Executive Summary

The *GitLab's Bot* is a Retrieval-Augmented Generation (RAG) chatbot designed to provide accurate, context-based answers by leveraging document data. It is tailored for efficient interaction with a wide variety of document formats, including PDFs, markdown files, HTML, and images containing text. The chatbot processes and indexes these documents, allowing users to ask questions and receive responses directly based on the document content.

At the core of this chatbot is a RAG approach, which combines a language model with a retriever to ensure that responses are grounded in relevant document sections. Text data is extracted, chunked, and transformed into embeddings using OpenAI's embedding model, which captures semantic meaning. These embeddings are stored in a vector database (Chroma) for fast similarity-based retrieval. Additionally, the chatbot utilizes memory management, enabling it to maintain conversational context within a session, thereby enhancing user experience through a more natural and coherent dialogue flow.

The chatbot's user interface, developed with the Panel library, features an intuitive multi-tab layout. Users can easily navigate through conversation, database information, chat history, and configuration tabs, which allow for document uploads and session resets. Feedback options embedded in the interface enable users to rate responses, helping to assess chatbot performance in real-time.

To evaluate the effectiveness of *Chatbot*, the quantitative metrics are employed. Quantitative metrics include answer relevance scores and response accuracy, while qualitative feedback will be added later which will include user ratings and session-end surveys. These evaluation methods will allow us for continuous improvement, ensuring that the chatbot remains accurate, responsive, and aligned with user expectations.

In conclusion, *Chatbot* exemplifies a document-driven, contextually aware chatbot that enhances information accessibility. By combining powerful retrieval methods, language generation, and a user-friendly interface, this chatbot provides a robust solution for knowledge retrieval and interaction, meeting the demands of users seeking reliable, document-based responses.