

Assignment Task for Interview Candidates: Machine Learning Role

Objective:

Design and implement a pipeline to build a contextual chatbot using data scraped from the website [GenZMarketing](#). The chatbot should be capable of answering user queries with accurate and relevant information derived from the website.

Task Description

You are required to:

1. Data Scraping

- Scrape the content from the website [GenZMarketing](#).
- Extract and organize the data in the following format:
 - `title`: The title of the webpage/article.
 - `url`: The URL of the specific page.
 - `content`: The main content of the page.

Deliverable: A CSV or JSON file containing the scraped data in the specified format.

2. Data Storage

- Process the scraped content and convert it into a vector representation suitable for retrieval.
- Store the vectorized data in a vector database like ChromaDB or any other database of your choice.

Deliverable: A Python script or notebook demonstrating the vectorization process and the data storage mechanism.

3. Retriever Design

- Use **LangChain** to design a retriever that can search through the vectorized data and provide the most relevant responses to a given query.
- Ensure the retriever is optimized for contextually accurate results.

Deliverable: A retriever module implemented with LangChain, including code and documentation.

4. Chatbot Implementation

- Implement a chatbot that uses the retriever to answer questions about the website.
- The chatbot should provide contextualized and relevant responses based on the content from the website.

Deliverable: A chatbot application (web-based or command-line interface) that demonstrates query answering.

Submission Guidelines

- **Code:** Provide all scripts in a well-documented format (preferably in Python).
 - **Documentation:** Submit a README file explaining:
 - Steps to run the code.
 - Tools and libraries used.
 - Any assumptions or design choices made.
 - **Evaluation:** Be prepared to discuss your approach, challenges faced, and potential improvements during the interview.
-

Expected Outcomes

- A functional chatbot capable of answering queries based on the content from [GenZMarketing](#).
- Demonstrated understanding of:
 - Web scraping techniques.
 - Data processing and storage in vector databases.
 - LangChain integration and retriever design.
 - Contextual chatbot implementation.

Good Luck!