

User:

Hello

Assistant:

Hello! Thank you for taking the time to speak with me today about the Entry Level RAG Engineer role. I'd like to start by asking you a couple of questions about your experience with the latest Large Language Models. Could you tell me about your familiarity with recent LLMs like GPT-4 or Claude?

User:

I'm sure I've used both GPT-4, that is GPT-4o mini and GLOD 3.5 sonnet which are the latest LLM, latest and largest LLM models produced till date. GPT-4o mini is a pruned down version of GPT-4o, whereas GLOD 3.5 sonnet is highly regarded as the most intelligent LLM till date.

Assistant:

I see. Could you describe a specific project or task where you've applied these LLMs in a RAG (Retrieval-Augmented Generation) context?

User:

I'm sure. So I've used Claude 3.5 Summit and its impressive reasoning capabilities to design an automated screening interview for candidates where they can enter the specified role and the job description. And based on that Claude Summit 3.5 comes up with questions to ask the candidate and at the end evaluates the candidate based on their answers and the job description.

Assistant:

Interesting. Can you explain how you implemented the retrieval aspect of RAG in this project? What sources did you use to augment the LLM's knowledge?

User:

I'm sure so I've used Claw 3.5 Sonnet as the base LLM and used Lang chain and Lang graph to build a rag pipeline and using that I web script topics based on the answers provided by the candidates and loaded it onto the context of the LLM namely Claw 3.5 Sonnet. I've also used other rag chains like Lama Index and other LLM models such as Lama 3.1 which has I believe an 8 billion token context.

Assistant:

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