

AVIJIT PRATAP SINGH

Software Developer • Full-Stack / Backend Engineering

Email: avijitpratapsin@gmail.com | Contact: +91 9630938987 | [GitHub](#) | [LinkedIn](#) | [Portfolio](#) | [LeetCode](#)

SUMMARY

Software Developer with hands-on experience building AI-driven and full-stack web applications using React, Node.js, AWS and cloud-native architectures. Skilled in REST API design, LLM integration, system scalability, distributed caching and performance optimization, with experience delivering low-latency, production-ready systems with measurable impact.

EDUCATION

B.Tech in Computer Science and Engineering

Vellore Institute of Technology - AP

Expected Graduation: July 2026 | CGPA: 8.58

PERSONAL PROJECTS

AI-Driven Healthcare Symptom Checker | React.js, Node.js, AWS Lambda, API Gateway, Redis, LLaMA

GitHub - github.com/Avijit12/healthcare-symptom-checker | Live - hsc.avijitsingh.dev

- Built a full-stack AI symptom-checker using React.js and serverless Node.js on AWS Lambda/API Gateway; processed 1,000+ real world symptom queries with **~85-90%** context-aware diagnostic accuracy.
- Integrated Hugging Face Inference Router (OpenAI-compatible, LLaMA-3.1-8B) with production resiliency, including automatic retries, timeout handling, HTTP-429 backoff, and model-load polling.
- Implemented Redis caching with deterministic keys and 24-hour TTL, reducing repeat-query latency and upstream LLM calls under concurrent workloads.
- Delivered secure, server-isolated REST APIs with secret isolation and Vercel proxy routing; achieved **<300 ms** average response time while preventing CORS and key-exposure risks.

AstraQA - Selenium Script Generator | Next.js, Node.js, PostgreSQL, Prisma, Supabase S3, LLaMA

GitHub - github.com/Avijit12/AstraQA | Live - astra.avijitsingh.dev

- Built an AI-driven Selenium script generation platform using Next.js, Node.js, and PostgreSQL with Prisma ORM, reducing manual test authoring effort **by ~60-70%**.
- Designed Supabase S3-compatible storage ingestion to process 100+ page knowledge bases per build, improving LLaMA-powered test-case relevance **by ~40%** versus prompt-only generation.
- Developed modular backend APIs for knowledge-base construction, RAG-assisted LLaMA test generation, and Selenium script synthesis, sustaining **<50 ms** average PostgreSQL query latency under load.
- Deployed on Vercel with Google/GitHub authentication and background processing, delivering scalable performance under heavy workloads with future multi-LLM extensibility support.

SKILLS

Programming Languages Java, JavaScript, TypeScript, SQL

Frameworks & Libraries HTML, CSS, React.js, Next.js, Node.js, Express

Database & Caching PostgreSQL, MySQL, MongoDB, Redis

Cloud & DevOps Tools AWS, GCP, Docker, Kubernetes, Git, GitHub, Postman, Vercel, Render

Backend & APIs RESTful APIs, Authentication/Authorization, CI/CD Pipelines, Integration Testing

CS Fundamentals Operating Systems, Computer Networks, DBMS, OOP, Algorithms & Data Structures

Technical Interests Scalable backend architectures, LLM & RAG applications, DevOps engineering

ACHIEVEMENTS

- Solved 300+ [LeetCode](#) problems and 450+ problems across platforms ([Codolio](#)), strengthening algorithmic thinking, data structures, and problem-solving depth.
- Active [GitHub](#) contributor with multiple production-grade, cloud-hosted full-stack and DevOps projects.

CERTIFICATIONS

AWS Cloud Architecting - AWS (2024) | ([Link](#))

DevOps, Agile and Design Thinking - IBM (2025) | ([Link](#))

Generative AI Certification - Oracle (2025) | ([Link](#))