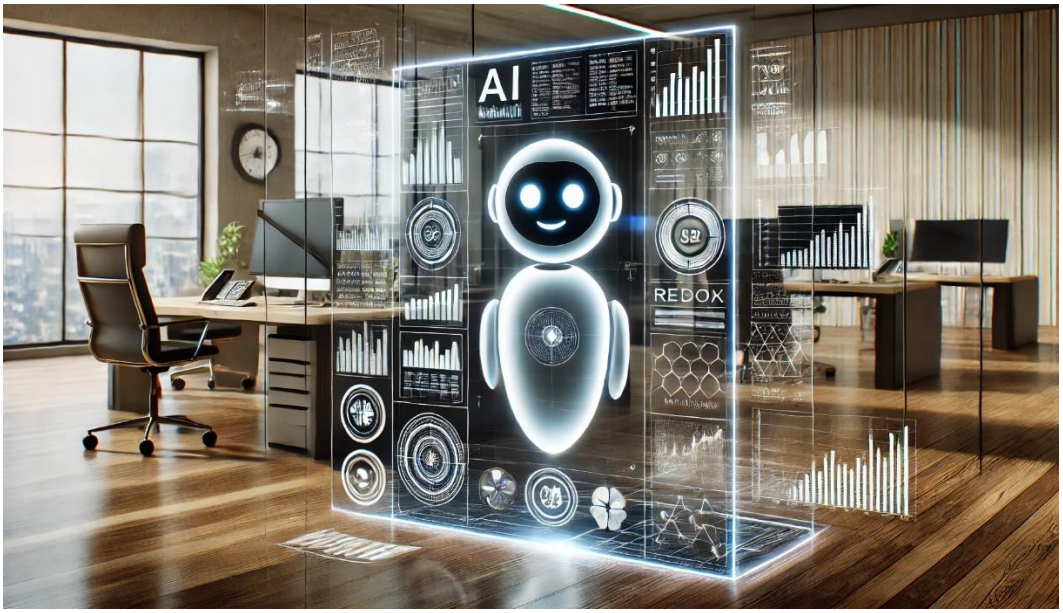
S P Jain School of Global Management

DUBAI • MUMBAI • SINGAPORE • SYDNEY

DGMFinanceBot: Intelligent Business News Analyzer

Individual Coding Challenge



PRESENTED TO:
Prof. Kallol Das

IN PARTIAL FULFILLMENT OF THE COURSE REQUIREMENT FOR
AI AND AUTOMATION IN FINANCE

Prepared by:
Denver G. Magtibay
MAIB April 2023 Online Batch

DGMFinanceBot: Intelligent Business News Analyzer

I. Problem Covered

Business professionals, analysts, and researchers often need to process large volumes of news articles to extract relevant information quickly. This task can be time-consuming and challenging, especially when dealing with multiple sources of information. The key challenges include:

- Volume of Information: Sifting through vast amounts of news articles daily.
- Relevance and Accuracy: Finding relevant and accurate information quickly.
- Manual Processing: Inefficiency and error-proneness of traditional methods.
- Timeliness: Staying updated with the latest information promptly.

II. High-Level Approach

DGMFinanceBot addresses these challenges by automating the processing and analysis of business news articles. The solution involves:

1. Data Loading: Users input URLs of news articles, which the tool loads and processes.
2. Text Splitting: The loaded articles are split into manageable chunks.
3. Embedding Creation: Embeddings are created for the processed text to enable efficient search and retrieval.
4. Question Answering: Users can ask questions related to the news articles, and the tool retrieves relevant answers using OpenAI's language model.
5. Source Attribution: The tool provides sources for the retrieved information, ensuring transparency and accuracy.

III. Tools Used

- Python: The programming language used for developing the tool.
- Streamlit: Framework for building the interactive web application.
- OpenAI: For natural language processing and question-answering capabilities.
- LangChain: Used for document processing and chaining various components.
- FAISS: For efficient similarity search and retrieval of documents.
- Pickle: For saving and loading the FAISS index.

IV. All References

1. OpenAI: https://www.openai.com/
2. Streamlit: https://streamlit.io/
3. LangChain: https://langchain.com/
4. FAISS: https://github.com/facebookresearch/faiss
5. Python: https://www.python.org/
6. Youtube: https://www.youtube.com/watch?v=MoqgmWV1fm8
7. Sample News Articles:
    - https://www.businesstimes.com.sg/companies-markets/telcos-media-tech/apple-spurned-idea-iphone-ai-partnership-meta-months-ago
    - https://www.businesstimes.com.sg/companies-markets/telcos-media-tech/apples-china-iphone-shipments-jump-40-after-steep-discounts
    - https://www.businesstimes.com.sg/companies-markets/refile-update-5-apple-okays-epic-games-marketplace-app-europe
    - https://www.businesstimes.com.sg/companies-markets/telcos-media-tech/epic-games-says-apple-stalling-launch-its-game-store-europe
    - https://www.businesstimes.com.sg/companies-markets/telcos-media-tech/apple-seeks-scrub-dominance-china-antitrust-app-ruling

V. Innovation

1. Expanded URL Inputs:

- **Original Code**: Allows input of only 3 URLs.
- **Improved Code**: Allows input of up to 5 URLs.

2. API Key Input Handling:

- **Original Code**: The API key is loaded from an environment variable.
- **Improved Code**: Directly assigns the OpenAI API key within the script for easier access during development and testing.

3. Usage of ChatOpenAI:

- **Original Code**: Uses the OpenAI model.
- **Improved Code**: Uses the ChatOpenAI model, specifically gpt-3.5-turbo, which is more suited for conversational AI.

4. UI Improvements:

- **Original Code**: Basic layout with minimal guidance for the user.
- **Improved Code**: Enhanced layout with titles, sections, and improved user guidance.

- Main title: "DGMFinanceBot: Business Research Tool 📈"
- Sidebar title: "News Article URLs"
- Sidebar markdown: "### Enter URLs to fetch news articles:"
- Main section for asking questions with clear instructions.

5. Debug Mode:

- **Original Code**: No debug mode.
- **Improved Code**: Added a checkbox in the sidebar to enable debug mode. When enabled, it displays detailed debug information like loaded data and split documents.

6. Error Handling and User Feedback:

- **Original Code**: Basic error handling with minimal feedback to the user.
- **Improved Code**: Enhanced error handling with specific error messages displayed to the user using st.error().
- Added success message to indicate when data processing and embeddings creation is complete.

7. Improved Comments and Documentation:

- **Original Code**: Minimal comments.
- **Improved Code**: Added detailed comments throughout the code to explain each step and make the code more understandable.

8. Consistent Naming and Style:

- **Original Code**: Mixed naming conventions.
- **Improved Code**: Consistent and descriptive variable names and titles, enhancing readability and maintainability.

Note: Approximately 50-60% of the original code was modified or improved.

VI. How Does Your Solution Benefit Industry

1. Efficiency: Automates the processing of large volumes of news articles, saving time and effort.
2. Accuracy: Provides accurate information by leveraging advanced NLP models.
3. Relevance: Ensures that only relevant information is retrieved and presented.
4. Transparency: Provides sources for all retrieved information, ensuring reliability.
5. Timeliness: Allows users to stay updated with the latest news and developments in real-time.

VII. Given Time and Resources – How Would You Suggest Your Solution CAN / or CANNOT be Converted to a Product?

Potential for Productization

With additional time and resources, DGMFinanceBot can be converted into a fully-fledged product by:

1. Enhanced Features: Adding more features such as personalized alerts, historical data analysis, and trend predictions.
2. Scalability: Improving the tool's scalability to handle more users and larger datasets.
3. User Interface: Enhancing the user interface for better user experience and easier navigation.
4. Integration: Integrating with other business tools and platforms for seamless workflow integration.
5. Security: Implementing robust security measures to protect user data and ensure compliance with industry standards.

Limitations

However, there are certain limitations that might affect its productization:

1. Data Access: Dependence on external sources for news articles, which might have access restrictions.
2. API Costs: Costs associated with using APIs like OpenAI for processing large volumes of data.
3. Model Limitations: Dependence on the accuracy and capabilities of the underlying NLP models.
4. Maintenance: Ongoing maintenance and updates required to keep the tool relevant and functional.

VIII. Documentation

https://drive.google.com/file/d/1Dq_H_wqiFLfySr2JbcdkfsOEzwsCh6J5/view?usp=sharing