Jason Yan

Email: jasonyan@umich.edu — Phone: +1 (917) 828-7884

Google Scholar: Click Here

Address: 105 S State St, Ann Arbor, MI 48109

Research Interests

Natural Language Processing, Computational Social Science, Machine Learning, Causal Inference, Political Communications, Education Policy

Education

University of Michigan, Ann Arbor, MI   *2022 – Present*
Ph.D. in Information Science
Advisor: Sabina Tomkins

University at Buffalo, Buffalo, NY   *2018 – 2022*
Bachelor of Science in Computer Science and Mathematics, Minor in Statistics
GPA: 3.97/4.00

Publications

- **Yan J.***, Lin T.*, Jurgens D., Tomkins S. (2024). *Tab2Text - A framework for deep learning with tabular data.* Findings of the Empirical Methods in Natural Language Processing (**EMNLP**) 2024. (* Equal contribution)

- **Yan J.**, Milita K., Krupnikov Y., Tomkins S. *What Do You Learn about Hard News from Soft News Outlets - A Case Study from People Magazine Online.* Currently under Revise&Resubmit for International Conference on Web and Social Media (**ICWSM**).

- Ananthasubramaniam A.*, Chen H.*, **Yan J.***, Alkiek K.*, Pei J.*, Seth A.*, Dunagan L.*, Choi M.*, Litterer B., Jurgens D. (2023). *Exploring Linguistic Style Matching in Online Communities: The Role of Social Context and Conversation Dynamics.* The 1st Workshop on Social Influence in Conversations (**SICon**) @ACL. (* Equal contribution) **Best Paper Award**

- **Yan J.**, Hall S.F., Sage M., Du Y., and Joseph K. (2024). *A Computational Social Science Approach to Understanding Predictors of Chafee Service Receipt.* **Child and Youth Services Review**.

Posters

- **Yan J.***, Lin T.*, Jurgens D., Tomkins S. (2024). *Tab2Text - A framework for deep learning with tabular data.* International Conference on Computational Social Science (**IC2S2**). (* Equal contribution)

- **Yan J.**, Lin T., Krupnikov Y., Tomkins S. *I've Seen That Before! Towards Understanding Hard News Exposure from Soft News Outlets.* Digital State Sponsored Disinformation and Propaganda: Challenges and Opportunities for Researchers (@ICWSM).

- **Yan J.**, Du Y., Sage M., and Joseph K. (2021). *What can Administrative Data tell us about Service Allocation to American Foster Youth?* International Conference on Computational Social Science (**IC2S2**).

- **Yan J.** and Miecznikowski J. (2020). *Plasma Metabolites and Diabetes.* Biostatistics, Epidemiology, and Research Design (**BERD**).

Research Experience

Graduate Student Research Assistant   *Jan 2024 – Nov 2024*
University of Michigan, Ann Arbor, MI

PI: Sabina Tomkins, David Jurgens

- Accepted to **EMNLP Findings 2024**.

- Proposed and developed Tab2Text, a deep learning framework designed to transform tabular data into textual representations using generative language models (gpt-4, claude, gemini, and llama-2-7b-chat). Tab2Text creates both descriptive and narrative text representations of tabular features to construct more contextually meaningful embeddings to improve predictive accuracy in downstream tasks, resulting in an average increase of 1.5% in prediction accuracy compared to tabular-only models.

Graduate Student Research Assistant   *Jan 2024 – Present*
University of Michigan, Ann Arbor, MI
PI: Sabina Tomkins, Yanna Krupnikov

- Currently under Revise&Resubmit for the **International Conference on Web and Social Media** (ICWSM).

- Developed a novel definition of "hard news," designed and conducted a large-scale annotation study, and implemented supervised learning models (DeBERTa, RoBERTa, DistilBERT) to classify political content in entertainment media, achieving an F1 score of 0.827

Undergraduate Research Assistant   *Jan 2020 – May 2022*
University at Buffalo, Buffalo, NY
PI: Kenneth Joseph

  - Accepted to **Child Youth and Services Review** (CYSR)

  - Applied predictive modeling (Random Forest, Gradient Boosted Trees) and advanced statistical methods to analyze National Youth in Transition Database (NYTD) data, identifying systemic disparities in service allocation among foster youth; findings revealed inequities by age, geography, and care duration, providing actionable insights to improve equity.

Teaching Experience

Graduate Student Instructor   *Jan 2025 – April 2025*
University of Michigan, Ann Arbor, MI

- SI206: Data-Oriented Programming

Undergraduate Teaching Assistant   *Aug 2019 – May 2022*
University at Buffalo, Buffalo, NY

- Assisted in courses: CSE250 Data Structures, CSE191 Discrete Structures, and CSE199 How the Internet Works.

Awards, Grants, and Scholarships

- **Rackham Conference Travel Grant**, University of Michigan (2024)

- **SI Conference Travel Grant**, University of Michigan (2023, 2024; 2x recipient)

- **UMSI Merit Scholar**, University of Michigan (2022)

- **Grace W. Capen Award**, University at Buffalo (2020)

- **Walker Family Experiential Learning Fund**, University at Buffalo (2021)

- **Grace W. Capen Fund**, University at Buffalo (2021)

- **Andrew and Virginia Craig Scholarship**, University at Buffalo (2021)

- **Anne Brody Scholarship**, University at Buffalo (2019–2020; 2x recipient)

- **Ellicott Club Scholarship**, University at Buffalo (2019)

Doctoral Coursework

Doctoral Foundation Seminar, Causal Inference in Education Policy Research, Bayesian Inference, Research Methods, Econometric Analysis 1, Econometric Analysis 2, Digital Public Goods, Machine Learning Theory

References

Sabina Tomkins
Assistant Professor
School of Information
University of Michigan
stomkins@umich.edu

Kenneth Joseph
Associate Professor
Department of Computer Science and Engineering
University at Buffalo
kjoseph@buffalo.edu

David Jurgens
Associate Professor
School of Information and
Department of Computer Science and Engineering
University of Michigan
jurgens@umich.edu