Xiaomeng (Tracy) Jin

Ph.D. Candidate, University of Illinois Urbana–Champaign
LinkedIn xjin17@illinois.edu

Research Interests

My research focuses on enhancing the performance and robustness of Large Language Models (LLMs) through advanced data synthesis strategies, such as adversarial attacks, data augmentation, and schema induction. I am also interested in improving the adaptability and reliability of LLMs using diverse approaches including retrieval-augmented generation and machine unlearning.

Education

University of Illinois Urbana–Champaign — **Champaign, IL**
Ph.D. in Computer Science — *Sep. 2021 – Mar. 2025 (Expected)*
Advisor: Prof. Heng Ji

Stanford University — **Stanford, CA**
Master of Science in Computer Science — *Sep. 2019 – Apr. 2021*
Specialization: Artificial Intelligence
GPA: 4.0 / 4.0

University of Toronto — **Toronto, ON**
B.A.&Sc. Specialist in Computer Science & Math Applications in Economics and Finance — *Sep. 2014 – Jun. 2019*
Focus: Artificial Intelligence
Minor in Statistics
GPA: 3.87 / 4.0

Publications

[16] *Unlearning as multi-task optimization: A normalized gradient difference approach with an adaptive learning rate*
Xiaomeng Jin, Zhiqi Bu, Bhanukiran Vinzamuri, Mingyi Hong, Heng Ji, Kai-Wei Chang.
Under Submission (https://arxiv.org/abs/2410.22086)

[15] *MUNCH: A Multitask Unlearning Benchmark for LLMs*
Anil Ramakrishna, Yixin Wan, **Xiaomeng Jin**, Kai-Wei Chang, Zhiqi Bu, Bhanukiran Vinzamuri, Volkan Cevher, Mingyi Hong, Rahul Gupta.
Under Submission

[14] *Efficient Adversarial Prompting using Attacker Large Language Models*
Xiaomeng Jin, Bhanukiran Vinzamuri, Heng Ji, Pradeep Natarajan.
Under Submission

[13] *Toward Universal Concept Recognition: Contrastive Visual Data Augmentation*
Yu Zhou*, Bingxuan Li*, Mohan Tang*, **Xiaomeng Jin**, Te-Lin Wu, Kuan-Hao Huang, Heng Ji, Kai-Wei Chang, Nanyun Peng.
Under Submission

[12] *ARMADA: Attribute-based Multimodal Data Augmentation*
Xiaomeng Jin, Jeonghwan Kim, Yu Zhou, Kuan-Hao Huang, Te-Lin Wu, Nanyun Peng, Heng Ji.
NLP for Wikipedia Workshop (EMNLP 2024).

[11] *MIRACLE: An Online, Explainable Multimodal Interactive Concept Learning System*
Ansel Blume, Khanh Duy Nguyen, Zhenhailong Wang, Yangyi Chen, M. Rothman, **Xiaomeng Jin**, Jeonghwan Kim, Zhen Zhu, Jiateng Liu, Kuan-Hao Huang, Mankeerat Sidhu Xuanming Zhang, Vivian Liu, Raunak Sinha, Te-Lin Wu, Abhay Zala, Elias Stengel-Eskin, Da Yin, Yao Xiao, Utkarsh Mall, Zhou Yu, Kai-Wei Chang, Camille Cobb, Karrie Karahalios Lydia Chilton, Mohit Bansal, Nanyun Peng, Carl Vondrick, Derek Hoiem and Heng Ji.
Proc. the 32nd ACM Multimedia Conference (ACMMM) Demo and Video Program Track, 2024.

[10] *Schema-based data augmentation for Event Extraction*
Xiaomeng Jin, Heng Ji.
The Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING), 2024.

[9] *Adversarial Named-Entity Recognition with Word Attributions and Disentanglement*

Xiaomeng Jin, Bhanukiran Vinzamuri, Sriram Venkatapathy, Heng Ji, Pradeep Natarajan.
The Conference on Empirical Methods in Natural Language Processing (EMNLP Findings), 2023.

[8] *Globally Consistent Event-Event Temporal Relation Extraction*
Xiaomeng Jin, Haoyang Wen, Xinya Du, Heng Ji.
Workshop on Matching From Unstructured and Structured Data (ACL), 2023.

[7] *Event Schema Induction with Double Graph Autoencoders*
Xiaomeng Jin, Manling Li, Heng Ji.
The North American Chapter of the Association for Computational Linguistics (NAACL), 2022.

[6] *RESIN-11: Schema-guided Event Prediction for 11 Newsworthy Scenarios*
Xinya Du, Zixuan Zhang, Sha Li, Pengfei Yu, Hongwei Wang, Tuan Manh Lai, Xudong Lin, Ziqi Wang, Iris Liu, Ben Zhou, Haoyang Wen, Manling Li, Darryl Hannan, Qi Zeng, Qing Lyu, Charles Yu, Carl Edwards, **Xiaomeng Jin**, Yizhu Jiao, Ghazaleh Kazeminejad, Rotem Dror, Zhenhailong Wang, Chris Callison-Burch, Mohit Bansal, Carl Vondrick, Jiawei Han, Dan Roth, Shih-Fu Chang, Martha Palmer, Heng Ji.
The North American Chapter of the Association for Computational Linguistics System Demonstration Track (NAACL), 2022.

[5] *Chemical-Reaction-Aware Molecule Representation Learning*
Hongwei Wang, Weijiang Li, **Xiaomeng Jin**, Kyunghyun Cho, Heng Ji, Jiawei Han, Martin D Burke.
The International Conference on Learning Representations (ICLR), 2022.

[4] *System and method for machine learning architecture with adversarial attack defence*
Weiguang Ding, Luyu Wang, Ruitong Huang, **Xiaomeng Jin**, Kry Yik Chau LUI.
US Patent App, 2019.

[3] *AdverTorch v0. 1: An adversarial robustness toolbox based on pytorch*
Weiguang Ding, Luyu Wang, **Xiaomeng Jin**.
arXiv Preprint, 2019.

[2] *On the Sensitivity of Adversarial Robustness to Input Data Distributions*
Weiguang Ding, Kry Yik Chau Lui, **Xiaomeng Jin**, Luyu Wang, Ruitong Huang.
The International Conference on Learning Representations (ICLR), 2019.

[1] *Gender, confidence, and mark prediction in CS examinations*
Brian Harrington, Shichong Peng, **Xiaomeng Jin**, Minhaz Khan.
The Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education, 2018.

Research Projects

Research Assistant at University of Illinois Urbana-Champaign **Champaign, IL**
School of Computer Science *Sep. 2021 – Present*
Advisor: Prof. Heng Ji
• Research topic: Enhance Model Performance and Robustness with Data Augmentation Strategies
• Designed a novel event schema induction method that depicts the common pattern of complex events. Utilized induced schema graphs to guide the data generation process for event extraction tasks
• Proposed a Multimodal Data Augmentation method via knowledge-guided manipulation of visual attributes
• Designed efficient adversarial attack methods on LLMs through few-shot learning strategies

Research Assistant at Stanford University **Stanford, CA**
Computer Science Department *Jan. 2020 – Aug. 2021*
Advisor: Prof. Jure Leskovec
• Research topic: Answer USMLE (United States Medical Licensing Exam) questions automatically with external medical knowledge graphs
• Constructed a medical knowledge graph and mapped entities in questions/choices to nodes in knowledge graphs. Achieved 50% accuracy on multiple-choice questions with 5 – 8 choices

Research Assistant at University of Toronto **Toronto, ON**
Computer Science Department *Aug. 2018 – Dec. 2018*
Advisor: Prof. Jimmy Ba
• Research topic: Explore and understand the factors which could cause the failure of agent move in a reinforcement learning setup
• Computed and ranked the influence function values of the training frames corresponding to the test frames at different training stages, explained how the influence values affected the agent to make the move

Research Assistant at University of Toronto **Toronto, ON**
Computer Science Department *Nov. 2017 – Feb. 2018*

Advisor: Prof. Brian Harrington
• Research Topic: Investigate the relationship between students' actual performance and their own predictions
• Implemented statistical software tool R language to analyze the effects of different factors (such as gender, year of study, programming skills level etc.) on the predicted and actual performance by using Spearman's rank-order coefficient and two-sample t-test

Internship

Applied Scientist Intern **Bellevue, WA**
Amazon Inc. *Jun. 2024 – Aug. 2024*
• Conducted in-depth research and exploration in LLM unlearning
• Formulated machine unlearning as a multi-task optimization problem
• Proposed an effective LLM unlearning approach that preserves high model utility

Applied Scientist Intern **Bellevue, WA**
Amazon Inc. *Jun. 2023 – Aug. 2023*
• Conducted in-depth research in improving LLMs' adversarial robustness
• Proposed an efficient prompt attack method to generate adversarial prompts

Applied Scientist Intern **Bellevue, WA**
Amazon Inc. *Jun. 2022 – Aug. 2022*
• Conducted in-depth research in improving NLP models' adversarial robustness
• Proposed a novel adversarial attack method with word attribution and disentanglement

Software Engineer Intern **Cupertino, CA**
Apple Inc. *Jun. 2020 – Aug. 2020*
• Designed and implemented a search proxy service in Java that provides an interface between apps and Solr search platform
• Deployed on Apple servers to be ready into production, will be used for all Zeus services and Apps in ACS Enterprise Team

Machine Learning Research Intern **Toronto, ON**
Borealis AI *Feb. 2019 – May. 2019*
• Designed and implemented AdverTorch, a Python toolbox for adversarial robustness research

Machine Learning Research Intern **Toronto, ON**
Borealis AI *Jan. 2018 – Aug. 2018*
• Explore the effects of the input data distribution on adversarial robustness on trained neural networks
• Performed standard neural networks training and projected gradient descent (PGD) based adversarial training on datasets with different scales of transformations, compared the sensitivity of robust accuracy to the input data distribution

Software Engineer Intern **Toronto, ON**
ScotiaBank *Sep. 2017 – Dec. 2017*
• Performed in-depth quantitative analysis on block trades from database and generated daily reports to both clients and traders
• Optimized ScotiaBank Order Tracker web application using Java, Javascript, added new features to the IOI posting board

Software Engineer Intern **Toronto, ON**
ParseHub *May. 2016 – Aug. 2016*
• Designed an Undo-Redo algorithm and implemented it via Angular.js that greatly improved productivity
• Tailored clients' needs by establishing the Importing External Template feature to achieve a high level of customer satisfaction

Last updated: Nov. 13, 2024