Alok

Roll No.: EE23MT001
M.Tech - CSPML
Indian Institute Of Technology, Dharwad

+91-7597694703
ee23mt001@iitdh.ac.in
aloklashok@gmail.com
GitHub | LinkedIn | Hugging Face

Professional Summary

AI/ML Researcher with expertise in **Generative AI, Large Language Models (LLMs), LLM Agents, Retrieval-Augmented Generation (RAG), and Computer Vision**. Proficient in **Transformers, Deep Learning, and Federated Learning**, with hands-on experience in **Vector Databases (Pinecone, Chroma), LangChain, and Knowledge Graphs**. Skilled in **fine-tuning, optimizing, and deploying scalable AI solutions** using **Azure, PyTorch, TensorFlow, and Hugging Face**. Passionate about **LLM-based applications, AI-driven automation, and knowledge retrieval**.

Education

•Indian Institute of Technology Dharwad *Aug.2023 - April 2025*
M.Tech.Communication Signal Processing and Machine Learning (CSPML) ***CPI: 8.07*** Dharwad, India

•Govt.College of Engineering and Technology, Bikaner *Aug.2016 - Dec.2020*
BTech Electronics and Communication Engineering (ECE) ***CPI: 9.0*** Bikaner, Rajasthan

Technical Skills

- **Programming Languages**: Python, C++
- **Web Development Technologies**: HTML, CSS, Flask, Streamlit, RESTful APIs
- **Deep Learning Frameworks**: PyTorch, TensorFlow, Keras, TensorFlow Federated
- **Libraries**: Scikit-Learn, Pandas, NumPy, Matplotlib
- **SOTA Architectures**: YOLOv5/YOLOv8, EfficientNet, ResNet, ViT, UNet
- **LLMs and Generative AI**: FastAPI, Flask, Hugging Face, LangChain, RAG, Vision Transformers, GANs, Diffusion Models
- **Cloud and DevOps Tools**: Azure ML, Databricks, Google Colab, Kaggle, Docker, Git, LaTeX
- **Advanced Topics**: Vector Databases (Pinecone, Chroma), Machine Unlearning, Blockchain-Federated Learning

Work Experience

•Machine Learning/ Generative AI Engineer Intern, Tipstat *May 2025 – Present*

– Developing intelligent agents using **LangGraph**, **LangChain**, and **Large Language Models (LLMs)**
– Building multi-agent systems with advanced tool integration including web search and research-oriented prompts to derive accurate result.
– Collaborating with cross-functional teams to deploy research-grade AI workflows for data-driven decision making.

•Teaching Assistant, Indian Institute of Technology Dharwad *August 2023 – April 2025*

– Helped students understand core concepts in the **Neural Networks & Deep Learning Lab** by guiding them through practical coding assignments and experiments..
– Supported the **Microprocessor & Microcontroller Lab** by assisting students with hands-on hardware and programming tasks.
– Graded assignments and provided feedback in the **Control System** Theory course to help students improve their understanding of key concepts.

M.Tech Thesis

•FORGET: Instance-Wise Data Unlearning for Pretrained Models- M.Tech Thesis *Sep. 2024 - April 2025*
Python, PyTorch, ResNet-18, ResNet-50

– Developed an approximate unlearning framework that removes specific training data from neural networks without requiring full retraining, preserving efficiency.
– Applied regularization and adversarial techniques to induce misclassification of targeted data and adjust model weights, ensuring both effective unlearning and retained model accuracy.

Projects

•PDF Summarizer AGENT *April. 2025*

n8n, Google Drive, OpenAI, Pinecone vector database

– Built a PDF Summarizing Agent lets you upload any PDF and instantly get summaries or answers to your questions. It reads the content, creates embeddings, and stores them in Pinecone for fast, smart search. Using OpenAI's models, it gives accurate, conversational replies. Ideal for quickly understanding long documents like reports, manuals, or research papers without reading everything.

•CrowdFlow: Real-Time People Counting and Tracking *Jan. 2025 - Feb. 2025*

Python, OpenCV, YOLOv8, SORT (Simple Online and Realtime Tracking)

– Developed a real-time system to detect and track people in video streams using YOLOv8 and SORT algorithm.
– Implemented custom counting logic for people moving in upward and downward directions.
– Enhanced visualization using overlays, bounding boxes, and entry-exit counters.
– Useful for crowd analytics in surveillance, public spaces, and event monitoring.

•LangBERT Summarizer *September 2024*

Python, FastAPI, LangChain, BERT Transformers

– Built an AI tool that summarizes long documents using LangChain and the BART(*facebook/bart-large-cnn*) model. It keeps key points clear and easy to understand. Used FastAPI to turn it into a web-friendly REST API for real-time use. Improved speed and scale with smart chunking and retrieval. Designed it to help with tasks like summarizing research papers, legal files, and articles for quicker understanding.

•Diffusion Model: Realistic Face Generation with DDPM *May. 2024 - Aug. 2024*

PyTorch, Built from Scratch

– Built a generative model based on the Denoising Diffusion Probabilistic Model (DDPM) architecture to synthesize realistic anime face image. Used a custom anime face dataset.The noise removal is done using SOTA U-Net.Trained the model to learn the reverse denoising process.

•DualWaveNet: A Multimodal Vision Transformer for Depression Classification *Jan. 2024 - March. 2024*

Python, Pytorch, Vision Transformer, Librosa, NumPy

– DualWaveNet is a deep learning model to classify depression in Depressed or Non-Depessed classes. EATD-Corpus dataset is used. Converting audio data to log-Mel spectrograms and tempograms, then feeding data to Multimodal Vision Transformer (ViT b/12) and applied early fusion to combine both modalities . Achieved improved accuracy by capturing both frequency and rhythm patterns in speech.

•FakeSpeechNet: DeepFake Speech Detection with CNN *Sept. 2023 - Nov. 2023*

Python, Pytorch, Librosa, CNN

– FakeSpeechNet is a CNN-based Deep learning model to detect fake speech audio. ASVspoof dataset is used. Converting the audio data to log-Mel spectrograms and then applying 2D-DCT to log-Mel Spectrogram. By doing so capture subtle spectro-temporal patterns.To extract low-dimensional features that capture both spectral and temporal modulations over longer time spans.We fed these 2D-DCt to our CNN model. Achieved superior results.

Coursework

- Linear Algebra
- Optimization Theory and Algorithms
- Probability Models & Applications
- Pattern Recognition and Machine Learning
- Neural Networks and Deep Learning
- Speech Processing
- DSA