Rupesh Ghimire

AI/ML Engineer & Researcher | rupeshghimire17@gmail.com | +977 9868 155 925

Portfolio | LinkedIn | Github | Medium

PROFESSIONAL SUMMARY

As a registered general Computer Engineer from Nepal, I have been working as AI/ML Engineer and Researcher with over a two-years of experience in creating, training, and evaluating advanced machine learning models, including large language models and multimodal models. I am familiar with techniques such as pretraining, instruction tuning, LoRA fine-tuning, reinforcement learning (DPO), and working with large-scale data for AI pipelines. With my strong forte as critical thinking, and systematic problem-solving, I have produced consistent, high-performing classification models in image, audio and text domains. I am increasingly drawn to examining the changing challenges at the intersection of AI efficacy, security, and transparency via more academic research.

RESEARCH INTEREST

AI, Machine Learning, Robotics and Automation, Model Quantization, Trustworthy AI, Adversarial ML, Cybersecurity-oriented AI Systems, Multimodal Learning, NLP & Knowledge Graphs

Education

Bachelors in Computer Engineering Oct 2019 – Apr 2024
Tribhuvan University, Institute of Engineering – Pashchimanchal Campus Percentage: 75.05%
Pokhara, Nepal

- **Coursework:** Computer Organization and Architecture, Theory of Computation, Artificial Intelligence, Digital Signal Analysis & Processing, Artificial Intelligence, Image Processing and Pattern Recognition, Software Engineering, Engineering Mathematics, Probability and Statistics

Micro Degree Jan 2023 – Jan 2024
Fusemachines - AI Fellowship Link
REMOTE, Nepal

- **Coursework:** Machine Learning, Time Series Forecasting, Reinforcement Learning, Deep Learning, Computer Vision, Natural Language Processing, Language Models, Transformers

High School - Science: STEM May 2017 – May 2019
Nepal Education Board(NEB) - Nepal Mega College GPA: 3.85/4.0
Kathmandu, Nepal

- **Coursework:** Physics, Chemistry, Mathematics, Computer Science

Academic Achievements

Fusemachine AI Fellowship 2023-2024
Golden Jubilee Scholarship Scheme 2019-20 2019-2023
Merit Based Scholarship for Undergraduate Studies - Indian Embassy

Research & Technical Skills

Machine Learning & AI: PyTorch, HuggingFace, Transformer Architectures, LoRA, DPO, RAG, Knowledge Graphs
Research Focus: Federated Learning, Model Quantization, Custom Model Design
Data & Experimentation: Data Preprocessing & Augmentation, STFT, LogMel, Evaluation Metrics, MLflow, Evidently AI, Literature Review
Backend & MLOps: Django, FastAPI, Docker, Nginx, Prefect, DVC, Server-based Pipeline Automation
Data Science & Tools: Pandas, NumPy, Scikit-learn, Matplotlib, LangChain, LlamaIndex, Vector Databases
Programming: Python, C, C++, SQL

Technical Reports

Gantavya – Landmark Recognition May 2024
Undergraduate Major Project Report, Tribhuvan University Link
Developed a landmark recognition system using YOLOv8, achieving **97% accuracy** on 15 classes; deployed as a mobile app with Django and React Native.

Research & Experience

North Star Developer's Village – AI/ML Engineer Aug 2024 – Present
Kathmandu, Nepal

- Designed and fine-tuned transformer-based image-to-text models using LoRA and PEFT methods for resource-constrained deployment.
- Built custom Transformer models in PyTorch for text generation and sequence-to-sequence tasks.
- Developed Federated Learning architecture for mobile and web clients, integrating model quantization for improved efficiency.
- Automated reproducible MLOps pipelines (MLflow, Prefect, DVC, Evidently AI) on company servers for experiment tracking and auditability.
- Containerized and deployed AI services (Django, FastAPI) as RESTful APIs with Docker and Nginx.
- Contributed to robotics R&D team through camera calibration and vision-based model development.

North Star Developer's Village – AI Research Intern May 2024 – Aug 2024
Kathmandu, Nepal

- Conducted literature reviews across NLP, audio processing, and multimodal learning to determine optimal architectures for company research.
- Performed prompt engineering for GPT-3.5, DALL·E-3, Whisper, and TTS APIs to produce accurate, context-aware outputs.
- Designed and trained a custom audio classification model using STFT and LogMel features, achieving **90.01% accuracy** on a curated dataset.

Panacea Solution Pvt. Ltd. - Data Science Mentor June 2023 – Feb 2024
Kathmandu, Nepal

- Taught 3 batches (50 students) a 3-month Data Science curriculum (Python, Statistics, ML, Visualization, DL).
- Designed hands-on projects using Kaggle datasets, guiding students through end-to-end workflows.

Selected Projects

Search Synthesizer 2024
RAG, LangChain, LlamaIndex, VectorDB, Web Scraping

- Built recursive LLM-based RAG pipeline with GoogleSearch API, content scraping, vector embeddings, and contextual media/sources.

Gantavya – Landmark Recognition 2024
YOLO, Transfer Learning, Django, React Native

- Created and augmented dataset for 15 landmark classes, achieving **97% accuracy** in detection/localization; deployed as mobile app with map navigation and JWT authentication.

Nepali Sentiment Analysis 2023
BERT, NLP, Flask API

- Preprocessed Nepali text (SpaCy/NLTK), fine-tuned NepBERTa to **76% accuracy** and deployed as real-time sentiment classification API.

Competitions & Awards

1st Runner Up – Datathon, Delta Tech Fest 2023
1st Runner Up – Hult Prize OnCampus, IOE-Pokhara Link – 2020

Volunteering & Outreach

ML Mentor – Innovative Computer Engineering Student's Society 2024
Conducted workshops on Python, Scikit-learn, Pandas, and guided student ML projects. Link

Django Mentor – Innovative Computer Engineering Student's Society 2023
Conducted training workshop on Python, Django, Database, ORM and RESTful APIs. Link