

# Siyu Chen

✉ [siyu.chen.sc3226@yale.edu](mailto:siyu.chen.sc3226@yale.edu)

📍 New Haven, CT, USA

🐟 FFishy-git

🎓 Google Scholar

## Research Interest

---

Deep Learning, Foundation Model, Various Aspects of Statistics, Reinforcement Learning

## Education

---

**Doctoral Student**     **Yale University**, Statistics and Data Science     Sep 2022 to now  
     • Advisor: Prof. Zhuoran Yang

**Bachelor of Science**     **Tsinghua University**, Electronic Engineering     Sep 2018 to May 2022

## Experience

---

**Yale University**, Teaching Assistant for Information Theory     CT, USA  
     Feb 2024 to May 2024

## Publications

---

- Chen, S., H. Sheen, T. Wang, and Z. Yang (2024). “Unveiling Induction Heads: Provable Training Dynamics and Feature Learning in Transformers”. In: *Advances in Neural Information Processing Systems*. Vol. 37, pp. 66479–66567.
- Chen, S., H. Sheen, X. Xiong, T. Wang, and Z. Yang (2025). “Taming Polysemanticity in LLMs: Provable Feature Recovery via Sparse Autoencoders”. In: *arXiv preprint arXiv:2506.14002*.
- Chen, S., M. Wang, and Z. Yang (2023). “Actions Speak What You Want: Provably Sample-Efficient Reinforcement Learning of the Quantal Stackelberg Equilibrium from Strategic Feedbacks”. In: *arXiv preprint arXiv:2307.14085*.
- Chen, S., Y. Wang, Z. Wang, and Z. Yang (2023). “A unified framework of policy learning for contextual bandit with confounding bias and missing observations”. In: *arXiv preprint arXiv:2303.11187*.
- Chen, S., B. Wu, M. Lu, Z. Yang, and T. Wang (2025). “Can Neural Networks Achieve Optimal Computational-statistical Tradeoff? An Analysis on Single-Index Model”. In: *The Thirteenth International Conference on Learning Representations*, (oral).
- Chen, S., J. Wu, Y. Wu, and Z. Yang (2023). “Learning to incentivize information acquisition: Proper scoring rules meet principal-agent model”. In: *International Conference on Machine Learning*. PMLR, pp. 5194–5218.
- Chen, S., D. Yang, J. Li, S. Wang, Z. Yang, and Z. Wang (2022). “Adaptive model design for Markov decision process”. In: *International Conference on Machine Learning*. PMLR, pp. 3679–3700.
- He, J., S. Chen, F. Zhang, and Z. Yang (2024). “From Words to Actions: Unveiling the Theoretical Underpinnings of LLM-Driven Autonomous Systems”. In: *International Conference on Machine Learning*. PMLR, pp. 17807–17841.
- He, J., X. Pan, S. Chen, and Z. Yang (2025). “In-Context Linear Regression Demystified: Training Dynamics and Mechanistic Interpretability of Multi-Head Softmax Attention”. In: *arXiv preprint arXiv:2503.12734*.
- Hu, X., F. Zhang, S. Chen, and Z. Yang (2024). “Unveiling the Statistical Foundations of Chain-of-Thought Prompting Methods”. In: *arXiv preprint arXiv:2408.14511*.
- Sheen, H., S. Chen, T. Wang, and H. H. Zhou (2024). “Implicit Regularization of Gradient Flow on One-Layer Softmax Attention”. In: *arXiv preprint arXiv:2403.08699*.

- Siyu, C., S. Heejune, W. Tianhao, and Y. Zhuoran (2024). “Training Dynamics of Multi-Head Softmax Attention for In-Context Learning: Emergence, Convergence, and Optimality”. In: *The Thirty Seventh Annual Conference on Learning Theory*. PMLR, pp. 4573–4573.
- Siyu Chen Theodor Misiakiewicz, I. Z. and P. Zhang (2025). “An Optimized Franz-Parisi Criterion and its Equivalence with SQ Lower Bounds”. In: *The Thirty-Ninth Annual Conference on Neural Information Processing Systems*, **(oral)**.
- Wu, J., S. Chen, M. Wang, H. Wang, and H. Xu (2024). “Contractual Reinforcement Learning: Pulling Arms with Invisible Hands”. In: *arXiv preprint arXiv:2407.01458*.
- Zhang\*, Y., S. Chen\*, Z. Yang, M. Jordan, and Z. Wang (2021). “Wasserstein flow meets replicator dynamics: A mean-field analysis of representation learning in actor-critic”. In: *Advances in Neural Information Processing Systems* 34, pp. 15993–16006.

## Skills

---

**Languages:** C++, C, Python, R