

Rithwik Swarnkar

San Jose, CA • Swarnkar.rithwik@gmail.com • 408-623-4797

<https://www.linkedin.com/in/rithwik-swarnkar/> • <https://github.com/rithwikswarnkar> • medium.com/@swarnkar.rithwik

US Citizen – No Sponsorship Required • Open to Relocation or Remote Roles

SUMMARY

AI/ML Developer specializing in building agentic and production-grade AI systems that integrate machine learning, deep learning, and generative intelligence. Experienced in developing multi-agent workflows using LangChain, MCP (Model Context Protocol), and RAG architectures for dynamic reasoning, retrieval, and automation. Skilled in end-to-end development, from data pipelines and model training to deployment across cloud and serverless platforms (Google Cloud Run, Vercel, Render). Proficient in Python, PyTorch, TensorFlow, and Hugging Face for computer vision, NLP, and applied GenAI. Combines a scientific foundation in neurobiology with a systems-level engineering mindset to deliver scalable, interpretable, and action-oriented AI solutions.

TECHNICAL SKILLS

Languages: Python, R, SQL

Machine Learning & Deep Learning: PyTorch, TensorFlow, scikit-learn, Keras, Hugging Face Transformers

Agentic & Generative AI: LangChain, OpenAI API, Model Context Protocol (MCP), Retrieval-Augmented Generation (RAG), Prompt Engineering, Multi-Agent Workflows

Web & API Development: FastAPI, Flask, RESTful APIs, Pydantic, CORS

Frontend Development: React, Vite, Tailwind CSS, Axios

Deployment & DevOps: Docker, Vercel, Render, Cloud Run, Cloud Build, Firebase Hosting, GitHub Actions

Cloud & Infrastructure: Google Cloud Platform (Vertex AI, Cloud Storage, IAM), Serverless & Containerized Architectures

Data Analysis & Visualization: pandas, NumPy, Matplotlib, seaborn, Plotly

Domain Expertise: Healthcare AI, Bioinformatics (Biopython), Scientific Imaging (DICOM)

Hardware Acceleration: GPU-optimized deep learning with NVIDIA CUDA 12.x and cuDNN

WORK EXPERIENCE

Applied AI/ML Developer & Founder

DeepNeuro.dev - San Jose, CA | April 2025 – Present

DeepNeuro.dev is an applied AI company developing intelligent agentic systems that combine ML, DL and GenAI to advance solutions in healthcare, bioinformatics, and real-world NLP applications.

- Lead AI product development, research, design, and deployment of production-grade AI systems.
- Architected and built end-to-end ML/DL pipelines for structured data and language modeling.
- Engineered agentic AI workflows using LangChain, OpenAI API, and Model Context Protocol (MCP) for contextual reasoning, retrieval, and multi-agent collaboration.
- Built and deployed RAG-powered, multi-agent backends on Google Cloud Run, Vercel, and Render, with automated CI/CD pipelines via Cloud Build and Firebase Hosting.
- Applied MLOps and observability principles for model lifecycle management, performance optimization, and scalable production inference.

Key Projects

Intelligent Sales Agent System

FastAPI, LangChain, PostgreSQL, pgvector, OpenAI API, Next.js, TypeScript, Render, Docker,

- Architected unified async backend with PostgreSQL, pgvector, and FastAPI using Pydantic v2 for semantic retrieval across product catalogs, company docs, and customer data.
- Developed a RAG pipeline with OpenAI embeddings for contextual reasoning, memory, and dynamic knowledge retrieval.
- Integrated LangChain for multi-agent orchestration, enabling agents to autonomously summarize, trigger follow-ups, and escalate issues.
- Designed a Next.js 14 / TypeScript frontend with Tailwind UI for live chat, sentiment tracking.

Emotion Analyzer API – End-to-End AI Application (FastAPI + React + GCP)

Python, FastAPI, Hugging Face Transformers, PyTorch, React, Vite, Tailwind CSS, Docker, Google Cloud Run, Firebase Hosting

- Built full-stack emotion classification app with DistilRoBERTa and FastAPI backend.
- Deployed app on Cloud Run with Docker, zero-idle scaling, and CORS control.
- Designed React frontend with Tailwind CSS, Axios, and CI/CD automation.

AI-Powered Brain MRI Classifier – Deep Learning for Medical Imaging

Python, TensorFlow, Keras

- Trained CNN for tumor classification with data augmentation and Grad-CAM visualizations.

Additional Projects (on GitHub):

- *Enzyme Classification*: Sequence-based protein classifier via amino acid composition features.
- *Fake News Detection (NLP)*: TF-IDF + Logistic Regression model with data leakage prevention.
- *Student Performance*: PCA and clustering (K-Means, DBSCAN) to identify academic patterns.
- *Diabetes Risk Predictor*: ML pipeline for disease risk prediction with class balancing, tuning.

Clinical Intern – Respiratory & Internal Medicine, NITRD

National Institute of Tuberculosis and Respiratory Diseases, New Delhi, India | Oct 2024 – Mar 2025

- Supported physicians in clinical data workflows and managing multi-department patient care.
- Applied Python, Excel to analyze survey data, building insights for healthcare-focused AI apps.

EDUCATION

Stanford University – Continuing Studies

TECH 152: A Crash Course in AI (Instructor: Ronjon Nag)

- Grade: A+ (2025). Covered neural networks, generative AI, ethical considerations of AI systems.

University of California, Davis – B.S. Neurobiology, Physiology, and Behavior

Graduated September 2024 | Dean's Honor List

- Relevant Coursework: Python, Data Analysis, Statistics, Neurobiology, Anatomy
- Research: Contributed to studies on metabolic inhibition of inflammation.

CERTIFICATIONS

- Deep Learning Specialization – Coursera (DeepLearning.AI)
- Applied Machine Learning for Healthcare – Coursera
- Clinical Training and Research – National Institute of Tuberculosis & Respiratory Disease