

GenAI / Hallucination / Evaluation

1. Could you tell me a little bit more about building an AI response evaluation and hallucination detection framework? Explain the hallucination detection framework.
2. What would be your sampling size for detecting hallucination? What sample size have you configured in your framework?
3. For a chatbot, if I ask a question and receive a response, how do you ensure whether the response is accurate or not?
4. Would 20–30% of queries cover all the questions in the domain?
5. How can you say that 30% of questions will cover the entire domain and that the LLM won't hallucinate around that particular domain?

Azure / Deployment / Infrastructure

6. Where exactly is your application deployed?
7. Are you aware of the infrastructure side—how your application and components are deployed on Azure?
8. How was your RAG application deployed on the Azure infrastructure side?
9. Where are you storing the document chunks?
10. Are both chunking and metadata stored in Azure AI Search, or how exactly are they stored?

RAG / Document Processing

11. Suppose you have a PDF containing heterogeneous data such as images, HTML, hyperlinks, and circular hyperlinks. What chunking strategy would you follow?

LangChain / LLM Integration

12. Do you need to write a program to integrate an LLM model using LangChain?
13. Can you write a simple piece of code that can connect to an LLM using LangChain?
14. Can you put that code inside a class and method/function so that it is reusable?

SOLID / OOP

15. What SOLID principles are you following in your coding?
16. If one class has nine functions, does that violate the Single Responsibility Principle?
17. If a class has nine functions, is that automatically a violation of SRP?

Design Patterns

18. Are you aware of any design patterns from your development experience?

API / FastAPI

19. Suppose `process_payment` is one function and you need to expose it as an API. What would be the basic API contract for the request and HTTP response code?
20. With POST, what do we need to return—successful code, status code, and so on?
21. Can you implement the status code?
22. What would the API contract look like without implementation?
23. If your application depends on an API Gateway and something goes wrong with the API Gateway, and you catch it using a custom exception, what HTTP status code would you recommend for that custom exception?