Task 2 : Detail two innovative techniques for optimising the RAG model developed in Task 1.

1. **Refining Vector Embeddings for Domain-Specific Insights**  
In my current workflow, I utilize a pre-trained model (deepset/roberta-base-squad) to generate embeddings, which are then stored in Pinecone for efficient retrieval. While this setup performs effectively for general-purpose queries, it can be further improved by fine-tuning the embeddings specifically for the domain of Yardstick Software Solutions. This customization would enable the model to gain a deeper understanding of the types of questions and answers related to the services and industries I aim to cover, ensuring greater relevance and accuracy in its responses.

2. **Enhancing Search with Query Expansion**  
Another technique I would adopt is Query Expansion to improve retrieval precision. At present, the model fetches answers based on an exact match between the user’s query and the stored embeddings. However, since users often phrase similar questions differently, it may occasionally struggle to find the most appropriate answer. Query Expansion addresses this issue by enriching the user's input with additional context, increasing the likelihood of retrieving the most relevant response regardless of how the question is phrased.