

Avishek Poudel

Aspiring Graduate Student in Machine Learning and AI

Phone: (+977) 984-592-3008 | Email: avishek.poudel.7777@gmail.com

SUMMARY

Recent graduate in Electronics, Communication, and Information Engineering with a strong focus on Machine Learning and Artificial Intelligence. Strong background in developing deep learning models, NLP pipelines, and AI-related applications using Python, Tensorflow, Scikit-learn, and Pytorch. Demonstrated sound leadership in projects like depth estimation pipelines and accident alert systems, as well as experience presenting projects at international conferences. Skilled in web development, hardware integration (Arduino, Raspberry Pi), and 3D modeling, supported by good communication abilities. Strong interest in applying AI and data-driven solutions to real-world problems and achieving impactful outcomes.

KEY SKILLS

- Machine Learning & AI: Strong theoretical and practical foundation; experience in model development and data analysis using Python (NumPy, Pandas, Scikit-learn), TensorFlow, Pytorch, etc.
 - Programming language: C, C++, Python, JavaScript.
 - Web Development: Skilled in HTML, CSS, Django.
 - Design Software & Tools: Adobe Photoshop, Adobe Illustrator, Canva.
 - Hardware & IOT: Arduino Uno, Raspberry Pi, GSM module, Accelerometer sensor, GPS module.
 - 3D Modeling & Simulation Software: Blender, AutoCAD.
 - Project Management & Process Development: Successfully led the team and managed ML projects from concept to completion, focusing on data handling, model building, training & evaluation, and optimization & tuning.
 - Microsoft Office: MS Word, MS Excel, MS PowerPoint, etc.
 - Soft Skills: Communication, Public speaking, Slide Presentation skills.
-

ACADEMIC PROJECTS

1. AI-Driven Monocular Depth & Texture Estimation System

- Led a team of 4 in designing a deep-learning-based system for depth estimation & texture extraction from RGB images.
- Implemented the project using Raspberry Pi and Camera module (Hardware) & Pytorch, Blender (Software).
- Developed our own custom datasets (synthetic, Blender-generated, experimental) due to the lack of complete PBR map datasets.
- Built a web interface to upload images, generate PBR maps, and successfully download a 3D model.

2. Music Recommendation App

- Built a Python-based application using Streamlit UI to recommend songs based on lyrics using TF-IDF vectorization and cosine similarity.
- Applied NLP techniques to analyze song lyrics and generate personalized recommendations.
- Utilized Pandas and Scikit-learn for data processing, analysis, and similarity computation.

3. Accident Alert System

- Guided a team in designing and building a hardware-based accident detection and alert system.
- Utilized Arduino Uno, GSM module, GPS module, accelerometer sensor to detect the accident, identify the location of the accident, and send an alert SMS automatically to a particular phone number.

4. Multi-Agent Research Assistant (AutoGen)

- Designed and implemented multi-agent LLM system with the help of AutoGen, in which specialized agents collaboratively perform literature exploration, summarization, etc.
- Created agent communication workflows and role-based reasoning pipelines to automate research tasks using LLM APIs.

5. Medical Intelligence Platform (Generative AI and RAG Pipeline)
 - Implemented the Retrieval-Augmented Generation (RAG) pipeline that enables LLMs to provide medical information from private PDF datasets.
 - Implemented neural semantic search using Pinecore vector database and HuggingFace to map medical terms into high-dimensional space.
 - Developed a full-stack interface using Flask and AWS to handle real-time queries from patients and utilized LangChain for document chunking and memory management.
 6. End-to-End Implementation, Training, and Quantization of a GPT-Based Small Language Model
 - Developed a 6-layer Transformer architecture from scratch in PyTorch by implementing self-attention and multi-head attention.
 - Optimized the training throughput using GPT-2 BPE tokenizer and memory-mapped binary storage for better data retrieval.
 - Improved convergence of the model using mixed-precision (FP16), gradient accumulation, and learning rate scheduling.
 - Implemented Saliency-Aware Quantization from 2-bit to 4-bit on the basis of weight sensitivity analysis to reduce the size of the model.
 7. Chronos-T5: Utilizing Large Language Model Architectures for Time Series Forecasting
 - Developed a time series forecasting system by using Amazon Chronos-T5 tiny model and achieved a Mean Absolute Percentage Error (MAPE) of 6.53% on the Airline Passengers dataset.
 - Implemented Rotary Positional Embeddings (RoPE) instead of absolute positional embeddings and created a custom tokenizer with quantization for better performance.
 - Created the complete end-to-end pipeline in Python and evaluated the results using MAE, MAPE, and RMSE metrics.
 - Compared the results of the base Chronos model against a custom fine-tuned implementation, which resulted in overfitting as the dataset used was small, and the project validated the effectiveness of foundation models for time-series tasks.
-

EDUCATION

Bachelor of Science in Electronics, Communication, and Information Engineering, Tribhuvan University, Nepal

Graduated: August 2025 | GPA: 3.6

Higher Secondary Education

Valmiki College, Nepal

Graduated: March 2021 | GPA: 3.38

Secondary Education Examination

Valmiki Shiksha Sadan Secondary School, Nepal

Graduated: July 2018 | GPA: 3.75

AWARDS AND ACHIEVEMENTS

- Best Performance Award, Fifth Young Scientists Summit (2020). Project: Testing Internet Applications – Terminology and Applicability in Rural Areas.
 - Best Bot Design, Battle for Speed, Robotics Club, TU, 2022.
 - Merit Scholarship in Bachelor of Science in Electronics, Communication, and Information Engineering.
-

ACADEMIC EXPERIENCE

- Team Leader, design and implementation of deep learning-based monocular depth estimation and texture extraction pipeline
- Team Leader, design and implementation of Accident Alert System integrating Arduino Uno, GPS module, GSM module, and accelerometer sensors.
- Oral Presenter, Gandaki University International Conference, organized by Gandaki University, January 3-5, 2025.
- Organizer, CODE WITH COFFEE 2023, organized by the Innovative Computer Engineering Student Society, TU.
- Organizing committee member, Battle for Speed 2023, organized by Robotics Club, IOE, TU.