

Harsh Dwivedi

New York, United States | +1 (680) 216-2032 | edwivediharsh@gmail.com | LinkedIn | GitHub | Portfolio

SUMMARY

AI Engineer and Backend developer with 3+ years of experience building real-time platforms, backend automation services, and production-grade systems across energy, fintech, and tech consulting. Skilled in **Python, FastAPI, Redis, SQL**, and distributed system design, with applied experience using **NLP, LLMs, embeddings**, and **RAG** architecture to enhance product features.

WORK EXPERIENCE

RELI Group Inc

Syracuse, NY (Remote)

Platform Engineering Intern

Jan 2026 – Present

- Built an **NLP-based** compliance matrix system using **spaCy, PySBD, and Sentence Transformers** to extract, categorize, and surface contractual requirements from SOWs, enabling proposal teams to review and track all compliance obligations.
- Designed and deployed an **acronym annotation and validation engine** using deterministic parsing logic and the **Schwartz–Hearst algorithm**, replacing a **\$10K/year third-party subscription** while improving proposal document quality and review efficiency.
- Developed an **RAG-based** capability matrix and organizational search engine using **FAISS, LangChain, and LLM based** reasoning to **semantically** retrieve past performance data and generate reasoning on organizational fit for new SOW requirements.
- Deployed document analysis and RAG services using **Azure App Service and Azure Functions**, storing documents and **persisted vector indexes in Azure Blob Storage**, while researching **data isolation, security boundaries, and responsible AI deployment** to enable future governed, agentic AI workflows.

iConsult Collaborative

Syracuse, NY

Team Lead & Full Stack Engineer

Mar 2025 – Dec 2025

- Led a five-member engineering team to design and ship **Gymin**, a real-time fitness and sensor analytics platform, improving sprint throughput by **20%** through structured planning, reviews, and delivery workflows.
- Built a scalable backend using **Python** and **FastAPI** to process continuous sensor and user activity streams, adding async execution, **Redis caching**, and a **Kafka-based event pipeline** to support sub-**100ms** API responses.
- Implemented structured logging, metrics, and health checks for backend services using **AWS CloudWatch**, improving production visibility and speeding up incident triage for API and pipeline failures.
- Developed and shipped the Gymin **web app** using **Next.js** and **React** with server-side rendering and optimized routing to improve load performance; integrated Cognito authentication and RBAC for secure multi-tenant access.
- Deployed containerized **CI/CD** workflows with **Docker** and **GitHub Actions** on **AWS EC2**, reducing deployment time by **40%** and enabling zero downtime releases with automated rollback.

Tata Consultancy Services (TCS)

Mumbai, India

Assistant System Engineer

Oct 2022 – Jul 2024

- Delivered backend IT operations for ENI Energy's EU infrastructure by integrating **automation scripts** into enterprise systems, improving uptime by **15%** and strengthening security compliance across **100+ servers**.
- Automated core backend workflows using **Python, REST APIs**, and **ServiceNow** integration, reducing manual workload by **4 hours per day** and accelerating ticket resolution by **30%**.
- Designed monitoring and alerting microservices with Python, PowerShell, and **Azure Monitor APIs** to track performance and failures across **200+ servers**, cutting incident detection time by **25%**.
- Improved system performance by refactoring **legacy Python** utilities into modular, reusable components and optimizing database queries and caching strategies, reducing API latency by **35%** under peak load.

EDUCATION

M.S. in Computer Engineering, Syracuse University, Syracuse, NY

Aug 2024 – May 2026

B.E. in Electronics Engineering, Mumbai University, Mumbai, India

Aug 2019 – May 2023

TECHNICAL SKILLS

Languages: Python, Java, C#, JavaScript/TypeScript, GO, SQL.

Backend Development: FastAPI, Flask, Node.js, REST APIs, Microservices Architecture, API Design, WebSockets, Celery, Redis, Kafka

Web and Frameworks: React, Next.js, Express.js, HTML5, CSS3, JavaScript, Frontend–Backend Integration

Cloud and Platform: AWS (EC2, S3, Lambda, DynamoDB), Azure (Functions, Monitor), Kubernetes, API Gateway, CloudWatch

Databases: PostgreSQL, SQL Server, MongoDB, DynamoDB, Databricks, Snowflake, ETL Pipelines, Data Modeling, Query Optimization

DevOps & Delivery: Docker, Git, CI CD, Grafana, Prometheus, System Design, Jira, Agile / Scrum, Jenkins

AI / ML & LLM Systems: RAG, LLMs, LangChain, Prompt Engineering, Hugging Face, PyTorch, scikit learn, Embeddings, Similarity Search, Vector DB, Text Classification, Model Evaluation, A B Testing, Rule Based Validation