

VIRUM RANKA

734-945-4256 || virum@umich.edu || [linkedin.com/in/virumranka](https://www.linkedin.com/in/virumranka) || [Portfolio Link](#) || [github](#) || Ann Arbor, MI

Master's student in Data Science building scalable software and AI systems

EDUCATION

M.S. Data Science – UNIVERSITY OF MICHIGAN, ANN ARBOR, MI Aug 2025 - Present

B.Tech Computer Science (Data Science Specialization)

UNIVERSITY OF MUMBAI, INDIA | **GPA: 3.89**

Aug 2021 - May 2025

Relevant Coursework: Machine Learning, Deep Learning, Cloud Computing, Natural Language Processing, Time Series Analysis, Big Data Engineering

WORK EXPERIENCE

Data Science & AI Researcher | Department of Anthropology, Michigan Nov 2025 - Present

- Designed end-to-end data preprocessing pipeline for **SAPPHIRE survey data** and developed a two-stage QC algorithm reducing measurement variability by **36-50%** while retaining **90%** of data.
- Engineered feature encoding system converting natural language responses into quantitative metrics, enabling statistical modeling of **353+ variables** reduced to **143-163** and conducted statistical analyses identifying relationships between sun exposure, melanin change and seasonal vitamin D decline.
- Collaborated on developing **Stage-0 evaluation benchmark** for LLM-based data analysis tools, designing **JSON schema** specifications for multi-file ingestion and cross-sheet relationship inference.

Machine Learning Intern | Sprect Pvt Ltd

Jul 2024 - Jan 2025

- Trained and fine-tuned Transformer models (**BERT, RoBERTa**) on **25K+** social media text samples, increasing the **macro-averaged recall score** by **~18%** for multi-class user-intent classification.
- Strategized a **recommendation** system to suggest personalized profile suggestions to new users, reducing the average profile setup time by up to **40%**.
- Built preprocessing pipelines for **data wrangling, tokenization**, automating daily ingestion of new text samples and reducing manual effort by at least **65%**.
- Automated data access for stakeholders with Python scripting, saving **7 hours** of manual effort weekly.

Generative AI Intern | White Turtle Studios, Trailer Park Group

Dec 2023 - Mar 2024

- Synthesized and inpainted images weekly using **Diffusion models**, fulfilling the team's content and asset requests, thus boosting creative team productivity by at least **30%**.
- Surveyed and benchmarked image generation models for **accuracy, inference speed** and recommended cost-effective options that lowered project expenses by nearly **15%**.
- Pioneered voice cloning initiative using Tortoise TTS achieving a **Mean Opinion Score(MOS)** increase of **0.8**, as the most junior member on the team, leading to improvements in perceived audio quality.

SKILLS & CERTIFICATIONS

Languages : Python, C/C++, Java, SQL **Data Science :** Pandas, Numpy, Matplotlib, SciPy, PySpark

Machine Learning: PyTorch, TensorFlow, , Scikit-learn, HuggingFace, LangChain, vLLM, OpenCV

Automation and Deployment:: AWS, Docker, Git, FastAPI, BS4, OpenAI API, AI Agents, CrewAI

Certifications: [AWS Cloud Foundations](#), [AWS Data Engineering](#)

SELECTED PROJECTS & RESEARCH

LLM-Powered Knowledge Retrieval System for Research Assistant

May 2024 - Jul 2024

- Developed a **RAG** pipeline using **LangGraph** to deliver grounded answers from research manuscripts.
- Integrated **open-weight LLM** with context retrieval and citation tracking using LoRA for efficient tuning.

EEG-based Machine Learning model for Autism detection [Link](#)

Jun 2024 - Feb 2025

- Designed a two-stage deep learning system using **CNN-LSTM** on EEG signals from 34 subjects.
- Achieved **91%** recall in autism detection and applied **unsupervised clustering** for severity analysis.