

Intelligent Sri Lanka IRD Tax Intelligence & Compliance Assistant

An AI-powered tax information assistant designed to answer Sri Lanka Inland Revenue Department (IRD) tax questions **strictly based on IRD-published documents**, without guessing or hallucinating.

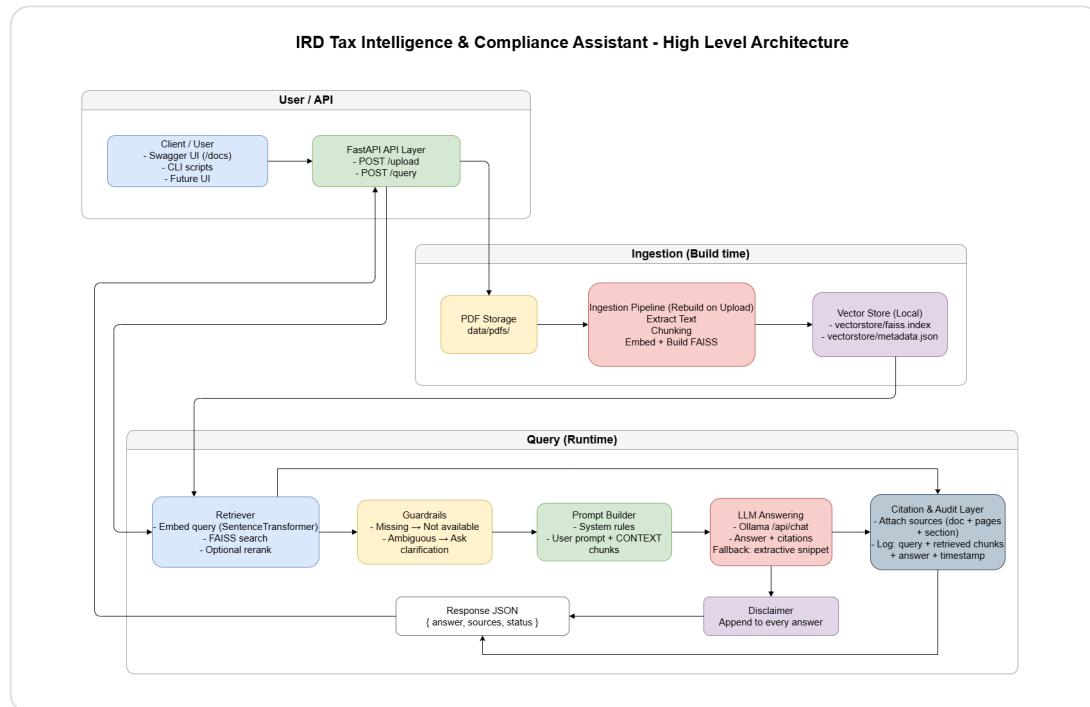
This system ingests official IRD PDFs, builds a local vector database, retrieves relevant sections, and optionally uses an LLM to generate safe, citation-backed answers with mandatory disclaimers.

This project includes a fully documented REST API for tax information retrieval.

📌 Key Objectives

- Answer tax questions **only using IRD documents**
- Prevent hallucinations, guessing, and unsupported inference
- Detect **missing** and **ambiguous** questions
- Provide **citations and sources**
- Append a **mandatory disclaimer** to every response
- Expose functionality via a **FastAPI API** and **CLI**
- Ensure deterministic behavior: no guessing, no inferred answers, no external knowledge

🏗️ High-Level Architecture



Main Components:

1. PDF Ingestion Pipeline
2. Vector Database (FAISS)

3. Retriever
 4. Guardrails
 5. Prompt Builder
 6. LLM (Optional – Ollama)
 7. FastAPI API Layer
 8. Disclaimer Injection
-

System Architecture Overview

- ◊ Ingestion (Build Time)
 - Upload IRD PDF documents
 - Extract text from PDFs
 - Chunk text into manageable sections
 - Generate embeddings
 - Store vectors in FAISS index with metadata
 - ◊ Query Processing (Runtime)
 - Embed user query
 - Retrieve top relevant chunks
 - Apply guardrails (missing / ambiguous detection)
 - Build safe prompts with context
 - Generate answer (LLM or extractive fallback)
 - Append disclaimer
 - Return structured JSON response
-

Features Implemented (Step-wise)

Step	Feature
Step 1–5	PDF ingestion, chunking, FAISS vector store
Step 6	Retriever with semantic search
Step 7	Prompt builder (system + user prompt)
Step 8	LLM answering (Ollama) + fallback
Step 9	Guardrails (missing / ambiguous detection)
Step 10	FastAPI API (/upload , /query)
Step 11	Mandatory disclaimer appended
Step 12	Architecture diagram

Technology Stack

Backend

- **Python 3.10+**
- **FastAPI**
- **Uvicorn**

AI / ML

- **SentenceTransformers** (`all-MiniLM-L6-v2`)
- **FAISS** (local vector database)
- **Ollama** (local LLM runner)
 - Model: `llama3.1:8b`

Data Processing

- **PDFMiner / PyMuPDF** (PDF extraction)
- **JSON metadata storage**

Tooling

- **Swagger UI** (API testing)
- **draw.io** (architecture diagram)

❖ Setup Instructions

1 Clone Repository

```
git clone <repository-url>
cd ird-tax-ai
```

2 Create Virtual Environment

```
python -m venv venv
```

Activate:

Windows

```
.\venv\Scripts\activate
```

Linux / macOS

```
source venv/bin/activate
```

3 Install Dependencies

```
pip install -r requirements.txt
```

4 Start Ollama (Optional – for LLM answers)

```
ollama serve  
ollama pull llama3.1:8b
```

If Ollama is unavailable, the system automatically falls back to extractive answers.

5 Run FastAPI Server

```
python -m uvicorn api.main:app --reload
```

Open:

- Swagger UI → <http://127.0.0.1:8000/docs>
-

💡 API Endpoints

❖ Upload PDFs

```
POST /upload
```

Uploads IRD PDFs and rebuilds the vector index.

❖ Query API

```
POST /query
```

Request

```
{  
    "question": "What is the personal relief for 2025/2026?",  
    "top_k": 5,  
}
```

```
"use_llm": true  
}
```

Response

```
{  
  "answer": "...",  
  "sources": [  
    {  
      "document": "PN_IT_2025-01.pdf",  
      "page_range": [1],  
      "section": "PN/IT/2025-01"  
    }  
  ],  
  "status": "ok",  
  "used_llm": true  
}
```

💻 CLI Testing

Run:

```
python -m llm.answer_with_llm
```

Example:

```
Enter question: What is the VAT rate for 2025?
```

```
This information is not available in the provided IRD documents.  
This response is based solely on IRD-published documents and is not professional  
tax advice.
```

⚠️ Guardrails Behavior

Scenario	System Response
Info not in PDFs	"Not available in provided documents"
Ambiguous question	Ask user to clarify
Low retrieval confidence	Treated as missing (score-based decision)
Valid question	Answer with citations

Scenario	System Response
LLM failure	Safe extractive fallback

Note: Missing-answer detection is based on retrieval confidence score thresholds, not real-world tax knowledge.
If relevant information is not found with sufficient confidence in the provided IRD documents, the system correctly responds as "not available".

⚠ Disclaimer

Every response ends with:

"This response is based solely on IRD-published documents and is not professional tax advice."

This is enforced at **code level** (Step 11).

⌚ Assumptions & Limitations

- Only IRD-published PDFs are considered trusted
 - No external tax knowledge is used
 - LLM is optional and local-only (Ollama)
 - No fine-tuning of LLM performed
 - Not a replacement for professional tax consultation
-

📈 Future Enhancements

- Web frontend UI
 - Multi-language support (Sinhala / Tamil)
 - Role-based access
 - Cloud vector store
 - Automated document updates
-

💻 Author

S.D.Weerathunga

Computer Engineering Undergraduate University of Sri Jayewardenepura Sri Lanka LK

- GitHub: <https://github.com/SawaniD-Weerathunga>
 - LinkedIn: <https://www.linkedin.com/in/sawani-weerathunga-507a55348/>
-