

# Birat Poudel

Phone: +977 9861281912 | Email: [poudel.birat25@gmail.com](mailto:poudel.birat25@gmail.com)

Research Portfolio: <https://biratpoudel.com.np>

## Education

Tribhuvan University, Institute of Engineering, Thapathali Campus

Thapathali, Kathmandu

Bachelor of Electronics, Communication and Information Engineering

November 2019 – April 2024

Percentage: First Division

**Relevant Courses:** Artificial Intelligence, Probability & Statistics, Mathematics, Algorithms, Discrete Structures and Big Data

## Research Papers

### 1. Fine-Tuning DialoGPT on Common Diseases in Rural Nepal for Medical Conversations

- Fine-tuned Microsoft DialoGPT-medium on a synthetically generated dataset of 1,000 doctor–patient dialogues covering ten common diseases prevalent in rural Nepal to create an offline-capable medical conversational AI system.
- Designed a two-stage data generation and validation pipeline using Gemini 2.5 Pro and Claude 4 Sonnet, followed by expert medical review, ensuring accuracy, empathy, and contextual relevance of dialogues.
- Conducted quantitative (perplexity, loss, token-level F1) and qualitative evaluations by healthcare professionals, demonstrating the model's ability to produce coherent, medically appropriate, and empathetic responses for low-resource healthcare settings.

Status: Completed - Manuscript under review ([View Draft](#))

### 2. Nepali Sign Language Characters Recognition: Dataset Development and Deep Learning Approaches

- Developed the benchmark dataset for Nepali Sign Language (NSL) with 36 gesture classes and 1,500 samples per class.
- Fine-tuned MobileNetV2 and ResNet50 architectures achieving classification accuracies of 90.45% and 88.78% respectively.
- Demonstrated effectiveness of transfer learning and fine-tuning for underexplored sign languages in low-resource settings.
- Contributed to accessibility research by creating systematic dataset and deep learning approaches for NSL recognition.

Status: Completed - Manuscript under review ([View Draft](#))

## Research Projects

Automobile License Plate Detection and Recognition | OpenCV, Convolutional Neural Network (CNN), Inception-ResNet-v2, YOLOv8, Google Tesseract and Flask

GitHub Link: <https://github.com/Birat-Poudel/Automobile-License-Plate-Detection-and-Recognition>

- Developed a comprehensive two-stage system: initial implementation with Inception-ResNet-v2, later enhanced with YOLOv8.

- Integrated state-of-the-art object detection with Tesseract OCR for robust license plate localization and text extraction.
- Achieved robust performance across diverse environmental conditions including varying lighting, angles, and image quality.
- Demonstrated practical applicability for real-time traffic monitoring and automated vehicle identification systems.

## Time Series Forecasting with Nepal Stock Exchange (NEPSE) Dataset | ARIMA, SARIMA, LSTM, Prophet

Kaggle Link: <https://www.kaggle.com/code/biratpoudelrocks/time-series-forecasting-nepse-dataset>

- Performed time-series forecasting on NEPSE (Nepal Stock Exchange) data, leveraging statistical and machine learning models to predict market trends and evaluate predictive performance.
- Cleaned, processed, and transformed raw financial data (e.g. handling missing values, normalization, feature engineering) to create robust input features for forecasting models.
- Compared and validated multiple forecasting approaches (e.g. ARIMA, SARIMA, LSTM, Prophet, etc.), evaluated accuracy using metrics (RMSE, MAE, etc.), and provided insights into optimal model selection for stock index prediction.

## Amazon Bedrock Foundational Models Evaluation Pipeline | LLM-as-a-Judge, Quality Metrics, Performance Metrics, Responsible AI Metrics

GitHub Link: <https://github.com/Birat-Poudel/AWS-Bedrock-Models-Evaluation>

- Implemented an end-to-end evaluation pipeline for Amazon Bedrock models (multi-region support) with performance (latency, throughput, time-to-first-token) and quality metrics (helpfulness, faithfulness, completeness, coherence, etc.).
- Used LangChain orchestration and an LLM-as-a-judge approach for automated, multi-dimensional quality and responsible-AI assessments (harmfulness, bias, refusal appropriateness), plus async processing for concurrent benchmarking.
- Produced structured evaluation outputs and human-readable reports to compare models, tune prompts, and inform safe deployment decisions.

## Deep Research Agent using Amazon Strands Agents | Agent Loop, Agent As Tools, Web Search and Extract Agent

GitHub Link: <https://github.com/Birat-Poudel/Deep-Research-Agent>

- Engineered an automated research agent pipeline that uses Amazon Strands Agents for domain-specific data gathering, filtering, and summarization, reducing manual research time by a significant margin.
- Implemented a multi-agent architecture combining "Agent Loop," "Agent-as-Tool," "Web Search," and "Extraction" modules to autonomously perform requirements analysis, competitive landscape research, and effort estimation.
- Incorporated iterative feedback loops and inter-agent coordination to refine project scoping, synthesize comparative analyses, and output actionable estimations and insights.

## Technical Skills

---

<b>Programming</b>	C/C++ · Java · Python · JavaScript · TypeScript
<b>Python Libraries</b>	NumPy · Pandas · Scikit-Learn · Matplotlib · Seaborn
<b>ML Frameworks</b>	Tensorflow · PyTorch
<b>Development</b>	Flask · Django · FastAPI · HTML/CSS/JS · ReactJS
<b>Database</b>	MySQL · PostgreSQL · MongoDB · Redis · Vector Databases (Pinecone, Qdrant)
<b>Evaluation</b>	Statistical Analysis · A/B Testing · Performance Metrics · LLM As A Judge

## Industry Experience

---

### Leapfrog Technology, Inc.

Kathmandu, Bagmati, Nepal

AI/ML Research Engineer

June 2024 - Present

- Built AI systems, including **IVR (Interactive Voice Response)** and **Conversational Voice AI** for medical patient follow-ups and referrals.
- Designed a **Foundational Models Evaluation Pipeline** that generates automated PDF reports for large-scale model benchmarking.
- Engineered and deployed **AI Agents and MCP Servers** to enable scalable, modular, and autonomous agent orchestration across workflows.

### Jobsflow.ai

Kathmandu, Bagmati, Nepal

Machine Learning Research Engineer

6 months | December 2024 - May 2024

- Built AI systems, including an **AI Voice Interviewer** and an intelligent **Chatbot** capable of making tool calls to over ten plus services like Google Calendar, Meet, Gmail, Zoom, etc.
- Developed algorithms for calculating **match score** of a particular applicant for a job based on job descriptions and applicant's resume and answers for the job related questions.
- Implemented **contextual searching, filtering and sorting** using embeddings to enhance candidate selection accuracy.

### Fusemachines

Kathmandu, Bagmati, Nepal

Machine Learning Research Engineer

3 months | September 2024 - November 2024

- Preprocessed and transformed datasets using **NumPy** and **Pandas**, applying advanced **feature engineering** techniques for time series forecasting and machine learning applications.
- Designed and implemented ML models, including **SARIMA**, **LSTM**, **Prophet**, and **XGBoost**, for time series forecasting and predictive analytics, achieving a 15% improvement over previous models.
- Enhanced **RAG-based** systems by optimizing vector storage and retrieval.

### Maven Solutions Pvt. Ltd.

Kathmandu, Bagmati, Nepal

Machine Learning Research Engineer

2 years | August 2022 - August 2024

- Worked on data preprocessing and feature engineering using libraries like **Numpy** and **Pandas** to prepare datasets for model training.
- Developed and implemented machine learning algorithms using libraries such as **Scikit-Learn** for tasks like classification, regression, and clustering achieving an accuracy improvement of 15% over previous models.
- Employed advanced **automation scripts** and conducted precise **web scraping** operations to streamline workflows and gather mission-critical data efficiently.
- Orchestrated the development and seamless integration of **backend APIs**, collaborating closely with cross-functional teams to enhance application functionality and performance.

## Certifications

---

### Machine Learning Specialization (DeepLearning.AI)

- Supervised Machine Learning: Regression and Classification, Advanced Learning Algorithms, Unsupervised Learning, Recommenders, Reinforcement Learning
- Certificate Link: <https://coursera.org/share/a4e925ef3acd5ea867b65ece719eedc5>

### Computer Networks and Network Security

- IP Addressing, Routing and Switching, Network Protocols, Network Security Techniques
- Certificate Link: <https://coursera.org/share/d7cb8cd9e101fc7863be3a34a8c0b749>