

YIGE YUAN

CAS Key Laboratory of AI Safety
Institute of Computing Technology, Chinese Academy of Sciences
No.6 Kezueyuan South Road Zhongguancun, Haidian District, Beijing, China

✉ yuanyige20z@ict.ac.cn
📄 github.com/yuanyige
🔍 [Google Scholar](#)

RESEARCH INTEREST

My research goal is to build **Trustworthy AI** that performs reliably across diverse scenarios. To achieve this, I worked on generalization and alignment within the domains of graph, vision, and language. My research includes:

- **Generalization & Robustness:** To make models generalize well and stably under domain shifts, noises, or perturbations.
- **Large Language Model Alignment:** To make large language models align with human preferences and safety requirements.

EDUCATION

Institute of Computing Technology, Chinese Academy of Sciences Sep 2020 - current
Ph.D. in Computer Software and Theory (Advisor: Prof. Xueqi Cheng & A.P. Bingbing Xu)
Xidian University, School of Cyberspace Security Sep 2016 - Jun 2020
B.S. in Information Security (Experimental Class, GPA: 3.8/4.0)

SELECTED PUBLICATIONS

C: conference, J: journal, W: workshop, P: preprint / * equal contribution

- [C1] TEA: Test-time Energy Adaptation
Yige Yuan, Bingbing Xu, Liang Hou, Fei Sun, Huawei Shen, Xueqi Cheng
IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**), 2024, Main Track, [CCF-A](#)
- [C2] PDE+: Enhancing Generalization via PDE with Adaptive Distributional Diffusion
Yige Yuan, Bingbing Xu, Bo Lin, Liang Hou, Fei Sun, Huawei Shen, Xueqi Cheng
AAAI Conference on Artificial Intelligence (**AAAI**), 2024, Main Track, [CCF-A](#)
- [J1] Towards Generalizable Graph Contrastive Learning: An Information Theory Perspective
Yige Yuan, Bingbing Xu, Huawei Shen, Qi Cao, Keting Cen, Wen Zheng, Xueqi Cheng
Neural Networks (**NN**), Volume 172, [CCF-B](#), Q1, IF=8.4
- [P1] Fact-Level Confidence Calibration and Self-Correction
Yige Yuan, Xu Bingbing, Hexiang Tan, Fei Sun, Teng Xiao, Wei Li, Huawei Shen, Xueqi Cheng
Annual Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL**), 2025, UnderReview
- [P2] MITA: Bridging the Gap between Model and Data for Test-Time Adaptation
Yige Yuan, Xu Bingbing, Liang Hou, Fei Sun, Huawei Shen, Xueqi Cheng
IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**), UnderReview
- [P3] SimPER: Simple Preference Fine-Tuning without Hyperparameters by Perplexity Optimization
Teng Xiao*, **Yige Yuan***, Zhengyu Chen, Mingxiao Li, Shangsong Liang, Zhaochun Ren, Vasant G Honavar
International Conference on Learning Representations (**ICLR**), 2025, UnderReview
- [C3] Calibrated Preference Optimization for Direct Language Model Alignment
Teng Xiao, **Yige Yuan**, Huaisheng Zhu, Mingxiao Li, Vasant G Honavar
Neural Information Processing Systems (**NeurIPS**), 2024, Main Track, [CCF-A](#)
- [C4] How to Leverage Demonstration Data in Alignment for Large Language Model? A Self-Imitation Learning Perspective
Teng Xiao, Mingxiao Li, **Yige Yuan**, Huaisheng Zhu, Chao Cui, Vasant G Honavar
Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2024, Main, [CCF-B](#)
- [C5] Augmentation-Aware Self-Supervision for Data-Efficient GAN Training
Liang Hou, Qi Cao, **Yige Yuan**, Songtao Zhao, Chongyang Ma, Siyuan Pan, et al.
Neural Information Processing Systems (**NeurIPS**), 2023, Main Track, [CCF-A](#)
- [C6] Negative as Positive: Enhancing Out-of-distribution Generalization for Graph Contrastive Learning
Zixu Wang, Bingbing Xu, **Yige Yuan**, Huawei Shen and Xueqi Cheng
International ACM Conference on Research and Development in Information Retrieval (**SIGIR**), 2024, Short Paper, [CCF-A](#)
- [C7] History Driven Sampling for Scalable Graph Neural Networks
Yang Li, Bingbing Xu, Fei Sun, Qi Cao, **Yige Yuan**, and Huawei Shen
International Conference on Database Systems for Advanced Applications (**DASFAA**), 2024, Reseach Track, [CCF-B](#)

HONORS & AWARDS

National Scholarship, *Ministry of Education of the People's Republic of China* 2024
First-Class Scholarship, *University of Chinese Academy of Sciences* 2022, 2024
Presidential Scholarship, *Institute of Computing Technology, Chinese Academy of Sciences* 2023
Outstanding Student Award, *University of Chinese Academy of Sciences* 2022
First Prize, *The 12th National College Students Information Security Contest* 2019

ACADEMIC SERVICES & TALKS

Conference Reviewer: NeurIPS 2024, EMNLP 2024, ICLR 2025, AISTATS 2025, AAAI 2025, COLING 2025, ACL Rolling Review
Journal Reviewer: IEEE Transactions on Knowledge and Data Engineering (TKDE), Applied Intelligence (APIN)
LOGS Webinar, "Partial Differential Equation-Driven Generalizable Neural Networks" Mar 2024
AITime Webinar, "TEA: Test-time Energy Adaptation" April 2024
WizSci Webinar, "PDE+: Enhancing Generalization via PDE with Adaptive Distributional Diffusion" Jan 2024