

YUNZHEN FENG

✉ yf2231@nyu.edu · 📞 917-8917732 · Homepage · Google Scholar

Throughout my PhD journey, I have transitioned from a focus on ML theory to its applied ML domains. This shift has equipped me with a unique blend of theoretical expertise and practical experience, positioning me to contribute innovative solutions and profound understandings to empirical challenges. Currently, my interests are centered around large language models (reasoning ability), data efficient learning, and multimodal learning.

EDUCATION

New York University (NYU), New York, U.S.

Sep'21 - present

- Ph.D. Student at the Center for Data Science.
- Advisor: Julia Kempe.
- Committee: Andrew Gordon Wilson, Qi Lei.

École Normale Supérieure de Paris, Paris, Fr

Jan'24 - present

- Visiting student at the Centre Sciences des Données.

Peking University (PKU), Beijing, China

Aug'17 - June'21

- B.S. (honor track) in Applied Mathematics.
- Advisor: Bin Dong, Di He.

PUBLICATIONS

1. **Yunzhen Feng***, Elvis Dohmatob*, Pu Yang, Francois Charton, Julia Kempe. A Tale of Tails: Model Collapse as a Change of Scaling Laws. *In submission*.
2. Elvis Dohmatob, **Yunzhen Feng**, Julia Kempe. Model Collapse Demystified: The Case of Regression. *In submission*.
3. Kai Yang, Jan Ackermann, Zhenyu He, Guhao Feng, Bohang Zhang, **Yunzhen Feng**, Qiwei Ye, Di He, Liwei Wang. Do Efficient Transformers Really Save Computation? *In submission*.
4. **Yunzhen Feng**, Ramakrishna Vedantam, Julia Kempe. Embarrassingly Simple Dataset Distillation. *In the Twelfth International Conference on Learning Representations (ICLR), 2024*.
5. **Yunzhen Feng**, Nikos Tsilivis, Tim G. J. Rudner, Julia Kempe. Attacking Bayes: Are Bayesian Neural Networks Inherently Robust? *In Fifth Symposium on Advances in Approximate Bayesian Inference (AABI), 2023*.
6. Chizhou Liu, **Yunzhen Feng**, Ranran Wang, Bin Dong. Enhancing Certified Robustness of Smoothed Classifiers via Weighted Model Ensembling. *ICML 2021 workshop on Adversarial Machine Learning, 2021*.
7. **Yunzhen Feng***, Runtian Zhai*, Di He, Liwei Wang, Bin Dong. Transferred Discrepancy: Quantifying the Difference Between Representations. *arXiv:2007.12446*.
8. **Yunzhen Feng**, Yue M. Lu. A Precise High-dimensional Analysis of Laplacian Semi-Supervised Learning, 2020.

RESEARCH EXPERIENCE

Math and Data group, Center for Data Science, NYU

Sep'21 - present

Research Assistant

New York, U.S.

- Advisor: Julia Kempe.
- Efficient Data Curation Frameworks: Develop and propose a simple yet effective method for dataset distillation, achieving state-of-the-art results across numerous benchmarks. Introduce a task-agnostic criterion for data selection, enhancing the fine-tuning efficiency of large-scale vision models.
- Understand the reasoning ability of transformers, in the perspective of chain of thought and in-context learning.

Marketplace Investment Team, Uber Technologies May'22 - Aug'22
Applied Scientist PhD Intern *California, U.S.*

- Mentor: Jason Dowlatabadi, Chen Xu, Allen Zhang.
- Team Advisor: Stefan Wager, Peter Frazier.
- Design doubly robust CausalML model for learning heterogeneous incentive response across cities with both observational data and experimental data.

Beijing International Center for Mathematical Research, PKU Jan'20 - Jul'21
Research Assistant *Beijing, China*

- Advisor: Bin Dong.
- Proposed an ensembling framework for certified robustness with improved robustness, less training time, and theoretical guarantees.
- Proved the existence of pure Nash Equilibrium in adversarial robustness and combined game-theoretic algorithms with adversarial training.

School of Engineering and Applied Science, Harvard University Aug'20 - Dec'20
Research Assistant *Remote*

- Advisor: Yue M. Lu.
- Investigated how unlabeled data help the classification for Gaussian mixtures and Laplacian regularization using high dimensional statistics.

Microsoft Research Asia Jan'20 - June'20
Visiting *Beijing, China*

- Advisor: Di He.
- Proposed to measure the difference between two networks' representation via their performances on a set of downstream tasks and investigate the effect of random initialization on the learned features.

OTHER EXPERIENCE

Ubiquant Investment Co., Ltd. Jul'21 - Sep'21
Quant Research Intern *Beijing, China*

- Deep learning for the stock market.
- Retrieved core patterns with designed pretrain tasks and improved IC with diverse ensembling.

TEACHING

Head TA for **Modern Topics in Statistical Learning Theory** 2023 Spring

SELECTED HONORS AND AWARDS

National Science Foundation Research Traineeship (NRT) Future Program	2021
Data Science Fellowship (\$ 190,000), New York University	2021
Outstanding Graduate Student, Peking University	2021
<i>Bronze Medal</i> , S.-T.Yau College Student Mathematics Contests (top 20 in China)	2020
<i>Meritorious Award</i> , Mathematical Contest in Modeling (top 9%)	2019
The Elite Undergraduate Training Program of Applied Mathematics (top 10%)	2018
<i>Gold Medal</i> , China Mathematics Olympiad (top 100 in China)	2016
<i>Ranked First in First Prize</i> , Chinese High School Mathematics League, Beijing (top 1)	2016

SPECIALIZED SKILLS

- **Programming:** Python (PyTorch), MATLAB, \LaTeX .
- **Language:** English (Fluent), Mandarin (Fluent), French (Beginner).

ACTIVITIES AND INTERESTS

- Successfully climbed the Luodui Mount (6010m) at Tibet, China.
- Rock-climber and outdoor hiker trained by committees from Chinese Mountaineering Association.
- One-year volunteer at Arthur M. Sackler Museum of Art and Archaeology, Peking University.
- Teaching Assistant for the *Outdoor Exploration* course at PKU.