Yichen Wang

Room 424, Paul G. Allen Center, 185 E Stevens Way NE, Seattle, WA 98195
yichen.wang@berkeley.edu

Education

Xi'an Jiaotong University (XJTU) — **Xi'an, China**
2020 - 2024 (Expected)

- B.S. in Computer Science
- **Computer Science Honors Program** (elite class for top 5% talented students)
- Overall GPA: 88.8/100 (**Ranking: 3rd/45**) , Junior Year: 95.0/100
- Outstanding Student (top 3%), Outstanding Undergrad Scholarship
- Advisor: Prof. Chao Shen and Prof. Xiaoming Liu, *Faculty of Electronic and Information Engineering, EECS*

University of California, Berkeley — **Berkeley, CA, US**
Jan. 2023 - July 2023

- Exchange Student in Computer Science, **GPA: 4.0/4.0**
- **A+** in CS288 NLP (Grad Level), A (rank top 5%) in CS188 Intro to AI
- **Intern in the Berkeley NLP Group, BAIR**
- Advisor: Prof. Dan Klein and Ph.D. Kevin Yang, *Computer Science Division, EECS*

Research Interests

Natural language processing (NLP) in general, including long-form generation, machine-generated text detection, and automatic evaluation, as well as relevant theories of machine learning (i.e., zero-shot learning, out-of-distribution detection), NLP security issues (i.e., watermark on LLM, backdoor attack), interpretation of LM, graph-based text representation, and multi-modality.

Publications

Improving Pacing in Long-Form Story Planning

- **Yichen Wang**, Kevin Yang, Xiaoming Liu, Dan Klein. [paper link]
- In submission to ***EMNLP23***. Finished at UC Berkeley NLP Group.

CoCo: Coherence-Enhanced Machine-Generated Text Detection Under Low Resource With Contrastive Learning

- Xiaoming Liu*, Zhaohan Zhang*, **Yichen Wang***(Equal Contribution), Hang Pu, Yu Lan, Chao Shen. [paper link]
- In submission to ***EMNLP23***. Assessed meta score **4/4** at rolling review.

Dialogue for Prompting: a Policy-Gradient-Based Discrete Prompt Optimization for Few-shot Learning

- Xiaoming Liu*, Zhengxu Licheng*, **Yichen Wang***, Yu Lan, Chao Shen. [paper link]
- In submission to ***AAAI24***.

Research Internship

University of Washington — **Seattle, WA, US**
June 2023 - Present

- Research Intern in the TsvetShop Group of Paul G. Allen School of CSE
- Advisor: Prof. Yulia Tsvetkov and Postdoc Tianxing He, *Paul G. Allen School of CSE*

University of Cambridge — **Cambridge, UK**
May 2022 - Nov. 2022

- Research Intern (remote), AI, Dept. of Computer Science and Technology
- Straight A's in Research Final Assessment (1st/8)
- Advisor: Prof. Pietro Lio', *Artificial Intelligence Group*

Research Experiences

Machine-Generated Text Detection with Coherence Graph Representation
Advisor: Prof. Xiaoming Liu, Prof. Chao Shen, EECS, XJTU — *April 2022 - Feb. 2023*

- Proposed a novel machine-generated text detector framework with graph-based coherence representation, applying a novel supervised contrastive learning method to handle imbalanced low data situations, a real-world scenario often neglected.
- Analyzed the distinguishability of our graph representation to evaluate the semantic coherence of generated texts.
- Constructed news-type detection-task datasets with GPT3.5-DaVinci mimicking various provenance and writing styles.
- Evaluated the detector's robustness against perturbation attack; interpreted detection mechanisms via integrated gradients.
- Personally took charge of coding (w/o any code base), all of the experiments, building open-source datasets, proposing novelty in coherence and graph, graph evaluation, perturbation attack, and co-writing manuscripts for the paper and the rebuttals.

Pacing – Enhancing Long-Form Story Planning
Advisor: Ph.D. Kevin Yang, Prof. Dan Klein, The Berkeley NLP Group *Jan. 2023 - June 2023*

- Proposed a reference-free zero-shot metric to evaluate pacing (i.e., level of concreteness) for generated long texts.
- Built hierarchical summarization datasets of story books via GPT3.5 and designed a dynamic pairing train procedure.
- Designed a novel adversarial concreteness evaluation benchmark and outperformed all baselines (including GPT4).
- Utilized the evaluator to build a novel pacing-controlled hierarchical generation framework based on outlines, proposing tree-structure outlining, vague-first expansion, and filtered story generation. Our final stories are favorable in human eval.
- Showed flexibility to adjust pacing with more sophisticated expansion strategies, which can account for the nebulous concepts of "engagingness" or "interestingness" on top of simply maintaining uniform pacing.

Toward Robust Detection: An Attack Benchmark on Machine-Generated Text Detection
Postdoc Tianxing He, Prof. Yulia Tsvetkov, Paul G. Allen School of CSE *Work in process*

- Aimed at fair comparison and comprehensive robustness on machine-generated text detection methods, proposed a benchmark with all aspects of invisible attack (e.g., from token- to paragraph-level, from prompt to post-generation edit).
- Presented the backdoors and shortage of all three mainstream detection categories, i.e., probability-based (like DetectGPT), model-based(like OpenAI-Detector), and watermark-based. Further, disclosed the interpretation of specific shortages.
- Proposed some out-of-box defense methods toward proposed attacks, showing outstanding performance.

Watermark on Large Language Model via Semantic and Syntax
Postdoc Tianxing He, Prof. Yulia Tsvetkov, Paul G. Allen School of CSE *Work in process*

- Priorly, proposed a sentence semantic watermark, showing better robustness to paraphrase. In submission to ICLR2024.
- Further, watermarking generated text on syntax structure instead of token selection, which minor the impact on semantics.
- Utilized the selection and order of entities as fingerprints to improve watermark robustness against simple paraphrasing.

Reinforcement Learning Discrete Prompt Optimization with Dialogue
Advisor: Prof. Xiaoming Liu, Prof. Chao Shen, EECS, XJTU *April 2023 - Sep. 2023*

- Proposed an efficient prompt-set generation method via pre-selected sample candidates to decrease collection cost, manual annotation cost, and better align prompts' distribution with the few-shot sample set via a designed multi-round dialogue.
- Utilized reinforcement learning to optimal prompt selection on the sample level, i.e., matching each sample with a prompt.

Graph-based Text Representation Static Analysis and Comparison
Advisor: Prof. Pietro Lio', Ph.D. Kehai Qiu, University of Cambridge *May 2022 - Nov. 2022*

- Improved entity graph's density to represent the semantic structure, inspired by research on relation extraction.
- Proposed a framework applying relation-aware graph neural network for text classification.
- Introduced static feature analysis of complex networks domain to our graph-based text representation for semantics.

Out-of-Distribution Text Detection in the Open World
Advisor: Prof. Xinyu Dai, NLP, Nanjing Univ. & Postdoc O.Yawen, Tsinghua Univ. *May 2022 - Sep. 2022*

- Survey existing out-of-distribution (OOD) detection models and reproduce them for evaluation on the same benchmark.
- Built a class-incremental learning model with regularization approaches and rehearsal approaches.
- Constructed an improved Stream Emerging New Class Problem (SENC) model architecture based on Manhattan distance.
- Awarded **Best Project Award (1/12)** of 2022.

Fast Question Answering via Novel Local Sensitive Hashing Sketch
Advisor: Prof. Pinghui Wang, Prof. Jing Tao, EECS, XJTU *Mar. 2021 - Feb. 2022*

- Approached question answering as retrieval problem via similarity and aimed at a sublinear-time algorithm. Inspired by our prior work, *Bidirectionally Densifying LSH Sketches with Empty Bins* on SIGMOD21.
- Took charge of applying sign random projections and local sensitive hashing on ColBERT.
- Personally received the annual **Rising Star Undergrad Researcher** award (top 5%) of 2021.

Honors and Awards

- **Social Enterprise Scholarship (first-class)**, *RaySilicon & XJTU.* *Sep. 2022*
- **Best Project Award (1/12)**, *NLP Group, Nanjing University.* *Sep. 2022*
- **Rising Star Undergrad Researcher (top 5%)**, *Lab for Intelligent Networks, XJTU.* *Jan. 2022*
- **Best Developer Award (top 3%)**, *Network Association, XJTU.* *Dec. 2021*

Leadership and Activities

- **Product Manager, AI Service Dialogue System**, *Network Association, XJTU.* *2021 - 2022*
- **Volunteer Leader, Bridge China Program**, *Bridge China Charitable Foundation, HK.* *2021 - 2022*
- **Organizer, Volunteer Service Center**, *XJTU.* *2020 - 2022*

Skills

- **Programming:** Python, C++, C, Java, MATLAB, JavaScript, NODE, HTML, PHP, RISC_V
- **Language:** Mandarin (native), English (advanced), German and Japanese (elementary)