

# Yashowardhan Singh Tomar

## RAG / LLM APPLICATION ENGINEER | RETRIEVAL EVALUATION | PYTHON

Indore, India | Open to remote and relocation | yashtomar10122@gmail.com | +91-97549-50809

linkedin.com/in/yashowardhansinghtomar | github.com/yashowardhansinghtomar

---

### PROFILE

AI engineer building practical LLM applications with a focus on RAG, document QA, retrieval evaluation, and local-first GenAI systems. Public portfolio includes a RAG evaluation workbench, PDF QA workflows, local Ollama assistants, and LLM evaluation tooling. Background combines applied Python development, LLM behavior evaluation, voice AI integration, and business-facing product experience.

RAG | LLM Applications | Retrieval Evaluation | Document QA | LangChain | FAISS | Ollama | Python

---

### EXPERIENCE

#### Jr. AI / Data Science Engineer | Engineer Master

Sep 2024 - Jan 2025

- Built and evaluated LLM workflows involving prompt design, structured model review, and failure analysis across Python and data-science domains.
- Contributed to RLHF pipelines for xAI and Meta Llama 4, giving direct experience with how LLMs fail on correctness, instruction following, and reasoning tasks.
- Integrated TTS and STT models into PreCall AI, a production AI calling workflow, strengthening end-to-end model integration experience.
- Translated business requirements into applied AI workflows, drawing on prior customer-facing experience in education technology sales.

#### Business Development Associate | Byju's

2020 - 2022

- Managed high-volume B2C sales of online education products with weekly revenue targets.
- Developed consultative discovery, stakeholder communication, and objection-handling skills across a large customer base.

---

### SELECTED PROJECTS

#### rag-evaluation-workbench

<https://github.com/yashowardhansinghtomar/rag-evaluation-workbench>

Python, BM25 retrieval, JSONL eval sets, Markdown/JSON reports

- Built a RAG evaluator that ingests markdown corpora, chunks documents, retrieves evidence, and scores answers for citation coverage and groundedness.
- Reports retrieval recall@k, citation precision/recall, required-fact coverage, grounded answer rate, and failure tags.

#### PDF-Question-Answering-System

<https://github.com/yashowardhansinghtomar/PDF-Question-Answering-System>

Streamlit, LangChain, FAISS, Groq

- Built a PDF ingestion and document-question-answering workflow using chunking, vector retrieval, and grounded response generation.

## **real\_estate\_chatbot\_ollama**

[https://github.com/yashowardhansinghtomar/real\\_estate\\_chatbot\\_ollama](https://github.com/yashowardhansinghtomar/real_estate_chatbot_ollama)

Python, Gradio, LangChain, FAISS, Ollama

- Created an offline real-estate assistant with local LLM execution, retrieval-backed FAQ behavior, and multi-modal routing concepts.

### **SKILLS**

---

- **RAG:** Document ingestion, chunking, retrieval, citations, grounded answer evaluation
- **LLM Apps:** LangChain, OpenAI API, Groq, Ollama, prompt workflows, local-first assistants
- **Data / Eval:** JSONL eval sets, failure tags, Markdown/JSON reporting, pairwise response review
- **Interfaces:** Streamlit, Gradio, CLI tools

### **EDUCATION**

---

#### **Prestige Institute of Management & Research**

Bachelor of Business Administration (BBA), Global Marketing & Brand Management | 2017 - 2020

### **CERTIFICATIONS**

---

Data Analyst - Edubridge | Generative AI - Debugshala