

Md Anish Ahamad

Greater Noida, India | +91 70339 35778 | mdanisahamad2003@gmail.com | github.com/anishraja04 |
linkedin.com/in/md-anish-ahamad | anishahamad.com

Backend Engineer | AI Evaluation and Agent Reliability

Summary

Backend engineer who builds agent systems and the harnesses that catch them being wrong. Built a two-agent debugging system that resolved 11 of 11 incidents where a single-prompt baseline resolved 0, by refusing any fix until the full test suite passes. Run two live products used by exam candidates across India. Currently paid to author the tasks that evaluate AI coding agents.

Skills

Languages: Python, Java, SQL, JavaScript

Backend: FastAPI, Spring Boot, Django REST Framework, REST API design, Spring Data JPA, JWT authentication

Data: PostgreSQL, MySQL, MongoDB, ETL

Infrastructure: Docker, Docker Compose, GitHub Actions, Terraform, AWS EC2, AWS VPC, Nginx, Linux, Git

AI and agents: AI evaluation, Benchmark authoring, Agent architecture, LangGraph, MCP, RAG, FAISS, Prompt engineering, Structured output

Work Experience

AI Coding Task Author

Handshake, Remote | Jul 2026 - Present

- Author realistic software engineering tasks used to evaluate AI coding agents, writing the issue description, a reference implementation and the automated tests that decide whether a fix is correct.
- Grade AI-generated code for correctness and adherence to engineering practice, distinguishing output that is right from output that only reads as right.
- Build and validate tasks against real repositories using Git and GitHub workflows.

CFD Engineer - AI Task Designer

Alignerr, Remote | Jul 2026 - Sep 2026

- Designed AI evaluation tasks in computational fluid dynamics with precise problem statements, reference solutions and deterministic checking, using OpenFOAM.
- Assessed AI-generated engineering answers for technical accuracy against defined criteria.

LLM Post-Training Intern

Ethara AI, Gurugram, India | Jan 2026 - Apr 2026

- Refined datasets and engineered prompts to improve large language model output quality.
- Analysed failure cases and validated model behaviour to assess deployment readiness.

Projects

exactsizetool.com - document tooling for exam forms

exactsizetool.com

- Built and run a suite of roughly 90 free browser-based tools that resize photos, signatures and documents to the exact pixel and kilobyte limits Indian government exam portals require.
- All processing runs client side, so candidate photos and signatures never leave the device and no server holds personal documents.
- Added an assistant that reads the raw instruction text from an exam portal and configures the correct tool and values automatically.

gateguide.co.in - GATE preparation books

gateguide.co.in

- Built and run a study business: six books covering CSE, ECE, Civil, Data Science and AI, Electrical and Mechanical, totalling 5,732 questions with worked solutions and 60 full mock tests.
- Derived topic weighting by analysing real past papers question by question rather than following syllabus order; published 80 free guides as the acquisition channel.

Incident Engineer - agentic root-cause debugging

github.com/anishraja04/incident-engineer

- Built a two-agent system that diagnoses and fixes software incidents: a triage agent forms hypotheses from logs, an investigator agent uses tools (read, grep, run-tests, sandboxed Python) to locate the cause and apply a fix.
- Designed a verification gate so a fix is only accepted once the full test suite passes, preventing the confident but incorrect patches single-prompt approaches produce.
- Measured against a deterministic harness over 11 incidents: 11 of 11 resolved versus 0 of 11 for the baseline, at roughly 5 minutes and 0.0079 USD per task.

KaamSathi - team task manager

github.com/anishraja04/team-task-manager

- Full-stack task manager on Django REST Framework and PostgreSQL with JWT authentication and a React single-page front end.
- Implemented role-based access control at two levels, platform wide and per project, so authorisation checks both permission and project membership.
- Deployed on Railway with the API, front end and database as separate services.

Cloud-Native DevOps deployment platform

github.com/anishraja04/cloud-native-devops

- Provisioned AWS infrastructure as code with Terraform: VPC, public subnet, security groups and an EC2 instance.
- Containerised the application behind an Nginx reverse proxy, with GitHub Actions building the image and deploying on every push.

Education

Bachelor of Technology (B.Tech), Computer Science and Engineering

Noida Institute of Engineering and Technology, Greater Noida, India | Jul 2022 - Jun 2026

- CGPA 8.42 out of 10, with Honors.

Certifications

Data Structures, University of California San Diego | Mar 2024

Introduction to Artificial Intelligence, IBM | Jan 2023

Frontier Engineering Challenge 2026 Participant, micro1 | 2026