MUKESH KUMAR

Computer Science & Engineering

@ sihagmukesh22@gmail.com   📞 +91-9079971790   ⭘ mukeshsihag   💼 itismukesh

EXPERIENCE

Delivery Engineering Manager

Turing | Python, Fastapi, Nextjs, Mongodb, Agentic AI, Postgres, GCP

📅 June 2025 – Present   📍 Remote, India

- Leading a team of 20+ developers across multiple AI initiatives, driving execution, quality, and timely delivery.
- Led development of UI and Non-UI gyms replicating real-world platforms (HR, ERP, Social Media, Salesforce, ServiceNow, ADP) to simulate production environments for AI model evaluation.
- Built and deployed MCP servers and delivered AI-driven applications directly to clients (including 7 AI Labs), working closely with stakeholders to gather requirements and ensure tailored solutions.
- Designed and launched Turing's internal LLM benchmarking platform to evaluate and compare model performance across UI and Non-UI environments.
- Developed an internal Slack monitoring bot to track client queries across channels and auto-alert relevant stakeholders, significantly improving response time and client satisfaction.

Senior Software Engineer

Sarvam.ai | Python, Fastapi, Bazel, Helm, Kubernetes, Postgres, Agentic AI, Azure, AWS, CLI

📅 Feb 2025 – June 2025   📍 Bengaluru, India

- Contributed to build India's Sovereign LLM Model, Sarvam-M.
- Designed and implemented Helm charts to streamline the deployment of microservices on Kubernetes clusters.
- Designed and implemented AI-driven retrieval systems using custom-built coder, retriever, and planner components to fetch contextually relevant answers from large datasets.
- Processed and analyzed large-scale datasets from various companies and government sources, leveraging vector databases, RAG techniques, and LLMs like GPT, DeepSeek, and Claude.

Founding Software Engineer

Eugenie.ai | Python, Django, Kafka, Mqtt, Docker, Azure, GCP, FastAPI, AWS, Kubernetes, ReactJs, Redis, Celery

📅 June 2021 – Jan 2025   📍 Bengaluru, India

- Developed backend APIs using Django, and improved existing APIs by 50-60% by optimizing queries, adding indexes, and refining logic.
- Developed and deployed a data streaming pipeline using Kafka, MQTT, and HAProxy which transfers lakhs of data records daily in a production environment.
- I implemented the CI/CD workflow with GitHub Actions and automated the deployment on the Kubernetes cluster.
- I designed the architecture and created a FastAPI microservice to trigger the ML pipeline, reducing the processing time from approximately 1 minute to less than a second.
- Configured and managed the ELK (Elasticsearch, Logstash, Kibana) stack to facilitate logging and monitoring of various systems and applications.
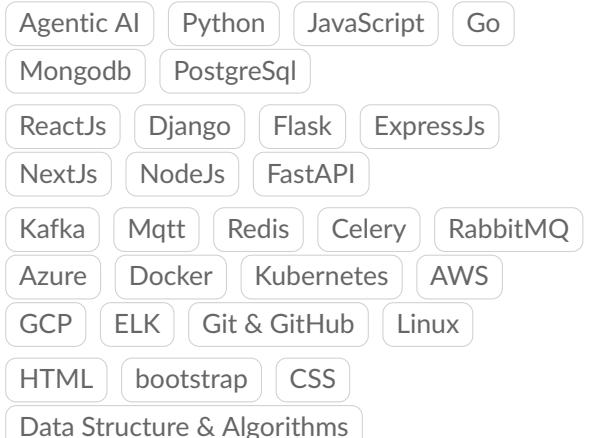
EDUCATION

M.Tech (CSE)

National Institute of Technology, Hamirpur

- CGPA: 9.62/10

SKILLS

Agentic AI | Python | JavaScript | Go | Mongodb | PostgreSql | ReactJs | Django | Flask | ExpressJs | NextJs | NodeJs | FastAPI | Kafka | Mqtt | Redis | Celery | RabbitMQ | Azure | Docker | Kubernetes | AWS | GCP | ELK | Git & GitHub | Linux | HTML | bootstrap | CSS | Data Structure & Algorithms

PUBLICATIONS

- Gupta, J., Gupta, N., Kumar, M., & Duggal, R. (2022). An Introduction to IoT-Based Posture Detection and Underlying Sensor Technology. In Smart and Secure Internet of Healthcare Things (pp. 31-48). CRC Press.

AWARDS

- 2nd prize in International IEEE Dataset Upload Contest. 100k+ samples of data is collected using esp32 sensors & Arduino Uno and submitted to the IEEE dataset contest.
- Winner of Hult Prize competition. My idea was chosen as the best startup idea among the total 53 teams. A total of 5 teams from India were selected, of which 1 was my team.

PROJECTS

Virtual Terminal | Python, Git

- I created a project called virtual terminal for safe experimentation with Linux commands, ensuring system stability and minimizing the risk of irreversible changes caused by incorrect command execution.