Sumit Lasiwa

Dekocha, Bhaktapur, Nepal
+977 9811459601 — lasiwasumit@gmail.com
linkedin.com/in/sumit-lasiwa-30ab86346 — https://github.com/Sumitlasiwa

Professional Summary

Final-year Computer Engineering student with strong interest in Artificial Intelligence and Deep Learning. Experienced in building and fine-tuning transformer models for NLP and computer vision tasks, particularly for low-resource languages like Nepali. Interested in applied AI research and developing scalable ML systems.

Education

B.E. in Computer Engineering Final Year
Khwopa College of Engineering, Libali, Bhaktapur

Technical Skills

- **Programming:** Python, C, C++
- **Machine Learning:** Deep Learning, NLP, Computer Vision, Transfer Learning, Fine-tuning
- **Frameworks:** PyTorch, Hugging Face Transformers, Scikit-learn
- **Data Tools:** NumPy, Pandas
- **Deployment:** FastAPI, Streamlit, Docker
- **AI Systems:** Retrieval Augmented Generation (RAG), LangChain, Vector Databases
- **Tools:** Git, GitHub, Weights & Biases, SQL, redis, MongoDB

Projects

- **Portfolio RAG Chatbot (Deployed)**
  Developed and deployed a Retrieval-Augmented Generation (RAG) chatbot using FastAPI, LangChain, and Hugging Face LLMs for portfolio-based Q&A. Built an end-to-end pipeline for document ingestion, embedding, and semantic retrieval using Pinecone and MongoDB. Integrated Redis for context-aware conversations and deployed the API on Render.
  Tech: Python, FastAPI, LangChain, Pinecone, MongoDB, Redis, Hugging Face
- **Nepali Grammar Error Correction using Transformer Models**
  Fine-tuned a multilingual T5 model for Nepali grammatical error correction using a synthetic parallel corpus and additional LLM-generated corrections. Fine-tuned a BERT as a reranker which further improved the score. Conducted experiments to improve model performance and deployed the system via FastAPI.
  Tech: PyTorch, Hugging Face Transformers, FastAPI
- **Nepali Image Captioning with Attention Mechanism**
  Built an encoder–decoder image captioning model using EfficientNet-B0 as encoder and LSTM with attention as decoder. Trained on a Nepali caption dataset derived from Flickr30k.
  Tech: PyTorch, Attention Mechanism, Computer Vision
- **RAG-based YouTube Question Answering System**
  Developed a Retrieval Augmented Generation system allowing users to query information from specific YouTube videos. Implemented document retrieval using vector embeddings and generation using LLM APIs.
  Tech: LangChain, ChromaDB, Vector Databases, LLM APIs
- **Dialogue Summarization using FLAN-T5**
  Fine-tuned a FLAN-T5 transformer model on the SAMSum dataset for dialogue summarization. Tracked experiments using Weights & Biases and deployed an interactive demo using Streamlit in a Docker container.
  Tech: PyTorch, Hugging Face, Docker, Streamlit

Certification

Data Science Training – DeepMind Creation Pvt. Ltd

Languages

English, Nepali, Newari, Hindi

Interests

Artificial Intelligence Research, Machine Learning Systems, Low-Resource NLP