

Table1: Comparison Matrice of Average Scores of LLMs

LLM	Relevance	Faithfulness	Context Recall	Context Precision
LLama3	5.10	3.70	4.10	5.30
Mistral	5.30	5.30	6.10	4.30
Gemma2	5.10	3.70	4.10	5.30
Phi3	4.30	4.10	5.20	4.00
Qwen2	6.80	4.90	5.70	5.80

Table 2: Highest Scores Matrice

Metric	Highest Score	LLM
Relevance	6.80	Qwen2
Faithfulness	5.30	Mistral
Context_Recall	6.10	Mistral
Context_Precision	5.80	Qwen2

Table 3: Lowest Score Matrice

Metric	Lowest Score	LLM
Relevance	4.30	Phi3
Faithfulness	3.70	Gemma2
Context_Recall	4.10	Gemma2
Context_Precision	4.00	Phi3

Table 5: Comparison Matrice of Overall Performance Ranking

Rank	LLM	Average Score
1	Qwen2	5.80
2	Mistral	5.25
3	LLama3	4.83
4	Gemma2	4.55
5	Phi3	4.40

Table 6: Comparison Matrice of Average of Overall Performance Ranking

Rank	Relevance	Faithfulness	Context Recall	Context Precision
1	Qwen2 (6.80)	Mistral (5.30)	Mistral (6.10)	Qwen2 (5.80)
2	Mistral (5.30)	Qwen2 (4.90)	Qwen2 (5.70)	Gemma2 (5.30))
3	LLama3 (5.10)	Phi3 (4.10)	LLama3 (5.70)	Mistral (4.30)
4	Gemma2 (5.10)	LLama3 (4.20)	Phi3 (5.20)	LLama3 (4.30)
5	Phi3 (4.30)	Gemma2 (3.70)	Gemma2 (4.10)	Phi3 (4.00)