KOSHISH SHRESTHA

Baudha, Kathmandu, Nepal

📞 9863336676 ✉ koshish62@gmail.com 🌐 koshishshrestha.com.np ⓞ github.com/koshishshresthaa

About Me

AI/ML Engineer specializing in deep learning, generative AI, and data-intensive software systems. Developing scalable, reliable, and human-centered intelligent solutions across domains including automation, analytics, and decision- support systems. Backed by a solid Computer Engineering foundation with a Bachelor's degree from Tribhuvan University, I aim to contribute impactful, research-driven innovations.

Education

Kathmandu Engineering College   **2019 − 2024**
Bachelor of Computer Engineering   *Kalimati, Kathmandu*
(Aggregate Percentage: 81.27%)

Skills

Frontend: HTML/CSS, JavaScript, ReactJS
Machine Learning: NumPy, Pandas, Scikit-Learn, FastAPI, PyTorch, LangChain, GenAI, MLFlow, AirFlow, LangFuse, TensorFlow, Streamlit, Google ADK
Language: C, C++, Python, SQL
Miscellaneous: Git, GitHub, BitBucket, Linux, Docker, Jira, Adobe Premier Pro, Adobe Illustrator, Figma
Soft Skills: Leadership, Team Work, Problem Solving, Self-Learner

Experience

Machine Learning Engineer II @ GritFeat Solutions   **August 2024 − Present**

- Developed and deployed time series forecasting models for business-critical applications, driving data-informed decision-making through accurate demand and trend predictions.
- Applied Large Language Models (LLMs) such as GPT in client-facing projects to solve complex, real-world challenges and demonstrate expertise in cutting-edge ML technologies.
- Collaborated cross-functionally with product, data, and engineering teams to align AI initiatives with strategic business objectives, improving system reliability and project efficiency.
- Mentored junior engineers and research fellows, providing technical guidance and fostering a culture of learning and innovation within the team.

Associate Machine Learning Engineer @ GritFeat Solutions   **July 2024 − July 2025**

- Developed and deployed end-to-end conversational AI and recommendation systems using LLM-based retrieval-augmented generation (RAG), advanced prompt engineering, and containerized microservices for scalable, context-aware interactions.
- Contributed to AI research and development initiatives by exploring emerging generative AI techniques, evaluating new frameworks, and assisting senior engineers in building proof-of-concept prototypes.
- Implemented robust model monitoring pipelines to track performance, response quality, and system reliability, ensuring stable production deployments and continuous optimization.

Machine Learning Fellow @ GritFeat Solutions   **April 2024 − July 2024**

- Studied end-to-end machine learning workflows, including data preprocessing, model development, evaluation, and deployment.

Activities

Student Lead @ PALS   **September 2023 − May 2024**

- Enriched my skills in leadership, teamwork, collaborations and work ethic under guidance of IIT Alumni Champions.
- Attended IIT Madras for special programs and events, demonstrating a commitment to expanding knowledge and networking opportunities.
- Collaborated with IIT alumni at PALS to promote responsible technology innovation among South Asian student networks.

President @ KEC IT Club   **2021 − 2024**

- Lead team of I.T Club in all sectors.
- Represented the club in different coding competitions, hackathons etc. organized by other colleges and organizations.

Content Creator @ YouTube   **2017 − Present**
- Travelling and creating contents on my own.

College Representative @ Engineers Vlogs
- Played a key role as a member of the organizing team for Orbit Engineering Expo 2022 and handled entire robotics section during the expo.

Publications

Comparison Analysis of Nepali News Classifier
- Authored and presented the paper on OKRP conference and published in the KEC Journal. Paper Link

Personal Projects

KaanunSathi 🎧   **RAG, RAPTOR, LangChain, LLM**
- Developed a law chatbot leveraging Advanced RAG using LangChain, integrating LLM to provide contextual and accurate responses from Nepali laws.
- Implemented state-of-the-art techniques like RAPTOR to ensure optimal context retrieval.

Nepali News Classifier 🎧   **Python, NLP, BERT, Transformers**
- Conducted a comparative analysis of machine learning models to optimize classification accuracy.
- Implemented data preprocessing, visualization, and analysis techniques to enhance model performance.
- Built the classifier in PyTorch using the NepBERT model from Hugging Face.

Automatic Video Captioning 🎧   **Python, FastAPI, Streamlit, OpenAI Whisper**
- Built a full-stack video captioning tool to automate subtitle generation and enable real-time transcription editing.

Certifications and Course Reviewed

Data Scientist Associate - DataCamp [Link]

AWS Cloud Quest: Generative AI Practitioner - Training Badge [Link]

AWS Cloud Quest: Cloud Practitioner - Training Badge [Link]

Resident Student Workshop: Think, Create, Engineer - IIT, Madras   **December 2023**
- Achieved first place in the implementation of a 'StormWater Drain Blockage Detection System' as part of a mini project. Led a team to design and execute the project, demonstrating problem-solving skills and technical competence.
- Engaged in a dynamic workshop at IIT Madras focused on fostering creative thinking and engineering skills. Explored System Thinking Sustainability, Design Thinking and various concepts.

HEx Genius Hackathon   **July 2023**
- Developed a web application that facilitates direct communication between local bodies and government entities.
- Implemented an issue resolution system where unresolved issues are automatically escalated to higher authorities. Also incorporated an upvote system and other features to enhance user engagement and feedback.
- Utilized the MERN stack for the project.

Python for Data Science, AI and Development - IBM(Coursera)

HarvardX: CS50's Introduction to Programming with Python - CS50P — by Prof. David J. Malan

AWS Artificial Intelligence Practitioner Learning Plan - AWS Training and Certifications

Advanced Learning Algorithms - DeepLearning.AI (Coursera)

Supervised Machine Learning: Regression and Classification - DeepLearning.AI (Coursera)

End-to-End Machine Learning - DataCamp