

Sachin Pokharel

AI/ML Engineer

Kathmandu, Nepal

Contact No: 9761727066, 9808033697

sachinwork33@gmail.com

<https://www.linkedin.com/in/sachin-pokharel-493716197>

<https://github.com/Sachin-Pokharel>

<https://www.sachin-pokharel.com.np>

Summary

AI/ML Engineer with 2+ years of experience specializing in Computer Vision, delivering production systems for object detection, document analysis, biometric verification, and image processing. Expert in detection models, OCR engines, liveness detection, and vision pipelines deployed via FastAPI and containerized microservices. Also experienced in Generative AI (RAG-based LLM agents) and NLP, with a track record of deploying vision solutions across fintech, health, and green energy. Strong in end-to-end ML lifecycle—from research and model development to production deployment and monitoring—with a focus on scalable, production-ready architectures.

Technical Skills

- **Programming Languages:** Python
- **ML & Deep Learning Frameworks:** PyTorch, Tensorflow/Keras, Scikit-learn, HuggingFace Transformers
- **Computer Vision:** YOLO (Ultralytics), OpenCV, Mediapipe, DeepFace, RoboFlow, Image Processing
- **Generative AI & LLM Tooling:** Langchain, LangGraph, LlamaIndex, OpenAI SDK, Pydantic Ai, CrewAI, MCP, Vector Databases (PineCone, Qdrant, Milvus), RAG Systems, Multi-Agent Orchestration
- **Backend & API:** FastAPI (Preferred), Django Rest Framework, Flask, RESTful API design, Microservices Architecture
- **Databases:** MongoDB, PostgreSQL, Neo4j Graph, Database Design
- **DevOps & Cloud:** Docker, Containerization, AWS, GCP, Oracle Cloud, Digital Ocean, CI/CD, GitHub Actions
- **MLOps & Monitoring:** Model Evaluation, LLM Monitoring, Performance Optimization, Production Deployment
- **Automations Tools:** n8n, Make
- **Version Control:** Git, GitHub

Work Experience

Senior ML Engineer

Nov 2025 – Present

ThirdFactor AI, Lalitpur, Nepal

- Architected and developed a production OCR engine for extracting information from all Nepali government IDs with multilingual support achieving >93% accuracy across document types and reducing manual processing time by 80%.
- Leading development of a passive liveness detection system to meet ISO standard iBeta requirements (Level 1 & Level 2), implementing state-of-the-art anti-spoofing techniques to enhance security and prevent fraud.
- Conducting research and development on human age estimation models and leading initiatives in forgery and deepfake detection, advancing the company's biometric verification capabilities.

AI/ML Engineer

Jan 2024 – June 2025

Aerawat Corp, Lalitpur, Nepal

- Developed and integrated face detection, document detection, and gesture recognition technologies to create a seamless KYC onboarding experience for digital wallet users, reducing onboarding time by 60% and improving user conversion rates.
- Collaborated with cross-functional stakeholders to ensure regulatory compliance and implemented robust security measures, including encryption and data privacy protocols, to protect sensitive customer information.
- Led internal research on generating personalized ABA therapy plans for children with autism using diagnosis and screening reports, exploring AI applications in healthcare to support therapy planning and outcomes.

Data/ML Engineer

Mar 2023 – Jan 2024

Khalti Pvt. Ltd, Lalitpur, Nepal

- Built a deep learning-based semantic search engine using NLP techniques and deployed via Django REST API, improving search relevance by 40% and enhancing user experience for 1M+ users.
- Developed a probabilistic record linkage model for duplicate detection on identity documents, achieving 99.5% accuracy and ensuring data integrity across the platform.
- Integrated Neo4j graph database to model user relationships and social networks, enabling personalized recommendations that increased user retention by 25%.

Freelance & Consulting

AI/ML Engineer (Contract)

Apr 2025 – June 2025

Logictive Solutions, Kathmandu, Nepal

- Developed a computer vision pipeline using high-resolution Google satellite imagery to detect rooftop solar panels, enabling solar adoption analysis and helping providers identify optimal installation locations.
- Fine-tuned YOLOv11 model for solar panel detection and deployed real-time inference API via FastAPI, achieving 96% detection accuracy and processing 1000+ images per hour. Containerized with Docker and deployed on AWS EC2 with 99.9% uptime.

Backend Developer (Part-Time)

June 2024 – Feb 2025

Yirifi.ai, Singapore

- Architected and developed Hybrid RAG-based LLM chat agents for regulatory management using LlamaIndex, serving clients in crypto, digital wallets, and blockchain ecosystems with specialized regulatory compliance support.
- Integrated MongoDB for persistent chat history and context-aware interactions, enabling seamless multi-turn conversations and improving user engagement by 50%. Implemented document QA system leveraging LlamaIndex and LlamaParse for parsing and indexing regulatory documents.
- Built and deployed scalable FastAPI services managing high-volume API interactions between users and AI agents, handling 5K+ requests daily.

ML Engineer (Contract)

May 2024 – July 2024

Swivt Technologies, Lalitpur, Nepal

- Developed a Retrieval-Augmented Generation (RAG) chatbot system for insurance companies in Hong Kong, enhancing customer service efficiency and reducing response time by 70% while maintaining high accuracy.
- Integrated Google-based search for real-time information retrieval and document upload capabilities, enabling users to chat with documents and improving client satisfaction scores by 85%.
- Provided mentorship to students in Data Science and AI courses at Swivt Education, guiding 15+ students through their final capstone projects and contributing to curriculum developments.

Professional Engagement

Mentor – IOE Final Year Students Project

Institute of Engineering, Pulchowk, Nepal (2025)

- Mentored final year engineering students on a Human Action Recognition (HAR) project, providing technical guidance on computer vision techniques, deep learning model development, and implementation strategies for action classification systems.

Judge Panelist – AI in Education Hackathon

Kings College Hackademia Event, Nepal (2023)

- Evaluated 20+ innovative AI solutions in education technology, assessing technical merit, scalability, and real-world impact. Selected winning projects that demonstrated potential for transforming educational outcomes through AI applications.

Education

- **MSc In Computer Science** **Sep 2019 – Nov 2021**
Blekinge Institute of Technology, Karlskrona, Sweden
Dissertation: **ARMAS: Active Reconstruction of Missing Audio Segments**
Thesis Report: <http://www.diva-portal.org/smash/get/diva2:1608536/FULLTEXT02.pdf>
- **B.Tech in Information Technology** **Jul 2013 – Jul 2017**
Indian Institute of Information Technology, Allahabad, India