

# RAG Project Documentation References

AI Document Assistant - Enterprise RAG System

Last Updated: December 2025

## Purpose

Central repository of all documentation links for building the AI Document Assistant. All URLs have been verified as current and active as of December 2025.

## Python & Core Libraries

### *Python Official*

[Python 3.13 Documentation](#)

[Python Standard Library](#)

## LangChain (Core RAG Framework)

### *Getting Started*

[LangChain Documentation Home](#)

[Installation Guide](#)

[Quickstart Tutorial](#)

### *Document Processing*

[Document Loaders \(API Reference\)](#)

[Text Splitters](#)

[Text Splitters \(API Reference\)](#)

### *Embeddings*

[Embeddings Overview](#)

### *Vector Stores*

[Vector Stores Overview](#)

[Chroma Integration](#)

### *Retrieval & RAG*

[Retrievers](#)

RAG Tutorial

## ***LLM Integration***

[Ollama Integration](#)

# ChromaDB (Vector Database)

## *Core Documentation*

[ChromaDB Homepage](#)

[Official Documentation](#)

[Getting Started Guide](#)

[Usage Guide](#)

[Deployment](#)

[Community Cookbook](#)

## *Operations*

[Collections](#)

[Querying](#)

[Embeddings](#)

## *Python Client*

[Python Client Documentation](#)

[GitHub Repository](#)

# Ollama (Local LLM)

## *Setup & Usage*

[Ollama Homepage](#)

[Official Documentation](#)

[Model Library](#)

[GitHub Repository](#)

[Python Library](#)

[JavaScript Library](#)

## *Popular Models (2025)*

[Llama 3.1/3.2](#)

[Mistral](#)

[Phi-3](#)

[DeepSeek-R1](#)

[Gemma 3](#)

# HuggingFace (Embeddings & Models)

## *Sentence Transformers*

[GitHub Repository](#)

[PyPI Package](#)

[HuggingFace Organization](#)

[Models Library \(17,000+ models\)](#)

[Popular Model: all-MiniLM-L6-v2](#)

[Official Announcement](#)

# Gradio (UI Framework)

[Gradio Documentation](#)

[Quickstart Guide](#)

[ChatInterface](#)

[GitHub Repository](#)

# Docker (Deployment)

[Docker Documentation](#)

[Dockerfile Reference](#)

[Docker Compose](#)

# Azure OpenAI (Enterprise Option)

[Azure OpenAI Documentation](#)

[Pricing](#)

# Git & GitHub

[Git Documentation](#)

[GitHub Guides](#)

# Quick Reference Commands

## *Ollama*

Pull a model: `ollama pull llama3.1`

Run a model: `ollama run llama3.1`

List models: `ollama list`

## *ChromaDB*

```
import chromadb
client = chromadb.Client()
collection = client.create_collection('docs')
```

## *LangChain Basic RAG*

```
from langchain_community.document_loaders import TextLoader
from langchain.text_splitter import RecursiveCharacterTextSplitter
from langchain_community.embeddings import HuggingFaceEmbeddings
from langchain_community.vectorstores import Chroma
```

## Notes

- Current Focus: Building Phase 1 - Document chunking and basic RAG
- Next Phase: Multi-model evaluation and cost tracking
- Deployment Target: Docker container with local models (compliance-focused)

Last Updated: December 17, 2025

Project: AI Document Assistant - Enterprise RAG System

All URLs verified and current as of December 2025