Project 4: Building a Multimodal AI Chatbot

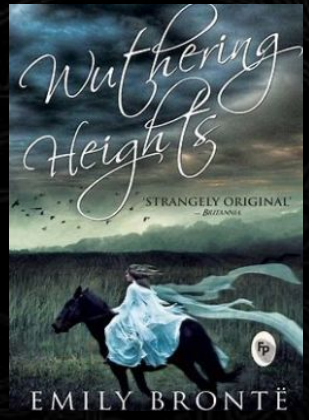
Saiqa Mehdi

Welcome to a realm where shadows breathe and every word holds the weight of forgotten sorrows & unspoken fears. Here the haunting beauty of gothic literature awaits, inviting you to wander the winding paths of words & unveil the secrets within.

Concept & Overview

A website for students or anyone passionate about gothic literature

- Discussion forum / book reviews
- Literary analysis of books & comparison
- Interaction with Characters



Website Mood Board

VEIL OF
SHADOWS

Technical Summary

Architecture

The solution is a multimodal AI chatbot designed to answer questions about *Wuthering Heights* based on YouTube video reviews.

- **Agents**: LangChain's Retrieval-Augmented Generation (RAG) agents are used for retrieving context and generating responses.
- **LLMs**: **Vector Database (Pinecone)**
- **Gradio Interface**:

Tools Used

- **LangChain**: Manages the text chunking, retrieval, and response generation processes.
- **Pinecone**: Stores vector embeddings for efficient retrieval.
- **OpenAI (GPT-4)**: Provides natural language responses based on retrieved content.
- **Whisper**: Transcribes audio input for the chatbot to process voice-based questions.
- **Gradio**: Used to create the user interface for both text and audio interactions.

Section	Component	Description
User Interaction	User (Top layer)	Two methods of interaction with chatbot: **Text Input** and **Audio Input**.
	Text Input	Directly connects to **Retrieval Agent**.
	Audio Input	Connects to **Whisper** for transcription, then sends the output to **Retrieval Agent**.
Text Processing	Retrieval Agent	Retrieves relevant text chunks from **Pinecone** based on the query.
	Retrieval Tool (retrieve_docs)	Fetches top-matching text chunks from **Pinecone**.
	Pinecone	Vector database for storing and retrieving relevant embeddings.
Q&A Agent	Conversation Buffer Memory	Stores user questions and generated answers to maintain conversational context.
	Save Conversation Tool	Saves the conversation history to **Memory** upon user command.
	Summarize Tool (optional)	Summarizes retrieved context if needed to fit within token limits.
	GPT-4	Generates answers based on the context retrieved from Pinecone.
Output Layer	Generated Text Answer	Arrows from **Q&A Agent** go to **Text Output** for displaying the answer.
	Text-to-Speech (TTS)	Converts the generated text answer to **Audio Output** with a haunting effect.
Gradio Interface	Gradio Box	Encloses **Text Input**, **Audio Input**, **Text Output**, and **Audio Output** to signify the user interface layer for all interactions.

Dataset

- Used the yt_dlp library to download audio directly from YouTube URLs
- FFmpeg issues
- OpenAI's Whisper model to transcribe audio
- Preprocessing

Q&A Agent (4 slides)

Agents

Retrieval Agent: This is used to fetch relevant document chunks based on user questions.

Q&A Agent: Initialized with LangChain to answer questions based on retrieved content, and is responsible for generating responses using GPT-4.

Tools

Transcribe_audio: -Whisper ASR model.

generate_haunting_audio: Converts text responses into haunting audio using gTTS and pydub.

retrieve_docs: Retrieves relevant text chunks based on the user's question by querying Pinecone.

save_conversation: Saves the conversation history to a file.

summarize_context (Optional): Summarizes long contexts to fit within token limits. This was set up in an earlier version of the code but may not be actively used in this Gradio code.

Memory Management:

ConversationBufferMemory: Used to store the chat history, which allows for follow-up questions by maintaining the context across interactions.

Giskard Evaluation

- **Data and Embedding Setup**
- **Giskard Dataset and Model Wrapping**
- **Detector Execution and Errors** : 7

Conclusion

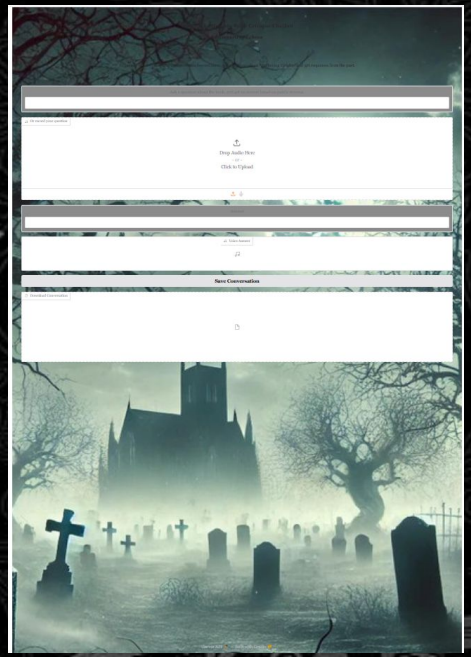
Enhanced Metadata: Adding timestamps or reviewer information for more targeted responses.

Prompt Engineering: Additional refinements to improve answer specificity and depth

Fine-Tuning & More Testing

Improve conversational interaction

Demo: Gradio Interface



Enter your question about Wuthering Heights

in wuthering heights, chapter 1, please give me a summary focusing on how literary devices are used

Or record your question

Drop Audio Here

- or -

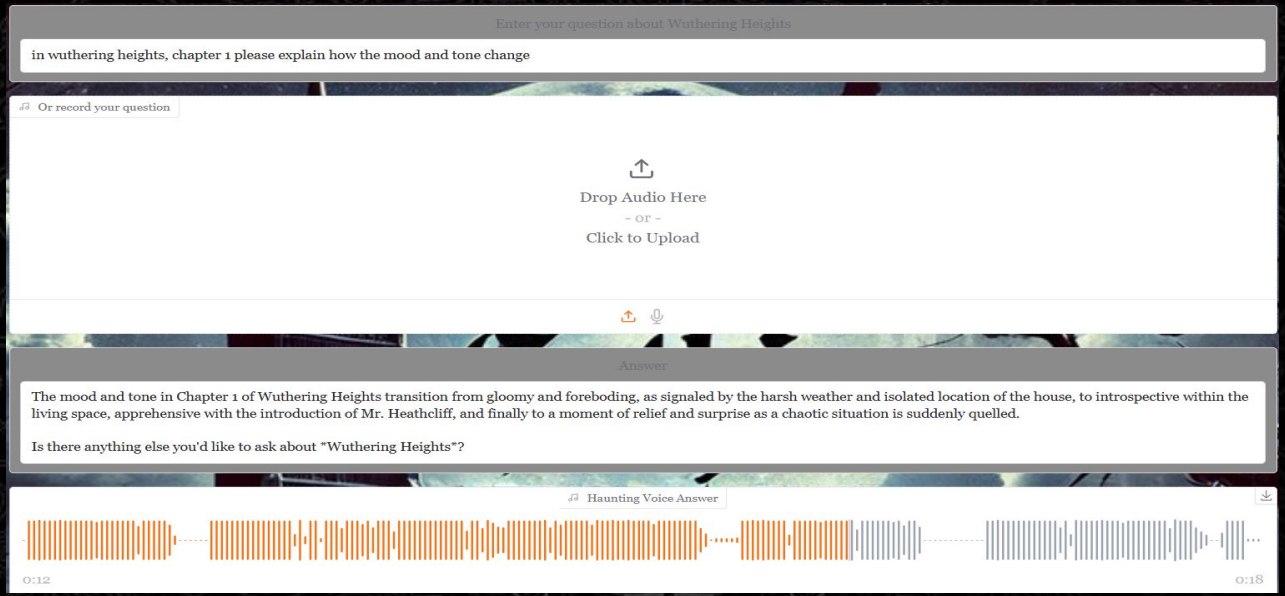
Click to Upload

Answer

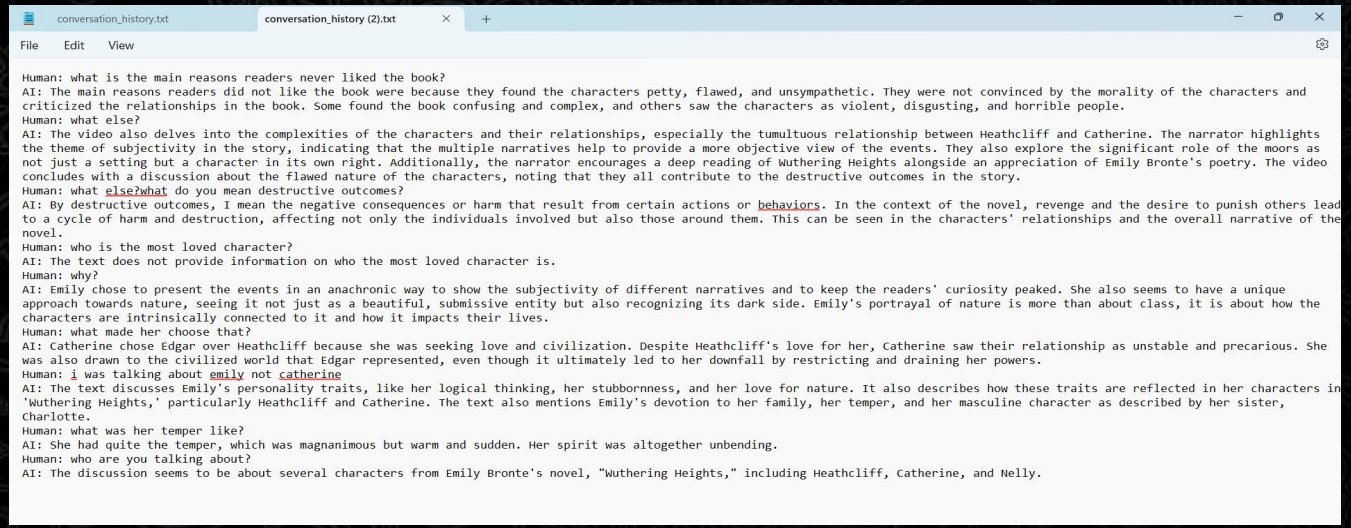
In Chapter 1 of Wuthering Heights, various literary devices are used. Imagery is used to vividly describe the house and its surroundings, setting the scene for the reader. Personification is used to describe characters, such as the dame who is likened to a storm and a sea. Contrast and irony are used to depict the character of Heathcliff, a man who doesn't fit into the typical representation of a farmer in the North of England. Finally, foreshadowing is used to hint at a future relationship between the narrator and Heathcliff.

Is there anything else you'd like to ask about "Wuthering Heights"?

Gradio Interface



Conversation Download



Thank you for you time

Any questions?