Wenxuan ZHANG

✉ wenxuan.zhang@kaust.edu.sa 🌐 wx-zhang.github.io
⌾ https://github.com/wx-zhang 🎓 Google Scholar in LinkedIn

Research Interest

○ Safety Alignment. Align language models with multifactorial human preference.

○ Efficient Finetuning. Enhance foundation models with emerging properties.

○ Continual Learning. Study the continual learning in the realistic scenarios.

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*.

Academic Experience

○ **Visiting student**, Torr Vision Group, University of Oxford, Oxford, United Kingdom.   2023.7 - 2023.11
Supervised by Dr. Adel Bibi and Prof. Philip Torr.

○ **Master thesis student**, LML group, Upenn, Philadelphia, United States.   2020.7 - 2021.12
Supervised by Prof. Eric Eaton.

○ **Research intern**, Vision Algorithm group, Xiaohongshu, Beijing, China.   2021.8 - 2021.11
Developed an efficient speaker verification system for video rating.

○ **Summer School**, College of William & Mary, Willimsburg, United States   2016.7 - 2016.8

Publications

● **Wenxuan Zhang**, P. Torr, M. Elhoseiny, and A. Bibi, "Bi-factorial preference optimization: Balancing safety-helpfulness in language models," In submission to ICLR 2025.

● N. Alballa, **Wenxuan Zhang**, Z. Liu, A. M. Abdelmoniem, M. Elhoseiny, and M. Canini, "Query-based knowledge transfer," In Submission to ICLR 2025.

● **Wenxuan Zhang**, P. Janson, R. Aljundi, and M. Elhoseiny, "Overcoming generic knowledge loss with selective parameter update," in *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 *ICLR*, 2024.

● B. Csaba*, **Wenxuan Zhang***, M. Müller, *et al.*, "Label delay in continual learning," in *NeurIPS*, 2024.

● D. Zhu, J. Chen, K. Haydarov, X. Shen, **Wenxuan Zhang**, and M. Elhoseiny, "Chatgpt asks, blip-2 answers: Automatic questioning towards enriched visual descriptions," in *TMLR*, 2024.

● **Wenxuan Zhang**, P. Janson, K. Yi, I. Skorokhodov, and M. Elhoseiny, "Continual zero-shot learning through semantically guided generative random walks," in *ICCV*, 2023.

- 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, **Zhang, Wenxuan**, and M. Elhoseiny, "Domain-aware continual zero-shot learning," *arXiv preprint arXiv:2112.12989*, 2021.
- H. Xu, **Wenxuan Zhang**, J. Fei, *et al.*, "Slamb: Accelerated large batch training with sparse communication," in *ICML*, 2023.

Academic Services

- **Conference reviewer**, ICLR, NeurIPs, CVPR, ICCV, TPMAI, CLAI Unconf
- **Teaching Assistant**, CS 326 Low Resource Deep Learning
- **Mentor**, KAUST Master Student Direct Research

Skills

- **Languages**: Strong reading, writing and speaking competencies for English and Mandarin Chinese.
- **Coding**: Python, LaTeX, MATLAB, CUDA, C++,

Awards

- KAUST Graduate Scholarship. 2022 - present
- First Class of Jingshi Scholarship, BNU. 2018
- Meritorious Winner, COMAP's Mathematical Contest in Modeling (MCM). 2018
- Athe Plan for Cultivating Top-notch Students of Basic Disciplines by Ministry of Education. 2015 - 2019