

Rohan Paul Richard

Hyderabad, India

📞 +91 91820 29325 • ✉️ rohanrichard.work@gmail.com
linkedin.com/in/rohanprichard • github.com/rohanprichard • rohanrichard.in

AI engineer with 3+ years of experience building and shipping production AI systems across agentic AI, voice, RAG, healthcare ML and full-stack applications, with five peer-reviewed publications. Experienced in owning systems end-to-end from architecture and model integration through backend development, deployment, evaluation and production infrastructure.

Professional Experience

- **AmalgamRx** **Wilmington, Delaware, USA (Remote)**
AI Engineer *May 2024–Present*
 - Designed, built and maintain an autonomous project-management agent used across the organization, turning meetings into decisions, de-duplicated action items and persistent follow-ups.
 - Sole builder and owner of the company's internal AI knowledge platform, combining a knowledge graph, semantic retrieval and MCP into a shared source of truth for employees and AI agents.
 - Built the backend and communication framework for a multi-tenant text and voice AI platform using FastAPI/WebSockets, AWS Bedrock, S3 and ElevenLabs.
 - Contributed hands-on to training Chiron, AmalgamRx's ~2M-parameter EHR-sequence transformer; its embeddings supported a ~34,200-patient screening study achieving 0.851 AUC and 94.7% NPV, which I co-authored.
 - Built production voice AI for healthcare and pharmaceutical applications using WebRTC/FastRTC, ElevenLabs and Deepgram across speech recognition, conversational inference and text-to-speech.
 - Developed a behavioral-AI research prototype progressing toward production, combining asynchronous user-model reflection, behavioral-science techniques, continuous learning and adversarial evaluation.
- **AmalgamRx** **Wilmington, Delaware, USA (Remote)**
Data Science Intern *Aug 2023–May 2024*
 - Built, deployed and load-tested a stateless AI chat service and multi-layer LLM safety stack for a patient-facing app that later ranked #1 in the App Store's Healthcare category; the work also contributed to a peer-reviewed publication.
- **Logituit** **Bangalore, Karnataka, India**
Software Trainee *May 2023–Jul 2023*
 - Built production Go REST APIs across multiple databases and a movie recommendation system for an OTT use case.

Technical Skills

Languages: Python, TypeScript/JavaScript, SQL, Swift, C++

AI / ML: LLM agents, MCP, RAG, knowledge graphs, transformers, PyTorch, scikit-learn

Voice: WebRTC/FastRTC, ElevenLabs, Deepgram, Whisper, Silero VAD

Systems: FastAPI, React/Next.js, Postgres/pgvector, Neo4j, Redis, Snowflake, AWS Bedrock, EC2, S3, Docker

Publications

- **Passive Screening for Obstructive Sleep Apnea via Value-Augmented EHR Token Embeddings** — HIMS / CSCE, Springer Nature, 2026
- **Taming Large Language Models for Healthcare** — CSCE, Springer CCIS, 2025
- **A Client-Server Based Educational Chatbot for Academic Institutions** — IEEE CONIT, 2024, First Author
- **Assistance for Visually Impaired People in Identifying Multiple Scenes Using Deep Learning** — Springer, 2024
- **A Deepfake Image Classifier System for Real and Doctored Image Differentiation** — IEEE ICICT, 2024

Selected Open Source

claude-overnight — Claude Code productivity tooling

Fantastic App — behavioral-science mobile app with 2,000+ MAU

Open-source Contributor

Honors and Awards

AI Breakthrough Award — "Overall Large Language Model of the Year" (AmalgamRx team), 2026

AI Breakthrough Award — "Best Large Language Model Application" (AmalgamRx team), 2024

Education

Karunya Institute of Technology and Sciences — B.Tech Computer Science and Engineering, 2024 • CGPA 8.07