Jyotsan Hamal

Baluwatar, kathmandu

• **E-mail** : hamaljyotsan@gmail.com • **Contact no** : +9779824595613

Experience

Intern at **Subiz Innovations** as Python Automation Engineer

- Jun 2023 - Nov 2023 · 6 months
  - **Data Extraction** : Made a python bot which extracts the Nepal stock exchange stocks data into a csv file.
  - **OCR** : Used a **PaddlePaddle** OCR which converts Bank statement images, pdf into excel sheet.
  - **Data Cleaning** : Extracted data from different sites which i was able to clean for ML model training

AI Engineer at **Palm Mind Technology**

- Dec 2024 - Present
  - **Hyundai Chatbot Revamp** : Migrated Hyundai chatbot from LangChain to LangGraph, integrated SQLite for memory, and optimized the RAG pipeline. Deployed and maintaining with improved performance and scalability.
  - **Sanima Bank OCR Integration** : Developed a FastAPI backend to integrate an OCR pipeline for extracting data from documents like application forms and KYC, and storing the structured output in MongoDB.
  - **OpenAI & Twilio Integration** : Integrated OpenAI APIs with Twilio to enable real-time user interaction and automation
  - **CG Motors Chatbot Optimization** : Refactored the entire CG Motors chatbot system to be fully asynchronous, significantly improving concurrency and response time. Enabled the bot to handle more users simultaneously with increased efficiency and scalability.

Projects

Phone Call Agent: Developed a real-time phone call agent using Twilio and OpenAI APIs. The agent can book, reschedule, and cancel appointments through natural conversation, leveraging WebSocket for real-time communication. **Github** : Link

Libraries used:
• WebSocket • FastAPI • Twilio

Build Llama.cpp for Custom GPU Architecture: Built and configured llama.cpp for custom GPU architectures.Compiled llama.cpp for GTX 1650 Ti Mobile and RTX 3060 (via Salad), enabling support for efficient inference.Successfully served the gemma-3 4B GGUF multi-model setup.**Github** : Link

Libraries used :
• Llama.cpp • Cmake • CUDA Toolkit • Multer • REST API

PPOCRv5 Fine-tuning Backend: Built a FastAPI backend to receive images and PDFs from a frontend dashboard, enabling users to easily fine-tune a PPOCRv5 model. The backend handles data preprocessing, configuration, and training, streamlining the fine-tuning process for document OCR tasks. **Github** : Link

Libraries used:
• Pytorch • FastAPI • PaddlePaddle • Datasets

ExStock : Developed ExStock to provide a risk-free platform for users to learn and practice stock trading using real-time NEPSE (Nepal Stock Exchange) data. Built features like a personalized AI tutor and pattern recognition learning to help users understand trading concepts and identify market trends. Simulated real trading scenarios with virtual currency to enable hands-on experience without financial risk.. **Github** : Link

Libraries used:
• BeautifulSoup • LangGraph • FastAPI • React • Socket

Training

AI Fellowship

- **Fusemachines** | Apr - Oct 2024
  - Engaged in a comprehensive AI fellowship focused on advanced machinelearning, deep learning, and computer vision
  - Applying technologies like face detection, gaze tracking, and head pose detection in Fellowship projects.

Education

Lincoln University,
Phoenix College of IT & Management
Maitidevi, Kathmandu
2022-2025

Bachelor of Information Technology
- BIT

Hope International College
Karkando , NPJ
2019-2021

High School