

Ishita Chauhan

☎ (413) 472-7137 ✉ ishita.chauhan2025@gmail.com 🔗 linkedin.com/in/chauhan-ishita/ 🐙 [Github](https://github.com)

EDUCATION

University of Massachusetts, Amherst

Master of Science Computer Science (Cumulative GPA: 3.76/4)

Amherst, MA, United States

September 2024 - May 2026

Relevant Coursework: Advanced Natural Language Processing, Systems for Data Science, Research Methods in Empirical CS

Vellore Institute of Technology

Bachelor of Technology Computer Science (Cumulative GPA: 8.73/10)

Vellore, India

August 2020 - August 2024

Relevant Coursework: Data Structures and Algorithms, Machine Learning, Statistics, Applied Linear Algebra

SKILLS SUMMARY

- **Languages:** Python, Java, SQL, Bash, JavaScript/TypeScript
- **AI/ML & Frameworks:** PyTorch, HuggingFace, TensorFlow, Scikit-learn, NumPy, Pandas, FAISS
- **Distributed Systems & Data:** Feature Engineering, EDA, Statistics, Apache Spark (PySpark, SparkML), Hadoop HDFS
- **Data Streaming & Storage:** Kafka, Parquet, PostgreSQL, MySQL, MongoDB, ChromaDB, Elasticsearch
- **Infra & Tooling:** AWS, Azure, Docker, Kubernetes, Git, MLflow, vLLM, LangChain, FastAPI, CI/CD, Linux

WORK EXPERIENCE

Socially Intelligent Media and Systems (SIMS) Lab - UMass Amherst

Graduate Student Researcher

Amherst, MA

July 2025 - January 2026

- Designed a **3-stage annotation taxonomy** and built an **LLM-based classification pipeline** using **Chain-of-Thought prompting**, achieving **92–98% accuracy** on relevancy detection on **10,000+ social media posts**
- Engineered a **multimodal ETL pipeline** using **OCR** and **Regex parsing** to structure raw data for downstream analysis
- Built a **human-in-the-loop labeling system** that identified a **1.4% estimation bias** over **281 location-specific** data points
- Benchmarked **precision, recall, and F1** across **GPT-4o** and **Gemini 2.5 Flash** on a domain-specific classification task

Affine Technologies

Machine Learning Intern

Bangalore, India

June 2024 - July 2024

- Built and evaluated a paid-service recommendation model across **collaborative filtering, KNN, and SVM** on **Azure** subscriber usage data, delivering a targeting strategy that translated directly into business revenue growth
- Improved model reliability for production use by engineering **login- and usage-frequency features**, lifting non-linear **SVM accuracy from 76.71% to 93.03%** — validated within the 90–95% band needed to trust predictions without overfitting

PROJECTS

Emotional Intelligence Assistant | GPT-4o-mini, LLaMA 3.2, LangChain, ChromaDB

- Engineered a **hybrid RAG + CAG memory architecture** for a therapy-style conversational assistant, combining **retrieval-augmented** session recall with **cache-augmented** static knowledge injection across **fine-tuned GPT-4o-mini** and **LLaMA 3.2 1B** pipelines
- Built a **ChromaDB-backed retrieval pipeline** over **1.2K** preprocessed multi-turn mental health conversations, using **Claude-3** and **OpenAI** summarization to preserve long-term user context across sessions
- Evaluated model variants via **semantic similarity** and **human expert ratings** (empathy, coherence, context awareness), improving **fine-tuned GPT-4o-mini conversation similarity from 63.07% to 67.51%** over baseline

Multi-Agent Climbing Coach | Python, LangGraph, Claude (Haiku, Sonnet)

- Designed a **multi-agent coaching system** end-to-end using a **LangGraph supervisor pattern**, reaching **90.8% routing precision** and **94.2% recall** against a 40-case labeled test set
- Built a **4-part, 77-case evaluation suite** (routing, tool-calling, LLM-judged conflict resolution, safety regression), hitting **96.3% recall** on safety-critical routing cases
- Identified and Resolved **2 escalation gaps** in a 12-case safety regression set, then applied **model tiering** (Haiku for routing, Sonnet for synthesis) to balance cost and quality

Real-Time Traffic Congestion Analytics Pipeline | Python, PySpark, Kafka, Parquet

- Engineered an end-to-end **PySpark Structured Streaming** pipeline that cleaned **1.09M** raw traffic records down to **381,894**, computing real-time congestion metrics usable for live routing decisions
- Benchmarked **6 streaming configurations**, sustaining batches of **5,000 rows** processed in under **0.1s**, then extended with **Kafka-backed** ingestion for fault-tolerant, horizontally scalable use

WORKSHOPS AND PUBLICATIONS

- **Parallelization of the Smith-Waterman Algorithm to Accelerate DNA Sequence Alignment.** 2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT) 🔗
- **Design and Development of an Efficient Risk Prediction Model for Cervical Cancer.** IEEE Access, Vol. 11, pp. 74290–74300 🔗
- **A Multi-Category Content Classification Approach Using DistilBERT Transformer and BiLSTM.** 2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) 🔗