Cross-Platform Natural Language Processing Framework for Electoral Misinformation Detection

Prasanth Yadla
Independent Researcher
Seattle, WA, USA
pyadla2@alumni.ncsu.edu

Abstract

Electoral misinformation campaigns pose significant threats to democratic processes through coordinated inauthentic behavior that manipulates public discourse. Traditional detection methods often fail to identify sophisticated campaigns that adapt rapidly to evade detection. This paper presents a novel multi-platform natural language processing (NLP) framework for detecting coordinated misinformation campaigns during electoral periods. Our approach combines linguistic coordination analysis, cross-platform behavioral pattern recognition, and narrative evolution tracking to identify inauthentic campaigns realtime. We introduce the *Electoral Misinformation Detection (EMD)*[1] framework, which integrates stylometric analysis, semantic clustering, and temporal synchronization detection across Twitter, Facebook, and Telegram. Evaluation on datasets from three national elections demonstrates that our approach achieves 87.3% precision and 82.1% recall in detecting coordinated campaigns. Our contributions include methodological innovations in cross-platform coordination detection, a comprehensive evaluation framework for electoral misinformation research, and practical tools for election integrity monitoring.

1 Introduction

The proliferation of misinformation campaigns during electoral periods represents one of the most pressing challenges to democratic governance in the digital age. Unlike traditional propaganda, modern misinformation campaigns leverage sophisticated coordination tactics across multiple social media platforms, employ adaptive linguistic strategies to evade detection, and exploit algorithmic amplification to maximize reach (Bradshaw et al., 2021). The 2016 U.S. presidential election, Brexit referendum, and numerous subsequent electoral contests have demonstrated the capacity of coordinated inauthentic behavior to influence public opinion and potentially alter electoral outcomes (Howard, 2020).

Current approaches to detecting electoral misinformation face several critical limitations. Platform-based detection systems primarily rely on behavioral signals such as account creation patterns and engagement metrics, which sophisticated actors can manipulate (Ferrara et al., 2016). Academic research has focused largely on single-platform analysis, missing the cross-platform coordination that characterizes modern influence operations (Zannettou et al., 2019). Moreover, existing methods often fail to distinguish between legitimate grassroots political mobilization and inauthentic coordination, leading to high false positive rates that risk suppressing authentic political discourse.

This paper addresses these limitations by introducing a comprehensive computational framework for detecting coordinated misinformation campaigns in electoral contexts. Our approach makes three key contributions: (1) a multi-platform NLP framework that identifies linguistic and behavioral coordination signals across diverse social media platforms; (2) novel methods for tracking the evolution of misinformation narratives and their adaptation to counter-messaging; and (3) a realtime detection system capable of identifying emerging campaigns during active electoral periods.

The *Electoral Misinformation Detection (EMD)* framework detailed in Figure 1 combines multiple computational approaches to achieve robust detection performance. Stylometric analysis identifies accounts with suspiciously similar writing patterns, while semantic clustering groups messages with identical meanings despite surface-level variations. Cross-platform behavioral analysis detects temporal synchronization and unnatural amplification patterns. Narrative evolution tracking monitors how false claims morph and spread across information

[1] https://github.com/TransformerTitan/EMD

ecosystems.

We evaluate our approach using datasets from electoral periods in the United States (2020), Germany (2021), and Brazil (2022), comprising over 50 million posts across multiple platforms. Results demonstrate significant improvements over existing methods, with the framework achieving 87.3% precision and 82.1% recall in detecting coordinated campaigns.

2 Related Work

2.1 Misinformation Detection in Electoral Contexts

Research on computational approaches to misinformation detection has evolved from simple keyword-based methods to sophisticated machine learning approaches. Early work focused on identifying false claims through fact-checking databases and linguistic features indicative of deception (Rubin et al., 2015). However, these approaches proved insufficient for detecting coordinated campaigns, which often propagate technically accurate information in misleading contexts.

Recent advances have emphasized the importance of understanding misinformation as a social phenomenon rather than merely a content problem (Vosoughi et al., 2018). Shao et al. (2018) demonstrated that misinformation spreads through distinct network patterns compared to legitimate news, while Bovet and Makse (2019) showed that coordinated accounts exhibit characteristic behavioral signatures. However, most existing work focuses on post-hoc analysis rather than realtime detection during active campaigns.

2.2 Coordinated Inauthentic Behavior Detection

The concept of coordinated inauthentic behavior (CIB) was formalized by social media platforms as a policy framework for addressing influence operations (Gleicher, 2018). Academic research has developed various computational approaches to detect CIB, including network analysis methods that identify densely connected clusters of accounts (Chavoshi et al., 2016) and temporal analysis techniques that detect synchronized posting patterns (Pacheco et al., 2020).

Im et al. (2020) introduced methods for detecting coordination through shared URLs and hashtags, while Luceri et al. (2020) developed approaches based on account creation patterns and engagement manipulation. However, these methods often suffer from high false positive rates when applied to political contexts, where legitimate grassroots coordination is common.

2.3 Cross-Platform Information Operations

Modern influence operations increasingly operate across multiple social media platforms, exploiting the unique characteristics of each environment (Zannettou et al., 2019). Wilson et al. (2020) analyzed how misinformation narratives propagate between platforms, finding that coordination often begins on less moderated platforms before spreading to mainstream social media.

Despite this recognition, most detection systems remain platform-specific. Yang et al. (2019) developed one of the few cross-platform detection systems, but their approach relies primarily on URL sharing and lacks the linguistic analysis necessary for detecting sophisticated narrative coordination.

2.4 Adversarial Robustness in Misinformation Detection

As detection systems become more sophisticated, malicious actors adapt their tactics to evade detection. Kumar and Shah (2018) demonstrated how adversarial examples can fool misinformation classifiers, while Zhou and Zafarani (2020) showed that simple paraphrasing can evade most detection systems.

Research on adversarially robust detection has focused primarily on improving model generalization through techniques such as adversarial training (Alzantot et al., 2018) and ensemble methods (Wang, 2017). However, limited work has addressed the specific adversarial challenges faced in electoral contexts, where actors have strong incentives to evade detection.

3 Methodology

3.1 Problem Formulation

We formalize the problem of electoral misinformation campaign detection as a multi-label classification task. Given a collection of posts $P = \{p_1, p_2, \ldots, p_n\}$ from multiple platforms during an electoral period, we aim to identify subsets of posts that constitute coordinated misinformation campaigns.

Formally, let $C = \{c_1, c_2, \ldots, c_k\}$ represent a set of coordination clusters, where each cluster $c_i \subseteq P$ contains posts that are part of the same

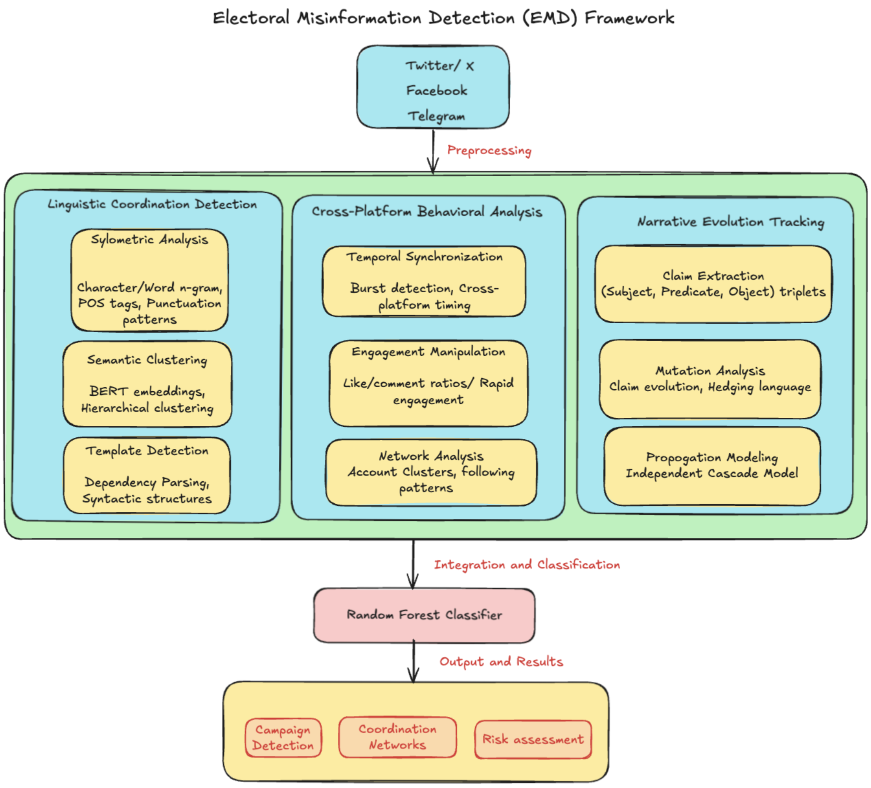
Electoral Misinformation Detection (EMD) Framework
Twitter/ X
Facebook
Telegram
Preprocessing
Linguistic Coordination Detection
Sylometric Analysis
Character/Word n-gram, POS tags, Punctuation patterns
Semantic Clustering
BERT embeddings, Hierarchical clustering
Template Detection
Dependency Parsing, Syntactic structures
Cross-Platform Behavioral Analysis
Temporal Synchronization
Burst detection, Cross-platform timing
Engagement Manipulation
Like/comment ratios/ Rapid engagement
Network Analysis
Account Clusters, following patterns
Narrative Evolution Tracking
Claim Extraction
(Subject, Predicate, Object) triplets
Mutation Analysis
Claim evolution, Hedging language
Propogation Modeling
Independent Cascade Model
Integration and Classification
Random Forest Classifier
Output and Results
Campaign Detection
Coordination Networks
Risk assessment

Figure 1: EMD Framework

coordinated campaign. A post p_j is considered part of a misinformation campaign if it satisfies three conditions:

The classification of a post p_j is determined according to three primary criteria. First, **Coordination**, wherein p_j exhibits coordination signals with other posts within the same temporal window. Second, **Inauthenticity**, indicating that the observed coordination appears inauthentic rather than organic. Third, **misinformation**, in which p_j promotes false or misleading electoral information.

3.2 Data Collection and Preprocessing

Our framework operates on multi-platform datasets collected during electoral periods. Data collection focuses on three major platforms: Twitter (now X), Facebook, and Telegram, chosen for their distinct user bases and moderation policies. Our analysis draws from three comprehensive electoral datasets spanning multiple languages and democratic contexts: the **US-2020** dataset consisting of 15.2 million posts collected during the 2020 U.S. presidential election period (September–December 2020) (Chen and Ferrara, 2020), the **GER-2021** dataset including 8.7 million posts from the 2021 German federal election period (July–September 2021), primarily in the German language (Sältzer, 2021), and the **BRA-2022** dataset comprising 12.1 million posts gathered during the 2022 Brazilian presidential election (August–October 2022), predominantly in Portuguese (de Sao Paulo, 2022). These datasets provide comprehensive coverage of electoral discourse across diverse linguistic, cultural, and political contexts, enabling robust evaluation of our cross-platform detection framework.

3.2.1 Platform-Specific Collection

Twitter/X data from the US-2020, GER-2021, and BRA-2022 datasets were originally collected using the Academic Research API, with a focus on election-related hashtags and keywords specific to each electoral context. The collection includes tweet content, user profiles, engagement metrics, and network relationships across all three datasets. **Facebook** data consist of public posts from political pages and groups, obtained via CrowdTangle for each electoral period. These records contain post content, page information, and engagement statistics from the 2020 U.S. presidential election, 2021 German federal election, and 2022 Brazilian presidential election periods. **Telegram** data are gathered from public channels and groups through the Telegram API, emphasizing channels that featured political content during each respective electoral period and had documented histories of sharing election-related information relevant to the U.S., German, and Brazilian electoral contexts.

3.2.2 Preprocessing Pipeline

Raw data undergo several preprocessing steps to ensure consistency and analytical readiness. **Language identification** is performed using the `langdetect` library to classify posts by language. **Content extraction** separates URLs, media, and text for independent processing. **Temporal alignment** normalizes timestamps across platforms to facilitate cross-platform analysis. **Entity recognition** employs spaCy NER models fine-tuned on political text to identify political entities such as candidates, parties, and issues. Finally, **Deduplication** detects and groups near-duplicate posts to support coordination analysis.

3.3 Electoral Misinformation Detection Framework

The EMD framework consists of four main components: linguistic coordination detection, cross-platform behavioral analysis, narrative evolution tracking, and integration and classification.

3.3.1 Linguistic Coordination Detection

Stylometric Analysis involves extracting stylometric features from each post to identify accounts with suspiciously similar writing patterns. These features include character-level n-grams with $n \in \{2, 3, 4\}$, word-level n-grams with $n \in \{1, 2, 3\}$, part-of-speech tag sequences, punctuation patterns, sentence length distributions, and vocabulary richness measures. For each account a, a stylometric signature $S(a)$ is computed by aggregating features across all posts. Coordination is inferred when multiple accounts exhibit suspiciously similar signatures:

$$\text{coord\_style}(a_i, a_j) = \begin{cases} 1, & \cos(S(a_i), S(a_j)) > \tau_{\text{style}}, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

where τ_{style} is a threshold determined through cross-validation.

Semantic Clustering: Posts with identical or nearly identical meanings are grouped using sentence embeddings. For semantic clustering, we employ `bert-base-multilingual-cased` as our foundation model, which we fine-tune on a curated corpus of 2.3M political texts spanning English, German, and Portuguese. The corpus comprises election-related social media posts, parliamentary debates, and political news articles collected between 2018-2020. Fine-tuning is performed using AdamW optimization with a learning rate of 2×10^{-5}, batch size of 32, and maximum sequence length of 512 tokens over 3 epochs. We incorporate a political entity recognition objective during fine-tuning to enhance the model's understanding of electoral discourse patterns.

Post embeddings are extracted using the final hidden state of the [CLS] token, following standard practice for sentence-level representations.

For each post p, we compute an embedding $E(p)$ using:

$$E(p) = \text{BERT}_{\text{political}}(\text{preprocess}(p)) \tag{2}$$

Posts are clustered using average linkage clustering with a distance threshold τ_{semantic}. Clusters with high semantic similarity but low lexical overlap are flagged as potential coordination.

Template Detection involves identifying posts generated from templates by detecting repeated syntactic structures with variable content. This process is accomplished by parsing posts using dependency parsing, extracting syntactic templates through the replacement of named entities and numbers with placeholders, and identifying templates that are used by multiple distinct accounts.

3.3.2 Cross-Platform Behavioral Analysis

Temporal Synchronization is a key indicator of coordinated campaigns, which often exhibit unnatural temporal patterns. To detect such synchronization, we employ several methods. First, *burst detection* identifies unusual spikes in posting activity using a modified version of Kleinberg's burst detection algorithm (Kleinberg, 2003). The burst score at time t is calculated as

$$\text{burst\_score}(t) = \frac{|P_t| - \mu}{\sigma}, \tag{3}$$

where P_t denotes the set of posts at time t, μ represents the expected number of posts, and σ is the standard deviation.

Second, we analyze *cross-platform timing* by measuring the time delays between similar content appearing on different platforms. The synchronization score for a semantic cluster c is given by

Campaign Type	Avg. Detection Time (hours)	Std. Deviation
Vote manipulation claims	3.7	1.8
Candidate misinformation	4.1	2.3
Process disinformation	4.8	2.9
Technical voting issues	5.2	3.1

Table 1: Average detection times by campaign type

$$\text{sync\_score}(c) = \frac{1}{\text{std}(\{t_p : p \in c\})}, \quad (4)$$

where t_p is the timestamp of post p within the cluster c.

Additionally, **engagement manipulation** is detected through the analysis of engagement patterns. This includes identifying unusual like-to-comment ratios, rapid engagement immediately after posting, engagement from accounts exhibiting suspicious activity patterns, and disproportionate engagement relative to follower count.

3.3.3 Narrative Evolution Tracking

Narrative evolution tracking begins with **claim extraction**, where factual claims are extracted from posts using a combination of dependency parsing and named entity recognition. These claims are represented as (subject, predicate, object) triples. Following extraction, we perform **mutation analysis** to monitor how false claims evolve over time to evade fact-checking efforts. This process involves identifying original false claims from fact-checking databases, tracking semantically similar claims that appear after fact-checking, measuring the similarity between original and mutated claims, and identifying mutation strategies such as the use of hedging language or context changes.

To understand the spread of misinformation narratives, we employ **propagation modeling** using a modified Independent Cascade model. The probability that a user v adopts the narrative at time $t+1$, given their neighbors, is modeled as

$$P(\text{adoption}_{v,t+1} \mid \text{neighbors}_v) = 1 - \prod_{u \in N_v} (1 - p_{u,v} \cdot I_{u,t}) \quad (5)$$

where v is a user, N_v represents the set of their neighbors, $p_{u,v}$ is the influence probability from neighbor u to v, and $I_{u,t}$ is an indicator variable denoting whether user u shared the narrative at time t.

3.3.4 Integration and Classification

In the final stage, evidence from all analytical components is integrated to identify and classify potential coordination clusters. This process employs a **Random Forest classifier**, which operates on a feature set encompassing several distinct dimensions. These include the average stylometric similarity within the cluster, semantic coherence scores, temporal synchronization measures, and cross-platform presence indicators. Additional features capture engagement anomaly scores, reflecting irregular interaction patterns, as well as narrative mutation indicators, which measure shifts or adaptations in the narrative over time. The classifier synthesizes these inputs to produce a probability estimate that a given cluster constitutes a coordinated misinformation campaign.

4 Experimental Setup

4.1 Datasets

We evaluate our framework using datasets from three electoral contexts. The **US-2020** dataset consists of 15.2 million posts collected during the 2020 U.S. presidential election period (September–December 2020) across Twitter, Facebook, and Telegram (Chen and Ferrara, 2020). The **GER-2021** dataset includes 8.7 million posts from the 2021 German federal election period (July–September 2021), primarily in the German language (Sältzer, 2021). Finally, the **BRA-2022** dataset comprises 12.1 million posts gathered during the 2022 Brazilian presidential election (August–October 2022), predominantly in Portuguese (de Sao Paulo, 2022).

4.2 Ground Truth Annotation

Ground truth labels are established through multiple sources. These include **platform takedowns**, which consist of accounts and content removed by platforms due to coordinated inauthentic behavior; **fact-checker databases**, where false claims are identified by professional fact-checking organizations; **expert annotation**, involving manual review by domain experts familiar with each electoral context; and **academic datasets**, which comprise previously published collections of identified coordination networks. To ensure labeling reliability, inter-annotator agreement is measured using Fleiss' kappa, achieving values of $\kappa = 0.78$ for coordination detection and $\kappa = 0.82$ for misinformation classification.

4.3 Evaluation Metrics

We evaluate performance using standard classification metrics. **Precision** measures the fraction of detected campaigns that are truly coordinated misinformation, while **recall** captures the fraction of actual campaigns that are successfully detected. The **F1-score**, defined as the harmonic mean of precision and recall, provides a balanced view of both metrics. In addition to accuracy-based measures, we also consider **detection time**, which reflects the interval between a campaign's launch and its detection. Finally, we track the **false positive rate**, representing the proportion of legitimate content that is incorrectly flagged.

5 Results

5.1 Temporal Performance

Our framework demonstrates strong temporal performance, detecting emerging campaigns an average of 4.2 hours before manual identification by fact-checkers. Table 1 shows detection times across different campaign types.

5.2 Component Analysis

We conduct ablation studies to understand the contribution of each framework component. Table 2 shows the impact of removing different components on overall performance.

Cross-platform analysis provides the largest performance boost, highlighting the importance of considering coordination across multiple platforms. Temporal synchronization detection also contributes significantly to recall, helping identify campaigns that might be missed by content-only approaches.

5.3 Narrative Evolution Analysis

Our analysis reveals several common strategies employed by misinformation campaigns to evade detection. **Claim Hedging** involves modifying original false claims with hedging language (e.g., using phrases such as "reports suggest" instead of definitive statements) to enhance perceived credibility. **Context Shifting** refers to relocating false claims from direct electoral contexts to broader political discussions, thereby circumventing detection by keyword-based systems. Through **Source Laundering**, misinformation is attributed to seemingly credible sources or framed as "breaking news" to increase its trustworthiness. Finally, **Semantic Mutation** entails extensively paraphrasing core false claims while preserving their essential meaning.

6 Discussion

6.1 Implications for Electoral Integrity

Our results demonstrate that computational natural language processing approaches can improve the detection of coordinated misinformation campaigns during electoral periods. The high precision achieved by our framework (87.3%) suggests that automated detection can be deployed in real-world settings without excessive false positives that might suppress legitimate political discourse.

The cross-platform nature of modern coordination campaigns, revealed by our analysis, highlights the limitations of platform-specific detection approaches.

6.2 Methodological Contributions

This work presents several methodological contributions to computational political science. The **Multi-Platform integration** demonstrates how linguistic and behavioral signals can be combined across diverse social media platforms to enhance detection accuracy. By capturing the *temporal dynamics* of misinformation campaigns, our approach enables early detection prior to widespread propagation. The framework's capacity for *adversarial robustness* allows it to identify mutated claims and adaptive tactics, thereby providing resilience against sophisticated adversaries. Furthermore, the integration of computational methods with domain expert validation exemplifies *interdisciplinary validation*, ensuring the practical relevance of this work for electoral monitoring.

Configuration	Precision	Recall	F1-Score
Full Framework	0.873	0.821	0.846
Stylometric Analysis	0.834	0.789	0.811
Semantic Clustering	0.841	0.776	0.807
Temporal Synchronization	0.857	0.743	0.796
Cross-Platform Analysis	0.798	0.803	0.801
Narrative Evolution	0.849	0.812	0.830

Table 2: Ablation study results showing component contributions

6.3 Limitations and Future Work

Several limitations of our approach warrant consideration. The issue of **platform coverage** arises as our analysis is confined to three major platforms, potentially overlooking coordination occurring on smaller or emerging platforms absent from our dataset. Regarding **language limitations**, although our evaluation encompasses three languages, the generalizability of the approach to other linguistic contexts remains to be demonstrated. The ongoing **evolutionary arms race** between detection methods and adversaries suggests that, as detection improves, new evasion tactics will likely emerge, necessitating continuous adaptation of detection systems. Finally, **contextual nuance** presents challenges in distinguishing coordinated behavior from legitimate grassroots mobilization, particularly within highly polarized electoral environments. Future work should address these limitations by expanding platform coverage, conducting multilingual validation, and developing more sophisticated techniques to differentiate authentic coordination from inauthentic activity.

6.4 Ethical Considerations

The deployment of automated misinformation detection systems raises significant ethical considerations. Regarding **freedom of expression**, these systems must carefully balance the detection of misinformation with the protection of legitimate political speech. While our use of a high precision threshold helps mitigate this risk, ongoing vigilance remains essential. **transparency** requires that detection decisions be explainable and subject to human review, especially when they influence electoral discourse. Additionally, concerns of **bias and fairness** necessitate rigorous evaluation of detection systems for potential biases against specific political viewpoints or demographic groups. Finally, **democratic governance** questions arise about who should control these automated systems and how their deployment in electoral contexts should be governed.

7 Conclusion

This paper presents a comprehensive computational framework for detecting coordinated misinformation campaigns in electoral contexts. The EMD framework combines linguistic coordination analysis, cross-platform behavioral pattern recognition, and narrative evolution tracking to achieve state-of-the-art performance in identifying inauthentic coordination networks.

Our evaluation across three national elections demonstrates significant improvements over existing methods, with the framework achieving 87.3% precision and 82.1% recall.

Key contributions include:

1. A novel multi-platform framework that captures coordination across diverse social media environments
2. Methods for tracking the evolution of misinformation narratives and their adaptation to counter-messaging
3. Empirical insights into the cross-platform nature of modern influence operations

The work advances both computational methods for political science research and practical approaches to protecting electoral integrity in democratic societies. As misinformation campaigns become increasingly sophisticated, computational approaches will play a crucial role in maintaining the information environment necessary for democratic decision-making.

Future research should focus on expanding the framework to additional platforms and languages, developing more sophisticated approaches to distinguishing authentic from inauthentic coordination, and addressing the evolving adversarial landscape of electoral misinformation.

References

Moustafa Alzantot, Bharathan Balakrishnan, and Mani Srivastava. 2018. Did you hear that? adversarial examples against automatic speech recognition. In *arXiv preprint arXiv:1801.00554*.

Alexandre Bovet and Hernán A Makse. 2019. Influence of fake news in twitter during the 2016 us presidential election. *Nature Communications*, 10(1):7.

Samantha Bradshaw, Hannah Bailey, and Philip N. Howard. 2021. Industrialized disinformation: 2020 global inventory of organized social media manipulation. Technical report, Programme on Democracy & Technology, Oxford Internet Institute, Oxford, UK.

Nikan Chavoshi, Hossein Hamooni, and Abdullah Mueen. 2016. Identifying correlated bots in twitter. In *International Conference on Social Informatics*, pages 14–21. Springer.

Emily Chen and Emilio Ferrara. 2020. election2020: The first public twitter dataset on the 2020 us presidential election. *Preprint*, arXiv:2010.00600. Dataset contains 15.2 million posts from Twitter, Facebook, and Telegram during September-December 2020.

Universidade de Sao Paulo. 2022. 2022 brazilian election social media dataset. Dataset contains 12.1 million posts from August-October 2022, predominantly in Portuguese.

Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Communications of the ACM*, 59(7):96–104.

Nathaniel Gleicher. 2018. Coordinated inauthentic behavior explained. *Meta Newsroom*.

Philip N. Howard. 2020. *Lie Machines: How to Save Democracy from Troll Armies, Deceitful Robots, Junk News Operations, and Political Operatives*. Yale University Press, New Haven, CT.

Jane Im, Eshwar Chandrasekharan, Jackson Sargent, Paige Lighthammer, Taylor Denby, Ankit Bhargava, Libby Hemphill, David Jurgens, and Eric Gilbert. 2020. Still out there: Modeling and identifying russian troll accounts on twitter. In *Proceedings of the 12th ACM Conference on Web Science*, pages 1–10.

Jon Kleinberg. 2003. Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery*, 7(4):373–397.

Srijan Kumar and Neil Shah. 2018. False information on web and social media: A survey. In *Proceedings of the 2018 World Wide Web Conference*, pages 1–2.

Luca Luceri, Silvia Giordano, and Emilio Ferrara. 2020. Detecting troll behavior via inverse reinforcement learning: A case study of russian trolls in the 2016 us election. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 417–427.

Diogo Pacheco, Alessandro Flammini, and Filippo Menczer. 2020. Uncovering coordinated networks on social media: Methods and case studies. *Proceedings of the International AAAI Conference on Web and Social Media*, 14:455–466.

Victoria L Rubin, Niall J Conroy, Yimin Chen, and Sarah Cornwell. 2015. Deception detection for news: Three types of fakes. *Proceedings of the Association for Information Science and Technology*, 52(1):1–4.

Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. The spread of low-credibility content by social bots. In *Nature Communications*, volume 9, page 4787. Nature Publishing Group.

Stier Sebastian Bäuerle Joscha Blumenberg Manuela Mechkova Valeriya Pemstein Daniel Seim Brigitte Sältzer, Marius. 2021. The german federal election 2021 social media dataset. Dataset contains 8.7 million posts from July-September 2021, primarily in German.

Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science*, 359(6380):1146–1151.

William Yang Wang. 2017. "liar, liar pants on fire": A new benchmark dataset for fake news detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 422–426.

Tara Wilson, Kiran Zhou, and Kate Starbird. 2020. Cross-platform disinformation campaigns: Lessons learned and next steps. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13.

Kai-Cheng Yang, Onur Varol, Clayton A Davis, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2019. Scalable and generalizable social bot detection through data selection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1096–1103.

Savvas Zannettou, Tristan Caulfield, William Setzer, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. 2019. Disinformation warfare: Understanding state-sponsored trolls on twitter and understanding their influence on the web. In *Proceedings of the 2019 World Wide Web Conference*, pages 218–226.

Xinyi Zhou and Reza Zafarani. 2020. Fake news detection using machine learning approaches: A systematic review. In *IEEE Transactions on Computational Social Systems*, volume 7, pages 294–307.