

TaxGPT

Project Documentation

OVERVIEW

TaxGPT is an AI-based question-answering system designed to assist individuals and small business owners with tax-related queries. Built using a Retrieval-Augmented Generation (RAG) approach, it combines the power of large language models with vector-based document retrieval to provide accurate, context-specific responses. If relevant information is not found in the internal database, the system optionally performs a web search to ensure completeness.

Objectives

- To build an AI assistant capable of answering Indian tax-related questions.
- To create a user-friendly frontend for seamless interaction.
- To implement a dual-retrieval pipeline using vector search and fallback web search.
- To deploy the solution using Docker and host it on AWS with a custom domain.

Tech Stack

- Frontend: Streamlit
- Backend: FastAPI
- Language Model: OpenAI GPT 3.5 Turbo
- Vector Database: FAISS

- Google Tavily API
- Deployment: Docker, AWS EC2, Nginx

Features

- Tax-specific RAG system.
- Intelligent fallback to web search when internal retrieval fails.
- Streamlit-based user interface.
- FastAPI for modular design.
- Real-time query handling and response generation.
- Dockerized services hosted on AWS with a public domain.

Architecture Overview

Frontend (Streamlit UI)

→ Sends user queries to backend via API.

Backend (FastAPI)

→ Routes the query to the RAG pipeline:

- Check vector database (FAISS)
- If no relevant data found → fallback to Google Search API (Tavly)
- Pass context to the language model (e.g., Gpt 3.5 turbo)

Vector Database

→ Stores embeddings of Indian tax documents.

LLM Model

→ Generates answers using context retrieved.

Deployment

Deployed 3 Docker containers: frontend, backend, and vector DB.

Hosted on **AWS EC2** instance (IP: 3.110.209.247:8501)

Domain: **taxgpt.services** routed via Nginx reverse proxy.

HTTPS enabled with SSL.

Performance

Retrieval accuracy on internal tax dataset: **~91%**

Query response time: **< 3 seconds**

Fallback trigger rate: **12%**

Testing & Validation

Tested 100+ tax-related questions (e.g., GST, ITR, 80C deductions).

Compared AI-generated answers with official tax portal and CA recommendations.

Verified fallback search quality and relevance.

Future Enhancements

Integrate multilingual support (Malayalam/Hindi).

Add voice-based query input.

Enable session-based context retention.

Include document upload and analysis (e.g., ITR forms).

Acknowledgments

OpenAI (Gpt 3.5 turbo as LLM)

Government of India Income Tax and ICAI for open content.

GitHub community and Stack Overflow for development help.