

Quick Start Guide

Get your RAG System up and running in 5 minutes!

Quick Installation

```
# 1. Navigate to the project
cd /home/ubuntu/rag_system

# 2. Run setup (creates venv and installs dependencies)
./setup.sh

# 3. Download Llama 3 model
source venv/bin/activate
huggingface-cli download TheBloke/Llama-3-8B-Instruct-GGUF \
  llama-3-8b-instruct.Q4_K_M.gguf \
  --local-dir models --local-dir-use-symlinks False

# 4. Start the server
./run.sh
```

Access the Application

Once the server starts, open your browser:

- **Web Interface:** <http://localhost:8000>
- **API Documentation:** <http://localhost:8000/docs>

First Steps

1. Upload a Document

- Click on the “Upload PDF” tab
- Choose a PDF file or drag & drop
- Wait for processing to complete

2. Ask a Question

- Go to the “Ask Questions” tab
- Type your question
- Click “Ask Question”
- View the answer and sources

3. Try Web Scraping

- Navigate to “Scrape URL” tab
- Enter a URL (e.g., https://en.wikipedia.org/wiki/Artificial_intelligence)
- Click “Scrape & Index”

Example Questions

After uploading documents, try:

- “What is the main topic of the document?”
- “Summarize the key points”
- “What are the conclusions?”

Documentation

- **README.md** - Complete documentation
- **INSTALL.md** - Detailed installation guide
- **EXAMPLES.md** - API usage examples
- **PROJECT_OVERVIEW.md** - Technical architecture

Troubleshooting

Server won't start?

- Check if model file exists: `ls models/`
- View logs: `tail -f logs/rag_system.log`

Out of memory?

- Edit `config.yaml` and set `n_gpu_layers: 0`
- Use a smaller model variant

Import errors?

- Activate venv: `source venv/bin/activate`
- Reinstall: `pip install -r requirements.txt`

Tips

- Start with small documents to test
- Use Ctrl+Enter to submit queries quickly
- Check the “Manage Documents” tab to see indexed content
- Monitor system stats on the main query page

Next Steps

1. Read the full [README.md](#) (README.md)
2. Explore the [API Documentation](http://localhost:8000/docs) (<http://localhost:8000/docs>)
3. Try the [examples](#) (EXAMPLES.md)
4. Customize [config.yaml](#) (config.yaml)

Need Help? Check the troubleshooting section in README.md