Wenxuan Zhang

wenxuan.zhang@kaust.edu.sa wx-zhang.github.io
https://github.com/wx-zhang Google Scholar LinkedIn

Research Interest

- Alignment. Aligning foundation models with multifactorial emerging properties.
- Efficient Finetuning. Efficient algorithms for the post-training stages of large models.

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

- **Research Intern**, Meta AI, London, United Kingdom. 2025.08 - 2025.12
  Supervised by Dr. Wei Wen.
  Topic: Post-training algorithms for Diffusion-LLM.
- **Research Intern**, Samsung Research America, Mountain View, United States. 2024.10 - 2025.1
  Supervised by Dr. Suren Kumar.
  Topic: Merging algorithms for large vision-language models.
- **Visiting student**, Torr Vision Group, University of Oxford, Oxford, United Kingdom. 2023.7 - 2023.11
  Supervised by Dr. Adel Bibi and Prof. Philip Torr.
  Topic: Alignment algorithms to improve the safety of LLM.
- **Master thesis student**, LML group, Upenn, Philadelphia, United States. 2020.7 - 2021.12
  Supervised by Prof. Eric Eaton.
  Topic: Theoretical gounded factorized lifelong learning.
- **Research intern**, Xiaohongshu, Beijing, China. 2021.8 - 2021.11
  Topic: Efficient speaker verification system for video rating.
- **Summer School**, College of William & Mary, Willimsburg, United States 2016.7 - 2016.8

Publications

Safety and Privacy

- **Wenxuan Zhang**, P. Torr, M. Elhoseiny, and A. Bibi, *Bi-factorial preference optimization: Balancing safety-helpfulness in language models*, 2025. **(ICLR Spotlight 2025)**.
- N. Alballa, **Wenxuan Zhang**, Z. Liu, A. M. Abdelmoniem, M. Elhoseiny, and M. Canini, *Query-based knowledge transfer for heterogeneous learning environments*, 2025 . **(ICLR 2025)**.

Multi-Modal Learning

- X. Shen, **Wenxuan Zhang**, J. Chen, and M. Elhoseiny, *Vgent: Graph-based retrieval-reasoning-augmented generation for long video understanding*, 2025 **(NeurIPS Spotlight 2025)**.

- **Wenxuan Zhang**, L. Zhou, and S. Kumar, *Towards a unified view of model merging for vision-language models*, Under Samsung internal review.
- **Wenxuan Zhang**, P. Janson, R. Aljundi, and M. Elhoseiny, *Overcoming generic knowledge loss with selective parameter update*, 2024. **(CVPR 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*, 2024. **(TMLR)**.
- **Wenxuan Zhang**, P. Janson, K. Yi, I. Skorokhodov, and M. Elhoseiny, *Continual zero-shot learning through semantically guided generative random walks*, 2023. **(ICCV 2023)**.

Efficient Fine-tuning and Continual Learning

- **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*, 2024. **(ICLR 2024)**.
- B. Csaba*, **Wenxuan Zhang***, M. Müller, *et al.*, *Label delay in continual learning*, 2024. **(NeurIPS 2024)**.
- H. Xu, **Wenxuan Zhang**, J. Fei, *et al.*, *Slamb: Accelerated large batch training with sparse communication*, 2023. **(ICML 2023)**.
- P. Janson, **Wenxuan Zhang**, R. Aljundi, and M. Elhoseiny, *A simple baseline that questions the use of pretrained-models in continual learning*, 2022.
- K. Yi, P. Janson, **Wenxuan Zhang**, and M. Elhoseiny, *Domain-aware continual zero-shot learning*, 2021.

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

- KAUST Graduate Scholarship. 2022 - present
- KAUST CEMSE Dean's List 2024-25 2025
- 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