

Zixuan Wang

Résumé

✉ wangzx@princeton.edu

📄 My Webpage: zixuan-wang-dlt.github.io/

🔗 Link to My Google Scholar

Research Interest

My research aims to uncover the **science of modern deep learning**, and design **more capable, efficient, and reliable language models and artificial intelligence with scientific principles**. I currently focus on **expressivity, optimization and scaling** of language models, with the broader goal of developing better model architectures and training algorithms for deep learning.

Education

2026–Present **EECS**, *University of California, Berkeley*.

Visiting Student Researcher. Supervisor: Prof. Jason D. Lee.

2023–Present **Electrical and Computer Engineering**, *Princeton University*.

Ph.D. in Electrical and Computer Engineering. Advisor: Prof. Jason D. Lee. GPA: 4.0/4.0

2019–2023 : **Institute for Interdisciplinary Information Sciences (IIIS)**, *Tsinghua University*, Special CS Pilot Class (**Yao Class**, established by Turing Award Laureate, **Prof. Andrew Chi-Chih Yao**). B.Eng. in Computer Science. GPA: 3.93/4.0

Internship

2026.5–8 **Center for Data Science**, *NYU Courant School*.

Junior Research Scientist, working on **LLM scaling laws**. Supervised by Alberto Bietti.

2025.6–12 **Seed Edge**, *Bytedance Inc., San Jose*.

Student Intern, working on **scaling looped transformers**. Supervised by Tianle Cai.

Publications

(*Equal contribution.)

- [1] **Zixuan Wang**, Xingyu Dang, Jason D. Lee, & Kaifeng Lyu. (2026). The Power of Power Law: Asymmetry Enables Compositional Reasoning. International Conference on Machine Learning (ICML), 2026, **Spotlight**.
- [2] **Zixuan Wang***, Xingyu Dang*, Rui-Jie Zhu, Zixin Wen, Hengyu Fu, Wenhao Chai, and Jason D. Lee. (2026). Learning What to Remember: Test-Time Training via Context Distillation. arXiv preprint arXiv:2608.01672 .
- [3] Hengyu Fu*, Tianyu Guo*, **Zixuan Wang***, Hanlin Zhu*, Jason D. Lee, Jiantao Jiao, Stuart Russell, Song Mei. (2026). DiscoLoop: Looping Discrete Embeddings and Continuous Hidden States for Multi-hop Reasoning. arXiv preprint arXiv:2607.00341.
- [4] Yixiao Huang, Hanlin Zhu, **Zixuan Wang**, Jiantao Jiao, Stuart Russell, Somayeh Sojoudi, Song Mei. (2026). Transformers Provably Learn to Internalize Chain-of-Thought. arXiv preprint arXiv:2605.28600.
- [5] Zhu, R. J.*, **Zixuan Wang***, Hua, K., Zhang, T., Li, Z., Que, H., ... & Eshraghian, J. (2025). Scaling Latent Reasoning via Looped Language Models. arXiv preprint arXiv:2510.25741.
- [6] Zheng, Z., Cheng, Z., Shen, Z., Zhou, S., Liu, K., He, H., ..., **Zixuan Wang**,... & Xie, S. (2025). LiveCodeBench Pro: How Do Olympiad Medalists Judge LLMs in Competitive Programming? *Neural Information Processing Systems (NeurIPS) 2025 Datasets and Benchmarks Track*.

- [7] **Zixuan Wang***, Eshaan Nichani*, Alberto Bietti, Alex Damian, Daniel Hsu, Jason Lee, Denny Wu. Learning Compositional Functions with Transformers from Easy-to-Hard Data. *Conference on Learning Theory (COLT)* 2025.
- [8] Noam Razin, **Zixuan Wang**, Hubert Strauss, Stanley Wei, Jason D. Lee, Sanjeev Arora. What Makes a Reward Model a Good Teacher? An Optimization Perspective. *Neural Information Processing Systems (NeurIPS)* 2025, **Spotlight**.
- [9] Jianhao Huang*, **Zixuan Wang***, Jason D. Lee. Transformers Learn to Implement Multi-step Gradient Descent with Chain of Thought. *International Conference on Learning Representations (ICLR)* 2025, **Spotlight**.
- [10] Humanity's last exam. arXiv preprint arXiv: 2501.14249.
- [11] Yunwei Ren*, **Zixuan Wang***, Jason D. Lee. Learning and Transferring Sparse Contextual Bigrams with Linear Transformers. *Neural Information Processing Systems (NeurIPS)* 2024.
- [12] **Zixuan Wang**, Ruocheng Wei, Daniel Hsu, Jason D. Lee. Transformers Provably Learn Sparse Token Selection While Fully-Connected Nets Cannot. *International Conference on Machine Learning (ICML)*, 2024.
- [13] Xingyu Zhu*, **Zixuan Wang***, Xiang Wang, Mo Zhou, Rong Ge. Understanding Edge-of-Stability Training Dynamics With a Minimalist Example. *International Conference on Learning Representations (ICLR)*, 2023.
- [14] Zhouzi Li*, **Zixuan Wang***, Jian Li. Analyzing Sharpness along GD Trajectory: Progressive Sharpening and Edge of Stability. *Neural Information Processing Systems (NeurIPS)* 2022.

Fellowships & Awards

- 2023 First-year fellowship for graduate student at Princeton University
- 2023 Yao Class Departmental summa cum laude (23 out of 94)
- 2022 Excellent Comprehensive Performance Scholarship, IIIS, Tsinghua University (top 2 in 30)
- 2021 Excellent Comprehensive Performance Scholarship, IIIS, Tsinghua University (top 5 in 30)
- 2018 Chinese Mathematical Olympiad in Senior, First Prize

Academic Services

- 2026 Conference Reviewer for ICML 2026. **Gold reviewer**.
- 2025 Conference Reviewer for NeurIPS 2025
- 2025 Conference Reviewer for ICLR 2025
- 2025 Co-organizer of Princeton Alg-ML Seminar
- 2024 Workshop Reviewer for NeurIPS 2024 Workshop on Mathematics of Modern Machine Learning.
- 2024 Conference Reviewer for NeurIPS 2024
- 2024 Conference Reviewer for ICLR 2024
- 2024 Workshop Reviewer for ICLR 2024 Workshop on Bridging the Gap Between Practice and Theory in Deep Learning.
- 2023 Conference Reviewer for NeurIPS 2023