

The image features a series of thin, black, overlapping lines that form various geometric shapes, including triangles and polygons, scattered across the upper and middle portions of the frame. These lines are not connected to form a single shape but rather create a complex, layered pattern.

**VOICE CHAT
APPLICATION USING
GEMINI PRO MODEL**

AGENDA

Introduction

Project Overview

Technologies Used

WhisperModel

Google Generative AI

Microsoft T5 Model

Implementation

User Interaction

Demo

Future Enhancements

Conclusion

Q&A

Thank You



INTRODUCTION

INTRODUCTION

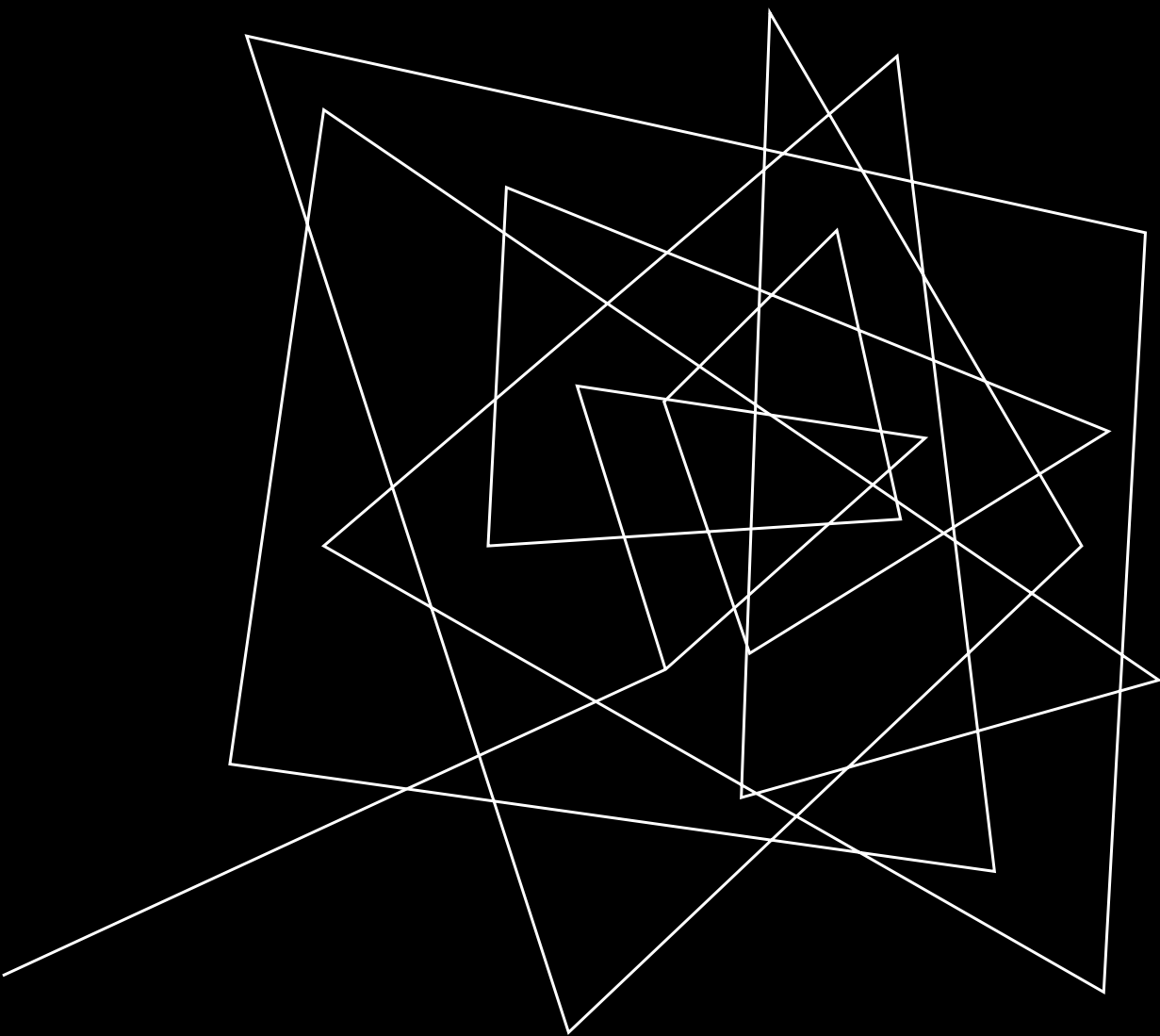
- Reshape the landscape of human-machine interaction.
- Voice Chat Application powered by the Gemini Pro Model
- Harnesses the capabilities of cutting-edge technologies to enable seamless conversations between users and an AI-powered bot
- Leveraging the WhisperModel for text generation and Google Generative AI for speech synthesis



TECHNOLOGIES USED

TECHNOLOGIES USED

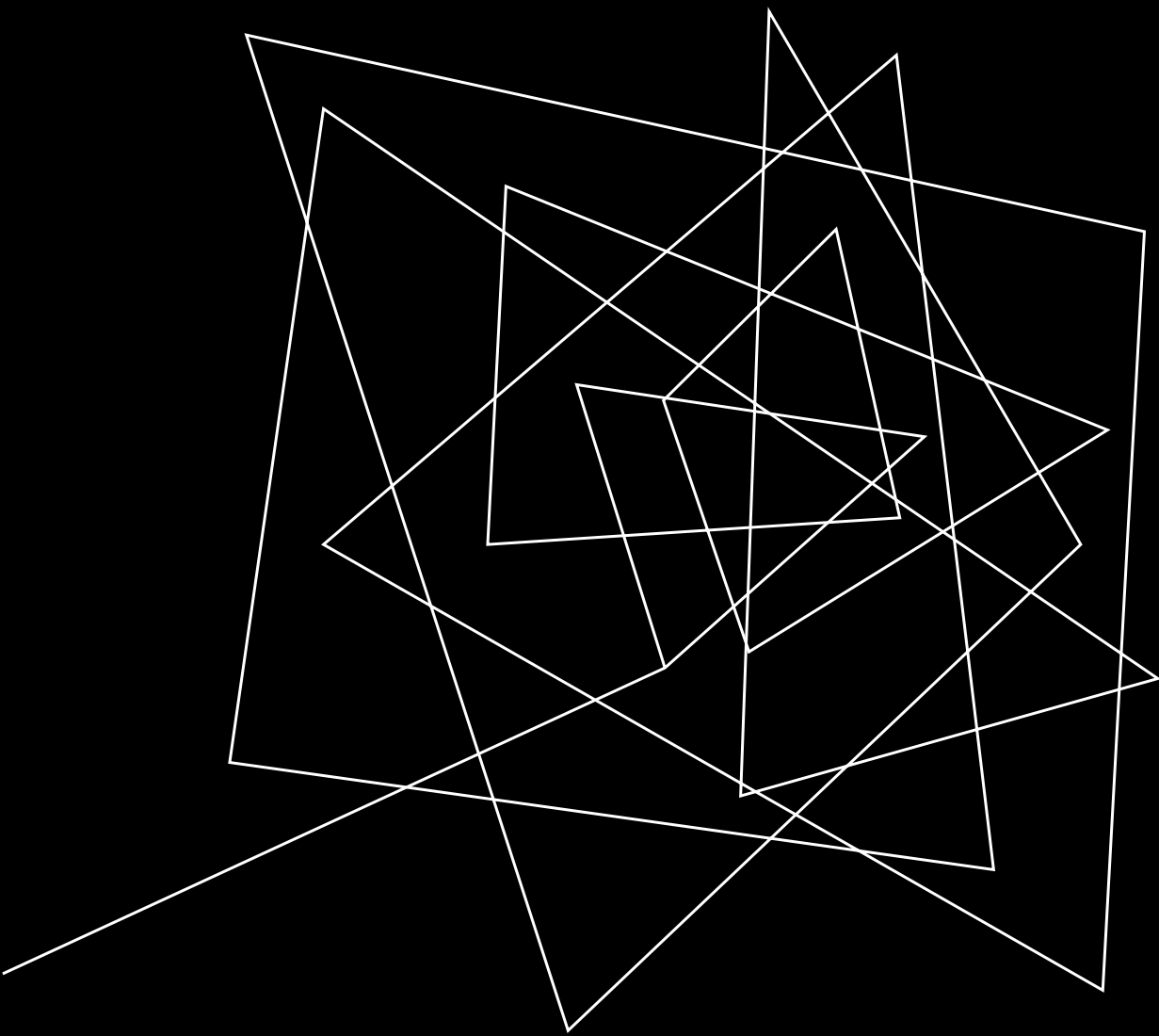
- Streamlit
- Whisper OpenAI for Speech to Text
- Microsoft Speech T5 for Text to Speech
- Gemini Pro Model for Conversation



WHISPER MODEL

WHISPER MODEL

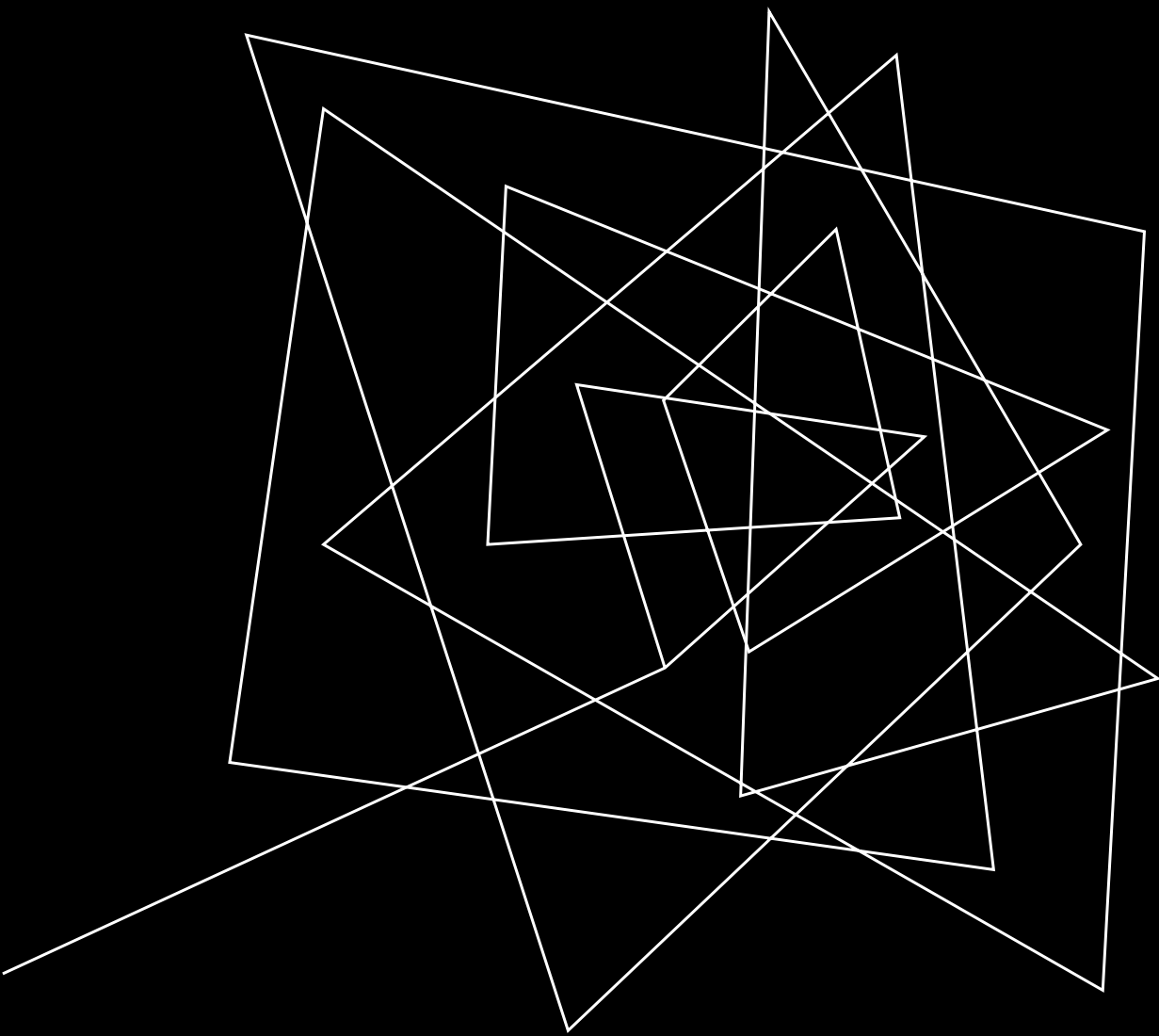
- WhisperModel is an advanced text generation model renowned for its proficiency in generating human-like responses.
- Leveraging state-of-the-art techniques, WhisperModel excels in understanding context and producing contextually relevant text.
- Its adaptive nature allows it to dynamically adjust its responses based on the conversational context, ensuring a natural and engaging dialogue experience.



GEMINI

GEMINI

- Google Generative AI stands at the forefront of speech synthesis technology, offering advanced capabilities in converting text into natural-sounding speech.
- With its cutting-edge algorithms and extensive training data, Google Generative AI excels in producing high-quality audio output with accurate intonation and pronunciation.
- As a crucial component of our application, Google Generative AI plays a pivotal role in enabling the AI bot to communicate with users through lifelike speech.



MICROSOFT T5

MICROSOFT T5 MODEL

- The Microsoft T5 Model for Text to Speech (TTS) is a state-of-the-art solution for converting textual input into synthesized speech.
- Developed by Microsoft, this model utilizes the Transformer architecture and large-scale training data to achieve remarkable accuracy and naturalness in speech synthesis.
- With its ability to generate speech in multiple languages and dialects, the Microsoft T5 Model offers versatility and scalability for various applications, including our voice chat application.

IMPLEMENTATION

Streamlit UI

1. Utilized Streamlit framework to create an intuitive and interactive user interface.
2. Leveraged Streamlit's prebuild api's for simplicity and flexibility to design a user-friendly environment.

Integration of Whisper Model and Gemini

- Integrated WhisperModel for text generation and Google Generative AI for speech synthesis into the application's backend.
- Implemented robust communication channels between the WhisperModel and Google Generative AI to ensure seamless interaction between text input and synthesized speech output.

IMPLEMENTATION

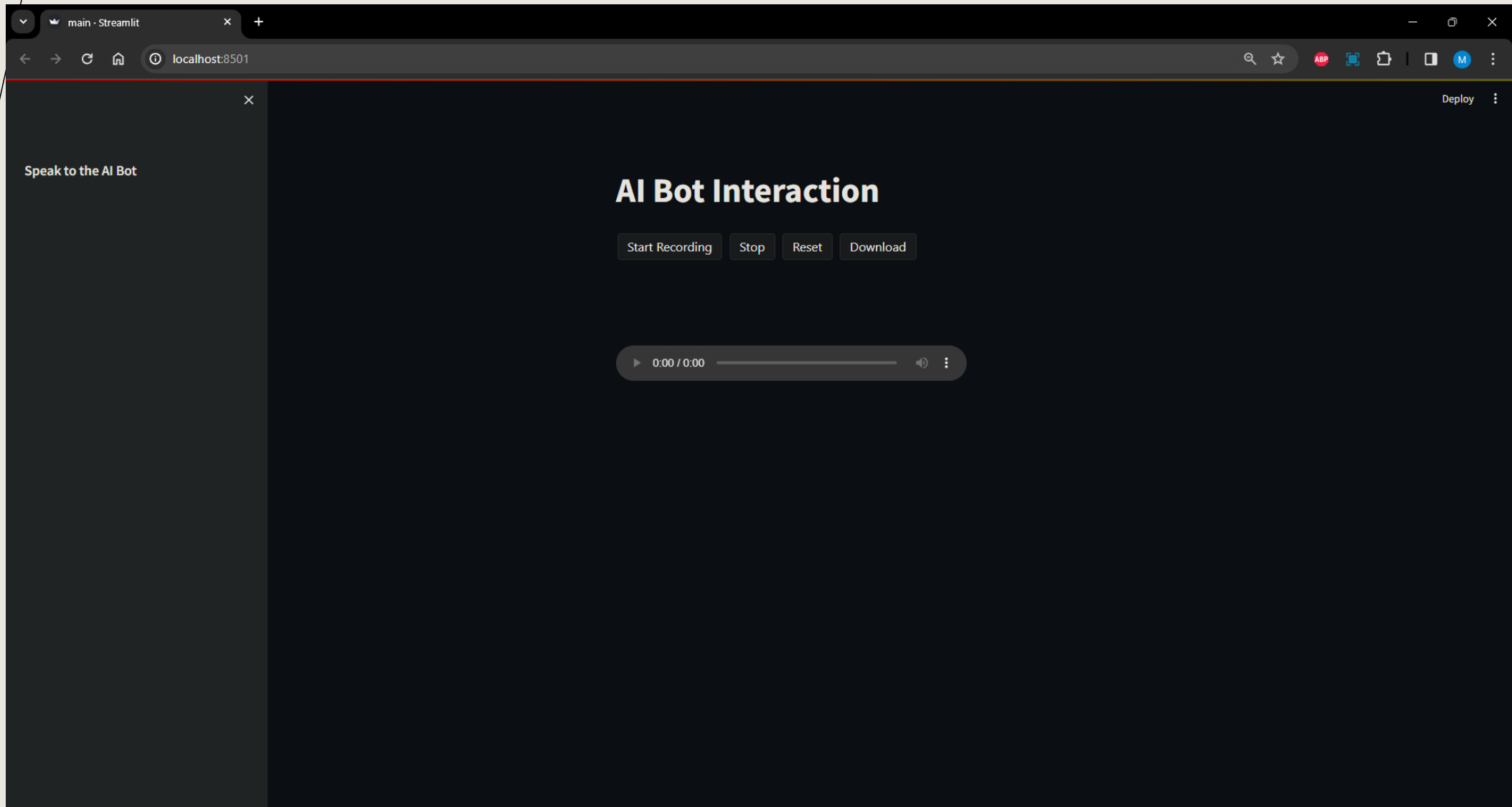
Audio Input Processing

1. Implemented functionality to process audio input from users.
2. Utilized libraries such as Streamlit Audio Recording Library to handle audio input processing tasks.

Response Generation and Presentation

- Developed algorithms to generate contextually relevant responses using the WhisperModel based on user input.
- Utilized Google Generative AI to synthesize text-based responses into natural-sounding speech for presentation to the user.

USER INTERACTION



DEMO

main · Streamlit

localhost:8501

Speak to the AI Bot

AI Bot Interaction

Start Recording Stop Reset Download

0:00 / 0:28

0:12 / 0:12



FUTURE ENHANCEMENTS

- Working on response quality, adding multi-language support, etc.
- Faster response generation and transcription using API's.

CONCLUSION

Closing Remarks

- In conclusion, our Voice Chat Application powered by the Gemini Pro Model represents a significant advancement in AI-driven communication technology.
- With its intuitive interface, advanced text generation capabilities, and lifelike speech synthesis, the application offers users a unique and engaging conversational experience

Q&A

Questions and Answers

- Open floor for any questions from the audience.
- Clarifications on the project or its implementation.



THANK YOU

Muneeb Ullah

USERNAME: moneebullah25

TEAM: MuneebUllah