KAMAL SHRESTHA

Machine Learning Engineer

kamalandshrestha@gmail.com | shresthakamal | shresthakamal
Bangalore, India | (+91) 7893887563 | shresthakamal.com.np

SUMMARY

- **Proficient in End-to-End Machine Learning and Deep Learning Pipelines:** Extensive experience across the full ML lifecycle, from data generation and preprocessing to system design, model training, evaluation, deployment and maintenance, with a proven record in deploying robust, scalable solutions that drive tangible outcomes.
- **Specialist in Applied NLP with Expertise in Transformer Architectures:** Skilled in solving diverse NLP challenges, from synthetic data generation, fine-tuning novel architectures for classification and generation to deploying advanced LLMs based RAGs and Agentic collaborations to address high impact business needs.
- **Proven Leader in Collaborative Projects and Technical Communication**: Recognized for strong leadership in team collaboration, technical documentation, and impactful presentations, supported by industry experience, academic research, and a track record of bridging technical expertise with clear, strategic communication.
- **Professional Career/Research Interests:** Intersection of applied NLP, DL, and ML Techniques for business application

WORK EXPERIENCE

Veritus AI — Hyogo, Japan (Remote)
Applied AI Consultant specializing in LLMs — *August 2024 – September 2024*

- Guided the development of RAG applications to automate search, review, and analysis in research workflows by implementing advanced techniques. customizing open-source solutions to streamline processes and reduce repetitive tasks—culminating in the successful launch of an innovative, automation-driven product that empowers researchers to focus on core innovations.

Research & Technology Center, BOSCH Global Software Technologies(BGSW) — Bengaluru, India
Machine Learning Engineer — *August 2023 – Present*

- Currently, working on a high-impact project centered on leveraging LLMs for fine-tuning and retrieval augmented generative text introducing interactive agents for streamlined workflows and user engagement aiming to enhance operational efficiency across business units. **Received the Bravo Award X3 for excellent rigor and engineering skills in successful completion of multiple applied Generative AI use cases.**
- Developed and implemented an novel classification model for long legal documents within the ProCodex team, enhancing compliance processes for different BOSCH products across multiple departments and jurisdictions around the world, **achieving an annual impact of €10 million savings from human efforts**.
- In parallel, **pursuing multiple research verticals** on advanced document processing, enhancement of core elements and architecture of LLM based RAG approaches like multimodal RAG, agentic RAGs, Graph RAG alongside fine-tuning open-source models like **Llama3** to make it compatible with understanding custom enterprise data

Machine Learning and Curriculum Engineer — Kathmandu, Nepal
Fusemachines — *July 2020 – Dec 2021*

- Designed an automated video recording, **Fuse Studio platform** for generating lecture videos from presentations and manual scripts to reduce extensive recording efforts and **increase production efficiency by 75%**
- Remodeled and optimized **Questions Answering and Difficulty Ranking Model** with better representations, semantic ranking, and recommendations for quizzes, assignments, and exams using custom approaches.
- Worked as a **lead curriculum engineer** to design, create, review, and refine materials (inclu. reading materials, slides, audio transcripts, graded assignments, hands-on implementations, and quizzes) for focused coursed like DL and NLP.
- **Represented Fusemachines as industry expert** in teaching CS concepts like AI, Python programming, and data analysis to students at Q.I. Roberts Jr-Sr High School in Florida, USA, and undergraduates at Herald International College.

EDUCATION

M. Tech. in Computer Science and Engineering, CGPA: 9.06/10 — *Aug 2021 – July 2023*
Indian Institute of Technology, Hyderabad (IITH) — *Hyderabad, India*
Advisor: Dr. Maunendra Sankar Desarkar, NLIP Lab
Area of focus: Recommendation Systems and Hostility detection on online social media conversation threads
Relevant Courses: NLP, Information Retrieval, DL, Fundamentals of Machine Learning, Software Engineering.

Bachelors in Computer Engineering, Percentage: 92.30% *Aug 2016 – Nov 2020*
Kathmandu University (KU) *Dhulikhel, Kavre, Nepal*
Relevant Courses: AI, ML, DSA, Algorithm and Complexity, Software Engineering, Probability and Statistics, Speech and Language Processing, C, C++, DBMS

COURSE WORK

1. **B. Tech.** 2016-2020
   Machine Learning, Artificial Intelligence, Data Structures and Algorithms, Algorithm and Complexity, Software Engineering, Probability and Statistics, Linear Algebra, Computer Architecture, Operating Systems, Human-Computer Interaction, Digital Signal Processing, Compiler Design, Speech and Language Processing, C, C++
2. **M. Tech.** 2021-2023
   Fundamentals of Machine Learning, Natural Language Processing, Information Retrieval, Deep Learning, Computer Vision, Network Security, Computer Networks, Recommendation Systems, Data Structure and Algorithms.
3. **Supplementary**
   - **Stanford course CS224N**: Natural Language Processing with Deep Learning 2022
   - **Hugging Face Course** (Datasets, Dataloaders, Transformers, NLP Tasks) 2022
   - DeepMind x UCL, **Introduction to Reinforcement Learning, 2015** 2019
   - **Technical Writing One**, Google Developers 2019

TECHNICAL SKILLS

Programming Languages	Python, C, C++, SQL
Libraries	Pytorch/Lightning, Transformers, Langchain/LanGraph, Microsoft Azure, VectorDB Pandas, Numpy, SciPy, Matplotlib, Flask/FastAPI, Docker, MLOps(Hydra, DVC)
Database	MySQL, MongoDB, Firebase, Elasticsearch
Management	Git, Github, JIRA, HRM Suite, Notion, Slack
Miscellaneous	Linux, Bash, Arduino, Latex

PUBLICATION

- Aditi Bagora, **Kamal Shrestha**, Kaushal Kumar Maurya, and Maunendra Sankar Desarkar. 2022. Hostility Detection in Online Hindi-English Code-Mixed Conversations. In Proceedings of 14th ACM Web Science Conference 2022 (WebSci '22). ACM, New York, NY, USA, 11 pages doi: 10.1145/3501247.3531579
- **Shrestha, K.** , Poudyal, P. , Karki, J. , Ranabhat, D. (2022). A Machine Learning Approach to Identify Fake News. Center for Project Management and Information Systems (PMIS) Review, 1–13. http://journal.pmis.du.ac.bd/journaldetails.php?pid=2203281648465920

RESEARCH EXPERIENCE

Natural Language and Information Processing Lab (NLIP) IIT Hyderabad
Academic 'C' Block, IITH *May 2022, July 2023*

- Developed personalized odd jobs recommendation engine based on heuristics and transformer-based learning approaches for a platform catering to differently able individuals with skills and training.
- Implemented SOTA models, TransferTransfo and DialoGPT, to enhance the representation of in-turn conversational history, resulting in improved accuracy, diversity, and human-like responses in dialogue systems.
- Introduced a novel hierarchical neural network architecture for threaded social media posts, comments, and replies in conversations that involve a mixture of Hindi and English languages compared with SOTA models.
- Conducted multiple human evaluation benchmarks to assess the performance and effectiveness of various trained models for text generation, recommendation analysis, similarity study, hostile text assessment, and more.

POSITION OF RESPONSIBILITIES

ACM - IIT Hyderabad Student Chapter - Machine Learning Moderator 2022-2023
Indian Institute of Technology Hyderabad *Telangana, India*

- Responsible for all in college **AI/ML/NLP** discussion groups, paper reading sessions, blogs, writing groups, implementation workshops.
- Organized session on "**Structuring Machine Learning Projects**" that aimed to provide insights and guidance for fellow students looking to excel in their ML projects for better reproducibility, collaboration and more with industry-standard tools like Poetry, Commit Hooks, Github, Tensorboard, MLFlow, Docker.

IT MEET v8.0 Documentation Lead and Marketing Representative 2020
Kathmandu University *Kavre, Nepal*

- **Lead of documentation team** and member of the marketing team of one of Nepal's premier annual IT events, IT Meet v8.0 with **25 different events** and participation of more than **50 tech. companies** with internships and full-time career opportunities.
- Directed **a team of 25** through multiple documentation stages for sponsorship (proposals, cost analysis, contracts, MoU) and collaboration with **10-20** companies to successfully bring in **8** different companies with total funding of **NPR 110,000**, **20%** of total funding.

IT MEET v7.0 Photography Event Organizer 1, 2 2019
Kathmandu University *Kavre, Nepal*

- Organized **all Nepal Photography Competition** (open theme), which was judged by the president of Nepal Photography Association (NPA) and the Dean of Engineering at Kathmandu University
- A collective prize pool of **NPR 25,000** was shared among the top three winners with scholarships to a photography workshop, 14*20 inch framed winning photos, and cash prizes.

Executive Board Member 2018 –2019
Kathmandu University Computer Club *Kavre, Nepal*

- An active student-run club of the Department of Computer Science and Engineering at Kathmandu University solely responsible for the majority of extracurricular activities like LTSP (Linux Terminal Server Project) and Software Freedom Day.
- Conducted **20+ workshops** (on varied topics) in collaboration with multiple guests from premium companies like **A Yomari Company**, **Fusemachines**, **Deerwalk**, F1Soft International, MIDAS and more every year.

Ambassador 2017
Em-Blood Android Application with Nepal Red Cross Society *Kathmandu, Nepal*

- Supervised a **team of ten volunteers among five different teams** in spreading awareness of the need for fresh emergency blood within different blood donation centers, hospitals, universities, schools, and blood banks.
- Involved in creating awareness posters, **an indexable database (using Google Firebase)** for health professionals, and managing help desks in every hospital **(in a distributed network of hospitals)** that patients can contact in need of fresh blood/blood donors.

PROJECTS

Inclusivity in Job Recommendation based on heuristic and learning approaches IIT, Hyderabad
M. Tech. Thesis, Patent Approved *May 2022 – July 2023*

- Developed a hybrid recommendation engine based on heuristics and transformer learning approaches for a personalized recommendation based on disability, skills, and preferences.
- Attained an impressive **F1 score of 0.9389** on the validation set and 65% accuracy on similar user analysis from human feedback with minimal space usage and low latency in recommendations

Hostility Detection in Online Hindi-English Code-Mixed Conversations IIT, Hyderabad
14^{th} *ACM Web Science Conference* 2022 *(WebSci'22,)* *June 2022*

- Proposed **a novel hierarchical neural network architecture** to identify hostile posts/comments/replies in online Hindi-English Code-Mixed conversations as a part of HASOC'2021.
- Adapted multilingual pre-trained models like **mBERT, XLMR, and MuRIL** to generate contextual representations for natural abstraction and selection of the relevant context by exploiting the hierarchy of the conversations. [URL], [PDF], [Presentation], [Video], [Code]

Natural Language and Information Processing (NLIP) Lab
IIT Hyderabad, Supervised by Dr. Maunendra Sankar Desarkar *Dec, 2022 - July, 2023*

- Actively involved in the information collection, design, development, documentation, and maintenance of the NLIP Lab at IITH, which is hosted at nlip.cse.iith.ac.in.
- Developed valuable collaboration skills by actively participating in project work with lab mates, focusing on repository management, agile workflow, work delegations, effective team communication, minimal front-end requirements, and structuring Jekyll websites

Zero Reference Low-Light Image Enhancement with Attention IIT Hyderabad
Dr. Sumohana Channappayya, Deep learning, AI5100 *2022*

- A low-light image enhancement task using a deep learning-based Zero-Reference Deep Curve Estimation(Zero-DCE). The idea is to use carefully formulated non-reference loss functions to convert the light enhancement as an image-specific curve estimation task.
- [Code], [Project Report], [Project Presentation]

Federated Semi-Supervised Medical Image Classification via Inter-Client Relation Matching IIT Hyderabad
Dr. C. Krishna Mohan, Visual Computing, CS6450 *April, 2022*

- Remodeled and evaluated **medical image classification** with the addition of Self Attention mechanism in every convolutional block: using CBAM to obtain better classification results.
- Outperformed the official implementation given a reduced dataset (only 2%) because of computational limitations
- Ranked with the best **Top 2%(A+) of the class** on the basis of two project presentations.
- [Paper], [Official Implementation], [Modification], [Presentation 1], [Presentation 2]

Cracking WPA2-PSK Wi-Fi Passphrase and Defenses IIT Hyderabad
Dr. Bheemarjuna Reddy Tamma, Network Security, CS6903 *May, 2022*

- Focused on de-authenticating and eavesdropping on the connection between an AP and clients to capture 4-way handshake messages used to brute force the passphrase using aircrack-ng tools
- Demonstrated comprehensive knowledge of handshake protocols, wireless MITM attacks, and proficient understanding of potential defense mechanisms. defenses.
- [Project Details], [Report],

Creating a two-way frewall using raw sockets IIT Hyderabad
Dr. Kotaro Kataoka, Network Security, CS6903 *May, 2022*

- Designed and implemented a bidirectional firewall system using raw sockets with extended rules set adaptable for all protocol layers, detection of DDoS attacks, and unbiased performance examination and evaluation
- [Project Details], [Report],

Secure chat communication with Openssl and Man-in-the-middle attacks IIT Hyderabad
Dr. Bheemarjuna Reddy Tamma, Network Security, CS6903 *April, 2022*

- Implemented and demonstrated **a secure peer-to-peer chat application using openssl** along with how evil Trudy(user) cab intercept the chat messages to launch various attacks(Downgrade Attack by rejecting the request for TLS Encryption and MITM attack with two TLS connections at either end and Fake Certifcates)
- [Project Details], [Application], [Interceptor]

Network Intrusion Detection System (NIDS) using Machine Learning Techniques IIT Hyderabad
Network Security, CS6903 *March , 2022*

- A machine learning approach to detect different anomalies and **attacks like DDoS, MITM, Probing attacks, and R2L**, in network systems using classical machine learning techniques like Support Vector Machine, Decision Tree, Random Forest, Naive Bayes, K-Means, and Neural Networks with sampling techniques like SMOTE to report weighted F1 score
- [Project Details], [Dataset], [Checkpoints], [Code]

Fake News Detection Kathmandu University
PMIS Review, Volume 1, No 1 *June 2020*

- Focused on applying NLP sentence classification to generate contextual sentence representations passed over classical machine learning classification heads to predict whether the provided sentence is fake or not with a certain degree of confidence.
- Evaluated using lexical/syntactical/grammatical/factual features based only on raw text and semantic features based on contextual representations with attentive weights.

Fuse Studio, Video Automation Kathmandu, Nepal
Fusemachines Nepal *July, 2020*

- Designed and Developed **a fully automated Video and Text-to-Speech(TTS) Generation System** for Fusestudio (an in-house project), which focuses on creating a complete lecture video with subtitles given google presentations slide and audio text transcript mimicking how a person presents in a virtual presentation.
- The best sounding Mozilla TTS models used were **Tacotron2, GlowTTs, and MelGAN** with different vocoders.
- [Slide], [Script] ,[Video]

A Machine Learning Approach to Identify Fake News Kathmandu University
Semester Project, Dr. Prakash Poudyal *June, 2020*

- Focused on applying NLP sentence classification to generate contextual sentence representations passed over classical machine learning classification heads to predict whether the provided sentence is fake or not with a certain degree of confidence.
- Evaluated using **lexical/syntactical/grammatical/factual features** based only on raw text and **semantic features** based on contextual representations with attentive weights.

A Machine Learning Approach to Detect Click baits in Online News Kathmandu, Nepal
Microdegree in Deep Learning, Fusemachines Annual Journal *2020*

- Characterization of the raw textual data using multiple hand-crafted attributes combined with the contextual word vector representations and modeled using RNN and LSTM with attention to the classification of click-bait headlines in online news portals.

Self Diagnosis, Computer-Aided Diagnosis (CAD) Kathmandu University
Semester Project, Dr. Dhiraj Shrestha *2019*

- A computer-aided diagnosis approach to detect potential diseases based on symptoms.
- Users were asked a series of dynamic questions (the next question depended on the previous answer) that were converted to a feature set for making inferences.
- Baseline models were Naive Bayes, multi-layered (four) deep neural networks, and Ensemble Techniques like Gradient Boosted Tree(XG-Boost) trained in detecting Tuberculosis and Hepatitis.
- [Proposal], [Code], [Report]

Automatic Obstacle Avoidance Four wheeler
Kathmandu University *2019*

- Designed and created **an obstacle avoiding self-driving car** that uses ultrasonic sound sensors directed motor modules for detection, navigation, and avoidance using Arduino.
- [Video], [Hardwares]

Generation of National Flags using GAN
Microdegree in Deep Learning, Fusemachines *2019*

- Scrapped 600 national fags of 60 different countries to train DC-GAN using Keras API for the generation of unique national fags of our own.

Simulation of the sorting algorithms using OpenGL
Kathmandu University *2018*

- Created a simple desktop application to visualize sorting algorithms like the **Bubble sort, Insertion sort, and Merge Sort** in C++ using SDL/SFML.

RentSpace, a rental solution Kathmandu University
Semester project, Dr. Gajendra Sharma *2017*

- An android application that acts as a mediator for customers (Customer to Customer approach) to address the need to rent, lease and sell available spaces like rooms, apartments, lands, hotels, conference halls, etc., online.
- [Proposal], [Presentation], [Report]

AWARDS AND ACHEIVEMENTS

Dr. Homi Jahangir Babha Scholarship Scheme-HJBSS. **Fully Sponsored by Ministry of External Affairs, Government of India** with EdCIL and provided by the Embassy of India, Nepal to study M.Tech in Computer Engineering at IIT, Hyderabad. **1 (in CSE) of** 3 **selected out of** 10,000+ **students per year** through embassy screenings and college interviews, worth **$15000 (self-sponsored)**. *2021 – 2023*

Golden Jubilee Scholarship (GJS). 1 of 200 out of 20,000+ recipients of the prestigious GJS awarded by Embassy of India for Nepalese students based on B.Tech academic certifications with a monthly stipend of **NPR 4,000 for 4 years.** *2016 – 2020*

Fuse Machines Artificial Intelligence Fellowship Program. 1 of 15 recipients in 2,000+ applications of Micro Degree™ in Artificial Intelligence continued to Micro Degree™ in ML and DL, worth **$500 each**[1][2]. Selections were based on knowledge in in-person exams, interviews, and coding sessions. *2019 – 2020*

Kathmandu University Merit-based scholarship (2x). 1 out of 60, awarded for securing the highest SGPA in the Computer Engineering in the 2^{nd} and 6^{th} semesters respectively each worth of **$500**. *2016, 2017*

CERTIFICATION

1. **Deep Learning Institute(DLI), NVIDIA**
   - Fundamentals of Deep Learning — April 6, 2022
   - Accelerating Data Engineering Pipelines — February 12, 2022
   - Fundamentals of Accelerated Data Science with RAPIDS — February 5, 2022
   - Accelerated Computing with CUDA Python — January 29, 2022
   - Accelerated Computing with CUDA C/C++ — January 22, 2022
2. **AWS Certified Machine Learning – Specialty**, Amazon AWS — August 31, 2021
3. **Complete Google Slides Course** -Create Stunning Slides, Udemy — May 23, 2021
4. **Machine Learning from Begineer to Advance**, Udemy — May 27, 2021
5. **Python for Machine Learning with Numpy, Pandas and Matplotlib**, Udemy — May 27, 2021
6. **How to win Data Science Competition: Learn from Top Kagglers**, Coursera — October 1, 2020
7. **Effective Client Communication**, Fusemachines, Nepal — July 23, 2020

PARTICIPATION

1. Symposium on Artificial Intelligence for Sustainable Development — January 29, 2022
2. Online Research Paper Writing Training conducted by NIRC — May 17, 2020
3. 3^{rd} National Workshop on Machine Learning and Data Science — 30 July - 3 August, 2020
4. Webinar on cyber security and cyber space organised by Oxford Stem, Code For Change — June 28-30, 2020
5. Arduino Workshop, K.U. Robotics Club — April 25-28, 2018
6. Prixa Excellence Award for Project RentSpace, Android Application — June, 2017
7. Effective Manuscript Writing, Ethics and Plagarism by ACS — 22 July, 2022

VOLUNTEERING

1. Worked as a medical volunteer for first aid in Inter College Basketball Tournament organized by Kathmandu University Student Welfare Council — 2017
2. Worked as a volunteer in all Nepal Counter Strike Competitions at IT MEET 2018, organized by Kathmandu University Computer Club — 2018
3. Worked as a volunteer in Annual General Meeting of Kathmandu University Youth Red Cross Circle (KUYRCC) — 2019