Dehong Xu

Center for Vision, Cognition, Learning, and Autonomy (VCLA), UCLA
Phone: +1(424)440-4146, Email: xudehong1996@ucla.edu

ACADEMIC BACKGROUND

University of California, Los Angeles (UCLA)
Ph.D. candidate in Statistics Expected graduation: 2025
M.S. in Statistics Sep 2019 - Jul 2021
Co-advised by Prof. Ying Nian Wu and Prof. Song-chun Zhu, GPA: 4.0 / 4.0

Beijing University of Posts and Telecommunications (BUPT)
B.Eng. in Software Engineering Sep 2015 - Jul 2019
GPA: 3.9 / 4.0; Ranking: 1 / 153

RESEARCH INTERESTS

Multi-modal LLM, LLM Alignment, Generative Models, Representation Learning

SELECTED PUBLICATIONS

(* denotes equal contributions)

Deqian Kong*, Minglu Zhao*, **Dehong Xu***, Jianwen Xie, Sirui Xie, Ying Nian Wu. "Variational Bayes Language Modeling with Latent Abstract Tokens and Fast-Slow Learning." *Preprint.*

Dehong Xu, Ruiqi Gao, Wen-Hao Zhang, Xue-Xin Wei, Ying Nian Wu. "An Investigation of Conformal Isometry Hypothesis for Grid Cells." *In submission to ICLR 2025.*

Deqian Kong*, **Dehong Xu***, Minglu Zhao*, Bo Pang, Jianwen Xie, Andrew Lizarraga, Yuhao Huang, Sirui Xie*, Ying Nian Wu. "Latent Plan Transformer for Trajectory Abstraction: Planning as Latent Space Inference." *Conference on Neural Information Processing Systems (NeurIPS)*, 2024.

Dehong Xu, Liang Qiu, Minseok Kim, Faisal Ladhak, Jaeyoung Do. "Aligning Large Language Models via Fine-grained Supervision." *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2024.

Yan Xu*, Deqian Kong*, **Dehong Xu***, Ziwei Ji*, Bo Pang, Pascale Fung, Ying Nian Wu. "Diverse and Faithful Knowledge-Grounded Dialogue Generation via Sequential Posterior Inference." *International Conference on Machine Learning (ICML)*, 2023.

Dehong Xu*, Ruiqi Gao*, Wen-Hao Zhang, Xue-Xin Wei, Ying Nian Wu. "Conformal Isometry of Lie Group Representation in Recurrent Network of Grid Cells." *Proceedings of the 1st NeurIPS Workshop on Symmetry and Geometry in Neural Representations*, PMLR 197:370-387, 2023 [Poster].

RESEARCH EXPERIENCE

Amazon Inc - Search M5 Team Jun 2024 - Sep 2024
Applied Scientist Intern, Advisors: Xiaohu Xie and Alejandro Mottini
Improving Instruction-following Capability of Multi-modal Embedding Models

- Developed a multi-modal, decoder-only framework for learning representations with instruction-following capabilities.
- Designed and implemented a two-stage training approach: a pre-training phase for modality alignment, followed by instruction fine-tuning.

- Our method achieved SoTA performance on multi-modal information retrieval benchmarks. (In submission to CVPR 2025).

Amazon Inc - Alexa AGI Team & Rufus Team Jun 2023 - Oct 2023
Applied Scientist Intern, Advisors: Liang Qiu, Puyang Xu and Yi Xu
Aligning Large Language Models via Fine-grained Supervision and Token-level RLHF **(Paper published in ACL 2024)**

- Developed a fine-grained data collection method for reward training via minimal editing, which pinpoints the exact output segments that affect user choices.
- Proposed token-level RLHF by training a token-level reward model with fine-grained supervision and incorporated it into PPO training.
- Our method outperformed LLaMA2-chat-7B and achieved the best performance on AlpacaFarm among all 7B models.

UCLA, Center for Vision, Cognition, Learning and Autonomy
Research assistant, Advisors: Prof. Ying Nian Wu and Prof. Song-chun Zhu
KokoMind: A Multifaceted Evaluation Dataset of Social Interactions

- Developed an evaluation dataset containing 150 complex multi-party social interactions with free-text questions and answers generated by GPT-4.
- For each social interaction, we ask various questions designed to assess multiple dimensions including Theory of Mind, social norms, emotion recognition, etc.

Sequential posterior inference models on knowledge-grounded dialogue generation

- We propose an end-to-end learning framework that efficiently selects knowledge-to-use and generates faithful, diverse dialogue responses by posterior inference.

SKILLS

Programming Skills: Python, PyTorch, TensorFlow, HuggingFace, Latex, C/C++

AWARDS

Summer Mentored Research Fellowship (SMRF), UCLA	Jun 2022
Doctoral Student Travel Award, UCLA	2019 - 2024
Cross-disciplinary Scholars in Science and Technology, UCLA	Jun 2018
First Prize Scholarship in BUPT (Top 1 in BUPT)	2016 - 2018

PROFESSIONAL SERVICES

Peer-reviewed Conferences Reviewer
Conference on Neural Information Processing Systems (NeurIPS)
The International Conference on Learning Representations (ICLR)
International Joint Conference on Artificial Intelligence (IJCAI)
International Conference on Artificial Intelligence and Statistics (AISTATS)
ACM Multimedia (ACM MM)

Journals Reviewer
Transactions on Machine Learning Research (TMLR)
IEEE Transactions on Neural Networks and Learning Systems (TNNLS)
IEEE Transactions on Image Processing (TIP)
The ISI's Journal for the Rapid Dissemination of Statistics Research (Stat)