JIAYU LIU

+1 217-731-3555 | jiayu13@illinois.edu | Homepage

EDUCATION

University of Illinois at Urbana-Champaign Aug. 2023 - May 2026 (Expected)
***B.S.** in Statistics, Minor in Computer Science* *GPA: 3.9/4.0*

Xi'an Jiaotong-Liverpool University Sep. 2021 - May 2023
Data Science and Big Data Technology *Transferred*

PUBLICATIONS

** indicates equal contribution*

[6] Haofei Yu, Chongrui Ye*, Jiayu Liu*, Yexin Wu, Zahaib Akhtar, Mona Pereira, Jiaxuan You. **Knowledge Debugger: Efficient Diagnosis of Knowledge Inconsistency with Multimodal Graph.** *preprint*.

[5] Chongrui Ye*, Yuxiang Liu*, Haofei Yu*, Jiayu Liu, Jiaxuan You. **Can Large Language Models Uncover the Structure of Social Opinions?** *preprint*.

[4] Jiayu Liu, David Jurgens. **MUNCH: Benchmarking Multimodal Humor Understanding in Visual Language Models through Meme Comprehension.** *preprint*.

[3] Jonathan Ivey*, Shivani Kumar*, Jiayu Liu*, Hua Shen*, Sushrita Rakshit*, Rohan Raju*, Haotian Zhang*, Aparna Ananthasubramaniam*, Junghwan Kim*, Bowen Yi*, Dustin Wright*, Abraham Israeli*, Anders Giovanni Møller*, Lechen Zhang*, David Jurgens. **Real or Robotic? Assessing Whether LLMs Accurately Simulate Qualities of Human Responses in Dialogue.** (Randomized Author Order) *preprint*. [arXiv]

[2] Jiayu Liu, Rongqian Ma, Keli Du. **Detecting "Parasitic Poems": Quantifying Poetic Style in Late Imperial Chinese Fiction.** *CHR 2025*. [ACH Anthology]

[1] Sarah Griebel, Becca Cohen, Lucian Li, Jaihyun Park, Jiayu Liu, Jana Perkins, Ted Underwood. **Locating the Leading Edge of Cultural Change.** *CHR 2024*. [arXiv]

RESEARCH EXPERIENCES

Synthesizing Diverse Reddit User Profiles and Posts Jun. 2025 - present
Advisor: Wei Xu and Alan Ritter *Remote*

- Extracted self-disclosure posts from Reddit and clustered inferred attributes to ensure diversity and quality across user profiles.
- Synthesized user profiles and evaluated them for realism, coherence, and privacy risk.
- Generated synthetic posts from the profiles to test quality and diversity in downstream tasks.

Connecting Social Opinions with LLMs Feb. 2025 - Jul. 2025
Advisor: Jiaxuan You *Champaign, IL*

- Introduced OPINIONBENCH, a benchmark constructed from Polymarket to detect the hidden structure within evolving social opinions, capturing correlations between evolving social beliefs across politics, elections, cryptocurrency, and sports.
- Developed a hybrid labeling framework combining time-series co-movement, semantic similarity, and human verification to establish graded correlation scores with strong inter-annotator agreement.
- Benchmarked state-of-the-art LLMs, showing they outperform baselines in detecting opinion correlations, inferring the underlying graph structure through edge prediction.

Detecting Inconsistencies in Multimodal Graphs Feb. 2025 - May 2025

Advisor: Jiaxuan You *Champaign, IL*

- Designed Knowledge Debugger, a graph neural network (GNN)-based framework that efficiently identifies and corrects inconsistencies in multimodal, structured knowledge sources such as research papers and Wikipedia pages.
- Constructed the Multimodal Knowledge Debugging Benchmark (MKDB), containing 699 Wikipedia pages, 10K+ research papers, and 10K+ debugging tasks across text, tables, and images, enabling systematic evaluation of inconsistency detection.
- Achieved state-of-the-art performance, surpassing retrieval-augmented generation (RAG) baselines by 11% on node-level bug detection while significantly reducing inference cost and improving scalability for real-world deployment.

Simulating Human Responses in Dialogue with LLMs Jul. 2024 - Oct. 2024

Advisor: David Jurgens *Ann Arbor, MI*

- Developed an evaluation framework using 21 linguistic metrics (lexical, syntactic, semantic, and stylistic) to compare LLM-simulated dialogue with real human responses from the WildChat dataset.
- Analyzed 828K simulated conversations across 9 LLMs, 50 prompts, and 3 languages (English, Chinese, Russian), revealing systematic gaps between LLM-generated and human dialogue styles.
- Identified key factors influencing human-likeness, showing that prompt design often impacts simulation quality more than model size, and that LLMs perform better in creative/storytelling contexts than in technical or structured tasks.

Understanding Multimodal Humor in Memes May 2024 – Feb. 2025

Advisor: David Jurgens *Ann Arbor, MI*

- Constructed MUNCH, a large-scale dataset of 127K memes curated from 886 templates and 275 semantic clusters, enabling systematic evaluation of multimodal humor understanding.
- Developed a visual question answering (VQA) benchmark where models must identify the correct meme caption, requiring humor, sarcasm, and commonsense reasoning across modalities.
- Benchmarked state-of-the-art VLLMs (e.g., GPT-4o, Gemini, Qwen, LLaVA), showing that fine-tuning significantly improves performance and generalization to unseen meme formats and semantics.

Exploring Poetic Style in Chinese Fictions Apr. 2024 – Jul. 2025

Advisor: Rongqian Ma and Keli Du *Remote*

- Developed a computational framework combining cosine similarity, mutual information, and large language model (LLM) prompting to detect "parasitic poems" in Qing-dynasty fiction.
- Constructed and annotated a dataset of 300 poem-context pairs from 18 fictions, enabling systematic evaluation of stylistic redundancy and narrative contribution.
- Demonstrated the potential of multilingual LLM prompts to approximate literary interpretation, achieving balanced performance compared to traditional proxy-based models.

Measuring leading edge of Culture Change Aug. 2023 – July 2024

Advisor: Ted Underwood *Champaign, IL*

- Utilized topic models, document embeddings, and word-level perplexity across three distinct corpora (literary studies, economics, fiction) to evaluate which works are "textually ahead of the curve".
- Found that both highly-cited and younger authors tend to produce texts that are ahead of broader trends, regardless of the representation method used.
- Showed that alignment with external social evidence (e.g., citations, cultural impact) is strongest when focusing on a text's top quartile of most forward-looking passages, suggesting that brief innovative moments matter more than sustained novelty.

PROJECTS

LLM Contemporary Etiquette Understanding and Analysis Aug. 2024 – Dec. 2024
Advisor: Ted Underwood *Champaign, IL*

- Conducted a study on LLMs ability to interpret historical versus contemporary etiquette, creating a dataset of scenario pairs from 19th-century conduct books.
- Implemented evaluation to test temporal generalization in LLMs, revealing distinct performance differences when time-period context is included.
- Identified and analyzed model biases in reasoning across historical contexts, contributing insights into temporal generalization and prompt design for AI applications in cultural and social norm analysis.

ACADEMIC AWARDS & HONORS

Dean's List, UIUC 2025
University Academic Excellence Award, XJTLU 2022
Academic Excellence Award for Year 1 Mathematics, XJTLU 2022

TECHNICAL SKILLS

Programming: Java, Python, C/C++, R, MATLAB, LaTeX
Framework: Pytorch, Tensorflow, Transformers, Deepseed, PEFT, NLTK, Scikit-Learn
Languages: Chinese (native), English (fluent)