

Shiv Sharan Kumar

📍 Gurgaon, Haryana, India ✉ shivsharan47@gmail.com ☎ 9122562491 🌐 in/shiv-sharan-kumar-93aa3219b 🖱 shivsharan.netlify.app/

SUMMARY

AI/ML Engineer with **4.5+** years of experience in **Python, FastAPI, React, PyTorch**, and **cloud deployment**, building end-to-end AI systems, computer vision, and **NLP** solutions, with a strong focus on **Generative** and **Agentic AI**.

EXPERIENCE

Senior Engineer

Motherson

April 2024 - Present, Gurgaon

- Built a full pipeline for microscopic blood smear analysis using YOLO, integrating camera hardware, Arduino controls, FastAPI, AWS GPU deployment, and a React dashboard.
- Optimized large-batch image inference and implemented cloud storage and processing with AWS S3.
- Developed a glucose prediction system using ML models with end-to-end data pipelines and real-time prediction APIs.
- Created MiBuddy, an agentic AI assistant, integrating OpenAI APIs for multi-step reasoning, task automation, and internal knowledge retrieval.

Software Engineer

Perceptiviti

August 2021 - April 2024, Gurugram, Haryana, India

- Developed and maintained 5+ web applications using Python, Django, React, MySQL, PostgreSQL.
- Optimized application performance by 20–30% through code and architecture improvements.
- Designed responsive UI for 15+ projects, improving engagement by 30%.
- Deployed applications on AWS EC2, reducing infrastructure cost by 25%.
- Implemented CloudFront optimizations, reducing load time by 40%.

PROJECT

MiBuddy

Motherson

- Developed MiBuddy, an internal AI agent used by 50+ users for research and task automation.
- Integrated OpenAI APIs, reducing manual research time by 40%.
- Implemented agent-style workflows, improving multi-step task efficiency by 30%.
- Built a knowledge assistant indexing 10,000+ internal documents.
- Designed scalable FastAPI backend supporting 100+ concurrent requests.
- Improved response accuracy by 25% through prompt engineering and conversation optimization.

Blood Smear Analysis System

Motherson

- Built an end-to-end microscopic blood smear analysis pipeline using YOLO, processing 500+ images per case.
- Integrated USB camera, Arduino controls, and AI inference, reducing manual effort by 40%.
- Developed FastAPI backend and deployed models on AWS GPU, achieving 3× faster inference.
- Managed 10,000+ images using AWS S3 with automated processing workflows.
- Designed React-based diagnostic interface, improving interaction speed by 30%.
- Optimized batch inference, cutting processing time from 20 minutes to under 8 minutes per case.

EDUCATION

Master of Computer Applications - MCA, Computer Application

Birla Institute of Technology • Mesra • 2021

SKILLS

Programming Languages: Python, JavaScript

AI/ML & Data Science: Scikit-learn, Pandas, NumPy, Matplotlib, OpenCV, YOLO, NLP, LLMs, Prompt Engineering

Generative & Agentic AI: OpenAI API, LLM Integration, AI Chatbots, Agentic Workflows, RAG (Retrieval-Augmented Generation), AGNO, n8n, crewai

Backend Development: FastAPI, Django, Flask, REST APIs, Celery

Frontend Development: React.js, Material-UI, HTML, CSS, Tailwind, Streamlit

Database Management: MySQL, PostgreSQL, Oracle

Cloud & Deployment: AWS (EC2, S3, CloudFront), Azure

DevOps & Tools: GitHub, AWS CodeCommit, Nginx, Gunicorn

Architecture & Systems: End-to-End AI Pipelines, API Design, Microservices Basics