

Zhili Feng

zhilif@andrew.cmu.edu

- EDUCATION**
- Ph.D. in Machine Learning** Aug 2019 - Present
Carnegie Mellon University
Advised by J. Zico Kolter
- Master of Science in Computer Science** Sept 2017 - May 2019
University of Wisconsin-Madison
Advised by Po-Ling Loh
Thesis: *On the Geometric and Statistical Interpretation of Data Augmentation*
- Bachelor of Science in Computer Science** Aug 2013 - May 2017
University of Illinois at Urbana-Champaign
Advised by Dan Roth
Thesis: *A Combination of Statistical Model and Rule-based Approach to Temporal Extraction and Normalization*
- RESEARCH**
- RESEARCH EMPLOYMENT**
- Research Intern** Jun 2024 - Aug 2024
Microsoft Research, Boston, MA
Hosts: Lester Mackey, David Alvarez-Melis, Nicolo Fusi
- Applied Scientist Intern** Jun 2023 - Aug 2023
Amazon, New York, NY
Hosts: Dhananjay Ram, Cole Hawkins
- Machine Learning Research Intern** Jun 2021 - Aug 2021
Bosch, Pittsburgh, PA
Hosts: Devin Willmott, J. Zico Kolter
- Research Intern** Sept 2020 - Dec 2020
NEC Lab America, remote
Hosts: Shaobo Han, Simon Shaolei Du
- PUBLICATION**
- * indicates equal contribution.
1. Yiding Jiang, Allan Zhou, Zhili Feng, Sadhika Malladi, and J. Zico Kolter. ADAPTIVE DATA OPTIMIZATION: DYNAMIC SAMPLE SELECTION WITH SCALING LAWS. 2024
 2. Zhili Feng, Tanya Marwah, Nicolo Fusi, David Alvarez-Melis, and Lester Mackey. ADAPTING LANGUAGE MODELS VIA TOKEN TRANSLATION, 2024
 3. Tanya Marwah, Zhili Feng, Lester Mackey, Nicolo Fusi, and David Alvarez-Melis. IMPROVING ZERO-SHOT REASONING USING DYNAMIC NON-VERBAL TOKENS. *In submission*
 4. Yiming Zhang, Zhuokai Zhao, Zhaorun Chen, Zhili Feng, Zenghui Ding, and Yining Sun. RANKCLIP: RANKING-CONSISTENT LANGUAGE-IMAGE PRETRAINING. *arXiv preprint arXiv:2404.09387*, 2024

5. Avi Schwarzschild*, Zhili Feng*, Pratyush Maini, Zachary C Lipton, and J Zico Kolter. RETHINKING LLM MEMORIZATION THROUGH THE LENS OF ADVERSARIAL COMPRESSION. *Advances in Neural Information Processing Systems*, 2024
 - Generative AI and Law workshop @ ICML 2024
 - **Best paper award**, Workshop on Data Contamination (CONDA) @ ACL 2024
 - **Oral presentation**, Red Teaming GenAI Workshop @ NeurIPS 2024
6. Pratyush Maini*, Zhili Feng*, Avi Schwarzschild, Zachary C Lipton, and J Zico Kolter. TOFU: A TASK OF FICTITIOUS UNLEARNING FOR LLMs. *Conference on Language Modeling*, 2024
 - **Spotlight presentation**, Workshop on Secure and Trustworthy Large Language Models @ ICLR 2024
7. Zhili Feng, Michal Moshkovitz, Dotan Di Castro, and J Zico Kolter. AN AXIOMATIC APPROACH TO MODEL-AGNOSTIC CONCEPT EXPLANATIONS. *arXiv preprint arXiv:2401.06890*, 2024
8. Zhili Feng, Dhananjay Ram, Cole Hawkins, Aditya Rawal, and Jinman Zhao. SEQUENCE-LEVEL LARGE LANGUAGE MODEL TRAINING WITH CONTRASTIVE PREFERENCE OPTIMIZATION. *In submission*, 2023
9. Zhili Feng, Anna Bair, and J Zico Kolter. TEXT DESCRIPTIONS ARE COMPRESSIVE AND INVARIANT REPRESENTATIONS FOR VISUAL LEARNING. *Transactions on Machine Learning Research*, 2024
10. Zhili Feng, Ezra Winston, and J Zico Kolter. MONOTONE DEEP BOLTZMANN MACHINES. *Transactions on Machine Learning Research*, 2023
11. Zhili Feng and J Zico Kolter. ON THE NEURAL TANGENT KERNEL OF EQUILIBRIUM MODELS. *arXiv preprint arXiv:2310.14062*, 2021
12. Zhili Feng, Shaobo Han, and Simon S Du. PROVABLE ADAPTATION ACROSS MULTIWAY DOMAINS VIA REPRESENTATION LEARNING. *International Conference on Learning Representations*, 2022
13. Zhili Feng, Fred Roosta, and David P Woodruff. NON-PSD MATRIX SKETCHING WITH APPLICATIONS TO REGRESSION AND OPTIMIZATION. *Conference on Uncertainty in Artificial Intelligence*, 2021
14. Jon Ergun, Zhili Feng, Sandeep Silwal, David P Woodruff, and Samson Zhou. LEARNING-AUGMENTED k -MEANS CLUSTERING. *International Conference on Learning Representations*, 2022
 - **Spotlight presentation**, ICLR 2022
15. Zhili Feng, Praneeth Kacham, and David Woodruff. DIMENSIONALITY REDUCTION FOR THE SUM-OF-DISTANCES METRIC. In *International Conference on Machine Learning*, pages 3220–3229. PMLR, 2021
 - **Oral presentation**, ICML 2021
16. Shashank Rajput*, Zhili Feng*, Zachary B. Charles, Po-Ling Loh, and Dimitris S. Papailiopoulos. DOES DATA AUGMENTATION LEAD TO POSITIVE MARGIN? In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, pages 5321–5330, 2019
 - **Oral presentation**, ICML 2019

17. Zhili Feng and Po-Ling Loh. ONLINE LEARNING WITH GRAPH-STRUCTURED FEEDBACK AGAINST ADAPTIVE ADVERSARIES. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 931–935, June 2018
18. Qiang Ning, Zhili Feng, Hao Wu, and Dan Roth. JOINT REASONING FOR TEMPORAL AND CAUSAL RELATIONS. In *Proc. of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 2278–2288, Melbourne, Australia, 7 2018. Association for Computational Linguistics
19. Qiang Ning, Ben Zhou, Zhili Feng, Haoruo Peng, and Dan Roth. COGCOMP TIME: A TOOL FOR UNDERSTANDING TIME IN NATURAL LANGUAGE. In *EMNLP (Demo Track)*, Brussels, Belgium, 11 2018. Association for Computational Linguistics
20. Daniel Khashabi, Mark Sammons, Ben Zhou, Tom Redman, Christos Christodoulopoulos, Vivek Srikumar, Nicholas Rizzolo, Lev Ratinov, Guanheng Luo, Quang Do, Chen-Tse Tsai, Subhro Roy, Stephen Mayhew, Zhili Feng, John Wieting, Xiaodong Yu, Yangqiu Song, Shashank Gupta, Shyam Upadhyay, Naveen Arivazhagan, Qiang Ning, Shaoshi Ling, and Dan Roth. COGCOMP NLP: YOUR SWISS ARMY KNIFE FOR NLP. In *Proc. of the International Conference on Language Resources and Evaluation (LREC)*, 2018
21. Qiang Ning, Zhili Feng, and Dan Roth. A STRUCTURED LEARNING APPROACH TO TEMPORAL RELATION EXTRACTION. In *EMNLP*, Copenhagen, Denmark, 9 2017

**TEACHING
EXPERIENCE**

Teaching Assistant Carnegie Mellon University	Jan 2022 - May 2022
<ul style="list-style-type: none"> • Teaching 10-707 Topics in Deep Learning 	
Teaching Assistant Carnegie Mellon University	Sept 2021 - Dec 2021
<ul style="list-style-type: none"> • Teaching 10-725 Convex Optimization 	
Teaching Assistant University of Wisconsin-Madison	Jan 2019 - May 2019
<ul style="list-style-type: none"> • Teaching CS525 Linear Programming Methods 	
Teaching Assistant University of Wisconsin-Madison	Jan 2019 - May 2019
<ul style="list-style-type: none"> • Teaching CS525 Linear Programming Methods 	
Teaching Assistant University of Wisconsin-Madison	Sept 2018 - Dec 2018
<ul style="list-style-type: none"> • Teaching CS525 Linear Programming Methods 	
Teaching Assistant University of Wisconsin-Madison	Jan 2017 - May 2017
<ul style="list-style-type: none"> • Teaching CS577 Introduction to Algorithms 	
Teaching Assistant University of Wisconsin-Madison	Sept 2017 - Dec 2017
<ul style="list-style-type: none"> • Teaching CS240 Introduction to Discrete Mathematics 	

Course Staff Aug 2016 - Dec 2016
University of Illinois at Urbana-Champaign
• Grading CS446 Machine Learning

PROFESSIONAL EXPERIENCE **Software Engineering Intern** May 2017 - Aug 2017
Facebook, Menlo Park, CA

Software Engineering Intern May 2016 - Aug 2016
Facebook, Menlo Park, CA

Software Engineering Intern May 2015 - Aug 2015
Marketo, San Mateo, CA

**HONOR &
AWARDS**

Dean's List
University of Illinois at Urbana-Champaign
• Fall 2013, Spring 2014, Fall 2014, Spring 2015, Spring 2016

Edmund J James Scholar
University of Illinois at Urbana-Champaign
• Academic Year 2013-2014