

User:

Hello?

Assistant:

Hello! Thank you for taking the time to speak with me today about the Entry Level RAG Engineer position. I'd like to ask you a few questions to learn more about your experience and skills. Could you tell me about your experience with retrieval-augmented generation (RAG) pipelines?

User:

Yes, sure. I have some experience in retrieval of augmented generation pipelines. Mainly I used RAG to do my final thesis project, which is an automated screening system. I used RAG for the verification process to put up-to-date information from the internet in the context of an LLM so that it can verify all the information that is available currently.

Assistant:

That's interesting. Can you elaborate on the specific challenges you faced while implementing RAG in your thesis project, and how you overcame them?

User:

Sure, so breaking down a candidate's answer into queries wasn't my first approach. I couldn't find a way to break down the answer or like provide a search query for getting information on the relevant topics and then through query decomposition I got relevant answers like by breaking down the string into sub queries and then searching for it. That expanded the search space so that is how I solved it.

Assistant:

Thank you for sharing that. Moving on, can you describe your experience with different large language models (LLMs) and how you've compared their performance for specific tasks?

User:

I'm sure I've worked with JadGPT, Claude, Sonnet and Lama. I would rank them as Claude Sonnet is the highest for reasoning, JadGPT comes in second and Lama comes in third.

Assistant:

Thank you for your time. We'll be in touch regarding the next steps in the interview process.