BISHNU K.C.

AI/ML Engineer | Generative AI • LLM Applications • RAG • NLP • PyTorch • MLOps

Kathmandu, Nepal | bishnu.kc7864@gmail.com | +977 9808893810 | linkedin.com/in/bishnu-kc-ai | github.com/Bcnukc

Professional Summary

AI/ML Engineer with a background in software engineering and applied machine learning, focused on large language model applications, retrieval-augmented generation, natural language processing, and low-resource language technologies. Experienced in end-to-end AI/ML development including data preparation, model training and fine-tuning, evaluation, API development, and deployment using Python, PyTorch, Go, and TypeScript. Currently completing a Master in Information Technology at Tribhuvan University, with research focused on parameter-efficient fine-tuning for low-resource Nepali speech recognition.

Technical Skills

Programming: Python, SQL, Go, JavaScript/TypeScript

Machine Learning: PyTorch, scikit-learn, Hugging Face Transformers, supervised learning, model evaluation, feature engineering, data preprocessing

Generative AI & NLP: Large Language Models, RAG, embeddings, semantic search, vector retrieval, prompt engineering, LLM evaluation, hallucination mitigation

Speech & Low-Resource NLP: Whisper, LoRA/PEFT, multilingual transformers, Nepali/Devanagari NLP, speech corpus preparation, ASR evaluation

Backend & Data: Go/Echo, Next.js, REST APIs, PostgreSQL, Python microservices, ETL pipelines

MLOps & Tooling: Docker, Git, GitHub, CI/CD fundamentals, Linux, experiment tracking, model versioning, FFmpeg, LaTeX

Professional Experience

AI/ML Engineer – Appware Technology Aug 2026 – Present

Remote • Full-time

- Develop AI/ML application features involving large language models, retrieval-augmented generation, and agentic workflows using Python.
- Build document-processing and retrieval pipelines covering text chunking, embeddings, semantic search, context retrieval, and grounded LLM responses.
- Integrate AI-powered application components through REST APIs and collaborate with software engineers on application integration.
- Evaluate LLM outputs for relevance, grounding, consistency, and response quality across application-specific use cases.

Software Engineer – KOSMOS Technology Limited 2023 – 2024

Kathmandu, Nepal

- Developed and maintained backend services and web application features across database, API, and client-facing application layers.
- Designed relational database schemas and wrote SQL queries supporting application features and reporting requirements.
- Built and documented REST APIs for internal application modules and external service integrations.
- Participated in debugging, code review, testing, and application deployment within a collaborative software engineering environment.

Software Engineer – SSB Technologies Pvt. Ltd. 2022 – 2023

Kathmandu, Nepal

- Developed application modules from requirements through implementation, testing, and deployment.
- Integrated third-party APIs and external services into existing software systems.
- Created Python-based automation scripts for repetitive data-processing and reporting workflows.

- Diagnosed and resolved application defects across backend and customer-facing functionality.

Research & Selected Projects

Low-Resource Nepali Automatic Speech Recognition – Master's Research 2026 – Present

- Investigating parameter-efficient fine-tuning using LoRA compared with full fine-tuning of Whisper for low-resource Nepali automatic speech recognition.
- Studying accuracy, computational requirements, trainable parameter count, and resource-efficiency trade-offs between fine-tuning strategies.
- Designing a spontaneous-speech evaluation benchmark for Nepali, including data collection, transcription protocols, dataset preparation, and Word Error Rate (WER) based evaluation.

Ghokante – Study Resource Platform for Nepali University Students 2025 – Present

- Developing a full-stack educational platform for organising course materials and learning resources for university students in Nepal.
- Designed an architecture using Next.js, Go/Echo, PostgreSQL, and Python-based machine-learning services for intelligent content processing and retrieval.

Nepali Sentiment Analysis with Multilingual Transformers 2025

- Evaluated multilingual transformer architectures including mBERT, XLM-RoBERTa, MuRIL, and NepBERTa for Nepali sentiment classification.
- Analysed tokenisation, model performance, and cross-lingual transfer behaviour for a low-resource Devanagari-language classification task.
- Documented experimental methodology, evaluation results, and model error analysis in research-paper format.

NANU BABU TV – Automated Nepali Media Production Pipeline 2024 – Present

- Developed Python and FFmpeg automation workflows for producing Nepali educational media, including video rendering, caption processing, audio integration, and post-processing.
- Implemented Devanagari text rendering using Pillow and HarfBuzz/RAQM-compatible text shaping for correctly rendered Nepali captions.
- Integrated generative media tools into a repeatable content production workflow to reduce manual editing requirements.

Education

Master in Information Technology (MIT) 2024 – 2026 (Expected)

Amrit Campus, Tribhuvan University • Kathmandu, Nepal

- Research focus: *Parameter-Efficient Fine-Tuning for Low-Resource Nepali Automatic Speech Recognition.*
- Relevant study areas: Machine Learning, Artificial Intelligence, Database Systems, Distributed Systems, Cybersecurity, and Research Methodology.

Bachelor of Computer Information System (BCIS) 2017 – 2022

Apex College, Pokhara University • Kathmandu, Nepal

Additional Information

Languages: Nepali (Native), English (Professional Working Proficiency), Hindi (Conversational)

Research Interests: Low-resource language technology, LLM reliability, retrieval-augmented generation, speech recognition, multilingual NLP, educational technology