

# Harsh Dwivedi

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

## EDUCATION

M.S. in Computer Engineering, Syracuse University, Syracuse, NY	Aug 2024 – May 2026
B.E. in Electronics Engineering, Mumbai University, Navi Mumbai, Maharashtra	Aug 2019 – May 2023

## TECHNICAL SKILLS

**Programming Languages:** Python, Java, C++, C#, TypeScript, JavaScript

**Web & Frameworks:** React, Node.js, Express, Flask, FastAPI, Spring Boot, Next.js, Tailwind CSS, Bootstrap

**Cloud & DevOps:** AWS (EC2, S3, DynamoDB, RDS), Azure, Docker, GitHub Actions, RBAC

**Databases:** PostgreSQL, MySQL, MongoDB, Firebase, DynamoDB

**AI & LLM Tools:** LangChain, OpenAI API, RAG pipelines, Prompt Engineering, XAI (Explainable AI)

**Tooling & Platforms:** Git, Postman, Poetry, VS Code, ManageEngine, Power BI, Power Automate, ServiceNow APIs

## WORK EXPERIENCE

### iConsult Collaborative | Syracuse, New York

<b>Team Lead &amp; Full-Stack Developer</b>	Mar 2025 – Present
• Led a 4-member Agile team, managing system architecture, backlog grooming, and code reviews, increasing delivery speed.	
• Developed and deployed a scalable, SEO-optimized web app using Next.js, React, and Tailwind CSS, achieving 45% faster load times and a Lighthouse SEO score of 95+.	
• Developed multi-floor, multi-date occupancy visualizations with Recharts to deliver real-time insights into gym equipment usage.	
• Architected client-side routing and state management using React Router and Context API to support seamless navigation.	
• Simulated backend services using AWS S3 and DynamoDB, and automated analytics data generation using Node.js scripts.	
• Implemented Git Actions CI/CD workflows, reducing deployment time by 40% and enabling zero-downtime production updates.	

### NEXIS AI Lab | Syracuse, New York

<b>AI Researcher</b>	Jan 2025 – Present
• Researching explainability in large language models by integrating token-level saliency into a retrieval-augmented generation (RAG) workflow.	
• Applied embedding-based semantic similarity to compare entities, enhancing RAG-driven document retrieval and relevance.	
• Integrating LLMs into CLI tools via Poetry and LangChain, enabling intelligent code workflows like code optimization.	
• Exploring emerging techniques in interpretable NLP (attention rollouts, attribution mapping) to improve LLM traceability.	

### Tata Consultancy Services (TCS) | Mumbai, India

<b>Assistant System Engineer</b>	Jan 2024 - Jul 2024
• Provided IT infrastructure support for ENI, managing user access, Azure resources, and enterprise systems across Europe.	
• Collaborated with DevOps and cloud teams to administer servers, optimize Azure resources, and ensure compliance with RBAC and data security protocols.	
• Developed Python scripts to automate ticket resolution and routine tasks via ServiceNow APIs, reducing resolution time by 30%.	
• Led the migration of over 1,500 user profiles, resources, and SharePoint data to new domains during organizational restructuring.	

### Picky with Picks | Mumbai, India

<b>Full-Stack Developer (Freelance)</b>	Nov 2022 – Jan 2023
• Delivered a responsive single-page application using ReactJS, TypeScript, Node.js, and Express.js, ensuring seamless client-server communication with minimal load times.	
• Designed and integrated RESTful APIs, improving backend maintainability and reducing client-side runtime errors by 40%.	
• Reduced load time by 35% through API throttling, lazy loading, and error boundary implementation.	
• Implemented form validation, session storage, and modular state logic to enhance UX and reduce submission bugs across workflows.	

### TCR Innovation | Mumbai, India

<b>Data Engineering Intern</b>	Jun 2021 – Aug 2021
• Built automated ETL pipelines using Python, Pandas, and Flask, reducing manual processing time by 50% and improving accuracy across financial datasets.	
• Developed real-time dashboards with Power BI, enabling KPI monitoring with under 2s latency across 10,000+ data points.	
• Designed RESTful APIs with modular service architecture to support secure and scalable multi-client data ingestion pipelines.	

## PROJECTS

<b>CodeFlow-AI</b>   Python, LangChain, OpenAI API, FastAPI, CLI	May 2025 – Present
• Developed a CLI + web-based AI pair programmer with contextual memory using FastAPI, LangChain agents, and RAG.	
• project multi-turn task planning and architecture suggestions via prompt chaining and user session tracking.	
• Designed FastAPI interfaces for local and remote collaboration, with customizable prompt templates for different dev use cases.	

<b>DevPack</b>   Python, LangChain, OpenAI API, Poetry, CLI	Mar 2025 – Present
• Built DevPack, a command-line toolkit for intelligent workflows like code analysis, test generation, and refactoring using LLMs.	
• Engineered modular CLI commands with Poetry and workflow YAMLS, automating coding tasks with GPT backend integration.	
• Reduced manual coding effort by 40% in testing by streamlining multi-step dev operations via configurable, explainable prompts.	