

Fazlul Karim Choudhury

Email: fazlul0127@gmail.com | Phone: +91 8486853823 | Location: India

LinkedIn: [linkedin.com/fazlul0127](https://www.linkedin.com/fazlul0127) | Github: github.com/FazlulKarimC | Leetcode: leetcode.com/u/fazlul_karim

Professional Summary

Software Engineer with 1 year of professional experience at ADP and a B.Tech in Computer Science, specializing in full-stack web development and AI-enabled systems. Proven track record building production-grade applications with React, Next.js, Node.js, and PostgreSQL, alongside hands-on expertise in RAG pipelines, LLM evaluation, and vector databases. Experienced in Agile teams, code reviews, and TDD with a strong focus on clean architecture, performance optimization, and scalability.

Work Experience

Automatic Data Processing (ADP) - Member Technical | Oct 2023 - Sept 2024

- Built an employee onboarding system with **React** and **Spring Boot**, enabling HR, managers, and new hires to access role-specific features via **JWT**-based authentication.
- Designed responsive, cross-platform, user-friendly dashboard interfaces using **React**, **Bootstrap**, and custom **CSS**, supporting core features like offer letter generation, email template automation, and document tracking.
- Implemented secure **RESTful APIs** with Spring Data **JPA** and implemented role-based access control via Spring Security, ensuring appropriate data access and smooth workflow management across user types.
- Collaborated in an **Agile** team environment to deliver new features end-to-end, improving onboarding efficiency and supporting 5,000+ employees with streamlined HR operations and reduced manual effort.

Projects

LlmForge | Jan 2026 - Present | LlmForge.vercel.app

- Built a full-stack LLM experimentation platform (**FastAPI**, **Next.js 16**, **PostgreSQL**, **Qdrant**) for systematically comparing Naive, **Chain-of-Thought**, **RAG**, and **ReAct** reasoning strategies across 10 datasets (factual QA, multi-hop reasoning, math, agent benchmarks, and 3 adversarial safety sets) with config-driven, version-controlled experiment management
- Engineered a multi-provider inference layer (**HuggingFace**, **OpenRouter**, **Groq**) with async-safe execution, **LRU prompt caching** (up to 10x speedup), **batched parallelism** for 2–3x throughput gains, exponential backoff with jitter, and auto-routing provider fallback via **Upstash Redis + RQ task queue**
- Designed a research-grade evaluation pipeline computing accuracy (**exact match**, **F1**), latency (**p50/p95/p99**), token cost, **LLM-as-Judge** quality scores, and a deterministic robustness scorer against prompt injection and jailbreak datasets with McNemar's chi-squared and bootstrap 95% CI for statistical significance testing across experiment pairs

PaperSight_AI | Jan 2025 - Present | Papersight.vercel.app

- Built an AI document platform (**Next.js 16**, **React 19**, **TypeScript**, Gemini API) with client-side PDF parsing via pdfjs-dist, token-by-token SSE streaming, five style-specific summary modes, and a cookie-based guest rate limiter with server-side validation.
- Engineered a visible **RAG** pipeline using **pgvector** cosine similarity on Neon **Postgres**; chat responses surface expandable source citation cards with retrieved chunks and relevance scores, with a three-stage UI animation telegraphing retrieval, ranking, and generation
- Implemented 500-word text chunking (50-word overlap), 768d Gemini embeddings, and vector search via raw SQL Prisma adapter for serverless Neon, with background indexing triggered post-upload to minimize cold-start latency

Brutalkart (E-Commerce) | May 2025 - Present | Brutalkart.app

- Built a full-stack e-commerce platform (**Next.js 16**, **React 19**, **TypeScript**, **Express**, **PostgreSQL**) with product listings, smart search, discount codes, and a 3-phase transactional checkout pipeline with concurrency guards and atomic rollback on failure.
- Implemented server-side guest-to-authenticated cart merging in a single Prisma transaction with inventory validation, session cleanup, discount logic (per-user limits, date-range validity, global caps), RBAC, and dual-layer Zod validation across a 615-line PostgreSQL schema (20+ models)

Technical Skills

- Languages:** Python, JavaScript (ES6+), TypeScript, Java, C/C++
- Frontend:** ReactJS, NextJS, HTML5, CSS3, Redux, Tailwind CSS, Jest, React Testing Library
- Backend:** NodeJS, ExpressJS, Spring Boot, NestJS, FastAPI, JWT/OAuth
- Architecture & Design:** Microservices, RESTful APIs, MVC Patterns, Object-Oriented Programming
- Databases:** SQL, NoSQL, PostgreSQL, MongoDB
- DevOps/Tools:** Git, AWS, GitHub, Docker, CI/CD pipelines, Jenkins, Linux/Unix
- AI & LLM:** RAG, Prompt Engineering, LangChain, HuggingFace, OpenAI APIs, Vector Databases, LLM Evaluation
- Familiar With:** GraphQL, Redis, Zustand, React Native, Tanstack Query, SEO, and Figma.

Education

- Bachelor of Technology in Computer Science | July 2019 - June 2023**
North-Eastern Hill University (NEHU)