ZHANHUI ZHOU

✉ zhanhui@berkeley.edu   ○ ZHZisZZ   🎓 Google Scholar   📍 Berkeley, CA, USA

EDUCATION

University of California, Berkeley   Berkeley, CA
Ph.D. in Computer Science   2025–

University of Michigan   Ann Arbor, MI
B.S.E. in Computer Science   2020–2022

Shanghai Jiao Tong University   Shanghai, China
B.E. in Electical and Computer Engineering   2018–2020

PROJECTS

- ○ dLLM, Simple Diffusion Language Modeling, *Lead* 2025

AWARDS

- Berkeley Fellowship for Outstanding Doctoral Students, *UC Berkeley* 2025–2026
- NeurIPS Scholar Award, *NeurIPS* 2025

SELECTED PUBLICATIONS

* Equal contribution   † Corresponding author

[1] **Zhanhui Zhou**, Lingjie Chen, Chao Yang, and Chaochao Lu. "VLMs Can Aggregate Scattered Training Patches". In: *Thirty-ninth Neural Information Processing Systems (NeurIPS)*. 2025.

[2] Zhichen Dong*, **Zhanhui Zhou***, Zhixuan Liu, Chao Yang, and Chaochao Lu. "Emergent Response Planning in LLMs". In: *Forty-second International Conference on Machine Learning (ICML)*. 2025.

[3] **Zhanhui Zhou**, Zhixuan Liu, Jie Liu, Zhichen Dong, Chao Yang, and Yu Qiao. "Weak-to-Strong Search: Align Large Language Models via Searching over Small Language Models". In: *Thirty-eighth Neural Information Processing Systems (NeurIPS)*. 2024.

[4] **Zhanhui Zhou**, Jie Liu, Zhichen Dong, Jiaheng Liu, Chao Yang, Wanli Ouyang, and Yu Qiao. "Emulated Disalignment: Safety Alignment for Large Language Models May Backfire!" In: *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL)*. 🏆 **Outstanding Paper Award ($<$ 1% of all submissions)**. 2024.

[5] **Zhanhui Zhou**, Jie Liu, Jing Shao, Xiangyu Yue, Chao Yang, Wanli Ouyang, and Yu Qiao. "Beyond one-preference-fits-all alignment: Multi-objective direct preference optimization". In: *Findings of the Association for Computational Linguistics: ACL (ACL Findings)*. 2024.

[6] **Zhanhui Zhou**, Man To Tang, Qiping Pan, Shangyin Tan, Xinyu Wang, and Tianyi Zhang. "Intent: Interactive tensor transformation synthesis". In: *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST)*. 2022.