

NAZRUL HUDA SHANTO

📍 New York, NY 📞 +1 405-564-6370 ✉️ mshanto@okstate.edu 🌐 dev.nazrulhuda.com
🐙 github.com/nazrulhuda 🔗 linkedin.com/in/nazrulhudashanto 🎓 Google Scholar

Summary

Software and ML engineer (M.S. CS, 3.82 GPA) with 3+ years of combined industry and research experience: building LLM agents and backend APIs, training and deploying ML models on AWS and Kubernetes, and turning large evaluation and log datasets into measurable results. First-author LLM-agent research; IEEE TCC paper (29+ citations).

Education

M.S. in Computer Science, Oklahoma State University, Stillwater, OK – **GPA: 3.82/4.0**

Jan 2024 – May 2026

B.S. in Computer Science, BRAC University, Dhaka, Bangladesh

May 2018 – May 2022

Technical Skills

Languages: Python, SQL, Java, TypeScript/JavaScript, Bash

ML & AI: PyTorch, TensorFlow, scikit-learn, Hugging Face Transformers, LSTM, LIME; LangGraph, LangChain, MCP, OpenAI API

Data Science: pandas, NumPy, Matplotlib, Seaborn, SPSS; feature engineering, regression, ablation studies, model evaluation

Backend & Web: FastAPI, Flask, Django, Spring Boot, Node.js, REST/OpenAPI, Next.js, React, JWT/JWKS, CSRF

Cloud & DevOps: AWS (Lambda, ECR, S3, Textract, Amazon MQ), Docker, Kubernetes, Istio/Envoy, Prometheus, Linux, Git

Databases: PostgreSQL/PostGIS, Redis, MongoDB, Supabase; SQL schema design, migrations, transactions, row locking

Experience

Graduate Research Assistant

Jan 2024 – May 2026

Oklahoma State University

Stillwater, OK

- Built an LLM agent (LangGraph, FastMCP, Redis, PostgreSQL, Docker) with **8 MCP tools** that lets non-experts run and verify zero-knowledge proofs of statistical tests on private clinical data; first-author paper under review (IEEE BigData 2026).
- Designed *skeptical tools*, **20 server-side checks** that validate, complete, or override LLM tool arguments; raised task completion from **70.2% to 93.0%** (Qwen 3 32B), about +23 points on all 3 model families; cut response time from **16.6 s to 12.4 s**.
- Designed the evaluation: an 82-query test suite and 5-configuration ablation (**6,000+ runs**); analyzed **6,371** result records, found 198 silently misrouted status checks, and built an LLM grader that matched human labels on **150 of 151** replies.
- Engineered six traffic features from Envoy/Prometheus telemetry (EMA-smoothed) over **4,705** labeled 20-second windows and trained per-service Random Forest, Gradient Boosting, SVM, and Decision Tree classifiers (**94.0–97.8%** accuracy).
- Deployed the detector on **Kubernetes/Istio** as one least-privilege pod with a JSON model registry and Isolation Forest fallback; parallel per-pod inference cut detection time by **55.9%** (37.4 vs. 84.8 ms), and traffic reroutes automatically.
- Fixed a silent train/inference feature-order mismatch; an earlier prototype analyzed **1.2M+** Envoy log entries across 6 services and explained each detection with LIME; also built LSTM forecasting of container resource exhaustion.

Junior Software Engineer

Feb 2023 – Dec 2023

Tirzok Private Ltd.

Dhaka, Bangladesh

- Led an ML document-validation system for CliQloan, a mortgage platform: classifiers sort and trim documents before AWS Textract extracts key fields; **45× faster** than manual validation (scikit-learn, Spring Boot, PostgreSQL, AWS Lambda/ECR).
- Built RESTful APIs (Node.js, MongoDB) for a government–citizen service portal serving **80,000+ citizens**; optimized aggregation pipelines to cut query latency by **40%** and designed JWT-based role management.
- Built a Flask currency-conversion API (5 currencies, PostgreSQL) for a payroll and tax system, enabling real-time multi-currency payroll; deployed services on Linux cloud servers.

Undergraduate Research Assistant

Oct 2020 – Jan 2023

BRAC University

Dhaka, Bangladesh

- Built a cloud image storage and retrieval framework with a compression module (up to 27% smaller files, **54%** less waiting time, validated with SSIM on COCO); corresponding author, IEEE TCC.
- Ran SPSS regression and descriptive analysis to find barriers to primary education in remote regions (ACM COMPASS 2026), and led fieldwork data collection for two digital-divide studies (73 participants, 15 villages).

AI Instructor, NSF Research Experiences for Teachers (RET)

May 2025 – Aug 2025

Oklahoma State University – taught Python, ML, and Transformers to 10 K–12 teachers

Stillwater, OK

Projects

FwdStar – Freight Marketplace Platform

Jan 2026 – Present

Technical Lead and Sole Architect (Next.js 16, TypeScript, FastAPI, PostgreSQL/PostGIS, Docker, Supabase)

Bangladesh

- Turned a client’s verbal needs into a **2,695-line** spec and shipped a freight marketplace in nine phases (sign-up, fleet management, load board, awards, GPS tracking, SMS alerts) with a bilingual frontend and **50+** REST endpoints.
- Designed a PostgreSQL/PostGIS schema (**20+ tables, 13 migrations**) with state machines and **SELECT FOR UPDATE** locks so a load is never awarded twice; layered security with JWT/JWKS, CSRF, OTP rate limiting, and audit logs.

Selected Publications

- IEEE Transactions on Cloud Computing** (2022; corresponding author; 29+ citations) · **IEEE BigData 2026** (first author, under review) · ICOIN 2027 (under review) · NSysS 2022 (oral) · ACM COMPASS 2026