

Charishma Reddy Yerra

Rochester, NY | +1 5853055260 | yerracharishmareddy@gmail.com | [LinkedIn](#) | [GitHub](#)

Summary

Data Science graduate student at RIT focusing on building scalable data pipelines. Hands-on experience designing ETL workflows, including a research project that cleaned and processed over 4 million raw comments into a curated dataset for stance-perception analysis. Proficient in Python, SQL, AWS/GCP, Airflow, and Docker to deliver reliable, high-performance data infrastructure. Eager to apply this background to develop robust data solutions that drive analytics and business insight.

Education

Rochester Institute of Technology | *Master's of Science, Data Science* (GPA: 3.710) **Rochester, NY** Aug 2024 - May 2026
RMK Engineering College | *B.E., Computer Science and Engineering* (GPA: 3.825) **Sep 2020 - May 2024**

Skills

- **Programming:** Python, R, C, Java
- **Visualization:** Matplotlib, Seaborn, Plotly
- **Databases:** SQL (PostgreSQL, MySQL), MongoDB
- **Software:** MS Excel (Advanced Functions)
- **Cloud:** Google Cloud, AWS
- **ML/AI:** Scikit-learn, XGBoost
- **ETL / Pipelines:** Airflow, SSIS
- **Version Control:** Git, GitHub, Docker
- **Statistics:** Hypothesis, A/B Testing, Regression Analysis
- **Big Data:** Spark, Hadoop

Experience

Rochester Institute of Technology

Graduate Research Capstone: LLM Stance Perception for Women's Safety Narratives

Aug 2025 - Present

- Developed an end-to-end research pipeline to process 4.3M+ raw Reddit and YouTube comments into a curated dataset of 351,501 comments across 16 women's-safety cases in India (2012–2024) for large-scale discourse and stance analysis.
- Conducted statistical hypothesis testing (Mann–Whitney U, chi-square, G-tests) to quantify stance distribution differences between politically implicated and non-political cases, revealing statistically significant patterns that guided feature selection for the classification model
- Adapted Qwen3.5-9B using LoRA for multi-class stance classification and evaluated the model on a 3,000-sample held-out set, achieving 0.7447 accuracy, 0.7450 macro-F1, and 0.6634 MCC; performed high-confidence error analysis to identify systematic model failure modes.
- Used RIT research-compute/SLURM workflows for model labeling, LoRA adaptation, validation scoring, and high-confidence error auditing.
- Co-authored the resulting research paper, accepted as a short paper at ASONAM 2026

GCCIS Technical Assistant

Aug 2025 - Present

- Supported faculty, staff, and students by maintaining reliable IT and computing systems across campus, ensuring seamless academic operations.
- Implemented and optimized academic software applications using Java and Docker, improving workflow efficiency and usability for faculty and students.
- Monitored lab resources and analyzed system performance using SQL, publishing short reports to ensure optimal utilization and resource management.

Rochester Institute of Technology – School of Interactive Games and Media

Research Assistant

Jan 2025 - Present

- Conducted data cleaning, integration, and analysis for Project 'Honeycrisp' using R and MongoDB, preparing visualizations that supported key findings and informed project decisions.
- Analyzed historical datasets using A/B Testing to surface patterns that informed instructional media research, leading to improved research methodologies.
- Contributed figures, tables, and summaries for papers and presentations based on quantitative results, enhancing the clarity and impact of research findings.

National Atmospheric and Research Laboratory

HPC Intern

Jun 2023 - Aug 2023

- Developed CUDA-accelerated Gauss–Jordan routines; reduced compute time for large matrices.
- Benchmarked kernels and documented performance insights for HPC workflows, leading to improved understanding and optimization of computational processes.
- Integrated compute improvements with networking team in distributed environments, facilitating enhanced system performance and collaboration.

Projects

SafeSpell – AI-Powered Harassment & Manipulation Detector

- Analyzed large volumes of text data to detect emotionally manipulative and abusive language, focusing on user safety insights.
- Designed NLP pipelines for keyword detection and severity scoring; summarized results into structured, interpretable outputs.
- Developed FastAPI backend (Python) and React/Tailwind frontend with dashboards to visualize flagged content and severity trends.

MindBill – Emotional AI Tracker for Daily Expenses

- Collected, cleaned, and analyzed micro-expense data with user-labeled emotions (regret, happy, neutral) to uncover spending trends.
- Applied machine learning (Logistic Regression, Random Forest) to predict emotional outcomes of purchases, providing actionable insights.
- Built interactive dashboards in React/Tailwind and Python to visualize regret trends, triggers, and emotional spending patterns.