SAUMYA BHANDARI

Experienced AI Engineer ● Deep Learning ● Computer Vision ● Innovator

📞 +977 9843866411 ● 📍 Mulpani, Kathmandu, Nepal ● @ saumyabhandary22@gmail.com
🌐 saumyabhandari.com.np ● in saumya-bhandari-558b8a210 ● �ured SaumyaBhandari

EXPERIENCE

Head of Research and Product

The Legalities

📅 Oct 2024 – Present   📍 Kathmandu, Nepal

- Developed an AI-powered compliance assistant (**SelfCompl.ai**) that auto-generates audit-ready documentation, reducing compliance preparation time from 3–6 months to just 20 days. Product based on RAG.
- Designed and deployed end-to-end AI-based document intelligence systems to extract and analyze structured data from unstructured legal files.
- Led development of LegalOps+, an AI-powered product for compliance automation and legal case tracking. Contributed to overall product strategy, UX, engineering, and growth roadmap of the product.

Core tech: Django REST APIs, Python, RAG LLMs, OCR, Pinecone, VectorDB, NLP, Computer Vision, MultimodelAI, Prompt Engineering, PostgreSQL

Machine Learning Engineer

Naamche (ReAlpha)

📅 Jan 2024 – Oct 2024   📍 Remote

- Fine-tuned SegFormer for internal room segmentation; generated virtual interiors by manipulating segmentation bitmaps, prompts, and diffusion-based inpainting. Applied research-backed strategies which is used by 86% of buyers for better property visualization to drive product design decisions. Leveraged AWS SageMaker and serverless GPU architectures for scalable experimentation and model deployment.
- Developed an AI-based virtual real estate agent (**Claire**) using RAG pipelines, OpenAI hooks, and Pinecone, guiding users on each step on their home buying journey.
- Conducted intensive data preprocessing, anomaly detection, and visualization using Pandas and SQL-based techniques to improve data quality.

Core tech: FastAPI, Python, RAG LLMs, OCR, Pinecone, VectorDB, NLP, GPT4-V, Multimodel AI, Prompt Engineering, Pytorch, Diffusion Models, Inpainting Models, Image segmentation, Finetuning SegFormer, AWS Sagemaker, PostgreSQL, AWS S3, AWS EC2, Pandas, Matplotlib.

Computer Vision Engineer

Treeleaf Technologies

📅 Oct 2023 – Dec 2023   📍 Lalitpur, Nepal

- Built, optimized and Reworked on development OCR and object detection microservices with gRPC.

Core tech: Python, gRPC, Protobuf, Microservices, OCR, YOLOv-8

Machine Learning Engineer

Wiseyak Inc.

📅 Sep 2022 – Oct 2023   📍 Kathmandu, Nepal

- Developed a Chest X-ray multi-label classifier (**AUROC > 0.91**) from scratch, covering 10+ diseases in a highly imbalanced dataset using weighted loss techniques and class balancing strategies. Integrated Grad-CAM based heatmap visualization and built a multimodal pipeline to generate radiologist-level reports using GPT-2.
- Earned **MIT**'s Data and Specimen Research certification upon nomination by Prof. Suresh Manandhar for excellence in ethical AI research.
- Co-designed and implemented a novel CNN-Transformer hybrid architecture for medical video processing, including autoencoder-based next-frame prediction and a fully convolutional transformer for medical image segmentation.

- Led the development of a domain-adversarial neural network (**DANN**) trained on a Nepali plant dataset curated in collaboration with (**GeoKrishi**) and professors from (**Madan Bhandari University of Science and Technology**). Scraped and organized a comprehensive plant disease Q-A dataset from Plantix forums to complement model training with real-world disease-symptom mappings. Using Meta's SAM for domain alignment techniques.

Core Tech: PyTorch, CNN, Transformers, Fully Convolutional Transformer, GradCAM-PCAM, LLMs, Flask, OCR, Domain Adversarial Neural Network (DANN), Segment Anything Model (SAM), RAG, Multimodal AI, Prompt Engineering, Temporal Modeling, Autoencoders, Next-frame Prediction, Video Denoising, Image Segmentation, Medical Image Analysis, Multilabel Classification, Imbalanced Data Handling, Data Annotation Pipelines, Data Augmentation, Transfer Learning, Domain Adaptation, Local Server Deployment, Ethical AI, Data Ethics Training (MIT), Knowledge Base Documentation.

ML Intern
Namespace.jp / Chulo Solutions

📅 May 2022 – Aug 2022　　　　　　　　　📍 Lalitpur, Nepal

- Worked on building YOLO3-inspired waste management model from scratch.
- Learned core Mathematics behind ML models including PCA, SVM, EVD, PseudoInverse and creating Neural Networks, CNNs, Transformers and even writing backpropagation and different optimization hacks like adagrad, adam from scratch - by using own function implementation.

EDUCATION

BSc. (Hons) Computer Science
Herald College Kathmandu

📅 2020 – 2023　　　　　　　　　📍 First Class Honors

Thesis: Image Segmentation for Autonomous Vehicle Navigation and Image Super Sampling
- DevCorps Leader (Leader of 3/6 student run communities), On Campus Hult Prize Winner, Student Academic Rep for 2 consecutive years, Planned and Ideation of Highest Altitude Hackathon 2023, Scholarship Achiever.

Higher Secondary
Uniglobe College

📅 2018 – 2020　　　　　　　　　📍 Mathematics + CS

- IT Club President, Hackathon Winner (2x), National Level Debate Finalist, Organized and won 15+ competitions including quiz, debates, essays and presentations.

Secondary School
The New Summit School

📅 2016 – 2018　　　　　　　　　📍 Kathmandu

- School CCA Captain, **Student of the Year**, Soft Skills Coach.

SKILLS

Deep Learning / Self Supervised Learning　　OpenCV / Computer Vision / Diffusion Models / CNNs
NLP / LLMs / Document Understanding / OCR　　PyTorch　　FastAPI / Flask / Django　　gRPC

SELF LEARNING AND GROWTH

Self Learning

Core Courses: Harvard Stats 110 || MIT 18.06 Linear Algebra (2006) || MIT 6.034 Artificial Intelligence (2010) || MIT Deep Learning Bootcamp (2020–2022) || Stanford CS231N Deep Learning for Visual Recognition (2017) || University of Waterloo CS480/680 Deep Learning (2021)

Extra Learning (Out of Passion): MIT 9.13 The Human Brain (Spring 2019) || Stanford Human Behavioral Biology (2011) || MIT 8.04 Quantum Physics I (Spring 2013) || NCCR SwissMAP Quantum Information Theory (2020)