Document Assistant
Setup and Installation Guide

1 Overview

This document describes the steps required to install, configure, and run the **Document Assistant** system on a local machine. The system allows users to upload PDF documents and interact with them using retrieval-based and large language model (LLM) based question answering.

2 Prerequisites

2.1 Required Software

- Python 3.10 or later

- Git

- A modern web browser

2.2 Optional Software

- Ollama (for offline LLM support)

- Google Gemini API access (for online LLM support)

3 Repository Setup

Clone the repository from GitHub:

```
git clone https://github.com/gnanadeep256/document-assistant.git
cd document-assistant
```

4 Virtual Environment

Create a Python virtual environment:

```
python -m venv venv
```

Activate the environment:

Windows

```
venv\Scripts\activate
```

Linux / macOS

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

5 Dependency Installation

Install required Python packages:

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

6 Gemini API Configuration (Optional)

Create a `.env` file in the project root directory with the following content:

```
GEMINI_API_KEY=your_api_key_here
```

This file must not be committed to version control.

7 Offline LLM Setup (Ollama)

Download and install Ollama from:

https://ollama.com

Pull the LLaMA model:

```
ollama pull llama3
```

Start the model:

```
ollama run llama3
```

8 Running the Application

Launch the Streamlit application:

```
streamlit run app.py
```

The application will be available at:

http://localhost:8501

9 Using the System

1. Upload a PDF document

2. The system extracts text and builds embeddings automatically

3. Select an inference mode:

   - Retrieval only (non-LLM)
   - Offline LLaMA (Ollama)
   - Online Gemini

4. Ask questions using the chat interface

10 Troubleshooting

10.1 Vector Database Permission Errors (Windows)

- Close all Python processes
- Restart Streamlit
- Re-upload the document

10.2 Gemini Errors

- Verify API key correctness
- Ensure correct model name is used

10.3 Ollama Issues

- Ensure Ollama is running
- Test using `ollama run llama3`

11 Data Storage and Privacy

All documents and embeddings are stored locally on the user's machine. No data is transmitted externally unless the online Gemini mode is explicitly selected.

12 Conclusion

The Document Assistant system is now fully configured and ready for use. Users can upload documents and interact with them through a unified, chat-based interface.