Utkarsh Gupta

9670279595 | Email | LinkedIn | GitHub | Portfolio

Professional Summary

AI Full-Stack Engineer with 1+ year of experience building production-grade Generative AI, RAG, and Agentic AI systems. Skilled in developing scalable AI applications processing 50K+ documents, serving 2K+ daily queries, and supporting 3K+ concurrent users. Proficient in Python, TypeScript, Node.js, LangChain, LangGraph, vector databases, and microservices.

Experience

Multibagg AI July 2024 – Present
AI Fullstack Engineer *Bengaluru, Karnataka*

Ask Iris – AI Financial Research Assistant

- Built a **multi-agent AI-powered financial research assistant** that enables investors to ask natural language questions across annual reports, concalls, regulatory filings, and company documents.
- Designed and scaled the end-to-end **RAG architecture**, enabling retrieval, semantic search, and source-cited responses across **50K+ financial documents**.
- Developed large-scale document ingestion pipelines, processing **50K+ PDF documents** through text extraction, chunking, embedding generation, metadata enrichment, and vector database indexing to build a production-ready financial knowledge base.
- Served **2,000+ investor queries daily** with an average response time of 20 seconds, significantly reducing manual research effort and accelerating access to financial insights.

Real-Time AI Pipeline for Union Budget 2026

- Built a **real-time AI pipeline** that analyzed the Union Budget speech and generated investor-focused insights within seconds.
- Developed automated workflows for live transcription, event extraction, content processing, and AI-powered summarization.
- Supported **3K+ concurrent users** during the Budget event and drove **5K+ SEO clicks** with **500K+ search impressions** in a single day.

Distributed Cron-Based System

- Developed a **distributed cron orchestration system** for running and monitoring large-scale data collection, document processing, and AI workflows.
- Automated hundreds of recurring jobs to sync and process financial data for **6K+ companies**.
- Improved pipeline reliability and scalability, enabling faster delivery of AI-driven insights.

RAG Evaluation Framework

- Built an automated evaluation framework for **RAG pipelines** serving **2K+ daily queries** across a **50K+ document** financial knowledge base, evaluating retrieval quality, response relevance, faithfulness, and citation accuracy.
- Implemented **LLM-as-a-Judge** and **rubric-based evaluation** for response quality, along with **deterministic evaluations** for tool selection and agent behavior, enabling automated regression testing for production AI workflows.

Projects

Chatbase GitHub, Live URL

- Built a SaaS platform that enables users to create custom AI chatbots from website content using **RAG, embeddings, and vector search**.
- Simplified chatbot creation by automatically extracting and organizing website content, allowing businesses to launch AI assistants without manually building a knowledge base.

UToob AI GitHub, Live URL

- Built an AI-powered SaaS platform that transforms YouTube videos into summaries, notes, quizzes, social media content, and interactive chat experiences using **OpenAI and Google Gemini** models.
- Helped users understand and repurpose long-form video content in minutes, eliminating the need to watch entire videos while making key insights instantly accessible.

Education

Vellore Institute of Technology – MCA – 9.3 CGPA, Chennai, Tamil Nadu *2023 – 2025*
Dr. RML Avadh University – B.Sc. Computer Science – 82.9%, Ayodhya, Uttar Pradesh *2019 – 2022*

Technical Skills

Languages: Java, JavaScript, TypeScript, Python, SQL, C, C++
Backend: Node.js, Express.js, REST APIs, FastAPI, Microservices, ETL Pipelines
Databases: MySQL, MongoDB, Redis, Vector Databases
Frontend: React.js, Next.js, HTML, CSS, Tailwind CSS, Bootstrap
AI/ML: LangChain, LangGraph, RAG, LLMs, Agentic AI, Prompt Engineering, Embeddings, AI Evaluation, Vector Search, Guardrails