

Tweeting the 2024 US Presidential Election

Longitudinal Social Media Dataset

Dataset Overview

- ▶ **Core dataset:** 2.1 million tweets from 900 U.S.-based X users
- ▶ **Sampling:** Quota-based approach for population representativeness
- ▶ **Longitudinal design:** Full tweet history over multi-year period
- ▶ **Data format:** CSV file with profile metadata and tweet-level variables
- ▶ **Scope:** Political expression, attitudes, and behavior during 2024 election cycle

Example Analysis Targets

- ▶ Stability and change in expressed political attitudes over time
- ▶ Event-driven shifts in political expression
- ▶ Construction and validation of survey-like variables from social media
- ▶ Uncertainty-aware LLM inference with confidence scores
- ▶ Relationship between disclosure patterns and inferred attitudes
- ▶ Geographic variation in political expression (national vs. state-level)

Potential Data Linkage & Enrichment

- ▶ **Survey data:** GSS, ANES, PEW-style questions for validation
- ▶ **Event data:** News and event timelines for temporal contextualization
- ▶ **Reference data:** Census, demographic, and geographic data
- ▶ **External APIs:** Real-time data enrichment and verification

Analytical Approaches

- ▶ **NLP:** Natural language processing and stance detection
- ▶ **LLM methods:** Survey-like variable construction and annotation
- ▶ **Time-series:** Panel-style analysis of longitudinal patterns
- ▶ **Topic modeling:** Unsupervised discovery of thematic content
- ▶ **Network analysis:** Interaction structures, co-mentions, and virality