Yuansheng Ni

Email: yuansheng.ni@uwaterloo.ca   Personal Website: https://yuanshengni.github.io

EDUCATION

University of Waterloo   Waterloo, CA
MS in Computer Science   Sept 2024-Present
● Advisor: Wenhu Chen

Zhejiang University   Zhejiang, China
B.E. in Computer Science and Technology   Sept 2018-Jun 2024

University of Waterloo   Waterloo, CA
Mathematics, Non Degree, Exchange   Sept 2023-Dec 2023

RESEARCH EXPERIENCE

TIGER Lab, University of Waterloo   Waterloo, CA
Research Assistant; Advisor: Wenhu Chen   Sept 2023-Present

ZJU NLP Group, Zhejiang University   Zhejiang, China
Research Assistant; Advisor: Ninyu Zhang   Jul 2023-June 2024

PUBLICATIONS

1. **SuperGPQA: Scaling LLM Evaluation across 285 Graduate Disciplines**
   M-A-P Team
   In NeurIPS D&B 2025. (Core Contributor)

2. **VisCoder: Fine-Tuning LLMs for Executable Python Visualization Code Generation**
   Yuansheng Ni, Ping Nie, Kai Zou, Xiang Yue, Wenhu Chen
   In EMNLP Findings, 2025.

3. **MMMU-Pro: A More Robust Multi-discipline Multimodal Understanding Benchmark**
   Xiang Yue*, Tianyu Zheng*, Yuansheng Ni*, Yubo Wang, Kai Zhang, Shengbang Tong, Yuxuan Sun, Ming Yin,
   Botao Yu, Ge Zhang, Huan Sun, Yu Su, Wenhu Chen, Graham Neubig
   In ACL 2025. (*: Core Contribution)

4. **MEGA-Bench: Scaling Multimodal Evaluation to over 500 Real-World Tasks**
   Jiacheng Chen, Tianhao Liang, Sherman Siu, Zhengqing Wang, Kai Wang, Yubo Wang, Yuansheng Ni, Wang
   Zhu, Ziyan Jiang, Bohan Lyu, Dongfu Jiang, Xuan He, Yuan Liu, Hexiang Hu, Xiang Yue, Wenhu Chen
   In ICLR 2025.

5. **II-Bench: An Image Implication Understanding Benchmark for Multimodal Large Language Models**
   Ziqiang Liu, Feiteng Fang, Xi Feng, Xinrun Du, Chenhao Zhang, Zekun Wang, Yuelin Bai, Qixuan Zhao, Liyang
   Fan, Chengguang Gan, Hongquan Lin, Jiaming Li, Yuansheng Ni, Haihong Wu, Yaswanth Narsupalli, Zhigang
   Zheng, Chengming Li, Xiping Hu, Ruifeng Xu, Xiaojun Chen, Min Yang, Jiaheng Liu, Ruibo Liu, Wenhao
   Huang, Ge Zhang, Shiwen Ni
   In NeurIPS D&B 2024.

6. **GenAI Arena: An Open Evaluation Platform for Generative Models**
   Dongfu Jiang, Max Ku, Tianle Li, Yuansheng Ni, Shizhuo Sun, Rongqi Fan, Wenhu Chen
   In NeurIPS D&B 2024.

7. **MMLU-Pro: A More Robust and Challenging Multi-Task Language Understanding Benchmark**
   Yubo Wang, Xueguang Ma, Ge Zhang, Yuansheng Ni, Abhranil Chandra, Shiguang Guo, Weiming Ren, Aaran
   Arulraj, Xuan He, Ziyan Jiang, Tianle Li, Max Ku, Kai Wang, Alex Zhuang, Rongqi Fan, Xiang Yue, Wenhu
   Chen.
   In NeurIPS D&B(**Spotlight**) 2024.

8. **VideoScore: Building Automatic Metrics to Simulate Fine-grained Human Feedback for Video Generation**
   Xuan He, Dongfu Jiang, Ge Zhang, Max Ku, Achint Soni, Sherman Siu, Haonan Chen, Abhranil Chandra, Ziyan Jiang, Aaran Arulraj, Kai Wang, Quy Duc Do, Yuansheng Ni, Bohan Lyu, Yaswanth Narsupalli, Rongqi Fan, Zhiheng Lyu, Yuchen Lin, Wenhu Chen
   In EMNLP Main, 2024.

9. **A Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI**
   Xiang Yue*, Yuansheng Ni*, Kai Zhang*, Tianyu Zheng*, Ruoqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, Cong Wei, Botao Yu, Ruibin Yuan, Renliang Sun, Ming Yin, Boyuan Zheng, Zhenzhu Yang, Yibo Liu, Wenhao Huang, Huan Sun, Yu Su, and Wenhu Chen
   In CVPR (**Oral**; **Best Paper Finalist**), 2024. (*: Core Contribution)

10. **EasyEdit: An Easy-to-use Knowledge Editing Framework for Large Language Models**
    Peng Wang, Ningyu Zhang, Bozhong Tian, Zekun Xi, Yunzhi Yao, Ziwen Xu, Mengru Wang, Shengyu Mao, Xiaohan Wang, Siyuan Cheng, Kangwei Liu, Yuansheng Ni, Guozhou Zheng, Huajun Chen
    In ACL SDT 2024.

PREPRINTS

1. **VisCoder2: Building Multi-Language Visualization Coding Agents**
   Yuansheng Ni*, Songcheng Cai*, Xiangchao Chen*, Jiarong Liang, Zhiheng Lyu, Jiaqi Deng, Kai Zou, Ping Nie, Fei Yuan, Xiang Yue, Wenhu Chen
   (*: Core Contribution) arXiv:2510.23642, 2025.

2. **PhyX: Does Your Model Have the "Wits" for Physical Reasoning?**
   Hui Shen, Taiqiang Wu, Qi Han, Yunta Hsieh, Jizhou Wang, Yuyue Zhang, Yuxin Cheng, Zijian Hao, Yuansheng Ni, Xin Wang, Zhongwei Wan, Kai Zhang, Wendong Xu, Jing Xiong, Ping Luo, Wenhu Chen, Chaofan Tao, Zhuoqing Mao, Ngai Wong
   arXiv:2505.15929, 2025.

3. **A Comprehensive Study of Knowledge Editing for Large Language Models**
   Ningyu Zhang, Yunzhi Yao, Bozhong Tian, Peng Wang, Shumin Deng, Mengru Wang, Zekun Xi, Shengyu Mao, Jintian Zhang, Yuansheng Ni, Siyuan Cheng, Ziwen Xu, Xin Xu, Jia-Chen Gu, Yong Jiang, Pengjun Xie, Fei Huang, Lei Liang, Zhiqiang Zhang, Xiaowei Zhu, Jun Zhou, Huajun Chen
   arXiv:2401.01286, 2024.

SELECTED HONORS & AWARDS

CVPR Best Paper Finalist	2024
Meritorious Winner, MCM/ICM, Consortium for Mathematics and Its Application	2022
Third Prize, CUMCM, Industrial and Applied Mathematics of P.R.China	2018
Second Prize, Collegiate Mathematics Competition, Department of Education of Zhejiang	2018

INVITED TALKS

MMMU: A Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI
Endless Summer School: Multimodal and Foundation Models, Vector Institute. Jul 2025

SERVICES

Reviewer: ARR'24-25; ACL'25; EMNLP'25; NeurIPS'25

Secondary Reviewer: ACL'24