CHENYU (MONICA) WANG

https://chenyuwang-monica.github.io | wangchy@mit.edu | (+1)617-902-8630
MIT Stata Center, 32 Vassar St. G418; Broad Institute, 75 Ames St. M1108, Cambridge MA 02139

EDUCATION BACKGROUND

Massachusetts Institute of Technology **Cambridge, MA**
Ph.D. Student in Electrical Engineering and Computer Science (EECS) | GPA 5.0/5.0 Aug. 2022-present
Advised by Prof. Tommi Jaakkola and Prof. Caroline Uhler

Tsinghua University **Beijing, China**
Bachelor of Economics, Minor in Data Science and Technology | GPA 3.99/4.0 (Ranking 1/192) Sep. 2018-Jun. 2022
Advised by Prof. Mingsheng Long, Prof. Mengdi Wang, and Prof. Cyrus Shahabi

University of California, Berkeley **Berkeley, CA**
Exchange Student, Department of Statistics (Instructed by Prof. Noureddine El Karoui) | GPA 4.0/4.0 Jan. 2021-Jun. 2021

RESEARCH INTERESTS

My research interests lie broadly in machine learning, representation learning, and AI for science. Recently my research focuses on multi-modal representation learning and perturbation modelling for drug discovery. I am also interested in foundation models for science and spatial-temporal modelling in system biology.

PUBLICATIONS & PREPRINTS

(: Equal Contribution)* Google Scholar

- Hannes Stark*, Bowen Jing*, **Chenyu Wang**, Gabriele Corso, Bonnie Berger, Regina Barzilay, Tommi Jaakkola. Dirichlet Flow Matching with Applications to DNA Sequence Design. In *International Conference on Machine Learning, **ICML 2024***. [link]
- **Chenyu Wang**, Sharut Gupta, Caroline Uhler, Tommi S. Jaakkola. Removing Biases from Molecular Representations via Information Maximization. In *International Conference on Learning Representations, **ICLR 2024***. [link]
- **Chenyu Wang***, Joseph Kim*, Le Cong, Mengdi Wang. Neural Bandits for Protein Sequence Optimization. In *56th Annual Conference on Information Sciences and Systems, **CISS 2022***. [link]
- **Chenyu Wang***, Zongyu Lin*, Xiaochen Yang, Jiao Sun, Mingxuan Yue, Cyrus Shahabi. HAGEN: Homophily-Aware Graph Convolutional Recurrent Network for Crime Forecasting. In *AAAI Conference on Artificial Intelligence, **AAAI 2022***. **(Oral Presentation.)** [link]
- Yang Shu*, Zhangjie Cao*, **Chenyu Wang**, Jianmin Wang, Mingsheng Long. Open Domain Generalization with Domain-augmented Meta-learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, **CVPR 2021***. [link]

RESEARCH EXPERIENCE

Removing Biases from Molecular Representations via Information Maximization **Cambridge, MA**
Advised by Prof. Tommi Jaakkola and Prof. Caroline Uhler, MIT EECS Aug. 2022-present

- Proposed InfoCORE to mitigate the confounding factors in multimodal molecular representation learning from multiple information sources, in particular the confounding batch effects in high-content drug screening data.
- Theoretically, InfoCORE maximizes the variational lower bound on the conditional mutual information of the representation given the batch identifier. It empirically outperforms various baselines on multiple downstream tasks.

Tree-Based Neural Bandits for High-Value Protein Design **Princeton, NJ**
Advised by Prof. Mengdi Wang, Department of Electrical Engineering, Princeton University Jun. 2021-Dec. 2021

- Proposed an MCTS-guided neural contextual bandits algorithm that utilizes a modified upper-confidence bound algorithm as guided by neural bandit and the Monte Carlo tree search process for accelerating the search for optimal protein designs.
- This approach finds a diverse and rich class of high fitness proteins using substantially fewer design queries.

Homophily-Aware Graph Convolutional Recurrent Network for Crime Forecasting **Los Angeles, CA**

Advised by Prof. Cyrus Shahabi, Department of Computer Science, USC Jan. 2021-Jun. 2021

- Presented a graph convolutional recurrent network with a novel homophily-aware graph learning module for crime forecasting.
- Utilized adaptive learning graph structure to capture the underlying high-order relationship between regions; constrained graph structure by designing homophily-aware loss to enhance the performance of graph neural network.

Open Domain Generalization with Domain-Augmented Meta-Learning **Beijing, China**

Advised by Prof. Mingsheng Long, School of Software, Tsinghua University Sept. 2020-Nov. 2020

- Utilized different ensemble model-based criteria including entropy, consistency, and cosine distance from class center to conduct outlier label recognition; introduced clustering loss into loss function to facilitate open-set recognition.
- Evaluated model performance with metrics including H-score and class average accuracy to guide parameter grid search.

Understanding Chinese Bond Yield Curve: Excess Return Prediction **Beijing, China**

Advised by Prof. Hao Wang, SEM, Tsinghua Jun. 2020-Aug. 2020

- Contributed one chapter in the book *Analyzing the Chinese Yield Curve, Hao Wang et al. (2021), Tsinghua University Press*.

HONORS & AWARDS

- **MIT EECS Great Educators Fellowship**, 2022
- **Outstanding Undergraduate in Tsinghua** (2% in Tsinghua), 2022
- Outstanding Undergraduate in Beijing, 2022
- **Chen Daisun Schorlarship** (3 in Tsinghua SEM), 2022
- Undergraduate Commencement Student Speaker of Tsinghua SEM, 2022
- Meritorious Winner in MCM/ICM Mathematical Contest in Modelling, 2021
- Chen Xiaoyue Scholarship, 2021
- Tang Lixin Scholarship (50 in Tsinghua), 2020
- **National Scholarship** (0.2% in China), 2019
- Athletics Excellence Scholarship of Tsinghua, 2019
- First Class Scholarship for Freshmen of Tsinghua, 2018
- **Gold medalist of 50th International Chemistry Olympiad** (4 in China, 6th place in the world), 2018
- Silver medalist of 15th China Girl's Mathematical Olympiad (50 in China), 2022

WORK EXPERIENCE

Jane Street Asia Limited **Hong Kong**

Quantitative Trading Intern (Return offer extended) Jun. 2021-Sept. 2021

- Produced predictive models for future market returns in two research projects; conducted mock trading.

WizardQuant Capital Management **Zhuhai, China**

Quantitative Research Intern, Quantitative Research Department Jun. 2020-Aug. 2020

- Built an alternative risk model based on equity research reports data supplemented to Barra model factors.

Techsharpe Quant Capital Management **Beijing, China**

Data Analyst Intern, Trading Department Jan. 2020-Feb. 2020

- Conducted research on futures rolling strategies of CSI500 index future and analyzed the advantage of rolling by open interest.

LEADERSHIP & ACTIVITIES

- Team Leader, Meritorious Winner in 2021 MCM/ICM Mathematical Contest in Modelling. Feb. 2021
- Co-president, Banking & Investment Mentor Program (A 10-year global non-profit organization). Feb. 2021-Feb. 2022
- Director of Department of Sports, Student Union of Tsinghua SEM. Mar. 2019-Sept. 2020

SKILLS & INTERESTS

- **Languages:** English (Proficient; TOEFL: 110/120); Mandarin (Native)
- **Technical Skills:** Python/C++/Matlab; Deep learning framework: PyTorch, Tensorflow; Basic knowledge of SQL and Linux.
- **Interests:** Sports (1st place in 4*400m; member of SEM basketball and soccer team), Chinese Zither (Amateur Certificate 9)