Andrea Wen-Yi Wang

✉ andreawwenyi@infosci.cornell.edu
andreawwenyi

Research Interests

My research interests lie at the intersection of Natural Language Processing, Data Science, and Computational Social Science. My works involve understanding the characteristics of large language models (LLMs) in two ways. First, I study the interpretability of multilingual large language models with the goal to improve the performance for low-resource languages. Second, I study both the opportunities and challenges that LLMs offer for social scientists with textual data. I have works in domains related to criminal justice, gendered studies, misinformation.

Education

2022-Present **Ph.D. in Information Science**, Cornell University
Advisor: David Mimno
Committee: Allison Koenecke, Karen Levy
Minor: Sociology

2017-2019 **MS Data Science**, New York University
Mathematics and Data track

2012-2016 **BA 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.

Jan 2019 - June 2022 **Data Scientist**
New York University Public Safety Lab
Supervisor: Anna Harvey
• Created Python programs to collect individual-level detainee records daily from around 1,000 U.S. county jail rosters.
• Designed operational pipeline involving Amazon Web Services, Github, and Airtable that improved the success rate by 80%.

Sep 2020 - Aug 2021 **Researcher**
Information Operations Research Group // Taipei, Taiwan
• Studied the temporal propagation patterns of false pandemic-related information on a social platform by developing an efficient classification-based clustering algorithm.

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.**
gov (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

Last updated: September 13, 2024 •