

Shiv Sharan Kumar

📍 India 📩 shivsharan47@gmail.com ☎ 9122562491 💬 in/shiv-sharan-kumar-93aa3219b

SUMMARY

Software Engineer with 4+ years of experience in building AI-powered applications using FastAPI, React, and AWS. Expert in deploying YOLOv8 models, real-time progress tracking, and distributed task execution across local and cloud systems. Skilled in WebSocket integration, Celery task management, and scalable architecture design. Developed RAG-based chatbots for internal document querying. Strong in full-stack development, Python, and AI/ML workflows.

EXPERIENCE

Engineer

Motherson

April 2024 - Present, Gurgaon

- Built a full-stack AI system with **React** frontend, local **FastAPI** backend for hardware tasks, and **EC2**-hosted backend for model inference.
- Enabled real-time status updates using **WebSocket** and progress bar integration.
- Designed step-based task flow (steps 1–3 local, 4–5 on **EC2**) for efficient **CPU/GPU** utilization.
- Optimized backend communication to simplify **EC2** status reporting to the frontend.
- Created a **RAG**-based **chatbot** for querying internal company documents.
- Trained and deployed **YOLOv8** model via **Roboflow** with 92% accuracy for blood cell classification.

Software Engineer

Perceptiviti

August 2021 -March 2024, Gurugram, Haryana, India

- Proficient in **Python**, **Django**, **React.js**, **MySQL**, and **PostgreSQL**.
- Built and maintained 5 web apps with 20% better performance and 30% higher user engagement.
- Deployed apps on **AWS EC2** with **CloudFront**, cutting server costs by 25% and load times by 40%.
- Optimized code to improve response times by 25% and reduce resource usage by 30%.
- Developed an **NLTK**-based search engine and an **analytics** tool.

PROJECT

AI-Powered Microscopic Blood Analysis Platform

Motherson • May 2024 - Present

- Built an end-to-end application with three components: **React frontend (local)**, **FastAPI backend for machine tasks (local)**, and **FastAPI+Nginx+Celery backend on AWS EC2 GPU instance**.
- Designed a 5-step processing workflow where steps 1–3 run locally and steps 4–5 execute on **EC2** for model inferencing using **YOLOv8**.
- Integrated real-time progress tracking using **WebSockets** from the backend to display step status on the frontend progress bar.
- Proposed and analyzed efficient methods for relaying **EC2** processing status to the frontend via the local backend.
- Developed a **chatbot** using **RAG** to answer product-specific queries from **dynamically** generated internal documents.
- Trained a **YOLOv8** model with Roboflow, achieving 92% accuracy in classifying **RBCs**, **WBCs**, **PLTs**, and subtypes from **microscopic** blood smear images.

EDUCATION

MCA

Birla Institute Of Technology • Mesra • 2021

SKILLS

- Programming Languages:** Python, JavaScript, SQL
- Web Development:** FastAPI, Flask, Django, React, Streamlit
- Front-End Technologies:** HTML, CSS, Tailwind CSS
- Frameworks & Libraries:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, NLTK, PyTorch, TensorFlow, OpenCV
- Data Science Tools:** Jupyter Notebook, Google Colab, MLflow, Weights & Biases, Roboflow, LabelImg
- Cloud Platforms:** AWS, Azure, Google Cloud Platform (GCP)
- Database Management:** MySQL, PostgreSQL, Oracle
- Web Infrastructure & Server Technologies:** Nginx, Gunicorn, Celery
- Version Control & CI/CD:** GitHub, AWS CodeCommit
- Other Skills:** Problem-Solving, Algorithm Design, Strong Analytical & Debugging Abilities, Excellent Communication & Team Collaboration