Akshita Gupta

akshita24680@gmail.com | (412) 214-2506 | https://www.linkedin.com/in/aks-gupta/ | https://github.com/aks-gupta

Education

Carnegie Mellon University *Pittsburgh, Pennsylvania*
Master of Computational Data Science (CGPA: 4.0/4.0) *December 2025*
Fall 2024: Foundations of Computational Data Science, Advanced NLP, Introduction to Machine Learning, Interactive Data Science
Spring 2025: Multimodal Machine Learning, AI Engineering/ Machine Learning in Production, Cloud Computing

R. V. College of Engineering *Bangalore, India*
Bachelor of Engineering, Computer Science and Engineering (CGPA: 9.27/10) *June 2023*
Selected Coursework: Data Structures, Advanced Algorithms, Artificial Neural Networks, AIML, Mathematical Modelling

Skills

Programming Languages: Python, C++, C, Java, SQL
Tools and Frameworks: Pandas, scikit-learn, Hugging Face, NLTK, OpenCV, Pytorch, Elasticsearch, Git, Flask, Tableau, Streamlit
Concepts: NLP, Multimodal Systems, CV, Deep Learning, Generative AI, Large Language Models, Affect Recognition

Experience

Carnegie Mellon University *Pittsburgh, Pennsylvania*
Teaching Assistant – Advanced Natural Language Processing (Spring 2025)

Cisco *Bangalore, India*
Software Engineer *August 2023 – July 2024*

- Optimized protocols within the L2 and slow path forwarding team for Cisco ASR9k switches.
- Investigated the application of eBPF for obtaining router packet path insights in ASR9k using Machine Learning techniques.
- Resolved over 15 critical sanity failures in the nightly test suite, improving nightly test pass rate by 12%.

Technical Undergraduate Intern *January 2023 – June 2023*

- Streamlined network telemetry data ingestion from Routing Information Base (URIB) and transmitted it to Elasticsearch via the Kafka Queueing service for efficient querying and storage.
- Designed optimized Elasticsearch queries to analyze over 100,000 entries with execution times under 1200 ms, enabling differential insights through AI models.

GlobalLogic *Bangalore, India (Remote)*
Artificial Intelligence Project Intern *October 2023 – November 2022*

- Developed a one-shot face recognition system through Siamese Neural Networks.
- Deployed a Flask application with OpenCV to automate real-time photograph capture and verification against passport images.

Samsung Research *Bangalore, India (Remote)*
Project Intern *June 2021 – January 2022*

- Translated the ATIS dataset with 4633 data points to "Hinglish" Code-Mixed Language for Intent Classification and Slot Tagging.
- Validated the dataset with ten-fold cross-validation and obtained an intent accuracy of 99.6%.

Projects and Publications

Debiasing Large Language Models through Casual-Guided Active Learning

- Streamlined the Active Learning approach to complex models such as LLaMA-3 and proposed metrics for bias evaluation.
- Improved accuracy by over 5% through prompt finetuning and mitigating position bias in MT-Bench and Chatbot Arena datasets.

Retrieval Augmented Generation (RAG) Model for CMU and Pittsburgh related data using LLMs

- Developed an end-to-end Question-Answering RAG Model on Llama 3.2 using Langchain and Ollama.
- Enhanced context retrieval through FAISS embeddings and BM25, obtaining an F1 score of 0.62 on unseen data points.

"Multimodality in Online Education: A Comparative Study." *arXiv (2023)* – Multimedia Tools and Applications (2024)

- Researched optimal classification models for multiple modalities including Facial Expressions, Posture, Speech, Eye-tracking.
- Compiled a dataset of over 4k images for posture recognition, achieving an accuracy of 95.96% on CNN and 93.7% on SVM.

"Joint Intent Classification and Slot Tagging on Agricultural Dataset for Indic Languages." *ICACCS* (2023)

- Created an Intent Based Dataset for Farmer Queries in the Agricultural Domain with approximately 2400 entries.
- Evaluated dataset performance with LSTM and BERT models, yielding an accuracy of 93.89% and 98.32% respectively.

"Comparison of Perplexity Scores of Language Models for Telugu Data Corpus in the Agricultural Domain." *ICICCS (*2024)

- Collected resources in "Telugu" for the agricultural domain through web scraping and trained language models on this data.
- Compared the perplexity scores of 114k unique tokens on n-gram, GPT-2, and LSTM models.

"**A Comparative Study on Storage Solutions for Analysis of Streaming Telemetry Data**". *ISDA (2023*)

- Evaluated the difference in AWS, Azure, and GCP based on pricing, ease, compute power, and efficiency.
- Proposed a hybrid model for hot storage in Time-Series Databases and cold storage in Cloud Systems like AWS.

Leadership and Activities

Student Placement Coordinator, RV College of Engineering *September 2022 – August 2023*
Member, Coding Club, RV College of Engineering *August 2019 – June 2023*
Director, Circle of Acting at RV College of Engineering *January 2023 – May 2023*