Assignment 14
Evaluation of LLMs & RAG

Part C — Building GenAI Applications

10 questions

★ **Must-know** — asked very frequently in GenAI / ML interviews (5 questions). Prepare these first.

1. How do you evaluate an LLM application? (Offline vs online, reference-based vs reference-free.) ★ · Asked: Infosys, Google
2. What is RAGAS? What metrics evaluate the retriever and the generator? ★ · Asked: Google
3. Explain faithfulness, answer relevancy, context precision and context recall. ★
4. What is LLM-as-a-judge? What are its biases and how do you reduce them? · Asked: Google
5. Explain BLEU, ROUGE, METEOR and perplexity. Where is each used?
6. What metric is used to evaluate speech-to-text (WER, CER)?
7. How do you build a golden evaluation dataset? · Asked: Google
8. How do you measure hallucination rate in production? · Asked: TCS, Google
9. How do you evaluate retrieval quality? Explain Precision@K, Recall@K, MRR (and nDCG). ★ · Asked: TCS
10. How do you compare two prompt versions, and what are regression testing and shadow testing for LLM apps? ★ · Asked: Google