Nirajan Paudel

paudelnirajan | Website | nirajan.paudel@colorado.edu | +1 (720) 595 8207

Education

University of Colorado Boulder 2025–2027
Master's of Science in Computer Science
Relevant Courses: Data Center Scale Computing, Object Oriented Analysis & Design
Institute of Engineering, Tribhuwan University 2020–2024
Bachelor of Engineering–Electronics, Communication & Information Engineering

Work Experience

Machine Learning Intern March 2025 – May 2025
PalmMind Technology

- Built scalable RAG chatbots for EV dealers and insurance clients using OpenAI models, LangChain, and LangGraph, leveraging orchestrator-worker workflows, routing, and parallelization for efficiency.
- Enhanced data retrieval and system performance by implementing web scraping, OCR (Tesseract), and Redis-based caching for faster, context-aware conversational AI.

Teaching Assistant Sep 2024 – Feb 2025
Pashchimanchal Campus, Tribhuwan University

- Taught C programming and Information Systems (theoritical concepts on cloud computing, distributed systems, neural networks, data mining).
- Designed lab activities, guided semester-end projects, and provided detailed feedback to enhance students' practical programming and problem-solving skills.

Projects

AI-Powered Question Assistant (IOE-GPT)
Personal Project

- Developed an AI-powered system for IOE students, leveraging LangChain and Groq LLM to retrieve and answer past programming questions with semantic and metadata-based search.
- Built scalable API endpoints using FastAPI and stored question embeddings in Milvus vector database, with Redis for conversational state ma nagement.
- Utilized sentence-transformer embeddings ('all-MiniLM-L6-v2') for efficient question retrieval and Docker for containerized deployment.

Voice RAG Assistant
Personal Project

- Developed a web application using FastAPI to enable intelligent question-answering from audio content, integrating OpenAI's Whisper for transcription and Groq's LLaMA-3.1-8B for real-time responses.
- Implemented Retrieval-Augmented Generation (RAG) with LangChain and HuggingFace sentence-transformer embeddings, optimizing text chunking and semantic search for efficient context retrieval.

digitalME–Personal-Chatbot
Personal Project

- Built an AI-driven conversational chatbot using Groq's LLM and LangChain, integrating Hugging Face sentence-transformer embeddings for semantic understanding and natural language processing.
- Implemented Pinecone vector database for efficient storage and retrieval of document embeddings, enabling scalable context-aware responses with file and SQL querying capabilities.

Nepali Image Captioning
Generating Coherent Paragraph-Length Descriptions Using Transformer
Undergraduate Major Project

- Developed a deep learning system for generating paragraph-length Nepali captions using Transformer architecture with Inception V3 feature extraction, trained on a curated dataset (20,350 pairs from Stanford Paragraph dataset translated/corrected to Nepali + 800 cultural heritage images).
- Compared Transformer (BLEU-1: 0.23, BLEU-2: 0.35, BLEU-3: 0.53, BLEU-4: 0.59) against LSTM with ResNet152, optimizing hyperparameters (batch size 32, learning rate 0.01, 8 attention heads, dropout 0.2) for Nepali's complex grammar.
- Preprocessed data with tokenization, vectorization (Keras TextVectorizer), and vocabulary building (14,022 unique words); integrated into a full-stack app (React frontend, Flask/Node backend); published in Journal of Soft Computing Paradigm (March 2024).

Skills

Languages	C, C++, Python, Java, SQL
AI/ML	Pandas, NumPy, Seaborn, TensorFlow, PyTorch, Scikit-learn, Keras, LangChain, LangGraph, sentence-transformers, CNNs, RNNs, Transformers, Linear/Logistic Regression, Clustering (K-means, DBSCAN), Natural Language Processing, Generative AI, Machine Learning Algorithms, Statistical Knowledge, ML Frameworks, NLP Techniques
Tools	Jupyter Notebook, Google Colab, Kaggle, Git, Selenium, pytesseract, FastAPI, Pydantic, Uvicorn, Milvus, Redis, Docker, Docker Compose, LangSmith
LLMs	Groq LLM
Soft Skills	Communication, Team Collaboration, Time Management

Publications

Subedi, N. et al. (Jan. 2024a). "Drowsiness and Crash Detection Mobile Application for Vehicle's Safety". In: *Journal of IoT in Social, Mobile, Analytics, and Cloud* 6.1, pp. 54–66. URL: https://doi.org/10.36548/jismac.2024.1.005.

– (Jan. 2024b). "Nepali Image Captioning: Generating Coherent Paragraph-Length Descriptions Using Transformer". In: *Journal of Soft Computing Paradigm* 6.1, pp. 70–84. URL: https://doi.org/10.36548/jscp.2024.1.006.