SAKIBUL HUDA ONNAY

Full-Stack Developer · AI-Integrated Systems · Django / React / FastAPI
Dhaka, Bangladesh | onnoy29@gmail.com | 01740175704
Portfolio | GitHub | LinkedIn | PyPI: voxrouter



OBJECTIVE

Full-stack developer shipping production ERP systems and AI-integrated routing infrastructure. Built VoxRouter, a hybrid LLM routing agent, for AMD's Developer Hackathon — including a self-graded benchmark suite and a published PyPI package. Seeking a full-stack or AI-integrated developer role to design and ship modern applications.

EDUCATION

BSc in Computer Science and Engineering *2021 – Present*
American International University-Bangladesh (AIUB) *Expected Graduation: 2026*

TECHNICAL SKILLS

Frontend: React.js, JavaScript (ES6+), HTML5, CSS3 (Tailwind, Bootstrap), Responsive Design, REST API Integration

Backend: Django, FastAPI, Node.js, Express.js, REST APIs, Server-Sent Events / Streaming

Databases: MongoDB, MySQL, SQL

AI / LLM Integration: Google Gemini API, Fireworks AI, Ollama (local model routing), LLM-driven complexity classification, prompt engineering

Languages: Python, JavaScript, C, C++, Java

ML & Data: NumPy, Pandas, regression models, YOLO, AlexNet

Infra & Tools: Docker, Docker Compose, Git, GitHub, CI-friendly deployments (Render)

EXPERIENCE

Junior Software Engineer Trainee *April 2026 – July 2026*
Team Solutions Bangladesh | Dhaka, Bangladesh

- Developed and maintained ERP system modules using Django, React.js, JavaScript, HTML5, CSS3, and MySQL.
- Implemented new features, bug fixes, and UI enhancements based on business requirements.
- Worked with backend logic, database operations, and REST APIs to support ERP functionalities.
- Communicated with clients to gather requirements and explain software workflows.
- Collaborated with senior developers using Git; participated in testing, debugging, and code reviews.

PROJECTS

VoxRouter — Hybrid Token-Efficient LLM Routing Agent (AMD Developer Hackathon ACT II)
Source Code · Live Demo · PyPI Package
FastAPI, Ollama, Google Gemini, Fireworks AI, React, Docker Compose

- Built a routing agent that classifies task complexity and dispatches to local (Ollama/ROCm) or remote (Gemini/Fireworks) models, cutting token cost by up to 75% vs an always-remote baseline.
- Designed a 4-layer complexity classifier with confidence-based escalation — low-confidence local answers automatically retry on a stronger remote model.
- Built a self-graded 50-task benchmark suite producing a reproducible accuracy / routing / cost score, and published the routing engine as a standalone pip-installable SDK.
- Shipped with live token streaming, a real-time React dashboard, and one-command Docker Compose deployment.

ZapShift — Full-Stack Delivery Management Application (MERN Stack)
Source Code · Live Site
MongoDB, Express.js, React.js, Node.js, Firebase, Tailwind CSS, RESTful APIs

- Built a full-stack delivery app with role-based access control for admin, user, and rider roles.
- Implemented Firebase authentication and real-time tracking with map services and RESTful APIs.
- Ran a production-readiness audit: fixed a price-tampering vulnerability, added auth middleware to unprotected routes, and rotated an exposed API key.

Personal Portfolio Website
Source Code · Live Site
HTML5, Tailwind CSS

- Designed a custom portfolio centered on a methodology-first narrative rather than a stack list, with a deliberate typography-driven design system.

COMPETITIVE PROGRAMMING

- Active on Codeforces (FhaLiFA) and CodeChef (onnoy) — 150+ problems solved across 14+ rated contests.
- Strong foundation in Data Structures and Algorithms using C++, with focus on pattern recognition for algorithmic technique identification.

RESEARCH & ACADEMIC WORK

- Gas Leakage Detection using IoT and Machine Learning (Under Review)
- QUIC-Flow Watermarking Under Connection Migration: Design and Preliminary Measurement

REFERENCES

Shamsad Zaman — Senior Software Engineer, IQVIA
shamsadzaman@yahoo.com | 01729-315771