

Firoj Paudel

Kathmandu, Nepal | <https://www.linkedin.com/in/firoj-paudel-17a66724b/> | firojpaudel.professional@gmail.com
<https://github.com/Firojpaudel> | <https://firojpaudel.com.np> | +977 9741661176

Education

Bachelor of Science in Computer Science and Information Technology
Madan Bhandari Memorial College

2022 – Present
Kathmandu, Nepal

Publications

- Fragile Mastery: On-Device Language Models* arXiv:2503.22698
- Proposed the **Generalized Edge Model (GEM)**, a framework for optimizing on-device language models across resource-constrained edge devices.
- Thinking About Thinking: Inverse Reasoning for Self-Aware Models* arXiv:2507.00092
- Proposed **Inverse Reasoning**, a framework for improving model interpretability and reasoning transparency.
- LoRA-Enhanced PaliGemma for Efficient VQA in Gastrointestinal Imaging* MediaEval 2025
- Adapted **PaliGemma** with LoRA for efficient visual question answering on gastrointestinal medical images.

Experience

Speech Engineer
WiseYak

December 2025 – Present
Kathmandu, Nepal

- Developing and fine-tuning **Nepali and Maithili Text-to-Speech (TTS)** systems through multilingual speech data curation and model optimization.
- Building multilingual **ASR** and **Machine Translation** (Indic ↔ English) systems for Nepali and Maithili.

Co-Founder and CRO
SAGEA

May 2025 – Present
Kathmandu, Nepal

- Developed **Helios**, an AI-powered KYC platform with passive liveness detection and document extraction, deployed by enterprise clients including **Khalti**.
- Proposed **Inverse Reasoning**, a framework for improving model interpretability and reasoning.

Research Scientist
NeuraGuard

June 2025 – June 2026
Kathmandu, Nepal

- Developed a multi-agent phishing website detection pipeline with **92% accuracy**, expanding to email filtering and deep web monitoring.

Data Science and Machine Learning Fellow
Fellowship.ai

April 2025 – July 2025
San Francisco, CA

- Developed a high-accuracy web agent for real estate and financial data extraction using **Playwright**, **Chromium**, and Pydantic Schemas (87+% task automation accuracy).
- Transitioned from SmolAgent MVP to a scalable production solution for efficient local deployment.

Skills

- Programming Languages:** Python, C++, C, JavaScript, CUDA.
- Technologies:** PyTorch, Transformers, LangChain, Keras, Scikit-learn, Numpy, Matplotlib, Seaborn, Pandas, SQL, Git, FastAPI, Docker.
- Languages:** Nepali, English, Hindi.

Leadership Experience

Co-Lead @ MBMC IdeaX 2025

2025

- Co-led the hackathon, managing and coordinating all organizing and participating teams.

General Manager @ MBM IdeaX 2024

June 2024 – Sep 2024

- Organized a nationwide hackathon in Nepal. My major task was to organize the hackers and my organizing team.

Notable Projects

100 Days of CUDA

A 100-day CUDA learning journey covering GPU programming fundamentals, parallel algorithms, CNN implementation, and Triton kernels.