KRISH PRASAD

+91-8789318322 | msgkrish192@gmail.com | linkedin.com/in/krishprasadd | github.com/gitKrishh

TECHNICAL SKILLS

Languages: Python, JavaScript, SQL
AI & LLM Systems: RAG Pipelines, Multi-Agent Workflows (LangGraph), Tool Calling, LangChain, LoRA/PEFT, Prompt Engineering, Vector Embeddings, Re-ranking
ML & Modeling: Transformers, CNNs, TensorFlow, Model Evaluation, Semantic Search, Vector Search
Backend & APIs: FastAPI, Node.js, Express.js, REST APIs, Async Processing, OCR Pipelines
Frontend: React.js, Tailwind CSS
Databases & Infra: PostgreSQL, MongoDB, FAISS, ChromaDB, Docker, Git

PROJECTS

AutoResearcher – AI Research Assistant | *LangGraph, Python, FastAPI, React* GitHub | Live

- Built a **multi-agent AI system** using **LangGraph** with specialized agents for research planning, paper retrieval, content analysis, and structured report generation – compressing multi-hour literature reviews into a single query, cutting research turnaround by ∼**70%**.
- Implemented a **RAG pipeline** with chunked PDF ingestion, **FAISS**-backed vector search, and citation-aware response generation; inline source references reduced ungrounded outputs by ∼**40%** vs. a vanilla LLM baseline.

ContextOS – RAG-Based Document Intelligence | *MERN Stack, OpenAI, FAISS, Redis* GitHub | Live

- Built a **document QA platform** with configurable chunking strategies, **embedding optimization**, and **cross-encoder re-ranking** – achieving a **30% improvement** in retrieval precision over baseline dense retrieval.
- Engineered an **async FastAPI** backend with **Redis** query caching, cutting average response latency to **under 1.5s**; added an eval dashboard tracking retrieval precision, hallucination rate, and p95 latency per dataset.

Saar AI – AI-Powered Document Assistant | *FastAPI, React, LangChain, LangGraph, PostgreSQL* GitHub | Live

- Designed an **end-to-end pipeline** that ingests bills, legal notices, and medical PDFs via **OCR**, classifies document type, then routes to **specialized LLM agents** to extract deadlines, flag risks, and generate plain-language summaries – processing documents in **under 8s** end-to-end.
- Built **contextual memory** over user document history using **ChromaDB** vector search, enabling **multi-turn Q&A** grounded in previously uploaded files without re-ingestion, supporting sessions spanning **20+ documents**.

EXPERIENCE

Freelance AI Developer 2025 – Present
Independent *Remote*

- Built and deployed **AISpark Blog** (blog.aispark.live), a **full-stack AI/ML** content platform (**React + FastAPI**) covering RAG, LLMs, and applied ML – handling production traffic with **sub-second page loads**.
- Delivered an additional **RAG-based automation** web app for a client; built **async FastAPI** backend handling **concurrent AI inference** requests with consistent low-latency responses under real workloads.

EDUCATION

B.Tech in Computer Science and Engineering 2024 – 2028
Madan Mohan Malaviya University of Technology *Gorakhpur, Uttar Pradesh*

ACHIEVEMENTS

ANRF AISEHack (Kaggle): Ranked in **top quartile** (Score: **19.088**) in India's premier **AI-for-Science hackathon** organised by ANRF in coordination with IBM and IIT Delhi – competing on flood detection and pollution forecasting tracks against **50+ finalist teams** nationwide.
Smart India Hackathon – College Round: Finished **6th place** in the internal college selection round for **SIH 2025**, competing among all shortlisted student teams at MMMUT.