Zhili Feng

zhilif@andrew.cmu.edu

EDUCATION

Ph.D. in Machine Learning   Aug 2019 - Present
Carnegie Mellon University

Master of Science in Computer Science   Sept 2017 - May 2019
University of Wisconsin-Madison
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
Thesis: *A Combination of Statistical Model and Rule-based Approach to Temporal Extraction and Normalization*

PUBLICATION

∗ indicates equal contribution.

1. Avi Schwarzschild∗, Zhili Feng∗, Pratyush Maini, Zachary C Lipton, and J Zico Kolter. Rethinking llm memorization through the lens of adversarial compression *Generative AI and Law workshop @ ICML 2024*
   Best paper award, *Workshop on Data Contamination (CONDA) @ ACL 2024*

2. 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*

3. 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

4. Zhili Feng, Dhananjay Ram, Cole Hawkins, Aditya Rawal, and Jinman Zhao. Sequence-level large language model training with contrastive preference optimization. *In submission*, 2023

5. Zhili Feng, Anna Bair, and J Zico Kolter. Text descriptions are compressive and invariant representations for visual learning. *Transactions on Machine Learning Research*, 2024

6. Zhili Feng, Ezra Winston, and J Zico Kolter. Monotone deep boltzmann machines. *Transactions on Machine Learning Research*, 2023

7. Zhili Feng and J Zico Kolter. On the neural tangent kernel of equilibrium models. *arXiv preprint arXiv:2310.14062*, 2021

8. Zhili Feng, Shaobo Han, and Simon S Du. Provable adaptation across multiway domains via representation learning. *International Conference on Learning Representations*, 2022

9. 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

10. 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*

11. 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*

12. 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*

13. Z. Feng and P. Loh. Online learning with graph-structured feedback against adaptive adversaries. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 931–935, June 2018

14. 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

15. Qiang Ning, Ben Zhou, Zhili Feng, Haoruo Peng, and Dan Roth. CogCompTime: A Tool for Understanding Time in Natural Language. In *EMNLP (Demo Track)*, Brussels, Belgium, 11 2018. Association for Computational Linguistics

16. 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. CogCompNLP: Your Swiss Army Knife for NLP. In *Proc. of the International Conference on Language Resources and Evaluation (LREC)*, 2018

17. Qiang Ning, Zhili Feng, and Dan Roth. A structured learning approach to temporal relation extraction. In *EMNLP*, Copenhagen, Denmark, 9 2017

PROFESSIONAL EXPERIENCE	**Research Intern** Microsoft Research, Boston, MA	Jun 2024 - Aug 2024
	Applied Scientist Intern Amazon, New York, NY	Jun 2023 - Aug 2023
	Machine Learning Research Intern Bosch, Pittsburgh, PA	Jun 2021 - Aug 2021
	Research Intern NEC Lab America, remote	Sept 2020 - Dec 2020
	Software Engineering Intern Facebook, Menlo Park, CA	May 2017 - Aug 2017
	Software Engineering Intern Facebook, Menlo Park, CA	May 2016 - Aug 2016
	Software Engineering Intern Marketo, San Mateo, CA	May 2015 - Aug 2015

**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