# Sindhura Kommu

+1 (779) 276 2823 | sindhura@vt.edu | https://sindhura-cs.github.io/ | linkedin.com/in/sindhura1997

github.com/sindhura-cs | Google Scholar

## Summary

Experienced researcher with **4.5+ years of industry experience**. Currently pursuing my thesis on single-cell **pre-trained foundation models** for genomic sequences. Expertise in developing **novel AI architectures** that integrate **transformer-based encoders** with **external domain knowledge**, with an emphasis on **multimodal AI, LLMs, and NLP**. Worked with **cross-functional teams** to conceptualize and deliver **big data solutions** as a Senior Data Scientist at a global AI company.

**Interest Areas:** Large Language Models (LLMs), Natural Language Processing (NLP), Multimodal AI

## Technical Skills

**Languages & Frameworks:** Python, C++, Java, Matlab, SQL, JavaScript, HTML, CSS, Neo4j, Docker, Kubernetes, Spark

**Deep Learning:** PyTorch, GPU, CUDA, NumPy, Pandas, Scikit-Learn, OpenCV, TensorFlow, Keras, Scipy, anndata

## Experience

**Research Assistant**, Virginia Tech – Blacksburg, USA                          Fall 2023 – ongoing

- Developed a novel **knowledge-aware fusion framework**, combining large-scale single-cell **pre-trained language models (PLMs)** and **graph encoders** for **causal network inference** by integrating prior biomedical knowledge.
- Achieved impressive performance improvements in network inference tasks, **outperforming state-of-the-art** by **5.4% in AUROC** score and **7.4% in AUPRC** score. Research paper accepted at **ICML 2024 workshop**.
- Currently training **network-regulated Large Language Models (LLMs)** for **multi-omics data** analysis. Secured a **prestigious award** from the **NSF NAIRR Pilot** program to support the training of these **single-cell foundation models**.

**Senior Data Scientist**, Fractal Analytics – London, UK                          Feb 2020 – Dec 2022

- Architected **distributed ML models** to analyze **viewing patterns** of over **15 million customers** across **3+ billion hours of content** from Sky Group's top TV channels using advanced **time series analysis** and **collaborative filtering** techniques.
- Reported actionable insights on customer churn, content preferences, and viewing habits, resulting in a **60% reduction in customer acquisition costs** and a **25% increase in viewer retention**. Tools used: **Apache Spark, Airflow, GCP, PyTorch**.
- Engineered a **scalable multimodal AI**-driven early warning system for Visa Europe. Leveraged an innovative fusion architecture that **integrated NLP-based insights** from news feeds and social media with **structured financial data**. Achieved **85% accuracy** in predicting merchant bankruptcies, surpassing traditional single-modality approaches.

**Data Scientist**, Fractal Analytics – Mumbai, India                          June 2018 – Jan 2020

- Developed and **fine-tuned CNN models** for **object detection** and **segmentation** using **YOLOv3** and **Fast-RCNN**.
- Applied these models to **classify P&G products** at the **SKU level** from **noisy retail shelf images**. Achieved over **90% validation accuracy** on **real-world datasets**, optimizing product identification and improving retail inventory management.
- Enhanced **AI-powered video surveillance** system for a leading drone manufacturer by automating the analysis of video footage for **threat detection**. Utilized **transfer learning** to adapt pre-trained models like **MMDetection** and **Detectron2**.
- The solution **increased precision by 30% outperforming human operators**, significantly reducing security costs.

**Software Intern**, Samsung R&D Institute – Bangalore, India                          Summer 2017

- Built usability extension in Tizen Browser for mirroring and hooking website regions, reducing scrolling by 15%.

## Education

**Virginia Tech**, M.S. (thesis) in Computer Science, **CGPA**: 4.0/4.0                          Jan '23 – May '25 (expected)

- **Advisor**: Dr. Xuan Wang, **Dept Scholarship:** 100% Tuition waiver with funding guarantee
- **Teaching Assistant:** Machine Learning, Intro to AI, Intro to Prog in Python, Computational Biology

**Indian Institute of Technology (IIT), Bombay**, B.Tech in Computer Science                          2014 – 2018

- **Teaching Assistant:** Computer Programming & Utilization (Spring 2018)

## Publications

- **Sindhura Kommu**, Yizhi Wang, Yue Wang, and Xuan Wang, "Gene Regulatory Network Inference from Pre-trained Single-Cell Transcriptomics Transformer with Joint Graph Learning", The Forty-first International Conference on Machine Learning AI for Science Workshop (**ICML'24 AI4Science**), July 26, 2024, Vienna, Austria
- **Sindhura Kommu**[*], Zhenyu Bi[*], Sajib Acharjee Dip[*], Daniel Hajialigol[*], Hanwen Liu[*], Meng Lu[*], Xuan Wang, "AI for Biomedicine in the Era of Large Language Models", arXiv preprint arXiv:2403.15673 (2024)