SHIYI (DENNIS) HU

+1 201-630-1978 | sh5701@nyu.edu | Website: https://dennis-who.github.io/

EDUCATION AND HONORS

New York University New York, NY
M.S. in Data Science | Cumulative GPA: **4.00**/4.00 Sep 2023 – May 2025
Relevant Coursework: Machine Learning, Big Data, Advanced Econometrics, Natural Language Processing, Probability and Statistics, Optimization and Computational Linear Algebra, Databases, Deep Learning, Information Visualization

New York University New York, NY
B.S. in Data Science, B.S. in Economics | Cumulative GPA: **3.93**/4.00 (**Summa Cum Laude**) Sep 2019 – May 2023

- Peng Yachao Global Excellence Scholarship ($17,500 awarded to top 2% of class)
- NYU Shanghai 2022 Recognition Award; Dean's List all eligible years

PROFESSIONAL EXPERIENCE

Academic Analytics, New York University New York, NY
Academic Assessment and Analytics Graduate Assistant (Tableau, Pandas, SQL) *Sep 2023 – Present*

- Led building the **ETL** pipeline of 300,000+ student records by employing **data warehousing** techniques and database management tools such as **Postgre SQL** and **Snowflake.**
- Improved efficiency of institutional research by 20% with **Tableau** Prep Builder, establishing a sustainable school-wide reporting infrastructure, managing access for 35 academic departments to access 24/7.
- Collaborated with the Academic Affairs Director to create sophisticated **data visualizations** using **Tableau** and **R**, effectively communicating complex findings to diverse audiences.

Craft Associates Hong Kong, China
Data Scientist Intern (Web crawling, Python, Excel, NLP sentiment analysis) *Jul 2022 – Jan 2023*

- Conducted **sentiment analysis** for 8 product releases, utilizing **Python** requests to extract and evaluate sentiment from over 100,000 web posts. This analysis contributed to a 20% increase in market PR tagging accuracy.
- Developed and optimized data pipelines, processing **over 20 TB of Meltwater data** to update author profiles, resulting in a 30% reduction in data maintenance time.
- Led the creation of more than 30 monthly market **insight presentations** for key tech clients (Qualcomm, MediaTek), delivering actionable insights contributing to a 10% increase in client satisfaction.

Applied Artificial Intelligence Center, Shanghai General Hospital Shanghai, China
Research Scientist Intern (Deep learning, Data ethics, Bioinformatics) *May 2021 – Nov 2021*

- Trained a large neural network on empirical patient data for arrhythmia detection, achieving an **F1 score of 0.813**.
- Enhanced diagnostic accuracy by 2.8% through the implementation of **leaky ReLU** activation
- Applied Grad-CAM techniques to enhance transparency and **interpretability** of **Convolutional Neural Network** (CNN) models, resulting in a 15% increase in user trust feedback.
- Delivered impactful insights through two comprehensive 3,000-word reports, addressing **ethical and social implications** of the project during the quarter conference, initiating first draft of institute policy guideline.

RESEARCH & PROJECT EXPERIENCE

Full-stack Data Engineering for Music Aphantsia Therapy New York, NY
NYU Fox Lab, Prof. Pascal Wallisch *Sep 2023 – Present*

- Developed a data storage pipeline enabling synchronous updates from 50,000+ MatLab entries
- Implemented a latent-factor based imputation algorithm for song data, turning sparse feedback data into dense matrices on which feature selection is performed
- Improved prediction accuracy by 5.8% over the vanilla collaborative filtering prediction with optimized model

Popularity Bias and Fairness in Recommendation Systems Shanghai, China
Capstone Project, Prof. Hongyi Wen *Mar 2023 – May 2023*

- Created a **nearest-neighbor-based hybrid continuation** model for personalized music playlist continuation problems
- Implemented user-specified reranker and BFGS optimization reranker, reducing **popularity bias** by 40%
- Enhanced user specificity in item-based **recommendation systems**, resulting in a 13% increase in user rating.

Factual Generalization Capabilities of GPT-3 Across Domains New York, NY
Center for Data Science, Prof. Samuel Bowman *Mar 2022 – Jun 2022*

- Cross-examined the falsehood patterns of large language models in six domains and explored truth-inducive prompts
- **Crowdsourced** 2000+ question-answer pairs to generate 200,000+ prompts with effectiveness evaluated on **GPT-3**
- Performed transfer learning on 36,000 samples to find that GPT-3 produces 43% less falsehoods with 5 more prompts

TECHNICAL SKILLS

- **Programming skills**: Python (Numpy, Pandas, Flask, Pytorch, Matplotlib), R, HTML, Bash, CSS
- **Tools and Frameworks:** ETL, Data Visualization, Excel, EDA, Linux, Git version control, Vim, Scikit-learn
- **Database:** SQL, Hadoop, Dask, Spark, Map-Reduce, MongoDB, AWS, Oracle
- **Machine Learning:** Deep learning, NLP, GNN, CNN, transformers, embeddings, GBMs
- **Data Visualization:** React, Tableau, PowerBI, Node.js