Wenjie Zhou Ph.D. Candidate

(+86)136-8342-9215 (WeChat)
zj4323005@gmail.com

Education

Institute of Computing Technology, Chinese Academy of Sciences. Beijing, China
Ph.D. in Computer Science, State Key Laboratory of AI Safety 2023–Present
- Research area: **Large Language Model Pre-Training**, Optimization Theory.

Beijing University of Posts and Telecommunications Beijing, China
B.S. in Computational Mathematics, Top 5% 2019–2023

Experiences

RedNote Dots Pre-Training Group. | Beijing, China 2026/10–Present
- Research Intern, with **RedStar/ACE** Talent Program.
- Contributed to Productive Overtarin Scaling Ladder.

Tencent Hunyuan Pre-Training Group. | Beijing, China 2026/04–2026/10
- Research Intern, with **Qingyun** Talent Program.
- Contributed to Productive Optimizers, Training Stability, and Scaling Laws.

Ant Group BaiLing Pre-Training Group. | Beijing, China 2025/11–2026/03
- Research Intern, with Pre-Training Algorithms Team.
- Contributed to Productive Optimizers and pre-training model merging.

Publications (Selected)

(* indicates equal contributions; † indicates correspondence.)

1. Zhanpeng Zhou*, **Wenjie Zhou***, Yufei Gu*, Juqiu Wang, Guanwei Zhang, Ruobing Xie, Xianbiao Qi, Yushi Bai, Kaifeng Lyu, Shunyu Yao. From LR to ELR: A Better Heuristic for Pretraining Dynamics. **Tencent Hunyuan Official Blog.**
2. **Wenjie Zhou**, Bohan Wang, Hongtao Zhang, Wei Chen, Xueqi Cheng. Extra-Merge: Tracing the Rank-1 Ridge of Model Merging in Language Model Pre-Training. **ICML 2026.**
3. Hongtao Zhang*, **Wenjie Zhou***, Wei Chen, Xueqi Cheng. The Stability of Singular Distribution: A Spectral Perspective on the Two-Phase Dynamics of Language Model Pre-Training. **ICML 2026.**
4. **Wenjie Zhou**, Bohan Wang, Wei Chen, Xueqi Cheng. BSFA: Leveraging the Subspace Dichotomy to Accelerate Neural Network Training. **EMNLP 2025 (Oral, Top 3%).**
5. Hongtao Zhang*, **Wenjie Zhou***, Wei Chen, Xueqi Cheng. When and Why Grouping Attention Heads Accelerates Muon Optimization. **ICML 2026 HiLD.**
6. **Wenjie Zhou**, Bohan Wang, Wei Chen, Zhiming Ma, Xueqi Cheng. Momentum Gradient Descent over Single Neuron Network: Rich Behaviors of Limiting Sharpness. **ICLR 2024 BGPT.**
7. Bing Liu, **Wenjie Zhou**, Chengcheng Zhao, Hongtao Zhang, Boao Kong, Felix Dangel, Wu Lin. How Bregman Divergences Shape Shampoo. **Under review.**
8. Bing Liu*, **Wenjie Zhou***, Chengcheng Zhao. Revisiting Bregman Matrix Divergences in Shampoo Optimization. **ICML 2026 HiLD.**

Projects

MuonH SOTA | Improves H with Minus-sqrt Root Schedule.
- Hyperball optimization SOTA on Track 3 Optimization of modded-nanogpt.

Group-Head Muon | Improves Muon with Grouping Attention Heads.
- Accelerates vanilla Muon on Track 3 Optimization of modded-nanogpt via Stochastic Grouping Attention Heads, compared with Per-head Muon.

Awards

- **National Scholarship (Top 0.2%)**, Ministry of Education 2025
- **Outstanding Student**, University of Chinese Academy of Sciences 2025