

1. **What is the difference between LLM Agents and RAG? Explain a simple RAG pipeline.**
2. **How do you route a request to an agent? What happens if none of the agents can handle the request?**
3. **If two agents can handle the same request, how do you decide which agent should handle it?**
4. **What is agent-to-agent handoff? How does it work?**
5. **Are you aware of AWS Bedrock? How does it compare with Microsoft AI / Azure OpenAI?**
6. **For a Senior AI Engineer position, how would you position yourself?**
7. **Suppose your platform has its own metadata standards, but a new AI agent comes from another platform and doesn't follow your standards. How would you integrate it?**
8. **Suppose there are two different agents from different platforms—one is responsible for planning and the other for execution. How would you design the communication between them?**
9. **If the execution agent silently fails, how would you catch the failure? What would be your design for detecting and handling it?**
10. **How would you decide which agent should handle which kind of request? How would you design agent selection/routing?**