Assignment 2 – Response

1. Self-Rating on AI, ML, Deep Learning, and LLMs

At this stage, I would say that my level in Machine Learning is a B. Currently, I am able to perform on simple data and write simple code related to machine learning with the help of others.

When it comes to Artificial Intelligence, Deep Learning, and Large Language Models, I would rate myself as C. I understand the concept of these subjects, but I have limited knowledge regarding these. I am actively learning to grow in these subjects.

2. High Level Architecture of an LLM-Based Chatbot

In simple words, a chatbot based on LLM involves processing a user query entered through an interface, and after processing, it sends this query to a language model so it receives a response based on this query.

Depending on whether further information is required by the chatbot, this information can be obtained from stored sources. The final reply is then formatted and displayed.

Total flow: User Input → Processing → LLM → Response → User

3. Vector Databases and Selection

A vector database stores data in numerical form. This will, in turn, enable the system to search for information based on similarity, rather than just word-to-word matching. This is particularly useful in chatbots and other search-related applications. Qdrant would be my vector database of choice for a simple chatbot case. I like the fact that it is open-source, works well with Python, and is good for learning and small projects.

Final Note

These responses reflect my current level of understanding and learning journey. I've tried to explain each concept clearly and honestly based on my learning thus far.

Name: Vishal Kumar