

## AI Solutions Architect: Capstone Project: Part 1

A large technology company wants to develop a state-of-the-art generative AI system capable of producing high-quality, coherent, and factual content on a wide range of topics on its products. The system, which will replace the search capability on the corporate homepage, should leverage AI retrieval-based methods to gather relevant information from various sources and produce human-like outputs. The key requirements include:

- 1) Build a comprehensive knowledge base by retrieving and indexing information from trusted sources, such as product documents, technical documentation.
- 2) Implement advanced natural language processing techniques to understand user queries and extract relevant context and intent.
- 3) Utilize retrieval algorithms to identify and retrieve the most relevant information from the knowledge base based on the user's query.
- 4) Incorporate large language models and generation techniques to synthesize the retrieved information into coherent, fluent, and contextually appropriate responses.
- 5) **Ensure factual accuracy** of the output by cross-checking generated content against the knowledge bases.
- 6) Implement techniques to control the style, tone, and formality of the generated content based on the user's preferences or the context of the query.
- 7) **Provide explainable and transparent outputs**, clearly indicating the sources of information used.
- 8) **Design a modular and extensible architecture that allows for easy integration of new data sources, language models, and generation techniques.**
- 9) Implement robust security measures to prevent the system from generating harmful, biased, or offensive content.
- 10) **Continuously monitor and evaluate the system's performance using human evaluations, automated metrics, and user feedback.**
- 11) **Ensure that the solution is only available to authorized users**
- 12) Consider the costs for the solution. How do they scale with increases or decreases in use by end-users, increases in data store size. Ideally, costs will scale with use.

Your assignment is to consider the above, draft a designs for the solution and answer the following questions.

1. Create a conceptual design of the solution, showing it in the context of the customer
2. Diagram the high-level processing-steps and data flow within the solution from user query input to solution output (Note: Not all aspects of the requirement need to be accounted for. The objective is to be able to explain to stakeholders how the main flow of solution works.)
3. Document 3 or more project specific risks and their mitigations
4. Consider notable 3rd party tools (e.g., models, frameworks, etc.) that you might use for the solution. Then, using the high-level processing-steps as reference, overlay the tools, to highlight you initial implementation decisions.
5. You need to return to the customer with the above artifacts before you can ask questions of specifics, as they will be incommunicado. You will need to make assumptions, based on the above and you industry knowledge. Are there questions that you would like to ask, so as to better ground you design? You quite possibly have many. For the assignment, briefly note the 3 most important questions, that you would ask the customer, when you meet with them to discuss you designs and progress.

Submit your answers in a single PDF document.