DIBIKA DAHAL

AI & ML ENGINEERING STUDENT

dibikadahal@gmail.com | +977-9860795577 | Kathmandu, Nepal

Dibika Dahal dibikadahal dibikadahal

SUMMARY

AI and ML engineering student focused on building reliable LLM systems, with hands-on experience in Retrieval-Augmented Generation (RAG) from an industry internship and independent projects. Experienced in designing hybrid retrieval pipelines that combine SQL and vector search, running LLMs fully on local hardware, and validating model outputs for numerical accuracy. Currently exploring model quantization and edge AI, with the goal of making capable models efficient enough to run on everyday, resource-constrained devices.

EDUCATION

BSc (Hons) Computing with Artificial Intelligence **2025 - Present**

Islington College
London Metropolitian University
Kamal Marg, Kamalpokhari, Kathmandu

SLC, NEB Board **2022 - 2024**

St. Mary's Secondary School
Jawalakhel, Lalitpur

SEE, NEB Board **2022**

St. Mary's Secondary School
Jawalakhel, Lalitpur

PROFESSIONAL EXPERIENCE

AI Intern **June 2026 - Sep 2026**

Pathway Technologies, Thapathali

- Designed RAG pipelines that route factual questions to direct database lookups and explanatory questions to semantic search.
- Built a framework-free RAG system from scratch using locally hosted models and statistical datasets, without relying on external libraries or APIs.
- Contributed to a cloud-based RAG system that answers questions over an agriculture document using an LLM through an API.
- Created an automated check to verify figures across district, province and national levels in a government report with dozens of tables, and used the findings to improve the accuracy of the company's RAG chatbot.

TECHNICAL SKILLS

Languages : Python, Java, SQL
LLMs and NLP: RAG, Vector Embeddings, Semantic Search, Hybrid Retrieval, Prompt Engineering, LangChain, ChromaDB, Ollama, Docling, Hugging Face
Data : NumPy, Pandas, Matplotlib, Seaborn; data cleaning, manipulation, visualization
Tools: Git/GitHub, Docker, AWS, Linux, Streamlit, FastAPI
Foundations: Data Structures & Algorithms, Linear Algebra, Calculus, Statistics & Probability
Currently Exploring: Model Quantization (PTQ, QAT, GGUF, GPTQ, AWQ), Edge AI

SOFT SKILLS

Team Collaboration • Problem Solving • Adaptability • Time Management • Self-Directed Learning • Communication

COMPLETED PROJECTS

Local RAG for Nepali Agricultural Statistics

Python, Qwen2.5:7B, Ollama, Docling, SQL, Vector Search, Streamlit

- Built a fully local RAG chatbot over Nepal's government agriculture statistics, running Qwen2.5:7B on Ollama with no cloud dependency.
- Designed a hybrid retrieval pipeline using exact SQL retrieval for numerical queries and semantic search for explanatory queries, with automatic routing between the two.
- Added source-page citations and refusal behaviour when an answer cannot be verified, prioritising trustworthy output over broad coverage.

Agriculture Assistant (A Cloud-based RAG System)

Python, Langchain, Google Gemini, ChromaDB, Streamlit

- Built a full-stack RAG application that answers questions on farming practices and agricultural statistics, grounding responses in source documents with page-level citations to eliminate hallucination.
- Engineered a vision-based extraction pipeline using Gemini's multimodal capabilities to accurately parse complex tables and bar charts from PDFs, unifying structured and unstructured data into a single ChromaDB knowledge base with cosine similarity search.
- Solved key data-representation challenges including row-linearization for table chunking and embedding-model consistency, then deployed the system via an interactive Streamlit interface.

Semantic Book Recommender

Python, LangChain, ChromaDB, Sentence Transformers, Hugging Face, Gradio

- Built a meaning-based book recommender using vector search, zero-shot genre classification (BART-MNLI) and emotion scoring (DistilRoBERTa).
- Resolved a CUDA compatibility issue on a Blackwell GPU by moving to PyTorch nightly builds with CUDA 12.8 support.

IOT-based Smart Irrigation System

Arduino, ESP32, Blynk, C++

- Built an IoT-based smart irrigation system using ESP32 and multiple sensors (soil moisture, ultrasonic, DHT11) to automate watering and monitor farm conditions in real time.
- Streamed sensor data to the Blynk cloud platform for remote monitoring, with threshold-based alerts for irrigation and water levels.

LEADERSHIP AND VOLUNTEERING

Tech Lead, Islington AIRO Community **Aug 2026 - Present**

- Core organising team for MITRA: Build the Agents of Tomorrow, a two-day agentic AI building competition with 60 to 70 participants; handled planning and hosted the event.

Volunteer, Robinhood Army Nepal **Jun 2024 - Present**

- Initiated and led an emergency donation drive for families affected by the Rasuwa Bhotekoshi floods.

CERTIFICATIONS

AWS

- AWS Academy Cloud Foundations
- AWS Academy Data Engineering
- AWS Academy Machine Learning for Natural Language Processing
- AWS Academy Machine Learning Foundations

OpenAI

- AI Foundations

WRITING

- LLM Quantization Explained: How do we Make Large Language Models Smaller and more Efficient? (Medium, 2026)
- 37 Days, One Stubborn PDF, and the AI Project That Almost Broke Me (Medium, 2026), on building a RAG system for complex tables

ADDITIONAL INFORMATION

- **Languages:** Nepali (native), English (Professional Working Proficiency)
- **Achievements:** 1st Place in build the Mini-Internet. Built a game portal using vanilla HTML/CSS/JS.

REFERENCES

Sunil Kumar Gharti
+977 9851067889
sunil@kavya.edu.np