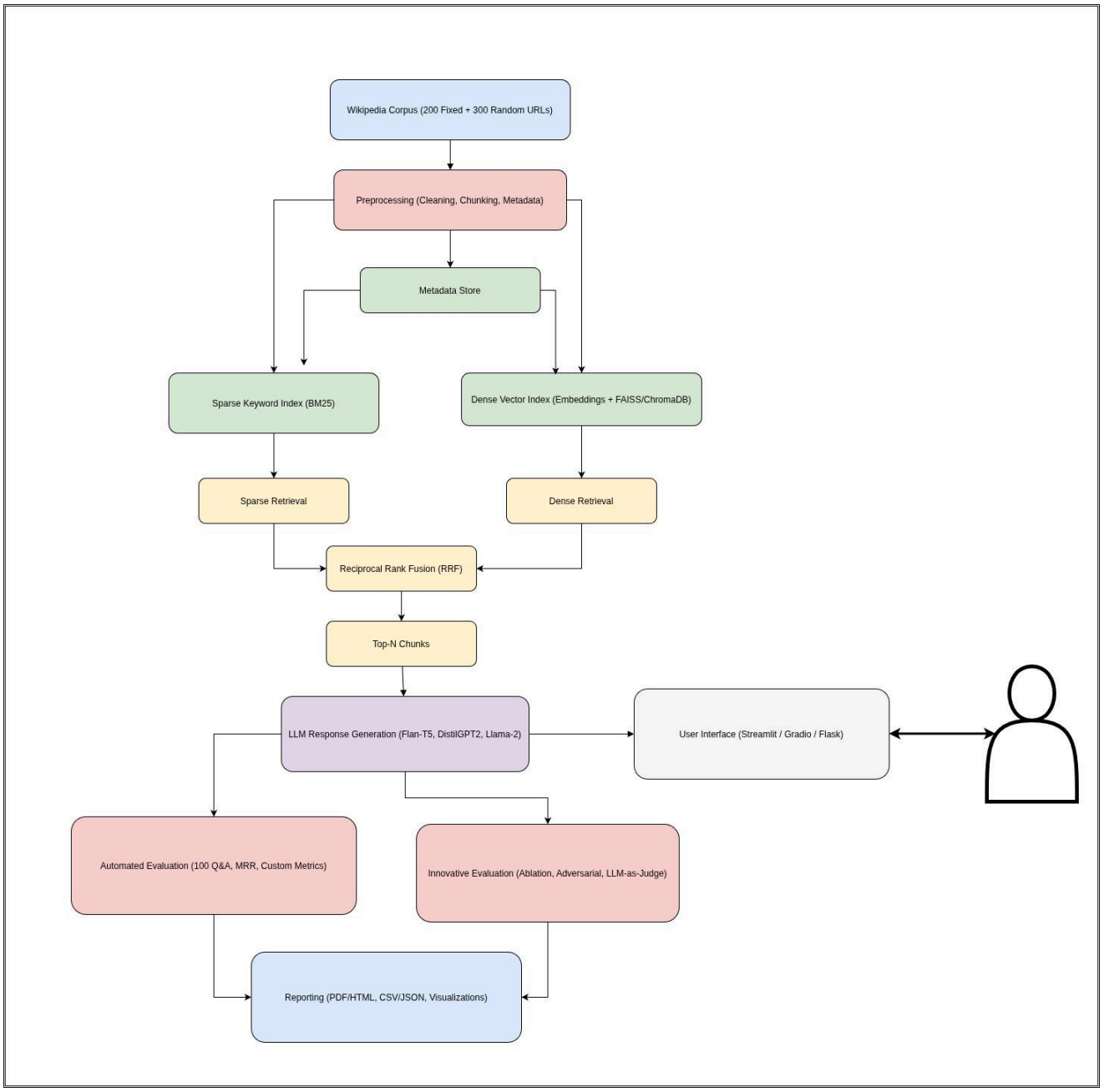
Name	Student ID	Contribution
NITESH KUMAR	2024AA05143	100%
LAWLESH KUMAR	2024AA05149	100%
VAIBHAV SAREEN	2024AA05923	100%
VIVEK TRIVEDI	2024AA05922	100%
LOGESH M	2024AA05163	100%

Git Hub Repo Link :

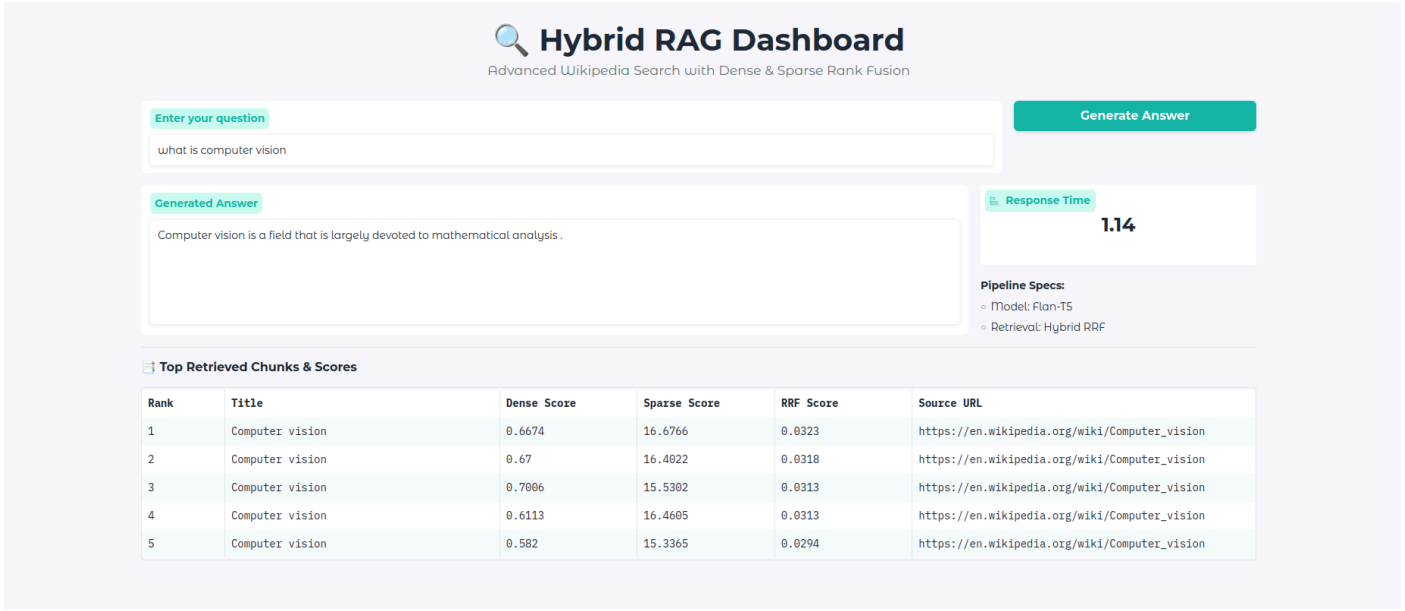
https://github.com/cseniteshkumar/bitsMtech_CAI_Assignment-2

Please move to the next page(s) for more details.

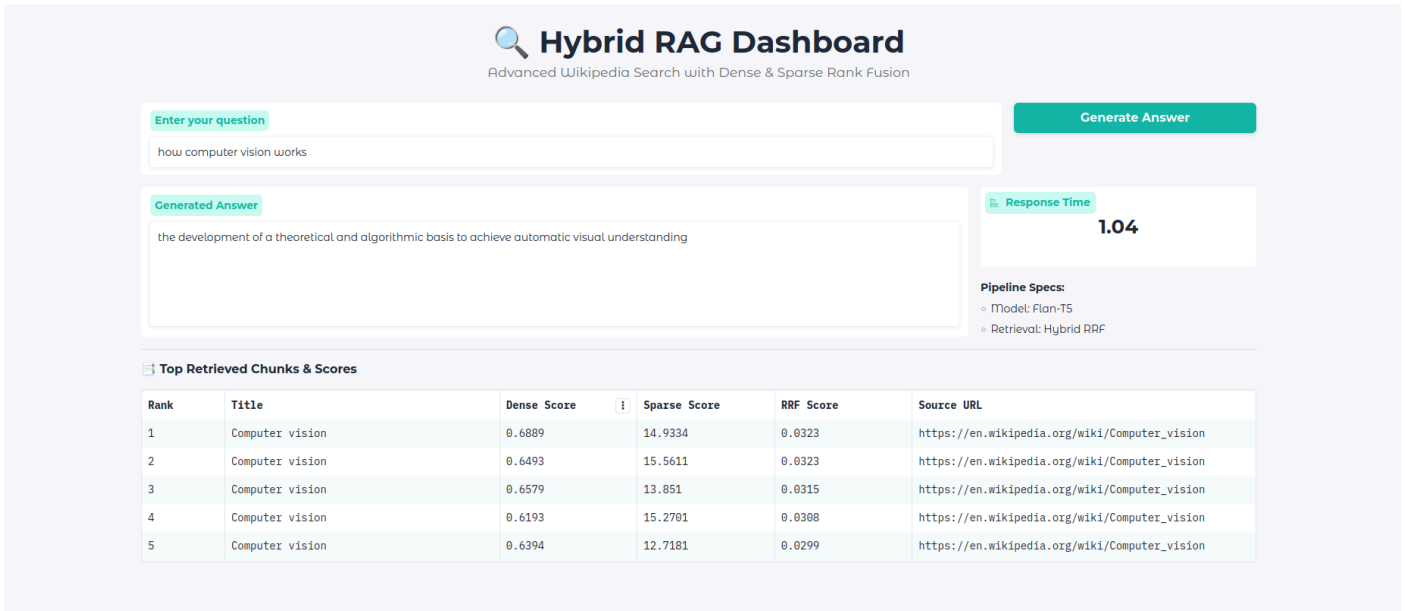
Solution Architecture and Approach :



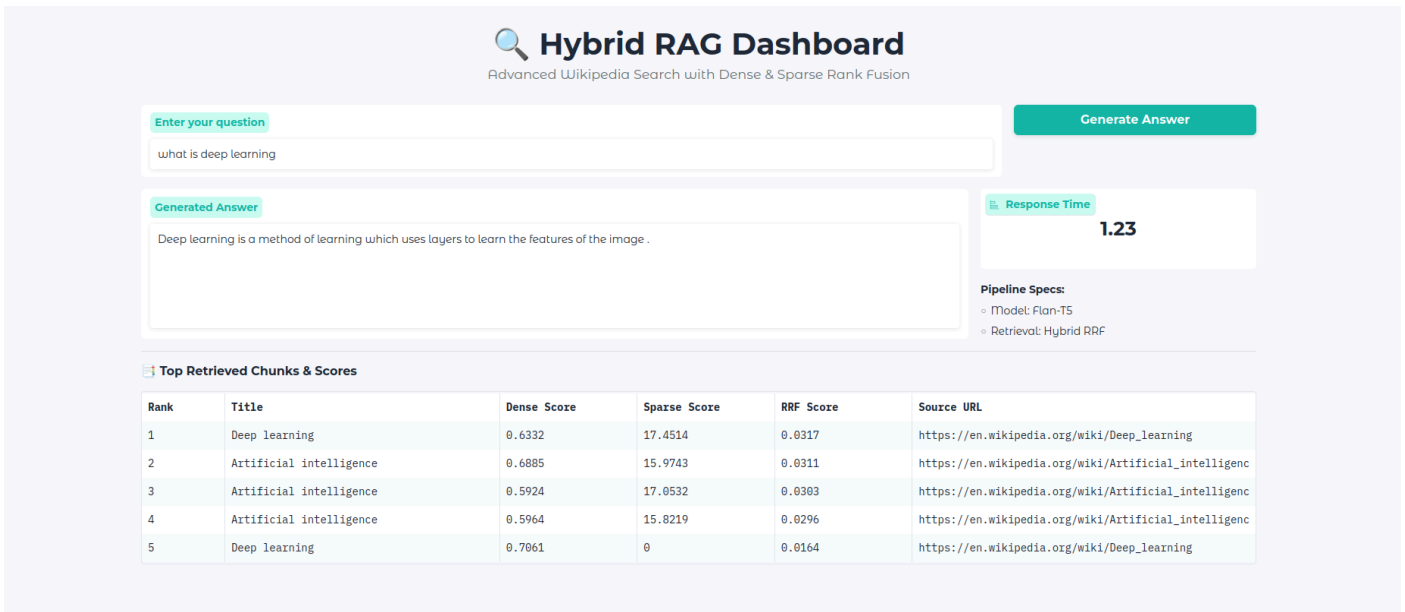
Gradio Interface SS:



Rank	Title	Dense Score	Sparse Score	RRF Score	Source URL
1	Computer vision	0.6674	16.6766	0.0323	https://en.wikipedia.org/wiki/Computer_vision
2	Computer vision	0.67	16.4022	0.0318	https://en.wikipedia.org/wiki/Computer_vision
3	Computer vision	0.7006	15.5302	0.0313	https://en.wikipedia.org/wiki/Computer_vision
4	Computer vision	0.6113	16.4605	0.0313	https://en.wikipedia.org/wiki/Computer_vision
5	Computer vision	0.582	15.3365	0.0294	https://en.wikipedia.org/wiki/Computer_vision



Rank	Title	Dense Score	Sparse Score	RRF Score	Source URL
1	Computer vision	0.6889	14.9334	0.0323	https://en.wikipedia.org/wiki/Computer_vision
2	Computer vision	0.6493	15.5611	0.0323	https://en.wikipedia.org/wiki/Computer_vision
3	Computer vision	0.6579	13.851	0.0315	https://en.wikipedia.org/wiki/Computer_vision
4	Computer vision	0.6193	15.2701	0.0308	https://en.wikipedia.org/wiki/Computer_vision
5	Computer vision	0.6394	12.7181	0.0299	https://en.wikipedia.org/wiki/Computer_vision

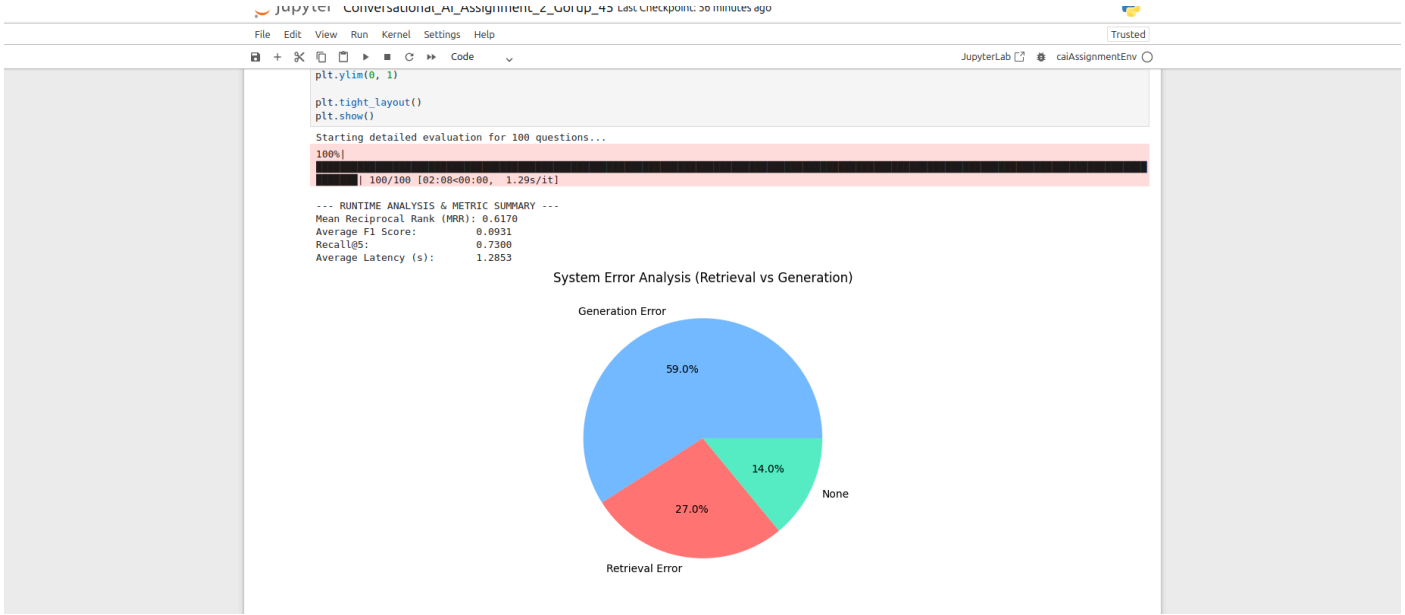


Rank	Title	Dense Score	Sparse Score	RRF Score	Source URL
1	Deep learning	0.6332	17.4514	0.0317	https://en.wikipedia.org/wiki/Deep_learning
2	Artificial intelligence	0.6885	15.9743	0.0311	https://en.wikipedia.org/wiki/Artificial_intelligenc
3	Artificial intelligence	0.5924	17.0532	0.0303	https://en.wikipedia.org/wiki/Artificial_intelligenc
4	Artificial intelligence	0.5964	15.8219	0.0296	https://en.wikipedia.org/wiki/Artificial_intelligenc
5	Deep learning	0.7061	0	0.0164	https://en.wikipedia.org/wiki/Deep_learning

Performance Comparison Table :

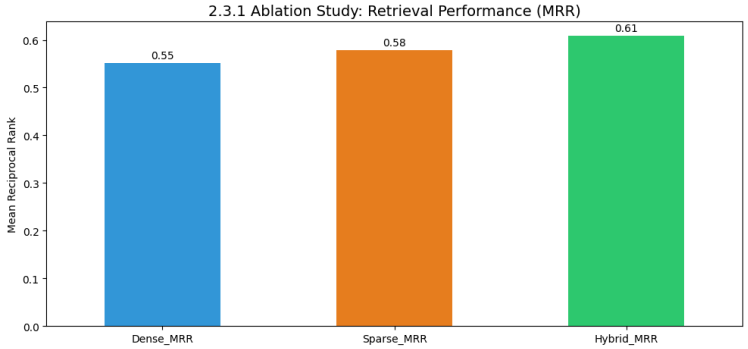
Method	Strength
BM25 Only	Fast, good for exact names/IDs
Dense Only	Handles synonyms well
Hybrid (RRF)	Best of both worlds
Hybrid + Re-rank	Maximum precision

Execution SS:

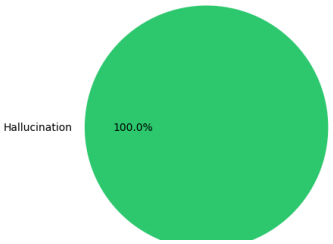


```
100%|
| 100/100 [02:07<00:00,  1.28s/it]
```

2.3.1 Ablation Study: Retrieval Performance (MRR)

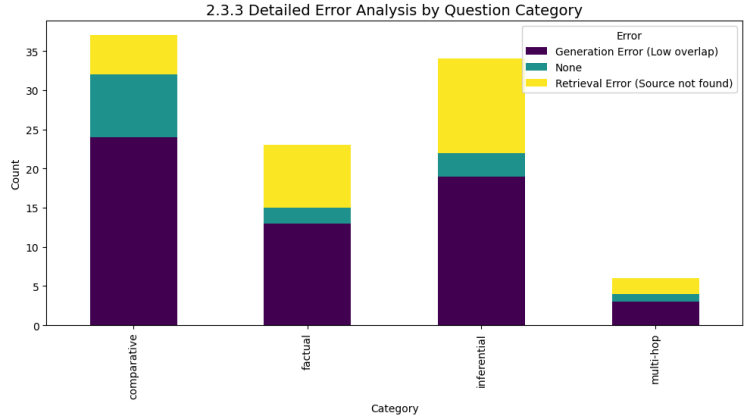


2.3.2 Hallucination Detection (Unanswerable Questions)



jupyter Conversational_AI_Assignment_2_Gorup_45 Last Checkpoint: 56 minutes ago
File Edit View Run Kernel Settings Help Trusted
Code JupyterLab caiAssignmentEnv

2.3.3 Detailed Error Analysis by Question Category




```
--- Hallucination Test Samples ---
```

	Question	Status	Answer
0	What is the secret recipe for Martian space-fuel?	Hallucination	The text is a list of references and bibliography of articles and books.
1	Who was the King of the United States in 2024?	Hallucination	The narrator believes that the French Revolution was a success .
2	How do you bake a cake using only binary code?	Hallucination	The narrator 's purpose in writing this passage is to provide information about the implementation of an instruction set architecture (ISA) .

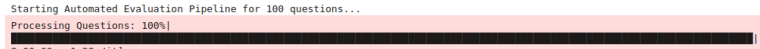
"**Hallucination Detection**" plot refers to measuring how effectively the RAG system handles adversarial or unanswerable questions.

jupyter Conversational_AI_Assignment_2_Gorup_45 Last Checkpoint: 56 minutes ago

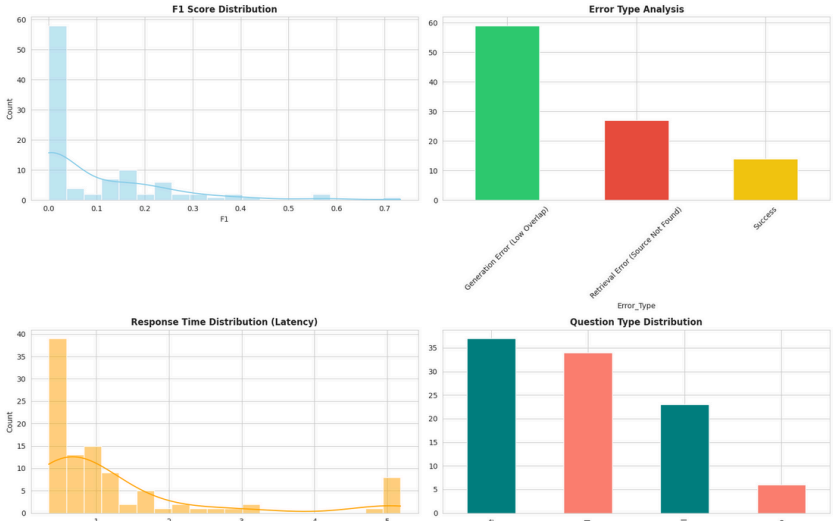
File Edit View Run Kernel Settings Help Trusted

Code JupyterLab caiAssignmentEnv

```
Starting Automated Evaluation Pipeline for 100 questions...
Processing Questions: 100%|
```



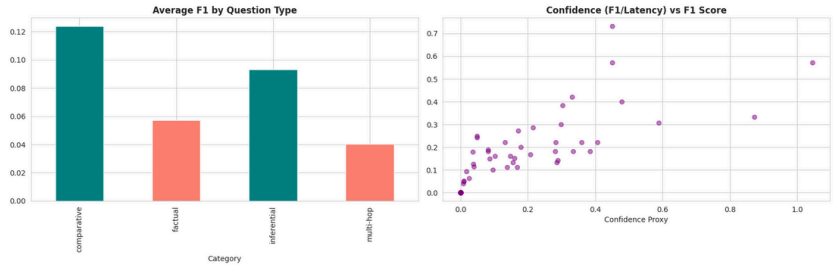
```
| 100/100 [02:08<00:00,  1.28s/it]
Files saved: 'evaluation_results_full.csv', 'evaluation_summary.json'

--- PERFORMANCE SUMMARY ---
Min Latency: 0.345s
Max Latency: 5.185s
Mean Latency: 1.280s
Median Latency: 0.800s
```

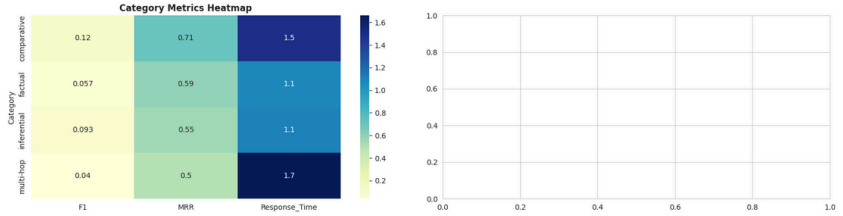


jupyter Conversational_AI_Assignment_2_Gorup_45 Last Checkpoint: 56 minutes ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab caiAssignmentEnv





[ ]:

[ ]:

[ ]:

[ ]: