Ishneet Kaur Chadha

Boulder, CO | github.com | ishneet.chadha@colorado.edu | linkedin.com

EDUCATION

University of Colorado - Boulder
Masters of Science in Computer Science *Expected 06/27*

- Coursework: Natural Language Processing, Datacenter Scale Computing, Introduction to Research

Maharaja Surajmal Institute of Technology (GGSIPU)
Bachelor of Technology with minor in AI/ML, CGPA- 9.06 *Graduated 06/25*

- Coursework: Data Structures & Algorithms, Operating Systems, Databases, Computer Networks, Object-Oriented Programming, Artificial Intelligence, Machine Learning, Probability and Statistics

SKILLS

Skills: Python, Java, C/C++ · TensorFlow, PyTorch, Scikit-learn, Keras · NumPy, Pandas, Matplotlib · Machine Learning, Deep Learning, Natural Language Processing, Algorithms, Feature Engineering · Git/GitHub, Linux, Jupyter Notebook

EXPERIENCE

Decision Tree Analytics & Services | ***Data Analyst Intern*** Jan - Jun 2025

- Developed real time analytics dashboards (Tableau) for for supply-chain monitoring, resulting in 25% reduction in delivery delays.
- Automated ETL pipelines and optimized SQL queries for large-scale logistics data, reducing query runtime by 20% and manual processing by 60%.
- Worked with structured/unstructured datasets, applying preprocessing and feature engineering techniques to improve system reliability, reducing missing/duplicate data cases by 15%.

Tech Mahindra | ***AI/ML Intern*** Jul – Aug 2024

- Fine-tuned ML/NLP models to optimize telecom field survey workflows, improving processing efficiency by 10%.
- Deployed speech-to-text and text-to-speech systems using Python, Streamlit, and AWS EC2, reducing manual reporting time by 30%.
- Explored and evaluated Large Language Models for automating survey summarization, increasing report generation speed by 40%.

Defense Research and Development Organization | ***Student Researcher*** Jul – Sep 2023

- Researched and implemented supervised ML models for a Crop Recommendation System, improving crop yield prediction accuracy by 22%.
- Analyzed large-scale agricultural datasets using Python, NumPy, Pandas, Scikit-learn, optimizing model selection and feature engineering.
- Developed a Django-based web application with SQLite for data visualization and experiment tracking, enhancing research accessibility for end users.

PROJECTS

Distributed Stream Processing System | ***Python, FastAPI, gRPC, Kafka, RocksDB, Kubernetes*** Link

∗ Designed and implemented a fault tolerant, exactly once distributed stream processing engine inspired by Apache Flink internals, supporting real-time stateful and stateless pipelines.

∗ Implemented Chandy-Lamport distributed snapshots for exactly-once semantics and built a self-healing control plane using heartbeat-based recovery, restoring failed components in under 60 seconds from S3/GCS-backed checkpoints.

∗ Deployed high-availability workloads on GKE with autoscaling and persistent state, achieving 10-50 ms end-to-end latency for stateless pipelines and 50-200 ms for stateful windowed processing, with full observability via Prometheus and Grafana.

Political QA Clarity Detection (SemEval 2026 Task 6) | ***Python, Scikit-learn, PyTorch, Transformers*** Link

∗ Built an end-to-end NLP system to detect clarity and evasiveness in political interview responses using question–answer pairs from the SemEval shared task dataset.

∗ Formulated the task as multi-class classification (Clear Reply, Ambivalent Reply, Clear Non-Reply), performed exploratory data analysis on class imbalance and answer-length effects, and evaluated lexical baselines alongside transformer-based models using Macro F1.

∗ Established a strong baseline with TF-IDF + Linear SVM, observed superior generalization from RoBERTa on hidden evaluation data, and found limited gains from naive ensembling without careful calibration.