Daiwei Chen

✉ daiwei.chen@wisc.edu 𝕏 @daiweichen99 in daiwei-chen
https://chendaiwei-99.github.io/

Research Interests

◇ My research aims to make LLMs more reliable by:

- **Pluralism**: Mitigate homogeneity and mode collapse by increasing diversity in preference alignment, cultural reasoning, and solution generation.
- **Hallucination**: Address weak metacognition by integrating statistical tools that provide calibrated confidence and formal reliability guarantees.
- **Generalization**: Study compositional generalization mechanisms and use synthetic data to enable weak-to-strong generalization.

Education

2023-Present ◇ **Ph.D. Student, University of Wisconsin-Madison**
Research Topics: *LLM Pluralistic Alignment, LLM Hallucination, W2S Generalization*.
Advisor: Ramya Korlakai Vinayak

2021-2023 ◇ **M.S. Electrical and Systems Engineering, University of Pennsylvania**
Research Topics: *Machine Learning Theory, PAC-Bayesian framework, Generalization*.
Advisor: Pratik Chaudhari

2017-2021 ◇ **B.S. Psychology, Zhejiang University**
Research Topics: *Visual Cognition Mechanism on Contrast Appearance*.
Advisor: Xiuying Qian, Yongchun Cai

Research Projects

08/2025-Now ◇ **Understanding Conformal Factuality Framework for LLM Hallucination**

- Integrate multi-agent systems with the statistical-grounded conformal prediction framework to mitigate LLM hallucination.

07/2025-Now ◇ **Self-Evolving Reasoning Pluralism: Align LLM reasoning with diverse human preference through evolutionary self-reflection**

- Develop a self-reflection LLM pluralism algorithm to align LLM Reasoning (usually "mode-collapsing") with pluralistic reasoning preferences.

12/2023-11/2024 ◇ **Pluralistic Alignment: Pluralistic alignment framework for learning from heterogeneous preferences**

- Developed the PAL framework to address AI pluralistic alignment using latent variables and mixture modeling techniques.
- Demonstrated that the PAL captures the diversity of user preferences while learning a shared latent preference space capable of few-shot generalizing to new users.
- Showcased PAL's competitive reward model accuracy in LLM tasks and image generation benchmarks, outperforming strong baseline models.

Research Publications

- **LinkedOut: Linking World Knowledge Representation Out of Video LLM for Next-Generation Video Recommendation**
  Haichao Zhang, Yao Lu, Lichen Wang, Yunzhe Li, Daiwei Chen, Yunpeng Xu, Yun Fu
  Under Review.
  arXiv, preprint 2025
- **PAL: Pluralistic ALignment Framework for Learning from Heterogeneous Preferences**
  Daiwei Chen, Yi Chen, Aniket Rege, Ramya Korlakai Vinayak
  *International Conference on Learning Representations (**ICLR**), 2025*
  *Behavioral ML workshop @ Neural Information Processing Systems (**NeurIPS**), 2024* **(Spotlight)**
  *MFHAIA workshop @ International Conference on Machine Learning (**ICML**), 2024* **(Oral)**
- **Unraveling The Impact of Training Samples**
  Daiwei Chen, Jane Zhang, Ramya Korlakai Vinayak
  *Blogpost @ International Conference on Learning Representations (**ICLR**), 2024*
- **Learning Capacity: A Measure of the Effective Dimensionality of a Model**
  Daiwei Chen*, Weikai Chang*, Pratik Chaudhari
  arXiv, preprint, 2023

Work Experience

05/2025-08/2025

- **GenAI Research Intern.** *Microsoft LinkedIn*, Sunnyvale HQ
  Developed a novel pretraining approach for LLM semantic ID embeddings using semantic alignment tasks, achieving state-of-the-art performance improvements in recommendation systems through the optimized token embedding space mapping.
  Mentor: Zhoutong Fu; Manager: Chengming Jiang

Service and Organization

Reviewer.	◇ NeurIPS 2026; ICLR 2026; CVPR 2026; EACL 2026.
TA.	◇ ECE 204 Data Science & Engineering, ***UW-Madison***.
	◇ ECE 532 Matrix Methods in Machine Learning, ***UW-Madison***.
	◇ CS 350 Software Design & Engineering, ***UPenn***.
	◇ ESE 542 Statistic for Data Science, ***UPenn***.

Skills

Coding	◇ Vibe Coding, PyTorch, DsPy
Language.	◇ English, Mandarin.

Awards and Achievements

2025	◇ **Research Grant Sponsorship**, Lambda.ai
2023	◇ **Outstanding Research Award**, University of Pennsylvania.
2021	◇ **Zhejiang University Scholarship**, Zhejiang University.
2020	◇ **Academic Excellence Award**, Zhejiang University.
2019	◇ **Title of School Outstanding Student**, Zhejiang University.