KRISHNA DADHICH

09521409593 | krsnadadhich@gmail.com

Applied AI Engineer | RAG Systems, Conversational AI & Computer Vision

linkedin.com/in/krsnadadhich | github.com/krsnadadhich | krsnadadhich.vercel.app

OBJECTIVE

AI/ML engineer seeking an entry-level role in GenAI, RAG systems, conversational AI, or computer vision, building AI systems that leave the lab and get used. CareerOS automates my own job search end to end, and every project below runs as a complete, working system tested with real users, including CMR University faculty and students.

PROFESSIONAL EXPERIENCE

AI/ML Intern, Medha Tech Solutions [LINK] Sep 2025 – Nov 2025

Bengaluru, India

- Developed and enhanced **MIRA**, an AI-powered domain-restricted chatbot using **Retrieval-Augmented Generation (RAG)** to deliver accurate, source-grounded responses from organization-specific knowledge bases, reducing hallucinations in an Agile environment.
- Engineered core backend features including knowledge base management and role-based administration using **PHP, MySQL, and REST APIs**, building an admin dashboard to manage chatbot instances and user access.

AI Automation Engineer, Deliverail [LINK] Dec 2025 – Present

Bengaluru, India

- Developed an AI-powered booking chatbot integrating **LLMs, Python, and REST APIs** to streamline customer interactions, automate booking workflows, and improve information processing across logistics operations.
- Engineered an administrative dashboard with **backend integrations, SQL-based data management, and workflow automation**, supporting booking management, operational data handling, and centralized administration while reducing repetitive manual tasks.

Open Source Contributor Aug 2026 – Present

- Publish and maintain independent AI/ML projects, including **RAG-based chatbots**, an **offline voice AI assistant**, and a **computer vision access-control system**, as public GitHub repositories with setup documentation for others to use, learn from, and build on.

PROJECTS

EMMA: AI Voice Receptionist [LINK] Python, faster-whisper, Ollama, Coqui TTS

- Built a fully **on-device voice AI agent** that handles reception calls end-to-end: speech-to-text (**faster-whisper**), reasoning via a local **Ollama** LLM, and **text-to-speech (Coqui TTS)**, with zero external API calls, zero per-request cost, and fully offline operation.
- Designed a **safety-first architecture**: a deterministic request-validation check and a knowledge-grounding check run in plain Python before the LLM is invoked, so the model can only respond within **RAG-grounded knowledge** or trigger a live **shipment booking via tool-calls**. Contracted and deployed for **Deliverail**, a logistics company, for shipment booking and delivery-status information over voice.

Kuro: Local Multi-Agent AI Assistant with RAG & Voice [LINK] React, FastAPI, Ollama, FAISS

- Built **Kuro**, a private, fully offline AI assistant: create unlimited assistants, each with its own knowledge base, system prompt, and settings, answering from PDF, DOCX, TXT, CSV, and Excel files with streamed, cited responses and a confidence score.
- Implemented **hybrid retrieval** (FAISS plus keyword search with reranking), a **voice interface** (faster-whisper STT, TTS playback), **CSV/Excel analytics** with KPI reporting, and user accounts with backup/restore, all running locally via Ollama; demoed and stress-tested with faculty and students at **CMR University**.

CareerOS: Autonomous Job-Search Platform [LINK] Next.js, React, TypeScript, Prisma, PostgreSQL, Ollama

- Built **CareerOS**, a personal AI command center integrating the **Gmail API** and **Adzuna Jobs API**: syncs and classifies job-related emails, scores job matches against my resume, and tracks every application through a pipeline, interview prep, and recruiter follow-ups across 10 live product sections.
- Designed a **human-approval-first architecture**: every AI-suggested action requires explicit confirmation before anything is sent or written, with no scraping or ToS-violating automation, backed by **196 passing tests** and a grounded Ask AI chat panel answering questions over my real tracked data.

KENNY: Autonomous AI Streamer [LINK] Python, Instagram API, Discord API, LLM

- Built **KENNY**, an AI streamer persona that replies live to real user comments and messages on **Instagram** and **Discord**, reasoning through an LLM-backed conversational pipeline running continuously in production.
- Actively engineering for reliability: building **retry logic** for failed replies, a **supervisor process**, and growing the evaluation set toward **50+ test cases**, alongside multi-platform content on Instagram and YouTube to grow the public audience.

TECHNICAL SKILLS

- **Languages:** Python, TypeScript, SQL
- **AI / ML Technologies:** TensorFlow, PyTorch, Scikit-learn, CNN, RNN, SVM, Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), LangChain, LangGraph, Prompt Engineering, Natural Language Processing (Intent Recognition, Dialogue Management), Knowledge-Base Grounded Chatbots, OpenCV, face_recognition, dlib, Model Evaluation
- **Web & Backend:** FastAPI, Next.js, React, Flask, PHP, REST APIs, Prisma, Auth.js, MVC Architecture
- **Automation & Integrations:** n8n (Workflow Automation), Gmail API, Adzuna Jobs API, Instagram API, Discord API, Tool-Calling / Function-Calling
- **AI Systems & MLOps:** RAG Pipelines, Offline/On-Device LLMs, Vector Databases (VectorDB), AI Chatbots, Enterprise Knowledge Retrieval, Docker, Microsoft Azure, MLflow
- **Databases & Tools:** PostgreSQL, MySQL, GitHub, Jira
- **Core Concepts:** Data Preprocessing, Exploratory Data Analysis (EDA), Feature Engineering, Database Design, Agile Development

EDUCATION

Master of Computer Applications (MCA), CMR University, AI & ML specialization, CGPA: 8 2024 – 2026

B.Sc. Mathematics, Subodh PG Autonomous College, 78% 2020 – 2023

CERTIFICATIONS

- AI Fluency, Anthropic
- Claude 101, Anthropic
- Deep Learning with TensorFlow, IBM
- Introduction to Generative AI, Google Cloud
- The Joy of Computing using Python, NPTEL, IIT Madras (Elite, 71%)