

AI-powered Voice Sales Agent

Project Overview:

This project focuses on building an AI-Powered Voice Sales Agent that can conduct automated voice conversations with potential customers to promote online courses. The system is designed to simulate a real phone call — capturing voice input from the user, interpreting it using AI and responding back with audio output all in real-time.

How It Works:

- **Voice Input:**
The user (mock caller) speaks into the microphone. The input is captured as audio.
- **Speech-to-Text (STT):**
The audio is transcribed to text using OpenAI Whisper, known for its high accuracy even in noisy environments.
- **Language Processing (LLM):**
The transcribed text is sent to GPT-4 (via OpenAI API) which acts as the brain of the agent. Based on the conversation stage and customer input, the model generates an appropriate response — including course recommendations or objection handling.
- **Conversation Memory:**
Each user input and GPT-generated response is stored in the conversation history. This history is used in every interaction to maintain context and ensure the agent responds naturally and coherently across multiple turns.
- **Text-to-Speech (TTS):**
The GPT-generated response is converted back to voice using gTTS (Google Text-to-Speech).
- **Audio Output:**
The response is played back to the user as audio, completing the voice-based interaction loop.

Backend API (FastAPI):

Three key endpoints have been developed:

- POST /start-call
- Initializes a new conversation session.

FastAPI - Swagger UI

:/default/start_call_start_call_post

default

POST /start-call Start Call

Parameters

No parameters

ExecuteClear

Responses

Curl

```
curl -X 'POST' \
  'http://127.0.0.1:8000/start-call' \
  -H 'accept: application/json' \
  -d ''
```

Request URL

```
http://127.0.0.1:8000/start-call
```

Server response

Code	Details
200	<div>Response body<pre>{ "message": "Conversation started. Memory cleared." }</pre></div> <div>Response headers<pre>content-length: 51 content-type: application/json date: Thu, 19 Jun 2025 15:06:15 GMT server: uvicorn</pre></div>

Code	Description	Links
200	Successful Response	No links

- POST /respond/
- Accepts audio input, processes it through STT → LLM → TTS and returns an audio

response.

Execute

Clear

Responses

Curl

```
curl -X 'POST' \
  'http://127.0.0.1:8000/respond' \
  -H 'accept: application/json' \
  -H 'Content-Type: multipart/form-data' \
  -F 'audio=@output.mp3;type=audio/mpeg'
```

Request URL

http://127.0.0.1:8000/respond

Server response

Code

Details

200

Response body

```
{
  "transcription": "Hello, I am Prone. Do you have AI course?",
  "response": "Yes, we do have a course on AI Development. The course explores core concepts of artificial intelligence including machine learning, neural networks, deep learning, and natural language processing. \n\nYou will learn about supervised and unsupervised learning, neural networks and deep learning, model evaluation and tuning, NLP basics and text preprocessing, and AI project deployment. \n\nThe course duration is 4 months and it costs 8000 Tk. However, we are currently offering a discounted fee of 6500 Tk.",
  "audio_file": "reply_20250619_235204.mp3"
}
```

Download

Response headers

```
content-length: 624
content-type: application/json
date: Thu, 19 Jun 2025 17:51:46 GMT
server: uvicorn
```

Responses

Code

Description

Links

- GET /conversation
- Retrieves the full conversation history..

GET /conversation Get Conversation

Cancel

Parameters

No parameters

Execute

Clear

Responses

Curl

```
curl -X 'GET' \
  'http://127.0.0.1:8000/conversation' \
  -H 'accept: application/json'
```

Request URL

http://127.0.0.1:8000/conversation

Server response

Code

Details

200

Response body

```
{
  "conversation": [
    {
      "type": "human",
      "content": "Hello, I am Prone. Do you have AI course?"
    },
    {
      "type": "ai",
      "content": "Yes, we do have a course on AI Development. The course explores core concepts of artificial intelligence including machine learning, neural networks, deep learning, and natural language processing. \n\nYou will learn about supervised and unsupervised learning, neural networks and deep learning, model evaluation and tuning, NLP basics and text preprocessing, and AI project deployment. \n\nThe course duration is 4 months and it costs 8000 Tk. However, we are currently offering a discounted fee of 6500 Tk."
    }
  ]
}
```

Download

Integration Components:

- **LLM Integration:**
Used OpenAI GPT-4 API known for its strong reasoning and context retention capabilities.
- **Course Database:**
A local courses.json file is used to simulate a course database. The LLM can search through this structured data to offer targeted course suggestions.
- **Vector Search:**
Implemented FAISS to support similarity-based retrieval from the course database — enhancing relevance in course pitching.

Conversation Flow Design:

- ❖ **Introduction:**
Bot welcomes the user and introduces the company and purpose of the call.
- ❖ **Qualification:**
Ask a few questions to gauge user interest, background or career goals.
- ❖ **Course Pitch:**
Based on responses, the bot selects a relevant course and provides a short pitch with features and benefits.
- ❖ **Objection Handling:**
Addresses common user concerns logically and politely, offering solutions or reassurances.
- ❖ **Wrap-up:**
Offers the user a follow-up or next steps (mocked).

Tech Stack Summary:

- ❖ **Language:** Python
- ❖ **LLM :** OpenAI GPT-4
- ❖ **Speech to text :** OpenAI Whisper
- ❖ **Text to speech :** gTTS (Google Text-to-Speech)
- ❖ **Vector Database :** FAISS
- ❖ **Framework :** Langchain, FastAPI