ZIJUN WANG

zijun714@gmail.com · asillycat.github.io · github.com/asillycat ·

Education

Zhejiang University, Hangzhou, China 2020.09 – Present

Undergraduate
Major in Computer Science and Technology, College of Computer Science and Technology
Minor in ITP (Intensive Training Program for Innovation and Entrepreneurship), Chu Kochen Honors College
GPA: 3.91/4.00 **Credits:** 204 / 172.5

Awards

- **National Scholarship (top 0.2% national-wide)** issued by Ministry of Education of the People's Republic of China
- **Provincial Government Scholarship (top 3%)** of Zhejiang Province
- **First-class Scholarship (top 3%)** of Zhejiang University

Reasearch Interests

AI Safety, Natural Language Processing (NLP), Multi-modal learning and their applications.

Experience

Visiting Research Intern Santa Cruz, CA

VLAA LAB, UC Santa Cruz *2023.08-Present*

- Under Supervision of Prof. Cihang Xie and Prof. Yuyin Zhou
- Worked on **Adversarial Attacks on LLMs & VLLMs**
- One paper in submission to **CVPR 2024**
- **Second Place** in both base large model subtracks of Red Teaming LLM@**NeurIPS 2023**, Torjan Detection Challenge**(Team leader)**.

Undergraduate Research Assistant Zhejiang

Zhejiang University *2023.01-Present*

- Under Supervision of Prof. Yang Yang
- Worked on **Genaralized Graph Pre-training**
- One paper preparing to submit to **ICML 2024**

Publications

How Many Unicorns Are in This Image? A Safety Evaluation Benchmark for Vision LLMs

Haoqin Tu, Chenhang Cui*, **Zijun Wang ***, Yiyang Zhou, Bingchen Zhao, Junlin Han, Wangchunshu Zhou, Huaxiu Yao, Cihang Xie (* represents equal contribution)*
In submission to *IEEE / CVF Computer Vision and Pattern Recognition Conference 2024* **(CVPR 2024)**
TL;DR: This work focuses on the potential of VLLMs in visual reasoning. Different from prior studies, we shift our focus from evaluating standard performance to introducing a comprehensive safety evaluation suite, covering both out-of-distribution (OOD) generalization and adversarial robustness.

GRAPHGENT: Foundation Model for Graph Pre-training

*Yifei Sun, **Zijun Wang**, Xiao Feng, Chunping Wang, Lei CHEN, Jie Tang, Yang Yang,*
Preparing to submit to *International Conference on Machine Learning 2023* **(ICML 2024)**
TL;DR: We design the GRAPHGENT(Graph Generalized pre-Training), a foundation model for generalized graph pre-training that leverages patch encoder and patch aggregator to learn transferable knowledge from different graphs.