Rajat Tiwari

🌐 therajattiwari.com ✉ rajattiwari1099@gmail.com 🔗 linkedin.com/in/rajat-tiwari-biz

Experience

SRSWTI **January 2024 − Present**
AI Engineer
- Designed and deployed **Srswti Axis**, a production-grade **multi-agent AI framework** using **LangChain and CrewAI**, featuring tool-use, memory, planning/re-planning, and agent orchestration patterns that reduced manual workloads by 55% for pilot clients across insurance and logistics domains.
- Prototyped, iterated, and productionized domain-specific AI agents on **AWS cloud infrastructure** (EC2, S3, Lambda), implementing end-to-end observability with **LangSmith** and custom monitoring dashboards tracking agent performance metrics (task completion rate, latency, token usage).
- Engineered scalable Python-based backend APIs (**Django REST Framework**) serving AI agent orchestration, integrating **OpenAI GPT-4, Claude, and open-source LLMs** with secure authentication, rate limiting, and high-availability architecture (99.9% uptime).
- Built robust data ingestion pipelines processing 200k+ daily records from 30+ sources (SFTP, REST APIs, Selenium), normalizing unstructured data into **PostgreSQL vector databases** for RAG-powered agent retrieval and context enhancement.
- Automated deployment lifecycle using **Terraform (IaC), Docker, and GitHub Actions CI/CD**, implementing blue-green deployments with automated testing (pytest, 90%+ coverage) and ensuring seamless agent updates with zero downtime.
- Collaborated cross-functionally with business stakeholders to translate requirements into technical AI solutions, conducting rapid prototyping cycles and iterating based on user feedback to optimize agent performance and business value delivery.

Weaftech **June 2022 − June 2023**
Software Engineer
- Developed high-performance **Python backend systems** with low-latency APIs (p99 < 40ms) serving real-time data to user-facing applications, demonstrating strong software engineering fundamentals and scalable architecture design.
- Designed microservices architecture on **Kafka** processing 5k msgs/sec with Redis caching and PostgreSQL persistence, implementing monitoring and alerting for system health and performance optimization.
- Applied **machine learning techniques** (scikit-learn, pandas) to analyze 500k+ user records, building predictive models that reduced churn by 10% and informed product strategy through data-driven insights.
- Built automated data extraction pipelines using **Python and Selenium**, converting unstructured documents into normalized datasets and saving 20+ hours/week of manual processing through intelligent automation.

Tüv Nord **Nov 2021 − June 2022**
Software Engineer
- Developed interactive dashboards using **React and Node.js** for real-time industrial sensor data visualization, demonstrating ability to translate complex technical data into actionable insights for non-technical stakeholders.
- Built RESTful APIs for live data ingestion and processing, working in agile cross-functional teams to deliver solutions that improved operational safety and reduced incidents by 5%.

Projects

AI Healthcare App (Capstone, Fidelity Investments) | *Python, React.js, PostgreSQL, LLM* **Aug 2024 − Dec 2024**
- Developed a full-stack price-transparency app that embeds **gpt-4-powered natural-language search** over a warehouse of **600 k+ charge entries from 1800 u.s. hospitals**, returning sub-second cost answers and cpt/hcpcs code explanations
- Optimised the backend with **django rest framework**, partitioned **postgresql** tables, **celery** async workers, and **redis** caches, cutting p95 api latency from 680 ms to 470 ms (**-31%**) while sustaining 2 000 rps under load
- Designed a modular **react.js** frontend featuring llm-assisted price-comparison widgets and explainability tooltips; analytics show an **18% lift in daily active sessions** and a 24% increase in quote-to-click conversions

Education

University of Texas at Dallas **Jan 2023 − Dec 2024**
Master of Science in Management Science
- Relevant Coursework: Data Structures and Algorithms, Cloud Computing, Machine Learning, Artificial Intelligence

Birla Institute of Technology and Science Pilani **Sep 2018 − June 2022**
Bachelor of Engineering in Computer Science
- Relevant Coursework: Operating Systems, Computer Networks, Object-Oriented Programming, Data Mining

Skills & Certifications

Languages & Core Frameworks: Python, Java, C++, SQL, Javascript, HTML, CSS, React.js, Node.js, Django, Nest.js
AI / ML toolchain: Pytorch, Tensorflow, scikit-learn, llms (gpt-4o, gemini, llama-3), NLP, Reinforcement learning, Prompt engineering, Fine-tuning, Sentiment analysis, Chatbot architecture
Data Engineering & Cloud: AWS (lambda, ec2, s3), Docker, Kubernetes, Pyspark, Neo4j, Redis, Kafka, Graphql, Rest Apis, Large-scale Data Pipelines, Pulumi/Terraform, CI/CD
Analytics & bi: Pandas, Numpy, Tableau, Power Bi, Looker, SAS, Predictive Modelling, Churn Analysis, Demand Forecasting
Certifications: IBM Data Science Professional Certificate, Microsoft AI & ML Engineering Certificate, Nvidia Deep Learning Specialisation
Soft Skills: Leadership, Problem Solving, Critical Thinking, Team Collaboration, Communication, Project Management