JOHNATHAN JIA, M.S.

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

SKILLS

Programming Languages
Python, R, Bash, SQL

Machine Learning Frameworks and Tools
PyTorch, TensorFlow/Keras, scikit-learn, CUDA, LangChain, Hugging Face Transformers, SHAP, LIME, RAG, Neural Networks

Statistical and Data Science Methods
Hypothesis & A/B Testing, Regression Modeling, Dimensionality Reduction, Model Evaluation & Cross-validation, Bayesian Optimization

Data Analysis and Visualization
NumPy, Pandas, SciPy, Polars, Matplotlib, Seaborn, tidyverse, ggplot2, dplyr

Workflow and Containerization
NextFlow, Docker, Singularity, Bioconductor

Cloud and HPC
Azure, AWS, Kubernetes, Slurm, OpenMPI

Version Control and Documentation
Git, GitHub, Markdown, LaTex, Jupyter

EXPERIENCE

2024 - Current   **AI Engineer | Baylor College of Medicine**

- Developed a graph-based multi-label classifier for histopathology images, boosting precision and recall by 20% over CNN baselines.
- Built a foundation model for Drosophila scRNA-seq annotation, batch correction, and in silico analysis, automating manual drug-discovery workflows.
- Created a Nextflow-powered Python/R pipeline for single-cell omics that reduced processing time by 50% through standardized, end-to-end automation.
- Implemented explainable AI for AI-MARRVEL, a rare disease prediction model, with SHAP and LIME to provide insights into model decision-making. Optimized the module by parallelizing explainability processes, reducing runtime by 80%.
- Led weekly team meetings and knowledge-sharing sessions; designed a proficiency assessment to streamline hiring and boost team productivity.

2023 - 2024   **Data Analyst Associate | Baylor College of Medicine**

- Built an LLM-powered search and chat interface using FAISS and LangChain to match students and faculty; deployed at the TCH Research Symposium.
- Analyzed 30+ single-cell omics samples across 10+ collaborative projects, generating insights adopted for publication in *Frontiers in Neuroscience*.
- Directed a bioinformatics team across 4+ collaborating groups, standardizing workflows and providing technical consulting to accelerate research deliverables.

2020 - 2022   **Graduate Research Assistant | UTHSC Graduate School of Biomedical Sciences**

- Developed a self-attention CNN in TensorFlow/Keras to predict viral transcription factor targets, boosting accuracy by 23% (published in *Patterns*).
- Performed single-cell RNA-seq analysis to delineate transcriptional programs driving COVID-19 severity, with findings published in *Genes*.

2018 - 2019   **Research Assistant I | MD Anderson Cancer Center, Department of Immunology**

- Prepared and QC-checked immunology assay samples, maintained SOP documentation, and supported downstream bioinformatics pipelines.

EDUCATION & COURSEWORK

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   **Medical School Coursework, UTHSC McGovern School of Medicine (no degree awarded)**
2016, Rheumatology Research Foundation Research Award

2011 - 2015   **B.S. Biology, B.A, 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*.