

AI Engineer Interview – All Questions

Rahul, here are all the questions covered in your interview conversation, in order.

1. LangGraph & AI Auto Assist

Show the flow from the user question to the final response.

2. Explain the end-to-end architecture of AI Auto Assist.

Can you explain LangGraph state management and the objects passed across nodes?

3. Did you use any conditional routing?

4. How was an unsupported question handled?

5. What strategy did you use for context management?

6. RAG & Retrieval

7. RAGAS and DeepEval

8. Discussed evaluation frameworks for RAG pipelines.

9. Suppose retrieval returns the correct document, but the LLM still gives the wrong answer. How will you debug it?

10. How do you decide whether to select Top-10, Top-12, or Top-50 documents?

11. How do you discriminate which documents are the most relevant when using cosine similarity or semantic search?

12. Python Coding & Cosine Similarity

13. Rate yourself on a scale of 1 to 5 in Python coding.

14. Implement cosine similarity to return the top two documents out of four documents.

15. What does the query embedding tell us, and what does each document embedding tell us?

16. What exactly is the output of cosine similarity based on?

17. Did you use built-in libraries in your code?

18. Can we use NumPy?

19. Azure Services

Which Azure services have you used?

Discussed Azure OpenAI, Azure AI Search, Blob Storage, and Azure AI Foundry.