

Thota Rahul

Vijayawada, Andhra Pradesh | +91 8328436890 | rahulthota21@gmail.com

 linkedin.com/in/rahulthota21 |  github.com/rahulthota21

Summary

Final-year CSE student with hands-on experience designing and deploying production-ready, AI-integrated backend systems. Proven track record of taking complex projects from initial architecture to final deployment, managing delivery timelines, and optimizing system outcomes. Adept at collaborating within or leading cross-functional teams with a strict focus on execution and shipping reliable software. Brings a pragmatic, results-driven mindset to engineering, ensuring scalable solutions move from concept to production. Authored three research papers on these implementations, published at IEEE conferences.

Education

Amrita Vishwa Vidyapeetham, Amritapuri Campus

B.Tech in Computer Science and Engineering

Cumulative GPA: 8.29 / 10.0

Kollam, Kerala

August 2026

Experience

AI/ML Intern - Infosys Springboard

Oct 2024 – Dec 2024

GlucSense: Diabetes Risk Prediction System

- Designed a classification pipeline in Scikit-learn to predict diabetes risk from an imbalanced Kaggle dataset; benchmarked Extra Trees, XGBoost, and SVM under consistent train/test splits and selected Extra Trees based on comparative test accuracy.
- Built the preprocessing layer using median imputation for missing values and correlation-based feature pruning to remove redundant columns; validated pipeline decisions through class-distribution and model-output plots in Matplotlib.

Projects

HealthATM: AI Lung Nodule Detection System | Team Project | GitHub

Jan 2026 – May 2026

- Built the backend for a lung nodule detection system for rural clinics - patients upload CT scans, receive AI-analyzed reports, and interact with a multilingual voice assistant with no radiologist needed on-site.
- Served a **0.32 MB** INT8-quantized MobileUNet3D (Dice **0.803**, AUC 0.800 on LIDC-IDRI; 120× smaller than teacher) via REST API - runs fully on-device without a GPU; research accepted at GCON-26 (IEEE).
- FastAPI + Redis/RQ for async scan queuing; LangGraph ReAct agent for clinical reasoning; Gemini 2.0 Flash Live for real-time voice in 10+ Indian languages; FHIR R4-compliant report export; Supabase with row-level security.

Mock'n-Hire: AI-Powered Resume Screener | Project Lead | GitHub

March 2025 – July 2025

- Built a recruiter-facing AI system that automates resume shortlisting, interview question generation, and candidate emotion analysis - each function running as an independent FastAPI service, configurable per job role.
- Tested the ranking model on 250 resumes across 5 roles against human recruiter judgments and hit **86.4% Precision@10**; end-to-end throughput clocked at 50 resumes per 3.2 minutes and ~2.5-second question generation per query.
- Ranking engine built with Mistral-8x7B and Sentence Transformers; all session data, video responses, and candidate records stored in Supabase (PostgreSQL + object storage) with JWT auth and role-based access tiers.

AffectLearn: Sentiment-Aware Intelligent Tutoring System | Project Lead | GitHub

Jan 2025 – Apr 2025

- Built a tutoring system that reads student emotion through text, voice, and webcam inputs in real time and adapts lesson content and quiz difficulty on the fly - no manual teacher input at any stage.
- Deployed a LoRA fine-tuned TinyLLaMA as the lesson reasoning backend on a standard 8GB GPU; a controlled student pilot returned **85.2% approval** on response clarity and tone.
- Indexed OpenStax textbooks with FAISS for grounded lesson generation; used Whisper for voice, CNN for facial sentiment; logged all scores and session history in Supabase to personalize future lessons per student.

Publications & Achievements

- HealthATM: 3D Lung Nodule Detection via Knowledge Distillation** - Co-Author, GCON-26 (IEEE).
- Mock'n-Hire: AI System for Automated Resume Screening** - 2nd Author, ICCNT 2026 (IEEE, Scopus-indexed).
- AffectLearn: Sentiment-Aware Adaptive STEM Tutor with LoRA-Tuned LLMs** - 2nd Author, IEEE TALE 2026.
- Top-6 Finalist, **Tredence Hackathon 2025** - built TravelSync, an AI itinerary planner with real-time travel API integration, in a 48-hour sprint.

Technical Skills

Languages: Python, SQL (PostgreSQL), MongoDB

Backend & Tools: FastAPI, REST APIs, Redis, Supabase, LangChain, Docker, Git/GitHub, Express Js, Mongoose, Prisma

ML & AI: Scikit-learn, XGBoost, Pandas, FAISS, Sentence Transformers, RAG Pipelines, Context Engineering

Core CS: OOP, DBMS, Operating Systems, Data Structures