

Siyi Liu

Email: siyiliu@seas.upenn.edu

Webpage: <https://liusiyi641.github.io/>

Research Interests

My primary research interest lies within the areas of natural language processing. Currently, I'm interested in studying various types of **conflicts** in the applications of Large Language Models (LLMs), including **hallucinations** (generations conflicting with contexts), **knowledge conflicts** (conflicts among contexts or between contextual and parametric knowledge), and also **perspectives** (conflicts among different sources of perspectives).

Education

University of Pennsylvania , Philadelphia, PA Ph.D. in Computer and Information Science Advisor: Dan Roth	Fall 2022 - Present
University of Pennsylvania , Philadelphia, PA M.S.E. in Computer and Information Science	Fall 2020 - May 2022
Boston University , Boston, MA B.A./M.S. Dual Degree in Computer Science	Fall 2016 - May 2020

Publications

- [1] **Siyi Liu**, Kishaloy Halder, et al.
“Towards Long Context Hallucination Detection”. in submission [\[Paper\]](#)
- [2] **Siyi Liu**, Qiang Ning, et al.
“Open Domain Question Answering with Conflicting Contexts”. in submission [\[Paper\]](#)
- [3] Nishanth Nakshatri, **Siyi Liu**, Sihao Chen, Daniel Hopkins, Dan Roth, Dan Goldwasser
“Using LLM for improving key event discovery: Temporal-guided news stream clustering with event summaries”. in Findings of EMNLP 2023 [\[Paper\]](#)
- [4] **Siyi Liu**, Hongming Zhang, Hongwei Wang, Kaiqiang Song, Dan Roth, Dong Yu
“Open-Domain Event Graph Induction for Mitigating Framing Bias”. 2023. [\[Paper\]](#)
- [5] Sihao Chen*, **Siyi Liu***, Xander Uyttendaele, Yi Zhang, William Bruno, Dan Roth
“Design Challenges for a Multi-Perspective Search Engine”. in Findings of NAACL 2022 [\[Paper\]](#)[\[Demo\]](#)
- [6] Anietie Andy, **Siyi Liu**, Daphne Ippolito, Reno Kriz, Chris Callison-Burch, Derry Wijaya
“Creating Multimedia Summaries Using Tweets and Videos”, 2022. [arxiv](#)
- [7] **Siyi Liu**, Sihao Chen, Xander Uyttendaele, Dan Roth
“MultiOpEd: A Corpus of Multi-Perspective News Editorials”, in NAACL 2021. [\[Paper\]](#)[\[Slides/Talk/Poster\]](#)

[8] Siyi Liu*, Ziang Leng*, Derry Tanti Wijaya.
“Learning To Mirror Speaking Styles Incrementally.” 2020. [arxiv](#)

[9] Siyi Liu, Lei Guo, Kate Mays, Margrit Betke and Derry Tanti Wijaya.
“Detecting Frames in News Headlines and Its Application to Analyzing News Framing Trends Surrounding U.S. Gun Violence.” In *Proceedings of Conference on Computational Natural Language Learning (CoNLL)* 2019.

[[Paper](#)] [[Slides](#)]

Industrial Experience

AWS AI Labs, NYC, New York **Summer 2024**
Title: Applied Scientist Intern
Host: Kishaloy Halder
Project: Long Context Hallucination Detection
Keywords: Hallucination Detection, Long Context, LLMs

AWS AI Labs, NYC, New York **Summer 2023**
Title: Applied Scientist Intern
Host: Qiang Ning
Project: Open Domain Question Answering with Conflicting Contexts
Keywords: Retrieval Augmented Generation (RAG), Conflicting Contexts, LLMs

Tencent AI Labs, Bellevue, WA **Summer 2022**
Title: Applied Scientist Intern
Host: Hongming Zhang
Project: Open Domain Neutral Event Graph Induction
Keywords: Framing Bias, Event Graph Induction

Demo

A Multi-Perspective Search Engine

We build a system that presents search results in clusters of perspectives using Google Search API. Empirically, 63% of the participants in our survey prefer our system’s organization to naive Google Search’s given the same search results and UI. [\[Demo\]](#)

Funded Projects

University of Pennsylvania Fall 2022 – Present
Advisor: Prof. Dan Roth
Project: Human Interpretable Attribution of Text Using Underlying Structure (HIATUS)

- This project is funded by Intelligence Advanced Research Projects Activity (IARPA) via 2022-22072200003 and is led by Raytheon BBN Technologies
- Develop explainable systems to identify authorship through linguistic representations of texts

University of Pennsylvania

Summer 2020 – Present

*Advisor: Prof. Dan Roth**Project: Information Pollution Project*

- *This project is funded by NSF award IIS-2135573*
- *A dataset with hierarchical structure of query, perspectives, evidence, article*
- *Study and analyze the argumentation structure of news articles*
- *Develop a demo website that searches in real time in a perspective-oriented manner*
- *Using LLMs to discover key events in news*

Boston University

Spring 2019 – Fall 2020

*Advisor: Prof. Derry Wijaya**Projects:*

- *This project is funded by NSF grant 1838193.*
- *Detect frames in news headlines and analyze news framing trends surrounding U.S. gun violence*
- *Capture and transform speaking styles in a low resource setting*
- *Create multimedia summaries using tweets and videos*