JOHNATHAN JIA, M.S.

HOUSTON, TX · JJIA.VERCEL.APP · JDJIA93@GMAIL.COM · GITHUB.COM/JJIA1

SKILLS

Program Languages
Python, R, BASH

Machine Learning Frameworks and Tools
PyTorch, TensorFlow, CUDA, LangChain, Transformers, FAISS, SHAP, LIME, RAG, Prompt Engineering, Neural Networks

Data Analysis and Visualization
Numpy, Pandas, Polars, Matplotlib, Seaborn, tidyverse, ggplot2, dplyr, Scanpy, Seurat

Bioinformatics and Workflow Tools
NextFlow, Docker, Singularity, Bioconductor

Cloud Computing and Infrastructure
Microsoft Azure, AWS, Kubernetes, Slurm, OpenMPI

Version Control and Documentation
Git, LaTex

EXPERIENCE

2024 - Current   **AI Engineer | Baylor College of Medicine**
- Initiated development of a novel machine learning model for classifying esophageal cancer endoscopic images, addressing a previously unmet need in automated diagnostics.
- Built a lightweight foundation model for automated annotation of Drosophila single-cell data, outperforming human-based methods and traditional machine learning models.
- Designed and implemented an automated pipeline for single-cell omics data analysis using Python, R, and NextFlow, significantly streamlining workflows (50% time reduction) by replacing manual processes with fully automated, standardized solutions.
- Developed an explainable AI module for AI-MARRVEL, a rare disease prediction model, integrating SHAP and LIME to provide insights into model decision-making. Optimized the module by parallelizing explainability processes, reducing runtime by 80%, and enhancing performance further with XGBoost integration.
- Led bioinformatics team in collaborations with over 4 labs, delivering high-quality, timely results to advance research objectives. Provided consulting and technical expertise to team members, collaborating labs, and broader organizational stakeholders to enhance workflows and drive efficiency.
- Cultivated a collaborative and results-driven work environment by organizing and leading weekly team meetings and knowledge-sharing sessions. Designed and implemented structured hiring processes, including the creation of a proficiency assessment, to enhance talent acquisition and evaluation.

2023 - 2024   **Data Analyst Associate | Baylor College of Medicine**
- Conducted single-cell RNA sequencing data analysis across 10+ collaborative projects, encompassing over 30 human and mouse samples. Delivered critical insights that supported experimental objectives, with recent findings approved for publication in Frontiers in Neuroscience.
- Developed a matching search engine and interactive chat interface for students and faculty using LLMs and Retrieval-Augmented Generation (RAG), improving efficiency by 80%.

2020 - 2022   **Graduate Research Assistant | UTHSC Graduate School of Biomedical Sciences**
- Designed a CNN with self-attention in TensorFlow/Keras to identify transcription factor targets for viruses, increasing accuracy by 23% (published in *Patterns*).
- Executed bioinformatic analysis on single-cell RNA data to uncover transcriptional programs driving COVID-19 severity, contributing to a publication in *Genes*.

2018 - 2019   **Research Assistant I | MD Anderson Cancer Center, Department of Immunology**
- Performed sample processing for immunological research studies, supporting downstream analysis workflows.

EDUCATION

2020 - 2022   **M.Sc. Quantitative Sciences, UTHSC Graduate School of Biomedical Sciences**
2022, Best Seminar Presentation, Quantitative Sciences Program
2022, Best Poster Presentation, Quantitative Sciences Program

2015 - 2018   **UTHSC McGovern School of Medicine**
2016, Rheumatology Research Foundation Research Award

2011 - 2015   **B.S. Biology, Emory University**

MISCELLANEOUS

Certifications
2024   Enhancing Data Science Outcomes with Efficient Workflow, NVIDIA

Languages
English (Native), Mandarin Chinese (Professional Working Proficiency)

SELECTED PUBLICATIONS

1. Belfort B, **Jia J***. (2025). Comparative Analysis of AAV Serotypes for Transduction of Olfactory Sensory Neurons. *Frontiers in Neuroscience* [Accepted for Publication].
2. Xu, H., **Jia, J.***, Jeong, H., Zhao, Z. (2023). Deep learning for detecting and elucidating human T-cell leukemia. *Cell Patterns*.
3. Jeong, H., **Jia, J.**, Dai, Y., Simon, L. (2021). Investigating cellular trajectories in COVID-19 severity. *Genes*.