

## Education Background

---

### University of Southern California (USC)

#### Ph.D. in Computer Science

- Advisor: Dr. Jay Pujara

*Los Angeles, USA*

*08/2023~Current*

### University of Southern California (USC)

#### M.S. in Computer Science

**Overall GPA: 3.9/4.0**

- Advisor: Dr. Filip Ilievski

*Los Angeles, USA*

*01/2022~12/2023*

### Southern University of Science and Technology (SUSTech)

#### B.S. in Computer Science

**Overall GPA: 3.78/4.0**

- Advisor: Dr. Xuan Song

*Shenzhen, China*

*09/2017~06/2021*

### University of California, Berkeley (UCB)

#### Exchange program in Computer Science

**Overall GPA: 4.0/4.0**

*San Francisco, USA*

*01/2020~05/2020*

## Experience

---

### Amazon Prime Video

Research Intern

*Seattle, USA*

*05/2025~Current*

### Hippocratic AI

Research Intern

*Palo Alto, USA*

*05/2024~08/2024*

### Information Sciences Institute, USC

Research Assistant

*Los Angeles, USA*

*05/2022~08/2023*

### Computer Science Department, USC

Teaching Assistant

*Los Angeles, USA*

*06/2022~12/2022*

### SUSTech-Tokyo Joint Research Center for Super Smart City

Research Assistant

*Shenzhen, China*

*01/2021~12/2021*

## Publication

---

- ORBIT: An Object Property Reasoning Benchmark for Visual Inference Tasks  
Abhishek Kolari, Mohammadhossein Khojasteh, Yifan Jiang, Floris den Hengst, Filip Ilievski  
[Preprint]
- SRPO: Enhancing multimodal llm reasoning via reflection-aware reinforcement learning  
Zhongwei Wan, Zhihao Dou, Che Liu, Yu Zhang, Dongfei Cui, Qinjian Zhao, Hui Shen, Jing Xiong, Yi Xin, Yifan Jiang, Yangfan He, Mi Zhang, Shen Yan  
[Preprint]
- A Comprehensive Survey in LLM (-Agent) Full Stack Safety: Data, Training and Deployment  
[Preprint]
- Engage for All: Making Ordinary Image Descriptions Appealing Again!  
Yuyan Chen, Yifan Jiang, Li Zhou, Jinghan Cao, Yu Guan, Ming Yang, Qingpei Guo  
[ICCV 2025] International Conference on Computer Vision, 2025.

- Recent Advances in Large Language Model Benchmarks against Data Contamination: From Static to Dynamic Evaluation  
 Simin Chen, Yiming Chen, Zexin Li, [Yifan Jiang](#), Zhongwei Wan, Yixin He, Dezhi Ran, Tianle Gu, Haizhou Li, Tao Xie, Baishakhi Ray  
 [EMNLP 2025] Conference on Empirical Methods in Natural Language Processing, 2025.
- RED QUEEN: Safeguarding Large Language Models against Concealed Multi-Turn Jailbreaking  
[Yifan Jiang](#), Kriti Aggarwal, Subhabrata Mukherjee, Tanmay Laud, Kashif Munir, Jay Pujara  
 [ACL 2025 Findings] The 63rd Annual Meeting of the Association for Computational Linguistics.
- COLUMBUS: A Visual Question Answering Benchmark to Evaluate COgnitive Lateral Understanding through Multiple-choice reBUSES  
 Koen Kraaijveld, [Yifan Jiang](#), Kaixin Ma, Filip Ilievski  
 [AAAI 2025] AAAI Conference on Artificial Intelligence, 2025.
- MARVEL: Multidimensional Abstraction and Reasoning through Visual Evaluation and Learning  
[Yifan Jiang](#)<sup>\*</sup>, Jiarui Zhang<sup>\*</sup>, Kexuan Sun<sup>\*</sup>, Zhivar Sourati, Kian Ahrabian, Kaixin Ma, Filip Ilievski, Jay Pujara  
 [NeurIPS DB 2024] Conference on Neural Information Processing Systems, 2024.
- The curious case of nonverbal abstract reasoning with multi-modal large language models  
 Kian Ahrabian<sup>\*</sup>, Zhivar Sourati<sup>\*</sup>, Kexuan Sun<sup>\*</sup>, Jiarui Zhang, [Yifan Jiang](#), Fred Morstatter, Jay Pujara  
 [COLM 2024] Conference On Language Modeling, 2024.
- ARN: Analogical Reasoning on Narratives  
 Zhivar Sourati, Filip Ilievski, Pia Sommerauer, [Yifan Jiang](#)  
 [TACL 2024] Transactions of the Association for Computational Linguistics, 2024.
- Semeval-2024 task 9: Brainteaser: A novel task defying common sense  
[Yifan Jiang](#), Filip Ilievski, Kaixin Ma  
 [SemEval@NAACL 2024] The 18th International Workshop on Semantic Evaluation, 2024.
- BRAINTEASER: Lateral Thinking Puzzles for Large Language Models  
[Yifan Jiang](#), Filip Ilievski, Kaixin Ma, Zhivar Sourati  
 [EMNLP 2023] Conference on Empirical Methods in Natural Language Processing, 2023.
- Transferring Procedural Knowledge across Commonsense Tasks  
[Yifan Jiang](#), Kaixin Ma, Filip Ilievski  
 [WNU@ACL 2023] ACL 2023 5th Workshop on Narrative Understanding.  
 [ECAI 2023] European Conference on Artificial Intelligence, 2023.
- FIRE: Food Image to REcipe generation  
 Prateek Chhikara, Dhiraj Chaurasia, [Yifan Jiang](#), Omkar Masur  
 [WACV 2024] IEEE/CVF Winter Conference on Applications of Computer Vision, 2024.
- TrafPS: A Shapley-based Visual Analytics Approach to Interpret Traffic  
 Zezheng Feng, [Yifan Jiang](#), Hongjun Wang, Zipei Fan, Yuxin Ma, Shuang-Hua Yang, Huamin Qu, Xuan Song  
 [CVMJ 2023] Computational Visual Media Journal, 2023.

## **Service**

---

- [2025] Reviewer for COLIING, COLING\_demo, ICLR, IEEE\_SSCI, Analogy-ANGLE Work shop II
- [2024] Reviewer for SemEval, LREC-COLING, Analogy-ANGLE Workshop(co-located with IJCAI)

## **Honors**

---

- The Honor Roll Merit Scholarship, Southern University of Science and Technology, 2018,2019,2020
- Computer Science Outstanding Graduate Award, Southern University of Science and Technology, 2021
- Computer Science Honor Program, University of Southern California, 2022
- ISI Distinguished Graduate Student Researcher Fellowship Award, University of Southern California, 2023