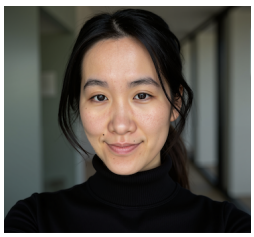
Xixuan Zhang



▪ Berlin, Germany ▪ +49-(0)15225127120 ▪ xixuan.zh@gmail.com
▪ linkedin.com/in/xixuan-zhang ▪ xixuanzhang2022.github.io/portfolio

Profile Summary

Coding: Python, SQL, Java, R, JavaScript **Framework:** React, Node.js, Vue

Toolkit: Git, Tableau, GCP, Docker, HPC **Language:** English (C1), German (C1), Chinese (Native)

Education

Freie Universität Berlin Berlin, Germany

- Ph.D. Candidate, Computational Social Science (Expected 2025) **2020 Present**
- M.A., Media and Political Communication **2016 2019**
- B.A., Media and Communication Studies **2012 2016**

Technische Universität Berlin Berlin, Germany

- M.S., Computer Science (Coursework completed) **2022 Present**
- B.S., Media Informatics **2017 2022**

Professional Experience (selected)

Data Scientist & Research Associate **March 2022 Oct 2025**

Research project "NEOVEX" at Freie Universität Berlin, Berlin, Germany

- Co-developed a graph-based dictionary expansion algorithm to discover domain-specific unknown keywords for data collection by fine-tuning the GloVe model
- Collected a large-scale corpus of 32 million conspiracy theory-related posts from social media, legacy media, and alternative media platforms over eleven years through scraping and API integration, and detected conspiracy theories by fine-tuning different BERT models
- Experimented with and analyzed the corpus using a wide range of NLP methods, including dictionary-based linguistic feature extraction, topic modeling, dependency parsing, unsupervised and supervised text classification, with and without the application of LLMs

Research Associate **May 2019 Sept 2022**

Research group "News, Campaigns, and the Rationality of Public Discourse" at Weizenbaum Institute – the German Internet Institute, Berlin, Germany

- Conducted in-depth analysis of research data using advanced statistical modeling and causal inference techniques
- Explored and integrated various computational methods (e.g., network analysis, machine learning, text mining, and NLP) into political communication research

PROJECTS (SELECTED)

Transnationality in Conspiracy Theories on Reddit | Python **Jan 2024 April 2025**

- Developed a pipeline to extract subject-verb-object triples from 235k Reddit posts and clustered fine-grained narrative patterns using a sentence-embedding-based clustering approach
- Investigated transnational convergence between cross-country narratives, measured cross-country narrative convergence via cosine similarity of narrative vectors, and grouped temporal patterns using dynamic time warping

Explainable AI: Unsupervised Concept Attribution of CNNs | Python **April 2024 Sept 2024**

- Extracted and aggregated activation maps from SimCLR's last layer to identify latent concepts in image features
- Assessed channel importance and feature disentanglement using PCA, tf-idf adaptations, and Shapley values in a linear attribution model
- Enhanced model interpretability and supported explainable AI in unsupervised settings

Tracking Climate Change Revisions in Wikipedia | Python & R **Jan 2022 July 2024**

- Parsed 930k sentence-level revision histories of 891 climate articles on Wikipedia and engineered features on content relevance, edit timing, and user types
- Created a codebook to classify four revision types, curated an annotated dataset using active learning and representative sampling, and fine-tuned BERT models for classification, resulting in a 30% improvement in average F1 score.
- Applied survival analysis, heterogeneity tests, and meta-regression to quantify the effects of content, context, and editor behavior on revision likelihood

Diffusion Dynamics of #FridaysForFuture on Twitter | Python **May 2019 Sept 2021**

- Constructed retweet and follower networks to map diffusion pathways, identified actor communities, and applied embedding clustering and topic modeling to detect frames
- Evaluated diffusion network dynamics using network metrics, measured virality, depth, size, and diffusion capacity of tweet cascades at the community level, and measured engagement and information flow through time-series analysis

PUBLICATIONS (SELECTED)

- Zhang, X. (2025). Decoding revision mechanisms in Wikipedia: Collaboration, moderation, and collectivities. *New Media & Society*. https://doi.org/10.1177/14614448251336418
- Buehling, K., Zhang, X., & Heft, A. (2025). Veiled conspiracism. Particularities and convergence in styles and functions of conspiracy-related communication across digital platforms. *New Media & Society*. https://doi.org/10.1177/1461444825131575
- Schindler, J., Jha, S., Zhang, X., Buehling, K., Heft, A., & Barahona, M. (2025). LGDE: Local Graph-based Dictionary Expansion. *Computational Linguistics*, 1-32.
- Zhang, X. (2023). Diffusion Dynamics and Digital Movement: the Emergence and Proliferation of the German-speaking #FridaysForFuture Network on Twitter. *Social Movement Studies.* http://doi.org/10.1080/14742837.2023.2211015