

# Er.Rebanta Aryal

Kathmandu, Nepal | [linkedin.com/in/rebanta-aryal/](https://www.linkedin.com/in/rebanta-aryal/) | [aryal.rebanta@gmail.com](mailto:aryal.rebanta@gmail.com) | +977- 9841723105 | [github.com/reban87](https://github.com/reban87)

## PROFESSIONAL SUMMARY

Experienced Machine Learning Scientist with a strong background in Generative AI and Data Science, focused on advancing research in LLMs and Generative AI. Skilled in driving innovation through scalable AI solutions and committed to contributing to cutting-edge advancements in recommendation systems to support industry needs.

## SKILLS HIGHLIGHT

- **Programming Languages:** Python, C++, C
- **Machine Learning and AI:** Tensorflow, PyTorch, Keras, Scikit-learn, Hugging Face, Transformers, Prompt Engineering
- **Natural Language Processing:** BERT, GPT, DistilBERT, Sentence Transformers, RASA,
- **Large Language Models:** Fine-tuning, Prompt Engineering, Few shot Learning, GPT, Mistral Embed, gemini, Grok, DeepSeek, Ollama, Groq, HuggingFace
- **Recommender Systems:** Collaborative Filtering, Content Based Filtering, Hybrid Approaches, Graph Neural Network (GNN)
- **Retrieval-Augmented Generation (RAG):** Basic RAG Implementation, Multi-vector Retrieval, Rerank-then-Read, Agentic RAG ,
- **Vector Database:** ChromaDB, Pinecone, Milvus DB
- **Knowledge Graphs:** Neo4j, Nebula DB, Networkx, knowledge graph construction and querying
- **Web Technologies:** FastAPI, Flask, Streamlit
- **Data analysis and visualization:** Numpy, Matplotlib, Seaborn, Pandas
- **Database:** SQL, MongoDB
- **Version control:** GitHub, GitLab
- **Framework and Tools:** Git, crewAI, LangChain, LangSmith, LangServe, LangGraph Docker, Docker Compose, Docanno, AWS, MCP
- **RPA Tools:** OpenRPA, UIPath
- **Languages:** English (Fluent), Nepali (Native)

## WORK EXPERIENCE

<b>AI Consultant</b> Amoeba Labs	<b>April - Present</b> Lalitpur, Nepal
-------------------------------------	---

- Leading architectural guidance on agentic AI systems, specializing in multi-agent orchestration and retrieval-augmented generation (RAG) pipelines.
- Designed and deployed chatbots leveraging website content and APIs to deliver contextual, domain-specific interactions.
- Developed agentic RAG applications for EdTech platforms, integrating OCR engines, graph-based retrieval, and custom retrieval algorithms.
- Contributed to the integration of MCP (Model Control Protocol) features into the development dashboard, enhancing developer control over model behaviors and monitoring

**Sr. AI Engineer | Team Lead**  
TechKraft Inc. Pvt. Ltd

**June 2024 - April 2025**  
Lalitpur, Nepal

### **Developed HR Bot with LLM for Recruitment Automation**

Created an AI-driven HR bot capable of handling recruitment tasks by integrating generative AI and automation features.

- Integrated Zoho recruitment for cv collection.
- Candidate Sorting: Automated candidate sorting based on job descriptions using NLP techniques.
- Candidate Filtering: Implemented intelligent filtering of candidates based on user queries for refined results.
- Resume Skills/Experience Extraction: Developed a system to extract specific skills or experience from a pool of resumes.
- General Query Handling: Enabled efficient response to general queries, enhancing user experience.
- Resume Delivery: Provided candidate CV details and attached full resumes in responses.

### **No-Code AI Chatbot Development**

- Explored Flowise, a no code framework for building drag-and-drop AI chatbots
- Researched and developed a web application that allows users to train and deploy chatbots within minutes

### **Graph-based AI System for Healthcare system**

- Scraped journal and author data from PubMed using BioPython for entity extraction.
- Designed and implemented a Neo4j knowledge graph to store and structure extracted data.
- Applied GraphRAG techniques and developed a chat interface to generate researcher profiles dynamically.

### **Virtual Try-On Research**

- Explored AI-driven virtual try-on models including Diffusion Models, CatVTON, and StableVITON for upper and lower body applications.

### **Led Research & Development on Retrieval-Augmented Generation (RAG) and Agentic RAG Systems**

- Leading research on developing an advanced RAG (Retrieval-Augmented Generation) agent for intelligent query processing and response generation by integrating cutting edge technologies such as LangChain and DSPy for LLM management, alongside vector database ChromaDB and Pinecone.
- Designing and implementing a system that stores user queries in a database, extracts specifications, and generates proposal requests using an AI agent.
- Conducting research on integrating RAG techniques with Knowledge Graph to enhance information retrieval and context understanding.

### **AI Education & Training**

- Actively contributed to designing the AI Bootcamp syllabus for TechKraft, focusing on LLMs, NLP, and AI-powered applications.

### **Lecturer (Part Time)**

MSc - Data Science and Computational Intelligence  
Softwarica College of IT and E-commerce

**August 2024**

**Instructed Advanced Machine Learning:** Delivered comprehensive lectures and hands-on practical sessions covering:

- Gaussian Processes & Dirichlet Processes: Explored the fundamentals of probabilistic modeling and Bayesian inference.
- Graphical Models: Taught techniques for structured probabilistic models, including Bayesian networks and Markov random fields.
- Fuzzy Sets and Systems: Explained the principles of fuzzy logic, fuzzy sets, and their application in uncertainty modeling.

- Adaptive and Hybrid Fuzzy Systems: Discussed the design and development of intelligent systems combining fuzzy logic with other machine learning techniques.
- Evolutionary Algorithms: Provided insights into bio-inspired algorithms such as genetic algorithms and evolutionary strategies.

**NLP Engineer | Client Team Engineering Lead**  
Riphey AI

**February 2022 - June 2024**  
Louisville, Colorado, United States

### **Research and Development**

- Conducted extensive experiments with various Large Language Models (LLMs) including OpenAI's GPT-3.5, GPT-Instruct, and GPT-3.5 Turbo 1106, comparing their performance and suitability for document based classification and entity extraction tasks.
- Developed advanced prompt engineering techniques maintaining the version and constantly evaluating against the test dataset to significantly improve accuracy in email classification and named entity extraction (NER) tasks.
- Utilized LangChain framework to integrate LLMs for processing and extracting entities from unstructured emails and documents.
- Researched and implemented context-based chunking techniques for extracting data from multi-page tables in complex documents.
- Fine-tuned DistilBERT and BERT multilingual base models for text classification & NER, achieving 90% and 88% accuracy respectively.
- Researched and fine-tuned BERT multilingual base model with a custom dataset and achieved an accuracy of 88%, serving customers speaking 7 different languages and integrating with the chatbot.
- Created a multilingual chatbot that increases conversation efficiency by 50% using RASA Framework, Python, spaCy, and Transformers.
- Fine-tuned Sentence Transformer model (all-MiniLM-L6-v2) for FAQ Model and achieved an accuracy of 95% by using a customized FAQ dataset.

### **Project Management and Leadership**

- Regularly presented model accuracy metrics and project progress to key stakeholders, effectively communicating complex technical information to non-technical audiences
- Led client engagements, assessing technical feasibility of AI solutions with a focus on NLP and Generative AI.
- Managed sprint planning and execution, breaking down user stories into tasks.
- Collaborated with cross-functional teams to align project outcomes with client expectations and business objectives

**AI Consultant**  
Bitcraft Technology

**November 2022 - July 2024**  
Kathmandu, Nepal

- Provided expert consultation on building a conversational AI agent using the RASA framework, enhancing client's customer interaction capabilities.
- Designed and implemented a recommendation model using feature embeddings for online products, improving personalization and user engagement.
- Collaborated with client teams to understand business requirements and translate them into effective AI solutions.
- Offered guidance on best practices for AI implementation, including model selection, data preparation, and performance optimization.

**AI Practitioner**  
Inspiring Lab

**March 2020 - July 2021**  
Kathmandu, Nepal

- Researched and experimented with new & early stage algorithms such as **LSTM, RNN, & CNN**.

- Applied image processing techniques and deep learning models like VGG16 and RESNET to solve image classification problems.
- Presented findings and recommendations to the team on a regular basis.

### **Community Leadership**

Community Leader - AI/ML for All

- Organize AI/ML workshops, hackathons, and mentorship programs for aspiring AI enthusiasts in Nepal.
- Guide community members by providing resources, learning roadmaps, and hands-on projects in AI/ML.
- Collaborate with industry experts and academic institutions to foster AI/ML education.
- Represent the community in AI panels, conferences, and discussions on emerging AI trends.

### **Lecturer**

National College of Engineering, Tribhuvan University

Full Time

**April 2018- April 2020**

Lalitpur, Nepal

- Supervised major and minor projects related to machine learning and deep learning for graduate students of **Computer Engineering**.
- Organized **Bootcamp** on Github, Python, Machine Learning, and node js for more than 100 students under the IT club.
- Instructed students with lectures on Data mining, C-programming, Embedded System Design, and Digital logic and mentored them with practicals.
- Worked closely with the management team to plan, develop, coordinate and execute technical strategies and projects on information technology.

## **KEY PROJECTS**

---

### **Laptop Recommendation System**

A chatbot developed using the RASA Framework to provide users with seamless laptop recommendations. The dataset was gathered through web scraping from Newegg.com, and recommendations are based on price range, specifications, laptop types, and store availability.

### **Humor Generation Using an Images**

A joke generator using a large pre-trained language model (GPT2) using a caption generated from an image using resnet50. We attempted to synthesize two language generation tasks into a single system - image captioning and joke generation. The image caption acts as a context from which the joke is generated. Overall, the model accomplishes the task of generating a joke from the given image to varying degrees of success in funniness.

### **Recruitment Bot: An LLM based Bot for HR professionals for screening the collected CVs**

This project is a Retrieval-Augmented Generation (RAG) system designed for recruitment. It uses advanced natural language processing techniques to answer queries based on ingested documents, providing accurate and context-aware responses based upon resumes.

#### **Features:**

Document ingestion and processing (PDF support)

Vector storage using Pinecone

Query interpretation using OpenAI's language models

Interactive CLI interface

RESTful API using FastAPI and LangServe

Monitoring and feedback collection with LangSmith

### **Air Quality Analysis and Prediction using ML**

Implemented different ML algorithms like Random Forest Regressor, Linear Regressor, and Decision Tree Regressors on time series air quality down COVID time. The analysis was done before the crisis, during the crisis and after the crisis. The real time data was collected from [US Embassy AQI Stations](#).

### **Traffic Symbol Detection**

Implemented Convolution Neural Network (CNN) and Support Vector Machine (SVM) for classifying traffic symbols from the dataset. Achieved more than 90% accuracy on CNN.

## **EDUCATION**

---

### **Master's Degree in GeoInformatics**

**2024**

Kathmandu University

Kavre, Nepal

Overall Grade: 3.91

- Areas of Interest: Natural Language Processing, Generative AI, Large Language Models (LLMs)
- Majors: Deep Learning, Geo data mining, Geo-statistics, Natural Language Processing
- Master Thesis: “[Disaster Analysis with Geo - Spatial Mapping and Visualization of Social Media Data Using Natural Language Processing](#)”
  - Supervisors:
    - Dr. Diana Contreras Mojica, Lecturer (Assistant Professor) in Geospatial Science, Cardiff University
    - Dr. Arun Kumar Pratihast, Senior Data Scientist, Earth Observation and Environmental Informatics Wageningen University and Research
  - Developed a RoBERTa based model for disaster classification from social media data.
  - Implemented a BERT based entity recognition (NER) model to extract key information such as location, disaster type and date from the text.
  - Conducted extensive experiments with various transformer models to optimize the performance
  - Integrated NLP results with geospatial analysis for real time visualization and mapping of disaster related information.

### **Bachelor's Degree in Electronics & Communication Engineering**

**2016**

Kathmandu Engineering College

Kathmandu Nepal

Overall Grade: 75.74%

### **High School**

**2012**

St. Xavier's College

Kathmandu Nepal

Overall Grade: 79.2%

## **PUBLICATIONS**

---

- [A Fire Hazard Assessment Using Sentinel Imagery; A Case Study Over Gippsland Australia](#) **July 2020**
- [Smoke Scene Detection From Satellite Imagery Using Deep Learning. Journal of Land Management and Geomatics Education, Volume \(IV\), Page Numbers : 22](#) **June 2022**
- [Comparative Analysis of Different Methods of Spatial Interpolation: Interpolating Annual Precipitation On the Koshi Basin, Journal of Land Management and Geomatics Education, Volume\(III\), Page Number: 41](#) **June 2021**

## **PERSONAL TRAITS**

---

- Adept self learning skills
- An amicable team worker and with a hard working ability

- Comfortable working in Windows, Linux or similar environment
- Good written and verbal communication skills both in English and Nepali

## REFERENCES

---

1. Santosh Koirala  
Executive Director  
Techkraft Inc  
santosh@techkraftinc.com
2. Dr Diana Contreras Mojica  
BSc, MSc, PhD FHEA  
Lecturer in Geospatial Sciences  
School of Earth and Environmental Sciences  
Cardiff University  
ContrerasMojicaD@cardiff.ac.uk  
+44 29208 74333  
Relationship: Master's Thesis Supervisor
3. Dr. Arun Kumar Pratihast  
Senior Data Scientist Lecturer  
Earth Observation and Environmental Geospatial Science  
Wageningen University and Research  
arun.pratihast@wur.nl  
Relationship: Master's Thesis Supervisor
4. Mr. Sujeet Rimal  
Infrastructure Manager  
Rippey AI  
1025 Cannon St Suite 2B,  
Louisville CO 80027  
sujeet@rippey.ai  
Relationship: Line Manager