

Wenxuan Zhang

 wenzhan@kaust.edu.sa
 <https://github.com/wx-zhang>

 wx-zhang.github.io
 [Google Scholar](#)  [LinkedIn](#)

Research Interest

- Fine-tuning and aligning foundation models with reinforcement learning.
- Efficient Training and Continual Learning. Parameter-efficient methods, mitigating forgetting, and learning under constrained computation.
- Agents. Building autonomous LLM/VLM agents with enhanced reasoning and tool-use capabilities.

Education

- **King Abdullah University of Science and Technology**, Thuwal, Saudi Arabia. 2022.1 – present
Ph.D., Computer Science, supervised by Prof. Mohamed Elhoseiny.
- **University of Pennsylvania**, Philadelphia, United States. 2019.8 – 2021.12
M.A., Applied Mathematics and Computational Science. GPA: 3.92/4.00
Thesis title: *Factorized lifelong machine learning on non-stationary tasks: An algorithm and analysis*.
- **Beijing Normal University**, Beijing, China. 2015.9 – 2019.6
B.S., Mathematics and Applied Mathematics. GPA: 90.5/100
Thesis title: *A hand gesture recognition module for medical robots*.

Experience

- **Research Intern**, Meta AI, London, United Kingdom. 2025.08 - 2025.12
Topic: Post-training algorithms for Diffusion-LLM.
- **Research Intern**, Samsung Research America, Mountain View, United States. 2024.10 - 2025.1
Topic: Merging algorithms for large vision-language models.
- **Visiting student**, Torr Vision Group, University of Oxford, Oxford, United Kingdom. 2023.7 - 2023.11
Topic: Alignment algorithms to improve the safety of LLM.

Publications

8 first-authored (5 top-published, 1 Spotlight), 5 second-authored (4 top-published, 1 Spotlight), and collaborations.

Post-Training and Alignment

- **Wenxuan Zhang**, Y. Wang, X. Shen, D. Jia, J. Ding, and M. Elhoseiny, “UniV: Mixed advantage with unified critic in agentic training for vision-language models,” In submission to ICML 2026.
- **Wenxuan Zhang**, L. Wu, C. Zhao, E. Chang, M. Zhuge, Z. Liu, D. Su, H. Huang, J. Chen, C. Zhou, R. Krishnamoorthi, V. Chandra, M. Elhoseiny, and W. Wen, “dTRPO: Trajectory reduction in policy optimization of diffusion large language models,” In submission to ICML 2026.
- **Wenxuan Zhang**, P. Torr, M. Elhoseiny, and A. Bibi, “Bi-factorial preference optimization: Balancing safety-helpfulness in language models,” in *The Thirteenth International Conference on Learning Representations*, 2025. (**ICLR Spotlight 2025**).
- H. Xu, **Wenxuan Zhang**, J. Fei, Y. Wu, T. Xie, J. Huang, Y. Xie, M. Elhoseiny, and P. Kalnis, “SLAMB: Accelerated large batch training with sparse communication,” in *The Fortieth International Conference on Machine Learning*, 2023. (**ICML 2023**).

Efficient Training and Continual Learning

- N. Alballa, **Wenxuan Zhang**, Z. Liu, A. M. Abdelmoniem, M. Elhoseiny, and M. Canini, “Query-based knowledge transfer for heterogeneous learning environments,” in *The Thirteenth International Conference on Learning Representations*, 2025 . (**ICLR 2025**).
- B. Csaba*, **Wenxuan Zhang***, M. Müller, S.-N. Lim, M. Elhoseiny, P. Torr, and A. Bibi, “Label delay in continual learning,” in *The Thirty-Eighth Annual Conference on Neural Information Processing Systems*, 2024. (**NeurIPS 2024**).
- **Wenxuan Zhang**, P. Janson, R. Aljundi, and M. Elhoseiny, “Overcoming generic knowledge loss with selective parameter update,” in *The IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. (**CVPR 2024**).
- **Wenxuan Zhang**, Y. Mohamed, B. Ghanem, P. Torr, A. Bibi, and M. Elhoseiny, “Continual learning on a diet: Learning from sparse labeled streams under constrained computation,” in *The Twelfth International Conference on Learning Representations*, 2024. (**ICLR 2024**).
- **Wenxuan Zhang**, P. Janson, K. Yi, I. Skorokhodov, and M. Elhoseiny, “Continual zero-shot learning through semantically guided generative random walks,” in *IEEE/CVF International Conference on Computer Vision*, 2023. (**ICCV 2023**).
- **Wenxuan Zhang**, L. Zhou, and S. Kumar, “Towards a unified view of model merging for vision-language models,” Under Samsung internal review.
- P. Janson, **Wenxuan Zhang**, R. Aljundi, and M. Elhoseiny, “A simple baseline that questions the use of pretrained-models in continual learning,” in *NeurIPS 2022 Workshop on Distribution Shifts: Connecting Methods and Applications*, 2022.
- K. Yi, P. Janson, **Wenxuan Zhang**, and M. Elhoseiny, “Domain-aware continual zero-shot learning,” *arXiv preprint arXiv:2112.12989*, 2021.

Agents

- X. Shen, **Wenxuan Zhang**, J. Chen, and M. Elhoseiny, “Vgent: Graph-based retrieval-reasoning-augmented generation for long video understanding,” 2025 (**NeurIPS Spotlight 2025**).
- L. Zhao, **Wenxuan Zhang**, M. Fan, I. Miller, O. Khattab, and M. Elhoseiny, “Cross-Reflect: Empowering multi-modal agents with joint reasoning across trajectories,” In submission to CVPR 2026.
- M. Zhuge, C. Zhao, H. Liu, Z. Zhou, S. Liu, Y. SUN, E. Chang, G. L. Lan, J. Fei, **Wenxuan Zhang**, W. Wang, Y. Xiong, Z. Liu, Z. Cai, Y. Yang, Y. Tian, Y. Shi, V. Chandra, and J. Schmidhuber, “Neural computers,” In submission to ICML 2026.
- M. Lu, R. Xu, Y. Fang, **Wenxuan Zhang**, Y. Yu, G. Srivastava, Y. Zhuang, M. Elhoseiny, C. Fleming, C. Yang, *et al.*, “Scaling agentic reinforcement learning for tool-integrated reasoning in VLMs,” *arXiv preprint arXiv:2511.19773*, 2025.
- D. Zhu, J. Chen, K. Haydarov, X. Shen, **Wenxuan Zhang**, and M. Elhoseiny, “ChatGPT asks, BLIP-2 answers: Automatic questioning towards enriched visual descriptions,” 2024. (**TMLR**).

Academic Services

- **Conference Organizer**, ICCV2025 SafeMM-AI Workshop
- **Conference Reviewer**, ICLR, NeurIPs, CVPR, ICCV, TPMAI, CLAI Unconf
- **Teaching Assistant**, CS 326 Low Resource Deep Learning
- **Academic Mentor**, KAUST Master Student Direct Research

Awards & Scholarship & Activities

- KAUST Graduate Scholarship. 2022 - present

Awards & Scholarship & Activities (continued)

- KAUST CEMSE Dean's List 2024-25 2025
- Research intern, Xiaohongshu, Beijing, China. 2021.8 - 2021.11
- Master thesis student, [LML group](#), Upenn, Philadelphia, United States. 2020.7 - 2021.12
- First Class of Jingshi Scholarship, BNU. 2018
- Meritorious Winner, COMAP's Mathematical Contest in Modeling (MCM). 2018
- Summer School, College of William & Mary, Williamsburg, United States 2016.7 - 2016.8
- Athlete Plan for Cultivating Top-notch Students of Basic Disciplines by Ministry of Education. 2015 - 2019