# JOHNATHAN JIA: SENIOR DATA ANALYST

I am a senior data analyst with experience in data science and machine learning. My interests are in NLP/LLMs, computer vision, multi-omics, and graphs.

# PROFESSIONAL EXPERIENCE

# 2023 Data Analyst Associate

**Baylor College of Medicine** 

Houston, TX

- Partnered with team to develop and deploy automated NGS analysis pipeline on local servers with Docker, Python, R, and Bash.
- Worked with software engineers and clinical geneticists to fine-tune and prompt engineer several open-source LLMs for precision diagnosis of rare Mendelian genetic disorders. I applied inference optimization techniques such as mini-batching and output clipping to greatly reduce the runtime of model inference.
- Collaborated with lab members via GitHub to develop the Python back-end for a mentormentee matching tool using GPT-4 (via API) and RAG (FAISS) for leaders of clinical departments. We deployed the tool using Microsoft Azure and delivered a working product demo with a JS front-end for multiple clinical department heads.
- Lead bioinformatics team in analyzing collaborator data and delivering high-quality results in timely fashion. Regularly provided consulting and technical expertise to team and other group members to improve overall group efficiency.
- Fostered collaborative work environment by regularly leading weekly lab meetings and journal clubs. Implemented structured interview process including proficiency exam for incoming analysts, filling vacancies with effective candidates.

2022 | 2020

## **Graduate Research Assistant**

UTHSC Houston Graduate School of Biomedical Sciences

Houston, TX

- Partnered with lab members to build a convolutional neural network with self-attention architecture in Python using Tensorflow/Keras for finding high-confidence transcription factor targets for viruses. Improved performance using a downsample-bootstrapping training method and 10-fold cross validation. Project was published in Cell Patterns and the code is publicly available on Github.
- Worked with team to run trajectory inference and pseudotime analysis using R packages Seurat and Slinghost on scRNAseq data to identify which transcription factor programs drive COVID-19 infection severity with the machine learning algorithms. Our work was published in Genes and the code is publicly available on Github.
- Fine-tuned ProteinBERT Transformer into a pan-virus MHC-epitope classification model for high throughput identification of vaccine targets for viruses
- Regularly lead group meetings and provided results to PI in timely manner. Helped foster a collaborative lab environment by helping to identify and solve problems encountered by all team members.

2020 | 2019

2022

#### Research Assistant I

MD Anderson Cancer Center

Houston, TX

# **EDUCATION**

# M.S. in Quantitative Sciences

UTHSC Houston Graduate School of Biomedical Sciences 

♥ Houston, TX

UTHSC Houston McGovern School of Medicine Phouston, TX

2015 2015

2011

2022

2018

#### B.S. in Biology, B.A. in Chemistry

Emory University

Atlanta, GA

#### **AWARDS**

#### 2022 • Best Quantitative Sciences Program Seminar Presentation

UTHSC Houston Graduate School of Biomedical Sciences 

♥ Houston, TX

#### First Prize Poster Presentation

UTHSC Houston Graduate School of Biomedical Sciences 

Phouston, TX

016 Rheumatology Research Foundation Research Award

UTHSC Houston McGovern School of Medicine

Houston, TX

# **CONTACT INFO**

github.com/jjia1

>\_ jjia-vscodeportfolio.vercel.app

# PROGRAMMING LANGUAGES

Python

R
Bash/Shell
LaTex
Next.js

# **CORE SKILLS**

Docker
NLP
LLM
LoRA
RAG
Prompt Engineering
Microsoft Azure
AWS
Kubernetes

# TOOLS/FRAMEWORKS

Tensorflow
PyTorch
Pandas
Numpy
CUDA
LangChain
Transformers
Matplotlib
FAISS

## **PUBLICATIONS**

Deep learning for detecting and elucidating human T-cell leukemia v... Cell Patterns (2023) H. Xu\* and J.Jia\*, H.Jeong, Z.Zhao Delineating COVID-19 immunological features using single-cell RNA s... The Innovation (2022) W.Liu, J.Jia, Y.Dai Investigating Cellular Trajectories in the Severity of COVID-19 and... Genes (2021) H.Jeong\* and J.Jia\*, Y.Dai, L.Simon What Controls the Acute Viral Infection Following Yellow Fever

### CERTIFICATES

Enhancing Data Science Outcomes With Efficient Workflow NVIDIA (2024)

Last updated on 2024-07-05.

Vacc... Bul. of Math. Bio. (2016) J.Moore, H.Ahmed, J.Jia