

Anuj Gaida

Bhaktapur, Nepal | gaidaanuj@gmail.com | +977-9861755117 | linkedin.com/in/anuj-gaida
github.com/Anuj-Gaida

Professional Summary

AI Engineer with 2+ years of experience building and deploying production-grade AI solutions. Skilled in AI agent development, LLM fine-tuning and quantization, geospatial AI, and end-to-end ML pipelines, with strong Python backend expertise using Django and FastAPI. Experienced in containerized deployments (Docker), cloud-native AI workflows (AWS), and local LLM serving using vLLM, llama.cpp, and Ollama. AWS Certified Solutions Architect with hands-on MLOps experience focused on scalable model deployment and maintainability. Computer Engineering graduate with active research experience, adept at translating research into deployable systems and collaborating with cross-functional teams.

Professional Experience

AI Engineer, Naxa Pvt. Ltd. – Kathmandu, Nepal

June 2024 – Present

- Built an end-to-end AI agriculture bulletin system using LangGraph agents, RAG pipelines, locally-served LLMs and Django to auto-generate municipality-templated PDF bulletins.
- Designed and deployed production-ready AI agents using Python, LangChain, and LangGraph with multi-step reasoning and tool integration.
- Built full-stack AI solutions with LLM backends, React frontends, and FastAPI/Django APIs.
- Served and optimized local LLMs using vLLM, llama.cpp, and Ollama for low-latency production inference.
- Fine-tuned open-source LLMs using LoRA/QLoRA and applied quantization (GGUF, GPTQ, AWQ) for resource-constrained deployment.
- Implemented geospatial AI pipelines using OSM, NetworkX, GeoPandas, and QGIS for spatial querying, analysis, and visualization.
- Developed semantic segmentation models for drone imagery analysis including road extraction, building footprint, and damage detection optimized for Nepalese contexts.
- Deployed AI models as scalable inference services using FastAPI, Docker, and AWS (SageMaker/EC2).
- Conducted R&D on deep learning-based flood forecasting for ungauged rivers, integrating HEC-RAS hydraulic models.
- Managed databases to support large-scale AI pipelines.
- Delivered AI and GeoAI contributions across **GenCV, DMAPS, DAASTA, Road Extraction, Building Damage Detection, Building Footprint Extraction**, and **Flood Forecasting R&D**.
- Supervised day-to-day tasks of junior team members, ensuring alignment with project goals, timelines, and engineering best practices.
- Participated in the hiring process for junior AI/ML engineers, conducting technical interviews and assessments to build and scale the team.

AI Automation Expert, Food And Agriculture Organization of UN – Kathmandu, Nepal

Nov 2025 – Dec 2025

- Developed agentic AI solutions for automated proposal writing and report generation using Copilot Studio, Power Automate, and Power Apps across standardized organizational templates.
- Integrated SQL, Dataverse, SharePoint, and OneDrive data sources to enable context-aware, dynamic document generation via AI agents.
- Automated repetitive tasks, approval workflows, and report formatting using AI-driven agents, improving organizational efficiency.
- Collaborated with FAO teams and stakeholders to tailor AI agents to specific operational and compliance requirements.
- Documented agent workflows, integration pipelines, and templates to ensure maintainability and scalability across the organization.

Technical Skills

Programming & AI Frameworks: Python, PyTorch, TensorFlow, Keras, LangChain, LangGraph, CrewAI, LangSmith, Hugging Face Transformers, Scikit-learn, OpenCV

LLM Development & Optimization: Fine-tuning (LoRA, QLoRA), Quantization (GGUF, GPTQ, AWQ), RAG Pipelines, Prompt Engineering, LLM Evaluation

Local LLM Serving: Ollama, vLLM, llama.cpp

LLM APIs & Platforms: OpenAI API, Anthropic Claude, AWS Bedrock, Groq, Hugging Face Inference API

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

Databases & Data Management: PostgreSQL, MySQL, MongoDB, Redis, Elasticsearch, SQLAlchemy, Pandas

Cloud & Deployment: AWS (EC2, ECR, ECS, S3, Lambda, SageMaker, Bedrock), Docker, Docker Compose

Microsoft Stack & AI Automation: Copilot Studio, Power Automate, Power Apps

Data Processing & Pipelines: Apache Airflow, Pandas, NumPy, GeoPandas, n8n

Geospatial & Visualization: QGIS, NetworkX, Matplotlib, Plotly, Folium, Seaborn

CI/CD & Automation: GitHub Actions, Jenkins, Bash, Python scripting

Version Control: Git, GitHub, GitLab

Operating Systems: Linux (Ubuntu, CentOS), Windows

Projects

AR Tour Companion – Augmented Reality Tour Guide App

- Developed a mobile app providing AR-based navigation, routing, monument detection, and interactive natural conversation with a virtual guide.
- Enabled immersive and context-aware user experiences for tourists using computer vision and AR technologies.
- Technologies: Kotlin, Python, TensorFlow, Azure, GitHub, Mapbox

GenCV – Automated CV Generation System

- Built a system to generate professional CVs in World Bank format, enabling business teams to create standardized resumes efficiently.
- Integrated AI agents using LangChain and AWS Bedrock for intelligent content generation, with a web frontend for easy access.
- Technologies: Python, Django, LangChain, AWS Bedrock, GitHub, React

Education

BE in Computer Engineering

Khwopa College of Engineering – Tribhuvan University

- Relevant Coursework: Data Structures, Algorithms, Machine Learning, Artificial Intelligence, Deep Learning, Cloud Computing, Database Systems, Software Engineering, Computer Networks, Operating Systems

Research Achievements

Paper Presentation at International Conference on Technologies for Computer, Electrical, Electronics Communication (ICT-CEEL 2023)

Khwopa College of Engineering

- Presented a research paper on Paddy leaf disease detection using a hybrid CNN-LSTM architecture, achieving improved accuracy on unseen data over traditional CNN-based models.

Research Works

- Conducted research and development using diverse geospatial datasets to deliver production-level models for garbage detection, road detection, building damage detection, and flood prediction systems.

Certifications

Python Certification – Khwopa College of Engineering

Machine Learning Scientist – DataCamp

AWS Certified Solutions Architect – Associate (SAA-C03) – AWS

AWS Certified Cloud Practitioner (CCP) – AWS

Community & Writing

Technical Writer – AWS & MLOps (Beginner)

- Started to write beginner-friendly articles on AWS, MLOps, AI/ML workflows, and practical Python-based solutions.

Community Educator – Python, AI & ML Classes

- Conduct workshops and classes at local level to teach Python programming, AI, and machine learning fundamentals to students and enthusiasts.