Xinliang (Frederick) Zhang

614-705-5429 | xlfzhang@umich.edu | https://web.eecs.umich.edu/~xlfzhang/ | *Google Scholar*

(Last updated: 03/17/2024)

Research Overview:

Structural Learning through the lens of **Event-Centric Information Extraction (IE)**: My study spans multiple facets of **events**: stance-based entity interaction [C6], partisan event detection [C8, C9], moral event extraction [C10], event argument extraction [P1] and complex-event script learning. I highlight knowledge augmentation for robust structural learning, utilizing graph neural network (GNN) and retrieval augmented generation (RAG). Beyond events, my recent interests stretch to **code generation** and **structural commonsense understanding**.

(Large) Language Model: I explore continual pretraining [C4], self-refinement [P1] and planning-as-an-agent.

Question Answering (QA): As a "full-stack" QA researcher, my work ranged from information retrieval [C2] to machine reading comprehension [C3; Best Paper]. My major attempt was to improve the model generalizability through synthetic query/question generations. I also have sufficient experience applying QA in clinical domain.

Computational Social Science: I investigate inherent annotation disagreements [C1, C7] to model human behavior. Several works of mine venture into social science topics like political bias [C4, C8, C9] and social norms [C10].

Education Background:

University of Michigan, Ann Arbor, MI, U.S.A. **Sept. 2021-present**

Ph.D. in Computer Science and Engineering (AI track) Dec. 2025 (Expected)

M.S. in Computer Science and Engineering (granted on my way to Ph.D. degree) Apr. 2023

Research Advisor: Dr. Lu Wang

The Ohio State University, Columbus, OH, U.S.A. **Aug. 2018-May 2021**

B.S. in Computer Science and Engineering (Specialization in AI) May 2021

B.S. in Industrial and Systems Engineering (Minor in Math) May 2021

Academic Performance: GPA: 3.99/4.0 (Summa Cum Laude)

Thesis: Towards More Robust Natural Language Understanding (Graduate w/ Honors Research Distinction)

Thesis committee: Dr. Huan Sun (Advisor), Dr. Marie-Catherine de Marneffe

Sichuan University, Chengdu, Sichuan, P.R.China **Sept. 2016-July 2018**

B.Eng. in Industrial Engineering June 2021

Academic Performance: GPA: 3.97/4.0 Weighted Average: 96.15/100 (STEM: 98.31/100)

Academic Performance and Comprehensive Performance both rank 1st (1/159)

Hangzhou Xuejun High School, Hangzhou, Zhejiang, P.R.China **Sept. 2013-June 2016**

High school degree in Natural Science May 2016

Won 3rd Provincial Prize in China Mathematics Olympiad; Excellent Student Award; Served as Class Leader

Work Experience:

Adobe Research, San Jose, CA, U.S.A. May 2024-Aug. 2024

Research Intern (Mentor: Dr. Victor Bursztyn; Supervisor: Dr. Eunyee Koh)

Bloomberg AI, New York, NY, U.S.A. May 2023-Aug. 2023

Research Intern (Mentor: Dr. Alakananda Vempala; Supervisor: Temma Choji)

Research Project: Unleashing open-sourced large language model's potential for event argument extraction task through "prompting-then-refinement" technique [P1] (submitted to *ACL 2024*).

Publications (*: Equal contributions; #: My mentees)**:**

Pre-print

[P1] **Xinliang Frederick Zhang**, Carter Blum, Temma Choji, Shalin Shah, Alakananda Vempala. 2023. ULTRA: Unleashing LLM's Potential for Event Argument Extraction through Hierarchical Modeling and Pairwise Refinement. Pre-print on *arXiv* & submitted to *ACL 2024.*

Peer-reviewed Conference Paper

[C10] **Xinliang Frederick Zhang**, Winston Wu, Nick Beauchamp, Lu Wang. 2023. MOKA: Moral Knowledge Augmentation for Moral Event Extraction. In *Proceedings of NAACL 2024.* [Code & Data]

[C9] Yujian Liu, **Xinliang Frederick Zhang**, Kaijian Zou, Ruihong Huang, Nick Beauchamp, Lu Wang. 2023. All Things Considered: Detecting Partisan Events from News Media with Cross-Article Comparison. In *Proceedings of EMNLP 2023* (Poster; acceptance rate: 23.3%). [Code & Data]

[C8] Kaijian Zou, **Xinliang Frederick Zhang**, Winston Wu, Nick Beauchamp, Lu Wang. 2023. Crossing the Aisle: Unveiling Partisan and Counter-Partisan Events in News Reporting. In *Findings of EMNLP 2023* (Poster; acceptance rate: 14.0%, findings acceptance rate: 19.5%). [Data]

[C7] Naihao Deng, **Xinliang Frederick Zhang**, Siyang Liu, Winston Wu, Lu Wang, Rada Mihalcea. 2023. You Are What You Annotate: Towards Better Models through Annotator Representations. In *Findings of EMNLP 2023* (Poster; acceptance rate: 23.3%, findings acceptance rate: 22.9%). [Code & Data]

[C6] **Xinliang Frederick Zhang**, Nick Beauchamp, Lu Wang. 2022. Generative Entity-to-Entity Stance Detection with Knowledge Graph Augmentation. In *Proceedings of EMNLP 2022* (Poster; acceptance rate: 20.9%). Also presented in *Text As Data (TADA) 2022* (Oral). [Code & Data]

[C5] Changyuan Qiu*,#, Winston Wu*, **Xinliang Frederick Zhang**, Lu Wang. 2022. Late Fusion with Triplet Margin Objective for Multimodal Ideology Prediction and Analysis. In *Proceedings of EMNLP 2022* (Poster; acceptance rate: 20.9%). [Code & Data]

[C4] Yujian Liu*, **Xinliang Frederick Zhang***, David Wegsman, Nick Beauchamp, Lu Wang. 2022. *POLITICS:* Pretraining with Same-story Article Comparison for Ideology Prediction and Stance Detection. In *Findings of NAACL 2022* (Poster; main conference acceptance rate: 21.0%, findings acceptance rate: 9.9%). [Code & Data | Huggingface | Demo]

[C3] Xiang Yue*, **Xinliang Frederick Zhang***, Ziyu Yao, Simon Lin and Huan Sun. 2021. *CliniQG4QA:* Generating Diverse Questions for Domain Adaptation of Clinical Question Answering. In *Proceedings of IEEE BIBM 2021* (Oral; acceptance rate: 19.6%). **IEEE BIBM 2021 Best Paper Award (1 out of 727 submissions).** Also presented in *Machine Learning for Health (ML4H) @ NeurIPS 2020.* [Code | Data]

[C2] **Xinliang Frederick Zhang**, Heming Sun, Xiang Yue, Simon Lin and Huan Sun. 2021. *COUGH:* A Challenge Dataset and Models for COVID-19 FAQ Retrieval. In *Proceedings of EMNLP 2021* (Poster; acceptance rate: 17.9%). [Data]

[C1] **Xinliang Frederick Zhang** and Marie-Catherine de Marneffe. 2021. Identifying Inherent Disagreement in Natural Language Inference. In *Proceedings of NAACL 2021* (Oral; acceptance rate: 22.6%). [Code & Data]

Thesis/Dissertation

[T1] **Xinliang Frederick Zhang**. 2021. Towards More Robust Natural Language Understanding. *Undergraduate Research Thesis. The Ohio State University.*

Talks:

Graph-Augmented Generative Entity-to-Entity Stance Detection Framework. *TADA 2022, New York, NY, USA*
Generating Diverse Questions for Domain Adaptation of Clinical Question Answering. *BIBM 2021 (virtual)*
Inherent Disagreement in Natural Language Inference. *NAACL 2021 (virtual)*

Research Experiences:

Script Learning & LLM planning

Advised by Dr. Lu Wang Dec. 2023 - present

- We study uncovering and inducing latent, non-sequential general structures (i.e., scripts) for complex events under various settings (e.g., daily activities and news articles).
- We investigate the temporal relation in complex-event script learning, and fill the gap in current paradigms of temporal graph extraction by complementing "interchangeable, alternative, optional" relations.
- We explore LLM's planning capabilities for a diverse array of goals (e.g., draft a skeleton for storytelling).
- This work will be submitted to EMNLP 2024.

MOKA: Moral Knowledge Augmentation for Moral Event Extraction [C10]

Advised by Dr. Lu Wang Sept. 2022 - Oct. 2023

- We define a new schema for moral events grounded in Moral Foundation Theory, and propose the task of moral event extraction: given unstructured text, detect morality-bearing event triggers, extract participating entities, and identify underlying moralities.
- We annotate a new dataset, MORAL EVENTS, consisting of 5,494 structured moral event annotations.
- We propose MOKA, a moral event extraction framework with Moral Knowledge Augmentation. Leveraging Retrieval Augmentation techniques, MOKA retrieves and integrates knowledge derived from different levels of granularity, lexical-based moral word and example-based moral scenario knowledge.
- MOKA outperforms competitive baselines (e.g., ChatGPT) across three moral event understanding tasks.
- This work is accepted to NAACL 2024.

ULTRA: Unleashing LLM's Potential for Event Argument Extraction [P1]

Mentored by Dr. Alakananda Vempala May. 2023 - Aug. 2023

- We study document-level event argument extraction (DocEAE) which presents a unique challenge of scattered arguments, and explore the capabilities of open-sourced LLMs, i.e., Flan-UL2, for this task.
- We propose ULTRA, a hierarchical framework that extracts event arguments more cost-effectively, through 1) sequentially reading text chunks, 2) dropping non-pertinent candidates through pairwise self-refinement, and 3) correcting boundary identification leveraging Learning From ERRors (LEAFER).
- ULTRA outperforms strong supervised models and ChatGPT by 9.8% under EM.
- This work is submitted to ACL 2024.

Generative Entity-to-Entity Stance Detection with Knowledge Graph Augmentation [C6]

Advised by Dr. Lu Wang Dec. 2021 - June 2022

- We introduce a new task, entity-to-entity (E2E) stance detection, which primes models to identify entities in their canonical names and discern stances jointly.
- We curate a new dataset, SEESAW, with 10,619 annotations labeled at the sentence-level from news articles of different ideological leaning.
- We present a novel generative framework to allow the generation of entities in canonical names as well as stances among them. We enhance the model with a graph encoder to summarize entity activities and external knowledge surrounding the entities. Our model outperforms strong comparisons by large margins.
- This work is accepted to EMNLP 2022. This work is also presented at Text As Data (TADA) 2022.

***POLITICS:* Pretraining w/ Same-story Article Comparison for Ideology Prediction & Stance Detection [C4]**

Advised by Dr. Lu Wang May 2021 - Jan. 2022

- We study the training of PLMs using novel ideology-driven pretraining objectives that rely on the comparison of articles that are on the same stories but written by media of different ideologies.
- We collect a large-scale dataset consisting of more than 3.6M political news articles for pretraining.

- Our model POLITICS and its variants outperform strong baselines on ideology prediction and stance detection tasks. POLITICS is especially good at understanding long or formally written texts, and is also robust in few-shot learning scenarios.
- This work is accepted to NAACL 2022 (Findings).

***COUGH:* A Challenge Dataset and Models for COVID-19 FAQ Retrieval [C2]**

Advised by Dr. Huan Sun July 2020 - May 2021

- We present a large challenging dataset, COUGH, for COVID-19 FAQ retrieval. We show that our dataset presents a great challenge for future research.
- We analyze COUGH by testing different FAQ retrieval models built on top of BM25 and BERT, among which the best model achieves 48.8 under P@5.
- This work is accepted to EMNLP 2021.

Identifying Inherent Disagreement in Natural Language Inference [C1]

Advised by Dr. Marie-Catherine de Marneffe May 2020 - March 2021

- We investigate how to tease systematic inferences (i.e., items for which people agree on the NLI label) apart from disagreement items (i.e., items which lead to different annotations).
- We propose Artificial Annotators (AAs) to simulate uncertainties in the annotation process by capturing modes in annotations. Our approach performs statistically significantly better than all baselines.
- This work is accepted to NAACL 2021.

***CliniQG4QA:* Generating Diverse Questions for Domain Adaptation of Clinical Question Answering [C3]**

Advised by Dr. Huan Sun July 2019 - May 2020

- We propose a simple yet effective framework, CliniQG4QA, which leverages a question generation (QG) system to synthesize QA pairs on new clinical contexts to overcome generalization issue faced by Clinical question answering (QA) systems.
- We observe that generating diverse questions is essential for training robust QA systems, so we further propose a Question Phrase Prediction (QPP) module to diversify the generation.
- Empirical results show that QPP helps achieve up to 8% absolute gain in terms of Exact Match.
- This work is accepted to IEEE BIBM 2021 and receives *Best Paper Award*. This work is also presented at Machine Learning for Health (ML4H) workshop at NeurIPS 2020.

Training Neural Network with Mixed Integer Linear Programing

Advised by Dr. Chen Chen May 2019 - Aug. 2019

- We leverage Mixed-Integer-Linear-Programming (MILP) to train Feed-forward Neural Network (FFNN) model for classification tasks, with the objective of minimizing 0/1 loss.
- We explore the feasibility (accuracy-efficiency tradeoff) of MILP training approach.
- Results show that MILP performs consistently better than gradient descent on small/medium-scale datasets with low/medium network complexity.

Research Training in Financial Futures & Quantitative Investment

Jointly advised by HuaXi Future Co. and Dept. of Math at Sichuan University Sept. 2017 - June 2018

- This research training mainly focuses on the field of quantitative investment on futures.
- We develop strategies for quant, backtest critical futures indicators by using existing historical data, and implement and optimize strategies through MATLAB.
- Developed strategies show satisfactory performances on Stock Market Simulator and Live Stock Market.

Awards & Honors:

Research Award & Scholarship

Rackham Conference Travel Grant	Apr. 2024
TADA 2022 Travel Grant	Oct. 2022
Outstanding Reviewer, NAACL 2022 (81 out of 2629)	July 2022
Rackham Conference Travel Grant	June 2022
Best Paper Award, IEEE BIBM 2021	Dec. 2021
CSE Undergraduate Research Award (Liu Scholarship)	Mar. 2021
CRA Outstanding Undergraduate Researcher Award, Honorable Mention [Media Coverage]	Dec. 2020
Lumley Engineering Fund Scholarship (Undergraduate research scholarship)	Nov. 2019

Academic Award & Scholarship

"Tomorrow Advancing Life" Merit Scholarship (Top 0.25% at SCU)	May 2019
National Scholarship (Highest Undergraduate Honor in China; 2017-18)	Sept. 2018
Best Academic Achiever Award (Academically Top 1; 2017-18)	July 2018
National Scholarship (Highest Undergraduate Honor in China; 2016-17)	Sept. 2017
Best Academic Achiever Award (Academically Top 1; 2016-17)	Sept. 2017

Academic Competition Award

Distinguished Performance in Mathematical Modeling Contest	Dec. 2018
Honorable Mention in Mathematical Contest in Modeling	Apr. 2018
Provincial 2nd Prize in Contemporary Undergraduate Mathematical Contest in Modeling	Sept. 2017
3rd Prize in Sichuan University Mathematical Contest	June 2017
2nd Prize in Sichuan University Mathematical Contest in Modeling	May 2017

Honor & Glory

Outstanding University Graduate of Sichuan Province (Highest Graduation Honor in China)	Mar. 2021
Outstanding University Graduate at Sichuan University	Sept. 2020
"Best One hundred" Top 10 Class Leader at Sichuan University (Highest Honor at SCU)	May 2018
Excellent Student at Sichuan University (2017-18)	Sept. 2018
Excellent Departmental Leader of Student Council at Sichuan University	June 2018
Excellent Student at Sichuan University (2016-17)	Sept. 2017
Excellent Class Leader at Sichuan University	Sept. 2017
Certified Volunteer in IET English Speech Contest (2017, 2018)	June 2017, June 2018
Recipient of Community Contribution Award (1 out of 260)	May 2017
2nd Prize in Sichuan University Video Competition of 120th Anniversary	Dec. 2016

Services:

Program Committee/Reviewer

ACL (2022, 2023, 2024)	Dec. 2021, Feb. 2023, Mar. 2024
EMNLP (2022, 2023)	July 2022, July 2023
NAACL (2022, 2024)	Feb. 2022, Jan. 2024
COLING (2024)	Dec. 2023
ACL Rolling Review (Nov.'21, Dec.'21, Jan.'22, Mar.'22, Sept.'22, Dec.'22, Dec.'23, Feb.'24)	
Bridging the Gap: from Machine Learning Research to Clinical Practice at *NeurIPS 2021*	Oct. 2021

Mentoring (@UM)

Xiaoyan Bai (UM UG; 2024 Winter-now; Experimental Narratives)

Yufan Wu (UM UG; 2024 Winter-now; Experimental Narratives)

Jahnavi Enaganti (UM UG; 2022 Fall-now; Moral Events Annotation & Experimental Narratives)

Changyuan Qiu (UM UG → UW PhD; 2021 Fall-2022 Summer; Multi-modal Media Bias Analysis [C5])

Teaching (UTA: Undergraduate Teaching Assistant)

UTA in CSE 3341 (Principles of Programming Languages)	2021 Spring
UTA in CSE 2331 (Foundations II: Data Structures and Algorithms)	2020 Summer
UTA/Lab Assistant in CSE 2231 (Software II: Software Development & Design)	2019 Autumn
UTA/Lab Assistant in CSE 2221 (Software I: Software Components)	2019 Spring
UTA/Lab Assist. in CSE 2112 (Modeling and Problem Solving w/ Spreadsheet & Database)	2019 Spring
Guest Lecturer in IE 1085 (Departmental Seminar: a 45-min seminar on *Intro to MCM*)	2018 Spring

University & Department Services

Volunteer for Statement of Purpose feedback program (SASP) @UM	Nov. 2023
AI Tea Coordinator @UM	Oct. 2022- May 2024
Peer Advisor @SCU	Sept. 2017-July 2018
Vice President of Finance in Student Council @SCU	June 2017-June 2018
Class Leader @SCU	Sept. 2016-Jan. 2018
Member of Editing Department in Science and Technology Association @SCU	Sept. 2016-June 2017

Skills & Languages:

Technical Skill

Programming*: Proficient:* Python, MATLAB, Java; *Intermediate:* C, JavaScript, Ruby

Machine Learning: Pytorch, Huggingface, Sklearn, Spacy, AllenNLP, PyG

Language Proficiency

Mandarin (Native), Wu (Native), English (Full professional proficiency), Japanese (Elementary proficiency)

References:

Lu Wang, Ph.D
Associate Professor
Computer Science and Engineering, EECS
University of Michigan, Ann Abor, MI
wangluxy@umich.edu
https://web.eecs.umich.edu/~wangluxy/

Huan Sun, Ph.D
Associate Professor and endowed CoE Innovation Scholar
Department of Computer Science & Engineering
The Ohio State University, Columbus, OH
sun.397@osu.edu
http://web.cse.ohio-state.edu/~sun.397/

Alakananda Vempala, Ph.D
Research Engineer
Bloomberg, New York, NY
avempala@bloomberg.net
https://alakanandav.bitbucket.io/

Marie-Catherine de Marneffe, Ph.D
FNRS Research Associate
Center for Natural Language Processing
UCLouvain, Belgium
marie-catherine.demarneffe@uclouvain.be
https://cental.uclouvain.be/team/mcdm/

Name Variants:

Chinese version: Legal Name: 张心亮; Name in Pinyin: Zhang Xinliang OR Xinliang Zhang (Western fashion)
English version: Preferred Name: Frederick Zhang; Name for publication: Xinliang Frederick Zhang

Japanese version: Chinese transliteration: 張(ちょう) 心亮(しんりょう); English transliteration: フレデリック・ジャン