

# Wanrong Zhu

Research Scientist  
Adobe Research

Email: [wzhu@adobe.com](mailto:wzhu@adobe.com)  
Homepage: <https://wanrong-zhu.com>

## Education

- Ph.D., Computer Science - University of California, Santa Barbara** Sep. 2019 - June 2024  
*Advisor: Professor William Wang.*
- B.S., Computer Science - Peking University** Sep. 2015 - July 2019  
*Advisor: Professor Ming Zhang.*

## Research Interests

Vision and Language; Multimodal Study; Text Generation

## Research Experience

- Research Intern at Adobe Research**, San Jose, CA June 2023 - Sep. 2023  
*Document Intelligence Lab - Hosts: Jennifer Healey, Ruiyi Zhang*
- Research Intern at AI2**, Seattle, WA June 2022 - Sep. 2022  
*MOSAIC Team - Hosts: Jack Hessel, Youngjae Yu*
- Research Intern at Google Research**, Remote June 2021 - Oct. 2021  
*GARCON Team - Hosts: Bo Pang, Ashish Thapliyal*
- Research Intern at Google**, Remote June 2020 - Nov. 2020  
*AdsAI Team - Host: Pradyumna Narayana*
- Research Assistant at UC Santa Barbara**, Santa Barbara, CA Sep. 2019 - Present  
*Natural Language Processing Group - Advisor: William Wang*
- Research Intern at Carnegie Mellon University**, Pittsburgh, PA May 2018 - Sep. 2018  
*Language Technologies Institute - Advisor: Zhiting Hu*
- Research Assistant at Peking University**, Beijing, China April 2017 - June 2019  
*Institute of Network Computing and Information Systems - Advisor: Ming Zhang*

## Publications

- [1] An Yan, Zhengyuan Yang, Junda Wu, **Wanrong Zhu**, Jianwei Yang, Linjie Li, Kevin Lin, Jianfeng Wang, Julian McAuley, Jianfeng Gao, Lijuan Wang. "List Items One by One: A New Data Source and Learning Paradigm for Multimodal LLMs". *The First Conference on Language Modeling (CoLM 2024)*.
- [2] Raphael Schumann, **Wanrong Zhu**, Weixi Feng, Tsu-Jui Fu, Stefan Riezler, William Yang Wang. "VELMA: Verbalization Embodiment of LLM Agents for Vision and Language Navigation in Street View". *The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI 2024)*.

- [3] **Wanrong Zhu\***, Jack Hessel\*, Anas Awadalla, Samir Yitzhak Gadre, Jesse Dodge, Alex Fang, Youngjae Yu, Ludwig Schmidt, William Yang Wang, Yejin Choi. “Multimodal C4: An Open, Billion-scale Corpus of Images Interleaved With Text”. *The Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track (NeurIPS D&B 2023)*. (\*equal contributions)
- [4] Yonatan Bitton\*, Hritik Bansal\*, Jack Hessel\*, Rulin Shao, **Wanrong Zhu**, Anas Awadalla, Josh Gardner, Rohan Taori, Ludwig Schimdt. “VisIT-Bench: A Benchmark for Vision-Language Instruction Following Inspired by Real-World Use”. *The Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track (NeurIPS D&B 2023)*.
- [5] Weixi Feng\*, **Wanrong Zhu\***, Tsu-Jui Fu, Varun Jampani, Arjun Reddy Akula, Xuehai He, Sugato Basu, Xin Eric Wang, William Yang Wang. “LayoutGPT: Compositional Visual Planning and Generation with Large Language Models”. *The Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS 2023)*. (\*equal contributions)
- [6] Xinyi Wang, **Wanrong Zhu**, Michael Saxon, Mark Steyvers, William Yang Wang. “Large Language Models Are Implicitly Topic Models: Explaining and Finding Good Demonstrations for In-Context Learning”. *The Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS 2023)*.
- [7] **Wanrong Zhu**, Xinyi Wang, Yujie Lu, Tsu-Jui Fu, Xin Eric Wang, Miguel Eckstein, William Yang Wang. “Collaborative Generative AI: Integrating GPT-k for Efficient Editing in Text-to-Image Generation”. *The 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023, Short)*.
- [8] **Wanrong Zhu**, An Yan, Yujie Lu, Wenda Xu, Xin Eric Wang, Miguel Eckstein, William Yang Wang. “Visualize Before You Write: Imagination-Guided Open-Ended Text Generation”. *The 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2023, Findings)*.
- [9] **Wanrong Zhu**, Xin Eric Wang, An Yan, Miguel Eckstein, William Yang Wang. “ImaginE: An Imagination-Based Automatic Evaluation Metric for Natural Language Generation”. *The 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2023, Findings)*.
- [10] Yujie Lu, Weixi Feng, **Wanrong Zhu**, Wenda Xu, Xin Eric Wang, Miguel Eckstein, William Yang Wang. “Neuro-Symbolic Causal Language Planning with Commonsense Prompting”. *The 11th International Conference on Learning Representations (ICLR 2023, Spotlight)*.
- [11] **Wanrong Zhu**, Bo Pang, Ashish Thapliyal, William Yang Wang, Radu Soricut. “End-to-end Dense Video Captioning as Sequence Generation”. *The 29th International Conference on Computational Linguistics (COLING 2022)*.
- [12] Yujie Lu, **Wanrong Zhu**, Xin Eric Wang, Miguel Eckstein, William Yang Wang “Imagination-Augmented Natural Language Understanding”. *The 2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2022, Oral)*.
- [13] **Wanrong Zhu**, Yuankai Qi, Pradyumna Narayana, Kazoo Sone, Sugato Basu, Xin Eric Wang, Qi Wu, Miguel Eckstein, William Yang Wang. “Diagnosing Vision-and-Language Navigation: What Really Matters”. *The 2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2022, Oral)*.
- [14] **Wanrong Zhu**, Xin Eric Wang, Tsu-Jui Fu, An Yan, Pradyumna Narayana, Kazoo Sone, Sugato Basu, William Yang Wang. “Multimodal Text Style Transfer for Outdoor Vision-and-Language Navigation”. *The 16th conference of the European Chapter of the Association for Computational Linguistics (EACL 2021)*.
- [15] **Wanrong Zhu**, Xin Eric Wang, Pradyumna Narayana, Kazoo Sone, Sugato Basu, William Wang. “Towards Understanding Sample Variance in Visually Grounded Language Generation: Evaluations and Observations”. *The 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020, Short)*.
- [16] Zhiting Hu, Haoran Shi, Bowen Tan, Wentao Wang, Zichao Yang, Tiancheng Zhao, Junxian He, Lianhui Qin, Di Wang, Xuezhe Ma, Zhengzhong Liu, Xiaodan Liang, **Wanrong Zhu**, Devendra Singh Sachan, Eric P. Xing. “Texar: A Modularized, Versatile, and Extensible Toolkit for Text Generation”. *The 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, (ACL 2019)*.

## Preprints & Technical Reports

- [1] Zekun Li, Xianjun Yang, Kyuri Choi, **Wanrong Zhu**, Ryan Hsieh, HyeonJung Kim, Jin Hyuk Lim, Sungyoung Ji, Byungju Lee, Xifeng Yan, Linda Ruth Petzold, Stephen D. Wilson, Woosang Lim, William Yang Wang. "MMSci: A Multimodal Multi-Discipline Dataset for PhD-Level Scientific Comprehension". *arXiv preprint*.
- [2] Xuehai He, Weixi Feng\*, Kaizhi Zheng\*, Yujie Lu\*, **Wanrong Zhu\***, Jiachen Li\*, Yue Fan\*, Jianfeng Wang, Linjie Li, Zhengyuan Yang, Kevin Lin, William Yang Wang, Lijuan Wang, Xin Eric Wang. "MMWorld: Towards Multi-discipline Multi-faceted World Model Evaluation in Videos". *arXiv preprint*.
- [3] **Wanrong Zhu**, Jennifer Healey, Ruiyi Zhang, William Yang Wang, Tong Sun. "Automatic Layout Planning for Visually-Rich Documents with Instruction-Following Models". *arXiv preprint*.
- [4] An Yan, Zhengyuan Yang, **Wanrong Zhu**, Kevin Lin, Linjie Li, Jianfeng Wang, Jianwei Yang, Yiwu Zhong, Julian McAuley, Jianfeng Gao, Zicheng Liu, Lijuan Wang. "GPT-4V in Wonderland: Large Multimodal Models for Zero-Shot Smartphone GUI Navigation". *arXiv preprint*.
- [5] Anas Awadalla, Irena Gao, Josh Gardner, Jack Hessel, Yusuf Hanafy, **Wanrong Zhu**, Kalyani Marathe, Yonatan Bitton, Samir Gadre, Shiori Sagawa, Jenia Jitsev, Simon Kornblith, Pang Wei Koh, Gabriel Ilharco, Mitchell Wortsman, Ludwig Schmidt. "OpenFlamingo: An Open-Source Framework for Training Large Autoregressive Vision-Language Models". *Technical Report*.
- [6] Yujie Lu, Pan Lu, Zhiyu Chen, **Wanrong Zhu**, Xin Eric Wang, William Yang Wang. "Multimodal Procedural Planning via Dual Text-Image Prompting". *arXiv preprint*.
- [7] An Yan, Jiacheng Li, **Wanrong Zhu**, Yujie Lu, William Yang Wang, Julian McAuley. "CLIP also Understands Text: Prompting CLIP for Phrase Understanding". *arXiv preprint*.
- [8] **Wanrong Zhu**, Zhiting Hu, Eric P. Xing. "Text Infilling". *arXiv preprint*.

## Awards and Honors

Rising Star in Machine Learning by University of Maryland	Nov. 2023
Regent's Fellowship, UC Santa Barbara	Sep. 2019
TP-LINK Scholarship, Peking University	Sep. 2018
Kuang-Hua Scholarship, Peking University	Sep. 2016

## Invited Talks

<i>University of Maryland</i> : "Towards Collaborative Generative AI for Vision-and-Language Studies"	Nov. 2023
<i>Apple AI/ML</i> : "Multimodal C4: Interleaving Text and Images for Advaced AI Capacities"	Aug. 2023
<i>Guest Lecture for UCSB CS190I</i> : "Introduction to Vision-and-Language"	Spring 2023
<i>Guest Lecture for UCSB CS165B</i> : "Introduction to Vision-and-Language"	Spring 2022

## Teaching

<b>Teaching Assistant</b> , UC Santa Barbara <i>CS190I: Introduction to Natural Language Processing</i>	Winter 2022
--	-------------

## Academic Service

### Organizing Workshops

Computer Vision in the Wild @CVPR 2024

### Reviewing for Conferences

AAAI, ICLR, ARR, ICML, CoLM 2024

NeurIPS, EACL, AAAI 2023

## Industry Experience

**Software Engineer Intern at Google, Beijing, China** April 2018 - June 2018  
*Mobile Harness Team*

**Engineering Practicum Intern at Google, Beijing, China** July 2017 - Sep. 2017  
*Search Infrastructure Team*

Last updated: July 10, 2024