Neil Fasching

Computational Social Scientist | Data Scientist | PhD Candidate
neilfasching@gmail.com | neilfasching.com | 3620 Walnut St. Philadelphia, PA 19104

PROFESSIONAL SUMMARY

PhD candidate at the University of Pennsylvania with advanced expertise in machine learning, statistical modeling, and large-scale data analysis. Experienced with experimental design, causal inference, and analyzing datasets with millions of data points. Published in avenues like ACL and Science Advances. Created pipelines for analyzing tens of thousands of podcasts and millions of hate speech sentences at scale.

EDUCATION

PhD, Computational Social Science Expected 2025/2026
University of Pennsylvania

Master of Science, Statistics and Data Science 2023
The Wharton School, University of Pennsylvania

Master of Science, Communication Science 2021
University of Amsterdam

PROFESSIONAL EXPERIENCE

Computational Research Fellow **Sep 2021 - Present**
University of Pennsylvania

- Collaborate with professors Dr. Yphtach Lelkes and Dr. Duncan J. Watts on projects into the influence of news media and social media on human behavior
- Analyze large-scale datasets to study the trends, patterns, and effects of news media and social media on human behavior
- Employ diverse data collection, data mining, and analysis techniques with emphasis on LLMs
- Develop ML models novel pipelines to classifying unstructured text as topics, segmentation, and toxicity
- One example: (mediabiasdetector.seas.upenn.edu/)

Co-Teacher, Modern Data Mining (PhD Level) **Jul 2022 - Present**
The Wharton School, University of Pennsylvania

- Teach PhD-level Data Science course in the Data Science department at Wharton
- Cover cutting-edge machine learning techniques including Boosted Trees, CNNs, RNNs, and LLMs
- Incorporate up-to-date case studies combining statistical theory with practical applications

RESEARCH PROJECTS

Leveraging Large Language Models to Evaluate Topics of Discussion, Misinformation, and Toxicity on Political Podcasts (Dissertation)

- Analyzed over 28,000 podcast episodes for topics, misinformation, and toxicity
- Developed two novel frameworks for assess the prevalence of misinformation and hate speech at scale
- Assessed LLM-based model performance for transcription, topic segmentation, misinformation identification and hate speech classification

Model-Dependent Moderation: Inconsistencies in Hate Speech Detection Across LLM-based Systems (Link to ACL Paper)

- Led evaluation of AI bias in hate speech detection across seven content moderation systems

- Created a synthetic dataset of over 1.3+ million sentences using a full factorial design
- Quantified model inconsistencies in content filtering decisions by different demographic groups
- Established new evaluation metrics for HSD detection in AI systems

Persistent polarization: The unexpected durability of political animosity around US elections (Link to Paper)

- Analyzed 66,000 cross-sectional and panel interviews to demonstrate that political animosity remains stable around elections
- Employed Interrupted Time Series (ITS) models to causally assess the impact of election proximity on animosity

Toxic Air in the Public Square: Quantifying Toxicity of Political Discourse on Twitter

- Analyzed 46.7 million tweets (2012-2022) for toxicity (including harassing, hate, and violent speech)
- Developed scalable pipelines for toxicity measurement using OpenAI and Mistral moderation systems
- Utilized advanced ML models to analyze difference in toxicity across time, demographics, and topics

TECHNICAL SKILLS

Programming Languages: R, Python, SQL, JavaScript

Machine Learning Libraries: PyTorch, scikit-learn, TensorFlow, Keras, XGBoost, Hugging Face Transformers, spaCy, NLTK, statsmodels

Data Processing: PySpark, PyArrow, Pandas, NumPy, Dask, dplyr/tidyverse, SparkR, arrow

Statistical Methods: Regression (Linear, Logistic, Multilevel), Neural Networks, Ensemble Methods, Time-Series Analysis, Causal Inference, Experimental Design

Platforms: AWS, Microsoft Azure, Google Colab, Posit Workbench, Git/GitHub

SELECT PUBLICATIONS

Fasching, N. and Lelkes, Y. (2025). **Model-dependent moderation: Inconsistencies in hate speech detection across LLM-based systems**. In *Findings of the Association for Computational Linguistics*

Fasching, N., Iyengar, S., Lelkes, Y., and Westwood, S. J. (2024). **Persistent polarization: The unexpected durability of political animosity around US elections**. *Science Advances*, 10(36), eadm9198.

Pangakis, N., Wolken, S., and Fasching, N. (2023). **Automated annotation with generative AI requires validation**. *arXiv preprint* arXiv:2306.00176.

Fasching, N., Wolken, S., and Dörr, T. (2025). **Toxic Air in the Public Square: Quantifying Toxicity of Political Discourse on Twitter**. (Under Review).

Fasching, N., Arceneaux, K., and Bakker, B. N. (2024). **Inconsistent and very weak evidence for a direct association between childhood personality and adult ideology**. *Journal of Personality.*

Fasching, N., and Lelkes, Y. (2024). **Ancestral Kinship and the Origins of Ideology**. *British Journal of Political Science.*

RELEVANT EXPERIENCE FOR ML

- Experience with large-scale data: 46.7+ million tweets, 28,000+ podcast episodes, 1.3+ million sentences
- Causal inference: Interrupted Time Series (ITS), Difference-in-Differences (DiD), Regression Discontinuity Design, Instrumental Variables, Propensity Score Matching
- Pattern recognition, Topic segmentation, and classification tasks using embedding models
- Human-in-the-loop validation frameworks for improving LLM-based annotations