

Andrea Wen-Yi Wang

✉ andreaawwenyi@infosci.cornell.edu

🌐 andreaawwenyi

Research Interests

My research focuses on developing NLP tools that promote personal perspectives. Through my work, I challenge the objectivity of computational models and think about how NLP tools, including LLMs, could empower domain experts and voices of marginalized communities, especially in contexts involving subjective decision-making.

Keywords: Natural Language Processing (NLP), Computational Social Science, Large Language Models (LLMs), Feminist Theory, Multilingual NLP

Education

- 2022 - Now **Ph.D. in Information Science**, Cornell University
Chair: David Mimno
Committee: Allison Koenecke, Karen Levy
- 2017 - 2019 **M.S. in Data Science**, New York University
Mathematics and Data track
- 2012 - 2016 **B.A. in Finance**, National Taiwan University

Publications

- 2024 [Automate or Assist? The Role of Computational Models in Identifying Gendered Discourse in US Capital Trial Transcripts](#)
Andrea W Wen-Yi, Kathryn Adamson, Nathalie Greenfield, Rachel Goldberg, Sandra Babcock, David Mimno, Allison Koenecke
AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES)
- 2023 [Hyperpolyglot LLMs: Cross-Lingual Interpretability in Token Embeddings](#).
Andrea W Wen-Yi and David Mimno.
Empirical Methods in Natural Language Processing (EMNLP)
- 2021 [The Evolution of Rumors on a Closed Social Networking Platform During COVID-19: Algorithm Development and Content Study](#).
Andrea Wang, Jo-Yu Lan, Ming-Hung Wang, Chihhao Yu.
JMIR Medical Informatics

Working
paper

How Chinese are Chinese Language Models? The Puzzling Lack of Language Policy in China's LLMs

Andrea W Wen-Yi*, Unso Eun Seo Jo*, Lu Jia Lin, David Mimno

Work/Research experience

Summer 2024

AI Researcher

Gena Co. // Seoul, Korea

Supervisor: Eunseo (Unso) Jo

- Text-to-SQL for Samsung SDI battery subsidiary.
- Conversation tagging classification for customer support copilot.

2019 - 2022

Data Researcher

New York University Public Safety Lab

Supervisor: Anna Harvey

- Developed a system for daily collection of detainee records from 1,000 U.S. county jail rosters using Python, AWS, GitHub, and Airtable for the [Jail Data Initiative](#).

2020 - 2021

Data Researcher

Information Operations Research Group // Taipei, Taiwan

Supervisor: Ming-Hung Wang, Chihhao Yu

- Developed an efficient classification-based clustering algorithm to study the temporal propagation patterns of false pandemic-related information on a social platform.

Posters, Talks, Presentations

July 2023

Seasonality Visualizations of Online Text

Andrea W Wang, Allison Koenecke, David Mimno

The International Conference for Computational Social Science (IC2S2)

July 2021

Information operations research as a data science research

Conference for Open Source Coders, Users & Promoters (COSCUP)

Dec 2020

Oarchive: an open source archiving system and open data for Taiwan online information space.

g0v (gov-zero) summit

Teaching Experience

Fall 2024

Graduate TA, [INFO 3300: Visual Data Analytics for the Web](#), Cornell University

Fall 2023

Graduate TA, [INFO 2950: Introduction to Data Science](#), Cornell University

Spring 2019

Graduate TA, [Introduction to Data Science](#), New York University

Fall 2018

Graduate TA, [Probability and Statistics for Data Science](#), New York University

Services

2023	EMNLP Student Volunteer
2022, 2023	Student Reviewer and Interviewer, Cornell Information Science PhD Admissions
2022, 2023	Mentor, Cornell InfoSci Student-Applicant Reading Program (SARP)