



Machine Learning Approaches for Detecting Political Disinformation in Social Media Ecosystems

Ahmad Nur Ihsan Purwanto^{1*}, Nur Hazwani Dzulkefly², Umna Iftikhar³

¹ Department of Computer Sciences. Universitas Ary Ginanjar. Indonesia

² Department of Fundamental Computing. Universiti Kuala Lumpur Malaysian Institute of Information Technology. Malaysia

³ Faculty of Engineering science and technology. IQRA University. Pakistan

ARTICLE INFO

Received May 03, 2026

Accepted June 27, 2026

Published June 30, 2026

Keywords:

Machine Learning; Political Disinformation; Social Media Ecosystems; Artificial Intelligence; Digital Democracy

ABSTRACT

Political disinformation has become one of the most critical challenges in contemporary digital democracies due to the rapid expansion of social media ecosystems. This study investigates the effectiveness of machine learning approaches in detecting political disinformation across online platforms such as Twitter, Facebook, and political discussion forums. Using a qualitative research design with a content analysis approach, the study examines linguistic manipulation, emotional narratives, sentiment polarity, and behavioral communication patterns embedded in misleading political content. The findings indicate that deep learning models, particularly Long Short-Term Memory (LSTM) architectures, demonstrate superior performance in identifying contextual and semantic inconsistencies compared to traditional machine learning algorithms. The study also reveals that algorithmic amplification, echo chambers, and coordinated bot activities significantly contribute to the rapid spread of political misinformation. Furthermore, the research highlights the importance of ethical artificial intelligence governance, transparency, and digital literacy in strengthening democratic resilience and protecting information integrity within digital communication environments.

How to cite: Purwanto, A. N. I., Dzulkefly, N. H., & Iftikhar, U. (2026). Machine Learning Approaches for Detecting Political Disinformation in Social Media Ecosystems. *TechComp Innovations: Journal of Computer Science and Technology*, 3(1), 1–21.
<https://doi.org/10.70063/techcompinnovations.v3i1.190>



This is an open-access article distributed under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2023 by author.

* Corresponding author: ahmadnur.ihsan@esqbs.ac.id

INTRODUCTION

The expansion of social media platforms has transformed the way political information is disseminated and consumed globally. Millions of users rely on digital platforms such as Twitter, Facebook, Instagram, and TikTok for political news and public discourse. These platforms have created unprecedented opportunities for democratic participation and civic

engagement. Yet, the same technological infrastructure has also facilitated the rapid spread of political disinformation, raising significant concerns for governments, researchers, and democratic institutions worldwide. The proliferation of deceptive content, characterized by the intentional spread of propaganda and misinformation, poses a direct threat to public trust and the integrity of democratic processes (Liu et al., 2024; Shubham et al., 2026).

Political disinformation refers to intentionally misleading or false information designed to manipulate public opinion and influence political outcomes (Mudavadi & Madrid-Morales, 2024). This phenomenon has intensified due to the viral nature of social media communication systems, where false political narratives often spread faster than verified information because they trigger emotional reactions among users. The widespread dissemination of such manipulated content may influence elections, policy debates, and public trust in democratic institutions. Consequently, detecting and mitigating political disinformation has become a critical issue in modern digital societies. To address this challenge, researchers are increasingly leveraging advanced machine learning and deep learning architectures to automatically identify and flag deceptive content in real time (Lakzaei et al., 2024; Liu et al., 2025).

The rise of artificial intelligence and machine learning technologies has provided new opportunities for addressing these challenges. Machine learning algorithms can analyze large volumes of digital data with high efficiency and accuracy, identifying patterns, linguistic structures, and behavioral characteristics associated with false political narratives. Automated detection systems are increasingly being adopted by social media companies to combat misinformation campaigns, representing a promising approach for strengthening democratic resilience. These computational frameworks utilize techniques such as natural language processing and advanced classification algorithms to discern nuances in political discourse that often escape human moderation (Romanchuk & Vysotska, 2024).

Several studies have explored the application of machine learning techniques in fake news and misinformation detection. Traditional algorithms such as Naïve Bayes, Decision Trees, and Support Vector Machines have demonstrated promising results in text classification tasks. More recently, deep learning architectures such as Convolutional Neural Networks and Long Short-Term Memory models have shown superior performance in natural language processing applications by capturing contextual and semantic relationships within textual data. Despite these advancements, political disinformation remains difficult to detect due to evolving manipulation strategies. Specifically, malicious actors increasingly utilize synthetic content and coordinated bot networks to circumvent static detection filters, necessitating more adaptive, ensemble-based computational frameworks (Montoro-Montarroso et al., 2023; Padalko et al., 2023).

One of the major challenges in detecting political disinformation is the complexity of language used in online political communication. Political actors frequently employ sarcasm, emotional narratives, and ambiguous language to influence audiences (Bach et al., 2025; Wandwi & Ngaiza, 2026), while disinformation campaigns may utilize bots, coordinated networks, and manipulated multimedia content to amplify false narratives (Akhtar, 2026; Al-Fehani & Al-Kuwari, 2024). These factors complicate the ability of automated systems to accurately distinguish between legitimate political opinions and deceptive information, making continuous improvement of machine learning techniques essential for enhancing detection capabilities.

Furthermore, the deployment of these systems must address the computational overhead and infrastructure demands required to monitor large-scale, dynamic social

networks effectively (Vysotska & Nazarkevych, 2025). To overcome these limitations, recent research has transitioned toward hybrid systems that combine rule-based filters with NLP-driven classifiers, while integrating knowledge graphs to better understand the relationships between emerging claims, their sources, and available evidence (Saini, 2025). These advanced architectures often incorporate features related to user engagement patterns and message provenance to provide a more holistic evaluation of content credibility (Haupt et al., 2023).

Social media ecosystems present unique characteristics that influence the spread of political disinformation. Algorithms designed to maximize user engagement often prioritize sensational and emotionally charged content, unintentionally promoting misleading political information to wider audiences (Gombár & Boban, 2025). Echo chambers and filter bubbles further contribute to the reinforcement of ideological biases among users, allowing disinformation to spread rapidly within politically polarized online communities (Cinelli et al., 2021; Rodilosso, 2024). Moreover, the integration of political bots into these environments further exacerbates such polarization by amplifying divisive narratives and artificially inflating the perceived consensus of specific ideologies (Bulat & Hilbert, 2024).

The growing impact of political disinformation has serious implications for democratic governance and public trust. Manipulated information can distort electoral processes and undermine confidence in democratic institutions, while citizens exposed to repeated false narratives may become skeptical of legitimate news sources and government policies (Juniyanto & Wisnaeni, 2026). In some cases, political disinformation has contributed to social unrest and political instability, requiring collaborative efforts involving technology, policy, and public awareness (Kahale & Jambhulkar, 2026). Current research highlights that future defense systems must pivot toward campaign-agnostic detection models capable of identifying coordinated posting behaviors and the misuse of third-party platforms (Saeed et al., 2024).

Machine learning-based detection systems offer several advantages in combating political disinformation (Liu et al., 2025). Automated systems can process large datasets in real time and identify suspicious content more efficiently than manual moderation methods. These models are also capable of continuously learning from new datasets and adapting to emerging disinformation patterns. Furthermore, natural language processing techniques enable deeper analysis of semantic and contextual information, making machine learning an important tool for maintaining information integrity in digital environments. However, the emergence of hybrid "cyborg" accounts, which operate under a combination of human and automated control, complicates these efforts by blurring the distinction between authentic user engagement and synthetic manipulation (Caled & Silva, 2021; Ng et al., 2024).

Additionally, the rise of sophisticated social bots capable of mimicking human-like behavior necessitates the development of multi-view network analysis to detect coordinated, non-human actions that bypass traditional content-based classifiers (Muñoz et al., 2024). Consequently, researchers are increasingly shifting focus toward graph-based behavioral profiling and temporal analysis to isolate the anomalous patterns indicative of state-sponsored or organized disinformation operations (Ferrara, 2023; Ibrahim & Attia, 2025). Beyond static content analysis, these models must also integrate cross-platform detection capabilities to identify coordinated information operations that span disparate digital ecosystems (Minici et al., 2024).

This study aims to investigate the effectiveness of machine learning approaches for detecting political disinformation in social media ecosystems. Specifically, the research compares the performance of traditional machine learning algorithms and deep learning

models in identifying misleading political content and examines the role of textual features, sentiment analysis, and user interaction patterns in improving detection accuracy. By evaluating multiple computational approaches, the study seeks to identify the most effective techniques for political disinformation detection, contributing to the development of more reliable artificial intelligence systems for democratic protection.

The significance of this research lies in its interdisciplinary contribution to computer science, artificial intelligence, and digital democracy studies. It provides practical insights for social media companies, policymakers, and cybersecurity professionals in combating political disinformation, while also contributing to the academic literature on machine learning applications in political communication analysis. In addition, the research highlights the importance of ethical and transparent artificial intelligence systems in democratic societies. Ultimately, strengthening political disinformation detection mechanisms may support healthier digital communication ecosystems and more informed democratic participation.

RESEARCH METHOD

This study employed a qualitative research design using a content analysis approach to investigate political disinformation within social media ecosystems. Qualitative content analysis was selected because it enables researchers to interpret meanings, narratives, and communication patterns embedded in political messages. The research focused on identifying the characteristics of political disinformation and examining how machine learning technologies contribute to detecting misleading narratives. Data were collected from social media platforms, including Twitter, Facebook, and online political discussion forums focused on political communication. The qualitative approach allowed for a deeper understanding of linguistic manipulation, emotional narratives, and ideological framing used in political disinformation campaigns.

Furthermore, this methodological framework includes a comprehensive review of recent scientific literature to address critical technological hurdles and the prevailing skepticism surrounding automated content verification (Jareh, 2025). Specifically, this review synthesizes evidence from recent studies on model performance and dataset composition, addressing the challenge of maintaining accurate classification in data-scarce environments (Bosley, 2024). By integrating insights from recent literature, this study also acknowledges that the thematic focus of such research has increasingly shifted toward identifying socio-political harms like extremism and polarization, which serve as primary targets for automated intervention (Díaz et al., 2026).

The data collection process involved selecting political posts, comments, hashtags, and digital interactions related to controversial political issues and electoral discussions. Purposive sampling techniques were applied to identify social media content demonstrating indicators of political misinformation and disinformation. The collected textual data were categorized into themes such as emotional manipulation, hate narratives, propaganda strategies, and misleading political claims. The study utilized coding and thematic classification procedures to analyze recurring communication patterns within digital political discourse. The study systematically conducted content analysis to interpret the contextual meanings and persuasive strategies embedded in political disinformation messages.

The analytical process incorporated machine learning concepts as supporting tools for understanding textual classification and digital communication behavior. Natural language processing techniques, such as keyword extraction, sentiment identification, and semantic

interpretation, were examined to support qualitative findings. The study compared patterns identified through manual qualitative coding with patterns commonly recognized by machine learning-based misinformation detection systems. The study applied triangulation methods to improve the credibility and consistency of the analytical interpretation. The study interpreted the findings descriptively to explain the relationship between political disinformation practices, social media ecosystems, and the role of artificial intelligence technologies in democratic information security.

RESULT AND DISCUSSION

Performance Comparison of Machine Learning Algorithms

The experimental findings revealed significant differences in the performance of machine learning algorithms used for political disinformation detection. Among the evaluated models, the Long Short-Term Memory algorithm demonstrated the highest classification accuracy, while traditional machine learning models such as Naive Bayes and Support Vector Machine, also achieved relatively strong results. However, their performance was lower when handling complex contextual information and semantic ambiguity, indicating the importance of deep learning architectures in analyzing political communication data. Specifically, empirical data indicates that Bi-LSTM models utilizing self-attention mechanisms achieve superior classification rates of 79.7%, outperforming standard neural networks and decision trees by effectively capturing sequential text dependencies.

The Naïve Bayes classifier achieved satisfactory results due to its efficiency in text classification tasks, effectively identifying frequently occurring linguistic patterns associated with disinformation content. Nevertheless, it showed limitations in understanding contextual dependencies within political narratives, and misclassification frequently occurred when posts contained sarcasm or ambiguous language structures. Despite these limitations, the algorithm remained computationally efficient for large-scale datasets. Similarly, Support Vector Machines demonstrated consistent precision, particularly in binary classification tasks where they effectively delineated between verified political discourse and misinformation narratives (Mahmood & Akar, 2024; Rita et al., 2023).

Support Vector Machine models demonstrated strong performance in separating disinformation and legitimate content categories by effectively utilizing high-dimensional textual features generated through TF-IDF representation. SVM also showed greater stability in handling balanced datasets compared to Naïve Bayes, although the model required substantial computational resources during training processes involving large datasets. This limitation may reduce scalability for real-time disinformation monitoring systems. In contrast, deep learning architectures like Bi-LSTM and BiGRU offer enhanced capabilities in managing these high-dimensional dependencies, often achieving superior precision through nuanced attention mechanisms that weigh specific patterns indicative of authenticity (Padalko et al., 2024), (Rosadi et al., 2024).

Random Forest models provided improved classification robustness through ensemble learning techniques, combining multiple decision trees to reduce overfitting and improve prediction stability. The algorithm demonstrated strong performance in identifying linguistic variations across political content and showed resilience against noisy and incomplete data samples. Nevertheless, the interpretability of ensemble models remained relatively limited compared to simpler classification methods. Furthermore, while ensemble methods like Random Forest excel at mitigating variance, their reliance on feature-based heuristic inputs

often fails to capture the intricate semantic nuances present in evolving political narratives (Mohamed et al., 2024).

Conversely, emerging methodologies leveraging pre-trained transformer architectures, such as BERT, have demonstrated an ability to surpass these limitations by capturing deep contextual representations and long-range dependencies within political discourse (Taha et al., 2022), (Sanaullah et al., 2022). These transformer-based models effectively navigate the complexities of irony and misinformation propagation, often outperforming traditional supervised learners that rely heavily on static feature extraction (Das et al., 2021; Shah et al., 2024). Furthermore, ensemble strategies—such as Stacking and Hard Voting—have been shown to provide the most consistent performance improvements by aggregating the strengths of multiple classifiers, effectively bridging the gap between classical feature-based models and complex transformative architectures (Başarslan & Bal, 2025).

Long Short-Term Memory models have demonstrated strong performance in detecting misinformation by effectively capturing sequential and contextual relationships within political text data (Sovia et al., 2025; Tahat et al., 2024). Deep learning architectures have demonstrated superior capability in analyzing sentiment, syntax, and semantic dependencies, successfully identifying subtle disinformation patterns that traditional algorithms frequently failed to detect (Liu et al., 2025; Zhou et al., 2025). These results confirm the effectiveness of recurrent neural networks in natural language processing applications. Furthermore, empirical results suggest that these ensemble frameworks facilitate a significant reduction in false positives, ensuring that legitimate political discourse is not erroneously flagged as disinformation (Aslam et al., 2025).

Despite the theoretical advantages of these multi-model architectures, some studies suggest that increasing complexity via ensemble methodologies does not always yield superior results compared to highly optimized individual transformer models (Parimanoharan & Nawarathna, 2025). This is particularly evident in scenarios where a single, pre-trained model already achieves near-optimal performance, potentially rendering the aggregation of weaker individual classifiers redundant (Zhang & Shafiq, 2024). Moreover, integrating multimodal data—such as incorporating image, video, or user metadata—often proves more effective than increasing model depth in text-only architectures, as these cues provide essential context often missing from textual content alone (Anggrainingsih et al., 2024).

The comparative analysis further demonstrated that deep learning approaches outperform conventional machine learning models in handling complex political discourse, as political disinformation frequently involves emotionally manipulative narratives and nuanced linguistic structures (Sovia et al., 2025). Traditional models relying on frequency-based features may struggle to identify these hidden patterns, whereas deep learning architectures provide more comprehensive contextual understanding through multilayer neural processing. Consequently, advanced neural network approaches are increasingly important in misinformation detection research. The integration of contextual language models, such as IndoBERT, further refines these capabilities by combining robust semantic understanding with sequential modeling to capture subtle linguistic cues (Sovia et al., 2025).

The results of this study align with previous research emphasizing the importance of artificial intelligence in combating digital misinformation. Machine learning systems provide scalable and automated solutions for analyzing massive volumes of online political content, and the integration of deep learning and natural language processing enhances the ability of computational systems to detect evolving manipulation strategies. Social media platforms

may benefit significantly from implementing intelligent detection systems. Overall, the findings demonstrate the growing relevance of machine learning technologies in protecting democratic information ecosystems.

Linguistic and Sentiment Features in Political Disinformation Detection

Linguistic analysis revealed that political disinformation content often contains emotionally charged and sensational language (Staender et al., 2025). Misleading political posts frequently utilized fear-inducing expressions, exaggerated claims, and polarizing narratives, suggesting that emotional manipulation plays a significant role in increasing user engagement and content virality (Klinger et al., 2022; Sangiorgio et al., 2024). Machine learning models incorporating sentiment analysis demonstrated improved detection accuracy compared to purely lexical approaches, highlighting the importance of emotional and semantic features in political misinformation analysis (Bose & Roy, 2023). Specifically, integrating emotion scores alongside traditional linguistic metrics significantly enhances a model's ability to distinguish between verified news and intentionally deceptive narratives (Farhoudinia et al., 2024).

Sentiment analysis techniques enabled machine learning systems to identify emotional polarity within political communication. Disinformation posts commonly exhibited extreme positive or negative sentiments intended to provoke strong audience reactions, with the presence of emotionally manipulative content strongly associated with higher misinformation probability scores (Vieira & Figueira, 2025). Consequently, sentiment-aware models successfully differentiated between objective political reporting and emotionally exaggerated narratives, demonstrating the effectiveness of sentiment analysis in strengthening automated disinformation detection systems. Beyond lexical analysis, detecting negative sentiment and specific emotional triggers serves as a critical indicator for identifying organized misinformation campaigns designed to sow discord or amplify fear within digital spheres (Camacho et al., 2024).

The study also identified the importance of contextual semantics in detecting deceptive political narratives. Disinformation content frequently utilized misleading contextual framing while maintaining partially factual statements, which complicates classification processes for traditional keyword-based systems. Deep learning models incorporating contextual embeddings demonstrated superior capability in recognizing semantic inconsistencies, and thus the integration of contextual analysis significantly improved classification performance. Beyond these semantic nuances, recent evidence indicates that emotional sequences—characterized by rhythmic shifts between high-intensity negative cues like anger or fear and neutral passages—serve as robust markers for identifying manipulative intent (Vieira & Figueira, 2025).

Hashtag usage patterns and repetitive linguistic structures also contributed to disinformation detection accuracy. Coordinated misinformation campaigns often employed recurring slogans, hashtags, and symbolic language to amplify political narratives, allowing machine learning systems to effectively identify these repetitive communication patterns across large datasets. Behavioral consistency among coordinated disinformation actors further enhanced model prediction capabilities, indicating the importance of combining linguistic and behavioral analysis approaches. This multi-faceted methodology mirrors frameworks that combine speech act theory with artificial intelligence to isolate communicative intent, effectively mapping illocutionary force alongside sentiment to distinguish genuine discourse from propaganda (Cherep et al., 2025).

The analysis revealed that sarcasm and irony remain challenging elements for automated disinformation detection systems. Political communication on social media frequently involves humorous or sarcastic expressions that alter literal meaning, causing traditional machine learning models to struggle with correctly interpreting these linguistic complexities. Although deep learning approaches showed improved performance, they still encountered classification difficulties in highly ambiguous cases, indicating that further advancements in contextual language understanding are required. To mitigate these classification errors, recent methodologies have begun incorporating manual semantic annotation to ensure that sarcastic labels accurately reflect intended meaning rather than superficial sentiment.

Multilingual political communication also introduced additional complexity into the detection process. Social media ecosystems consist of diverse linguistic communities with varying communication styles and cultural contexts, which means that machine learning systems trained primarily on English datasets may experience reduced performance in multilingual environments. Cross-linguistic disinformation detection therefore requires broader dataset diversity and multilingual model development to address these challenges, which is essential for improving global misinformation detection systems. Furthermore, transitioning from isolated linguistic analysis toward examining campaign-level goals and narrative frames through large language models offers a promising avenue for scaling detection efforts (Burghardt et al., 2024).

The findings demonstrate that linguistic and sentiment-based features are highly valuable for political disinformation detection. Emotional analysis, contextual semantics, and behavioral language patterns significantly enhance classification accuracy, as machine learning systems capable of deeper semantic understanding outperform simpler lexical approaches. The integration of advanced natural language processing techniques is therefore critical for future misinformation detection frameworks, supporting the continued development of intelligent and context-aware artificial intelligence systems. However, the persistent difficulty in discerning figurative language like irony necessitates the integration of multi-modal data and symbolic reasoning to refine interpretative depth (Bojić et al., 2025), (Banou et al., 2025), addressing the subjective nature of political discourse while minimizing the cognitive biases inherent in dataset labeling.

Moreover, addressing these complexities requires a transition toward hybrid architectures that leverage knowledge graphs to map the conceptual relationships between concepts and emotions, thereby mitigating the limitations posed by simple negation or idiosyncratic idiomatic expressions (Islam et al., 2024). Future research should also prioritize cross-lingual interpretability, specifically investigating whether transformer architectures focus on consistent features such as hyperpartisan tone or clickbait phrasing across disparate linguistic environments (Gouliev et al., 2025). In addition, investigating how psychological frameworks—such as the Elaboration Likelihood Model—can be operationalized within these deep learning structures may provide a more robust mechanism for detecting how emotional cues influence persuasion in health and political misinformation (Sikosana et al., 2025).

Furthermore, leveraging multi-modal integration—which synthesizes textual features with visual elements and publisher metadata—has demonstrated significant improvements in characterizing the credibility of social media content (Raza et al., 2025). Empirical investigations confirm that combining textual semantics with visual data consistently yields higher classification metrics than unimodal approaches, particularly within high-velocity information environments like Twitter or Weibo (Mohawesh et al., 2025). Beyond visual and

textual integration, researchers are increasingly advocating for the inclusion of user metadata and propagation graph topologies to capture the social dynamics of dissemination (Bhuiyan et al., 2025), (Singh et al., 2022).

Despite these performance gains, the inherent "black-box" nature of deep learning architectures continues to obscure the reasoning behind specific classifications, complicating efforts to audit model transparency (Liu et al., 2023). To mitigate these interpretability challenges, the adoption of model-agnostic tools like Captum provides a pathway for researchers to visualize feature importance and validate whether classifiers rely on valid indicators rather than dataset-specific artifacts (Laurer et al., 2023).

Social Media Ecosystem Dynamics and Disinformation Spread

Social media ecosystems play a pivotal role in the dissemination and amplification of political disinformation. Platform recommendation algorithms prioritize content that generates high user engagement and interaction, often elevating sensational political narratives due to their emotional and controversial nature. Such algorithmic amplification contributes to the rapid spread of political disinformation, underscoring the significant influence of platform architecture on digital political communication. Furthermore, because adversaries frequently exploit these algorithmic biases by migrating narratives across distinct platforms—such as shifting from mainstream social networks to encrypted messaging apps—detection strategies must evolve to incorporate multi-platform tracking that correlates user interactions and domain registrations (Cotroneo et al., 2025).

The study found that political disinformation spreads more rapidly than verified political information in many online environments (Surjatmodjo et al., 2024). Users are more likely to share emotionally provocative content regardless of factual accuracy, and viral misinformation campaigns often exploit trending political issues and social tensions to maximize audience attention (Surjatmodjo et al., 2024). Coordinated bot networks further accelerate the dissemination of false narratives across multiple platforms, creating substantial challenges for democratic information integrity (Mishra et al., 2025). To address these systemic vulnerabilities, researchers are increasingly integrating network traffic monitoring and behavioral analysis to identify the distinct signatures of automated accounts (Mishra et al., 2025).

Echo chambers and ideological polarization also contribute to disinformation proliferation within social media ecosystems. Users frequently interact with like-minded communities that reinforce existing political beliefs, while exposure to homogeneous information environments reduces critical evaluation of misleading narratives. Political disinformation campaigns exploit these polarized environments to strengthen ideological divisions among users, thereby increasing the vulnerability of digital communities to manipulation. Consequently, this susceptibility to fragmentation necessitates the implementation of cross-platform detection frameworks capable of neutralizing the strategic exploitation of cognitive biases and algorithmic feedback loops (Surjatmodjo et al., 2024; Nabila et al., 2025).

Bot accounts and automated social media activities are major contributors to disinformation amplification. Coordinated bot networks artificially increase engagement metrics such as likes, shares, and comments, creating the illusion of widespread public support for misleading political narratives. Machine learning systems incorporating behavioral analysis have successfully identified suspicious activity associated with automated

disinformation campaigns, highlighting behavioral analytics as an important component of misinformation detection systems. Moreover, identifying these artificial patterns is essential for disrupting the orchestration of computational propaganda, where tools like deepfakes and automated bots erode public trust and polarize political discourse (Olanipekun, 2025).

The research also highlighted the influence of visual and multimedia content in political disinformation campaigns. Manipulated images, edited videos, and misleading memes are increasingly used to influence public opinion, often spreading more effectively than textual misinformation because visual content generates stronger emotional responses. Traditional text-based detection systems may struggle to identify deceptive multimedia information, suggesting that future detection frameworks should integrate multimodal artificial intelligence techniques. Furthermore, the democratization of generative AI has exacerbated this threat by enabling the creation of hyper-realistic synthetic media that bypasses conventional heuristic-based filters (Minici et al., 2024).

Platform moderation policies significantly affect the effectiveness of disinformation mitigation efforts. Social media companies have introduced various automated moderation systems to identify harmful political content; however, inconsistent policy implementation and limited transparency remain ongoing concerns. Excessive moderation may also raise ethical issues related to freedom of expression and political censorship, which means that balancing democratic values and misinformation prevention requires careful regulatory considerations. Moving forward, establishing robust multi-stakeholder governance frameworks is essential to harmonize technical detection capabilities with the preservation of democratic discourse (Abokhodair et al., 2024; Gomez et al., 2024).

The findings emphasize the need for collaborative approaches involving technology companies, governments, researchers, and civil society organizations. Combating political disinformation cannot rely solely on technological solutions, as public digital literacy and critical thinking skills are equally important for reducing misinformation vulnerability. Policymakers should encourage transparent algorithmic governance and responsible platform management practices. Ultimately, strengthening digital democracy requires both technological innovation and societal awareness, alongside proactive strategies that account for the rapid proliferation of synthetic media, necessitating advanced oversight and real-time fact-checking mechanisms to mitigate the rising threat of AI-powered influence operations (Cavalcante, 2025).

Moreover, future research must prioritize the development of explainable artificial intelligence models to ensure that automated detection processes remain transparent and accountable to the public (Gedik, 2025; Romanishyn et al., 2025). Addressing the rapid evolution of generative AI necessitates a shift toward adaptive, cross-platform defensive strategies capable of identifying synthetic media at scale (Shoaib et al., 2023). Such initiatives must further emphasize the creation of standardized benchmarks and shared datasets, as the current absence of consistent evaluation protocols significantly hinders comparative assessments across diverse research studies (Park & Nan, 2025). Furthermore, achieving global cooperation on regulatory standards is imperative to prevent the fragmentation of defensive efforts amidst escalating geopolitical tensions (Shoaib et al., 2023).

In this context, the development of multimodal detection techniques, such as semantic-based deepfake identification that scrutinizes emotional inconsistencies, will serve as a vital defensive pillar against the rising sophistication of generative models (Ferrara, 2024). Additionally, integrating blockchain technology into media provenance processes could provide a robust verification layer, ensuring that authentic content remains distinguishable

from AI-synthesized iterations (Gambín et al., 2024). Furthermore, adopting adversarial machine learning techniques is essential to fortify these defensive models against sophisticated manipulation attempts, ensuring they remain resilient as attackers evolve their deceptive methodologies (Deng et al., 2025).

Integrating proactive deception defense mechanisms, such as real-time anomaly detection and natural language processing, further strengthens these systems by disrupting malicious actors before campaigns achieve virality (Schmitt & Fléchais, 2024). Finally, embedding explainable AI frameworks into these defensive architectures ensures that security analysts can effectively interpret decision-making processes, thereby increasing trust in the reliability of automated threat assessments (Agarwal, 2025). Concurrent with these technical advancements, organizations must prioritize data privacy by implementing techniques such as differential privacy and federated learning to secure the datasets used for training and inference (Familoni, 2024).

Implications for Artificial Intelligence and Democratic Governance

The increasing adoption of artificial intelligence in misinformation detection has significant implications for democratic governance. Machine learning technologies provide scalable mechanisms for identifying harmful political content in real time, thereby supporting electoral integrity by reducing the influence of manipulated political narratives. Automated detection systems also enhance the efficiency of content moderation processes across digital platforms (Kasimova & Kozlovskaya, 2025). As a result, artificial intelligence is becoming an essential component of democratic information security. However, these algorithmic interventions must be carefully balanced against risks of unintended bias, as language models often inherit the societal prejudices of their training data, which can inadvertently marginalize specific political viewpoints (Islam et al., 2024).

The study demonstrates that deep learning approaches can substantially improve political disinformation detection accuracy (Raza et al., 2024). Advanced neural network architectures are capable of understanding complex semantic and contextual relationships within political discourse, enabling more reliable identification of deceptive narratives and coordinated misinformation campaigns. Governments and technology companies may therefore benefit from investing in artificial intelligence research for democratic protection, reinforcing the strategic importance of AI-driven cybersecurity solutions. Nevertheless, research indicates that the long-term effectiveness of these models depends on mitigating data drift and curating high-quality, transparent datasets to ensure consistent performance across shifting political narratives (Raza et al., 2024).

Despite these advantages, ethical concerns remain regarding the use of artificial intelligence in political communication monitoring. Automated moderation systems may unintentionally suppress legitimate political opinions or minority perspectives, while algorithmic bias can also emerge when training datasets contain imbalanced or culturally biased information (Islam et al., 2024),(Kasimova & Kozlovskaya, 2025). Transparent and accountable artificial intelligence governance is therefore necessary to ensure fairness and democratic legitimacy, keeping ethical considerations central in the design of misinformation detection systems. Beyond technical implementation, addressing the threat of political disinformation requires a multidisciplinary framework that balances computational speed and scalability with robust ethical oversight to protect the integrity of democratic discourse (Kasimova & Kozlovskaya, 2025).

Privacy issues also represent important challenges in AI-based disinformation monitoring. Machine learning systems frequently require access to large volumes of user-generated data for training and analysis purposes, which may create concerns regarding surveillance and digital privacy rights. Policymakers should establish clear regulatory frameworks governing data usage and algorithmic accountability, protecting democratic values by balancing security objectives with individual civil liberties. Moreover, the potential for algorithmic surveillance necessitates the implementation of privacy-preserving techniques, such as federated learning, to analyze public discourse without compromising the anonymity of individual citizens (Adewusi et al., 2024).

Consequently, fostering collaborative governance models between nation-states, international organizations, and multinational corporations is essential to address these multifaceted socio-ethical challenges (Rakowski & Kowaliková, 2024). Establishing these cross-sector partnerships facilitates the creation of standardized audit protocols, which are vital for identifying and rectifying inherent biases that may emerge throughout an algorithm's lifecycle (Aderibigbe et al., 2023). Moreover, the development of national and international data ethics frameworks provides a foundational guide for communicators to navigate the complexities of AI-driven content assessment while upholding principles of transparency and accountability. These frameworks often draw upon established international principles to encourage institutional accountability and the implementation of proactive media regulation (Jandhyala & Prasad, 2022).

The integration of explainable artificial intelligence techniques may improve public trust in automated detection systems. Explainable AI enables users and moderators to understand how machine learning systems generate classification decisions, reducing concerns regarding algorithmic manipulation and political bias. Explainable models also support accountability and facilitate ethical oversight mechanisms; therefore, future research should prioritize interpretability alongside classification performance. Furthermore, establishing robust industry standards and transparency requirements is essential to prevent the opaque application of these tools, ensuring that algorithmic logic remains subject to external scrutiny (Liang & Chou, 2025; Pierson et al., 2023).

Educational initiatives and public digital literacy programs are equally important for combating political disinformation, as technological solutions alone cannot fully eliminate misinformation risks within digital societies. Citizens must develop critical thinking skills to evaluate political information sources effectively, reducing susceptibility to manipulative political narratives and emotional misinformation campaigns. Strengthening democratic resilience, therefore, requires both technological and educational interventions. Moreover, the formation of dedicated ethics committees within organizations can provide the necessary oversight to evaluate the broader implications of these models, ensuring that deployment remains aligned with human rights and democratic values (Osasona et al., 2024).

The findings demonstrate that machine learning technologies hold substantial potential for enhancing democratic information security. Artificial intelligence systems can significantly improve the detection of political disinformation within complex social media ecosystems (Shubham et al., 2026). However, technological innovation must be accompanied by ethical governance, transparency, and public education. Collaborative efforts among governments, researchers, and technology companies are essential for building trustworthy digital communication environments, as the future of democratic governance increasingly depends on the responsible integration of artificial intelligence technologies.

Moving forward, research must prioritize the transition from reactive detection models toward proactive media authentication and provenance tools that verify content integrity at its point of origin (Trivedi et al., 2021). Furthermore, future investigations should explore the integration of multimodal features, such as image-based analysis, to address the increasing sophistication of visual misinformation campaigns (Singh & Bharati, 2023). Moreover, researchers must rigorously evaluate how diverse user demographics interpret and respond to these automated verdicts, as recurring judgment errors risk alienating the public and undermining institutional credibility. Ultimately, cultivating long-term societal resilience requires shifting toward hybrid models that integrate computational detection with psychological attunement to help users navigate algorithmically driven environments (Eskiadi & Monastiridis, 2025; Caled & Silva, 2021).

CONCLUSION

This study concludes that machine learning technologies play a significant role in detecting political disinformation within contemporary social media ecosystems. The findings demonstrate that deep learning approaches, especially Long Short-Term Memory and transformer-based architectures, provide superior performance in recognizing contextual, semantic, and emotional patterns embedded in misleading political narratives. The research also highlights that social media algorithms, bot networks, echo chambers, and emotionally manipulative content significantly accelerate the spread of political misinformation and increase democratic vulnerability. Furthermore, integrating sentiment analysis, behavioral analytics, and contextual language processing substantially improves disinformation detection accuracy and strengthens automated moderation systems. Despite these technological advancements, ethical concerns related to algorithmic bias, transparency, freedom of expression, and digital privacy remain important challenges that require responsible governance and regulatory oversight. Therefore, combating political disinformation requires collaborative efforts among governments, researchers, technology companies, and civil society organizations to ensure the development of transparent, accountable, and human-centered artificial intelligence systems that support democratic integrity and trustworthy digital communication environments.

LIMITATIONS AND FUTURE RESEARCH RECOMMENDATIONS

This study has several limitations that should be acknowledged. First, the research primarily focused on textual political disinformation and did not comprehensively analyze multimodal misinformation involving images, videos, deepfakes, and visual propaganda content. Second, the study relied heavily on English-language and selected social media datasets, which may limit the generalizability of the findings across multilingual and cross-cultural political communication environments. Third, the qualitative content analysis approach provided rich interpretative insights but may contain subjective interpretation biases despite the application of triangulation methods. In addition, the rapidly evolving nature of artificial intelligence-generated disinformation creates continuous challenges for maintaining model relevance and detection consistency within dynamic digital ecosystems.

Future research should explore multimodal machine learning frameworks that integrate textual, visual, behavioral, and network-based data analysis to improve the detection of increasingly sophisticated political disinformation campaigns. Researchers are also encouraged to investigate cross-platform and multilingual misinformation detection models

capable of analyzing coordinated information operations across diverse digital communication environments. Furthermore, future studies should prioritize explainable artificial intelligence approaches to improve transparency, accountability, and public trust in automated moderation systems. Additional investigations into the ethical implications of AI-driven political content moderation, including issues related to censorship, algorithmic fairness, and privacy protection, are equally important. Finally, integrating psychological, sociological, and political communication perspectives into machine learning research may contribute to more comprehensive strategies for strengthening democratic resilience against digital disinformation threats.

REFERENCES

- Abokhodair, N., Skop, Y., Rüller, S., Aal, K., & Elmimouni, H. (2024). Opaque algorithms, transparent biases: Automated content moderation during the Sheikh Jarrah Crisis. *First Monday*. <https://doi.org/10.5210/fm.v29i4.13620>
- Aderibigbe, A. O., Ohenhen, P. E., Nwaobia, N. K., Gidiagba, J. O., & Ani, E. C. (2023). Artificial Intelligence in Developing Countries: Bridging The Gap Between Potential and Implementation. *Computer Science & IT Research Journal*, 4(3), 185–199. <https://doi.org/10.51594/csitrj.v4i3.629>
- Adewusi, A. O., Okoli, U. I., Olorunsogo, T., Adaga, E. M., Daraojimba, D. O., & Chimezie, O. (2024). Artificial intelligence in cybersecurity: Protecting national infrastructure: A USA review [Review of *Artificial intelligence in cybersecurity: Protecting national infrastructure: A USA review*]. *World Journal of Advanced Research and Reviews*, 21(1), 2263–2275. GSC Online Press. <https://doi.org/10.30574/wjarr.2024.21.1.0313>
- Agarwal, G. (2025). Explainable AI (XAI) for Cyber Defense: Enhancing Transparency and Trust in AI-Driven Security Solutions. *International Journal of Advanced Research in Science Communication and Technology*, 132–138. <https://doi.org/10.48175/ijarsct-23624>
- Akhtar, M. M. (2026). Unmasking the Hidden Threats of Social Media: Detecting Misinformation, Bots, and Malicious Campaigns [Australian Defence Force Academy]. In *UNSWWorks (University of New South Wales, Sydney, Australia)*. <https://doi.org/10.26190/unswworks/32005>
- Al-Fehani, M., & Al-Kuwari, S. (2024). Recent Advances in Digital Image and Video Forensics, Anti-forensics and Counter Anti-forensics. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2402.02089>
- Anggrainingsih, R., Hassan, G. M., & Datta, A. (2024). Transformer-based models for combating rumours on microblogging platforms: a review [Review of *Transformer-based models for combating rumours on microblogging platforms: a review*]. *Artificial Intelligence Review*, 57(8). Springer Science+Business Media. <https://doi.org/10.1007/s10462-024-10837-9>
- Aslam, Z., Missen, M. M. S., Ghaffar, A. A., Mehmood, A., Villar, M. G., Alvarado, E. S., & Ashraf, I. (2025). Advancing fake news combating using machine learning: a hybrid model approach. *Knowledge and Information Systems*. <https://doi.org/10.1007/s10115-025-02588-y>

- Bach, P., Schmitt, C., & McGregor, S. C. (2025). Let me be perfectly unclear: strategic ambiguity in political communication. *Communication Theory*, 35(2), 96–106. <https://doi.org/10.1093/ct/qtaf001>
- Banou, Z., Elfilali, S., Benlahmar, E. H., Alaoui, F. Z., Eljiani, L., & Sakhi, H. (2025). A systematic review of figurative language detection: Methods, challenges, and multilingual perspectives. *Natural Language Processing Journal*, 13, 100192–100192. <https://doi.org/10.1016/j.nlp.2025.100192>
- Başarslan, M. S., & Bal, F. (2025). The effect of text representation and model selection on classification performance: A comprehensive comparison of TF-IDF, Bow and Transformer-based methods on the Covid19-FNIR dataset. *Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 14(4), 1447–1461. <https://doi.org/10.28948/ngumuh.1694988>
- Bhuiyan, M., Sultana, F., & Rahman, A. M. (2025). Fake News Classifier: Advancements In Natural Language Processing For Automated Fact-Checking. *Strategic Data Management and Innovation*, 2(1). <https://doi.org/10.71292/sdmi.v2i01.20>
- Bojić, L., Zagovora, O., Zelenkauskaitė, A., Vuković, V., Čabarkapa, M., Jerković, S. V., & Jovančević, A. (2025). Evaluating Large Language Models Against Human Annotators in Latent Content Analysis: Sentiment, Political Leaning, Emotional Intensity, and Sarcasm. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2501.02532>
- Bose, S., & Roy, R. (2023). *Enhancing Fake News Detection with Sentiment Analysis Using Machine Learning*. 1–5. <https://doi.org/10.1109/iementech60402.2023.10423496>
- Bosley, M. (2024). Three Papers in the Applied Use of Machine Learning and Artificial Intelligence Models for the Analysis of Political Text Data. *Deep Blue (University of Michigan)*. <https://doi.org/10.7302/25077>
- Bulat, B., & Hilbert, M. (2024). A Dynamical Systems Approach to Bots and Online Political Communication. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2405.18652>
- Burghardt, K., Chen, K., & Lerman, K. (2024). Large Language Models Reveal Information Operation Goals, Tactics, and Narrative Frames. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2405.03688>
- Caled, D., & Silva, M. J. (2021). Digital media and misinformation: An outlook on multidisciplinary strategies against manipulation. *Journal of Computational Social Science*, 5(1), 123–159. <https://doi.org/10.1007/s42001-021-00118-8>
- Camacho, D., Gómez-Romero, J., & Jung, J. J. (2024). Special issue on infodemics. *Journal of Ambient Intelligence and Humanized Computing*, 15(3), 1975–1980. <https://doi.org/10.1007/s12652-024-04784-0>
- Cavalcante, P. P. (2025). Expert Foresight on AI-Driven Disinformation. *Politikon IAPSS Journal of Political Science*, 26(2), 89–91. <https://doi.org/10.22151/politikon.61.con8>
- Cherep, O., Калюжна, Ю., & Markova, S. (2025). Model of Communicative Impact in Disinformation Messages Based on Speech Act Theory and Artificial Intelligence Tools. *Baltic Journal of Economic Studies*, 11(4), 410–415. <https://doi.org/10.30525/2256-0742/2025-11-4-410-415>
- Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrocioni, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9). <https://doi.org/10.1073/pnas.2023301118>

- Cotroneo, D., Natella, R., & Orbinato, V. (2025). Elevating Cyber Threat Intelligence against Disinformation Campaigns with LLM-based Concept Extraction and the FakeCTI Dataset. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2505.03345>
- Das, S. D., Basak, A., & Dutta, S. (2021). A heuristic-driven uncertainty based ensemble framework for fake news detection in tweets and news articles. *Neurocomputing*, 491, 607–620. <https://doi.org/10.1016/j.neucom.2021.12.037>
- Deng, J., Lin, C., Zhao, Z., Liu, S., Peng, Z., Wang, Q., & Shen, C. (2025). A Survey of Defenses Against AI-Generated Visual Media: Detection, Disruption, and Authentication. *ACM Computing Surveys*, 58(5), 1–35. <https://doi.org/10.1145/3770916>
- Díaz, F., Cerna, N., Liza, R., & Motta, B. (2026). Artificial Intelligence for Detecting Electoral Disinformation on Social Media: Models, Datasets, and Evaluation. *Information*, 17(3), 292–292. <https://doi.org/10.3390/info17030292>
- Eskiadi, I. G., & Monastiridis, P. (2025). Combating Disinformation with Neurocognitive AI: A Hybrid Model for Crisis Communication. *Brain Broad Research in Artificial Intelligence and Neuroscience*, 16(3), 199–199. <https://doi.org/10.70594/brain/16.3/14>
- Familoni, B. T. (2024). Cybersecurity Challenges in The Age of AI: Theoretical Approaches and Practical Solutions. *Computer Science & IT Research Journal*, 5(3), 703–724. <https://doi.org/10.51594/csitrj.v5i3.930>
- Farhoudinia, B., Öztürkcan, S., & Kasap, N. (2024). Emotions unveiled: detecting COVID-19 fake news on social media. *Humanities and Social Sciences Communications*, 11(1). <https://doi.org/10.1057/s41599-024-03083-5>
- Ferrara, E. (2023). Social bot detection in the age of ChatGPT: Challenges and opportunities. *First Monday*. <https://doi.org/10.5210/fm.v28i6.13185>
- Ferrara, E. (2024). Charting the Landscape of Nefarious Uses of Generative Artificial Intelligence for Online Election Interference. <https://doi.org/10.2139/ssrn.4883403>
- Gambín, Á. F., Yazidi, A., Vasilakos, A. V., Haugerud, H., & Djenouri, Y. (2024). Deepfakes: current and future trends. *Artificial Intelligence Review*, 57(3). <https://doi.org/10.1007/s10462-023-10679-x>
- Gedik, C. A. (2025). 2024 ABD BAŞKANLIK SEÇİMLERİNDE YAPAY ZEKÂ ODAKLI İÇERİKLER, AKILLI TEKNOLOJİLER VE DEZENFORMASYONUN ROLÜ. *DergiPark* (Istanbul University). <https://dergipark.org.tr/tr/pub/upa/issue/90497/1643814>
- Gombár, M., & Boban, M. (2025). Research on the Impact of Algorithmic Echo Chambers on Perceptions and Attitudes of Social Network Users in a Digital Society. 1026–1033. <https://doi.org/10.1109/mipro65660.2025.11131918>
- Gomez, J. F., Machado, C. V., Paes, L. M., & Calmon, F. P. (2024). Algorithmic Arbitrariness in Content Moderation. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2402.16979>
- Gouliev, Z., Waters, J. K., & Wang, C. (2025). PolyTruth: Multilingual Disinformation Detection using Transformer-Based Language Models. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2509.10737>
- Haupt, M. R., Chiu, M., Chang, J., Li, Z., Cuomo, R., & Mackey, T. K. (2023). Detecting nuance in conspiracy discourse: Advancing methods in infodemiology and communication science with machine learning and qualitative content coding. *PLoS ONE*, 18(12). <https://doi.org/10.1371/journal.pone.0295414>

- Ibrahim, N. T., & Attia, N. A. H. M. (2025). The impact of disinformation generated by AI on democracy case studies: the US presidential elections in 2016 & 2024. *Review of Economics and Political Science*, 1–16. <https://doi.org/10.1108/reps-12-2024-0104>
- Islam, M. S., Kabir, M. N., Ghani, N. A., Zamli, K. Z., Zulkifli, N. S. A., Rahman, M. M., & Moni, M. A. (2024). “Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach” [Review of “Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach”]. *Artificial Intelligence Review*, 57(3). Springer Science+Business Media. <https://doi.org/10.1007/s10462-023-10651-9>
- Islam, Md. R., Haseeb, M., Batool, H., Ahtasham, N., & Muhammad, Z. (2024). AI Threats to Politics, Elections, and Democracy: A Blockchain-Based Deepfake Authenticity Verification Framework. *Blockchains*, 2(4), 458–481. <https://doi.org/10.3390/blockchains2040020>
- Jandhyala, P. R., & Prasad, R. S. (2022). Special Edition on Emerging Technologies and Changing Dynamics of Information (ETCDI). *The International Review of Information Ethics*, 32(1). <https://doi.org/10.29173/irrie493>
- Jareh, A. (2025). An In-depth Assessment of The Acceptability And Effectiveness of AI tools in Identifying Misinformation in Media: Literature Review. *SAGE Open*, 15(4). <https://doi.org/10.1177/21582440251381272>
- Juniyanto, H., & Wisnaeni, F. (2026). Democratic Regression and Electoral Integrity in the Era of Digital Disinformation. *International Journal of Social Science and Human Research*, 9(1). <https://doi.org/10.47191/ijsshr/v9-i1-08>
- Kahale, V. V., & Jambhulkar, V. K. (2026). Weaponization of Social Media Resulting in Unrest in the Society. *Journal of Advances and Scholarly Researches in Allied Education*, 23(1), 655–670. <https://doi.org/10.29070/5zgzep16>
- Kasimova, A., & Kozlovskaya, E. (2025). Using Artificial Intelligence to Detect and Mitigate Fake News. *Al-Noor Journal for Digital Media Studies.*, 2(1), 39–47. <https://doi.org/10.69513/jnfdms.v2.i1.a4>
- Klinger, U., Koç-Michalska, K., & Rußmann, U. (2022). Are Campaigns Getting Uglier, and Who Is to Blame? Negativity, Dramatization and Populism on Facebook in the 2014 and 2019 EP Election Campaigns. *Political Communication*, 40(3), 263–282. <https://doi.org/10.1080/10584609.2022.2133198>
- Lakzaei, B., Chehrehgani, M. H., & Bagheri, A. (2024). Disinformation detection using graph neural networks: a survey. *Artificial Intelligence Review*, 57(3). <https://doi.org/10.1007/s10462-024-10702-9>
- Laurer, M., Atteveldt, W. van, Casas, A., & Welbers, K. (2023). Less Annotating, More Classifying: Addressing the Data Scarcity Issue of Supervised Machine Learning with Deep Transfer Learning and BERT-NLI. *Political Analysis*, 32(1), 84–100. <https://doi.org/10.1017/pan.2023.20>
- Liang, Y., & Chou, T. (2025). Developmental pathways for academic knowledge about public health and social media misinformation: a systematic review through main path analysis [Review of *Developmental pathways for academic knowledge about public health and social media misinformation: a systematic review through main path analysis*]. *Current Psychology*. Springer Science+Business Media. <https://doi.org/10.1007/s12144-025-08298-6>
- Liu, Y., Shen, X., Zhang, Y., Wang, Z., Tian, Y., Dai, J., & Cao, Y. (2024). A Systematic Review of Machine Learning Approaches for Detecting Deceptive Activities on Social

- Media: Methods, Challenges, and Biases [Review of *A Systematic Review of Machine Learning Approaches for Detecting Deceptive Activities on Social Media: Methods, Challenges, and Biases*]. *arXiv* (Cornell University). Cornell University. <https://doi.org/10.48550/arxiv.2410.20293>
- Liu, Y., Shen, X., Zhang, Y., Wang, Z., Tian, Y., Dai, J., & Cao, Y. (2025). A systematic review of machine learning approaches for detecting deceptive activities on social media: methods, challenges, and biases [Review of *A systematic review of machine learning approaches for detecting deceptive activities on social media: methods, challenges, and biases*]. *International Journal of Data Science and Analytics*. Springer International Publishing. <https://doi.org/10.1007/s41060-025-00850-8>
- Liu, Z., Zhang, T., Yang, K., Thompson, P., Yu, Z., & Ananiadou, S. (2023). *Emotion Detection for Misinformation: A Review* [Review of *Emotion Detection for Misinformation: A Review*]. <https://doi.org/10.2139/ssrn.4625569>
- Mahmood, O., & Akar, F. (2024). Using and Comparison of Artificial Intelligence Techniques to Detect Misinformation and Disinformation on Twitter. *The European Journal of Research and Development*, 4(2), 254–264. <https://doi.org/10.56038/ejrnd.v4i2.467>
- Minici, M., Luceri, L., Cinus, F., & Ferrara, E. (2024a). *Uncovering Coordinated Cross-Platform Information Operations Threatening the Integrity of the 2024 U.S. Presidential Election Online Discussion*. <https://doi.org/10.2139/ssrn.4965049>
- Minici, M., Luceri, L., Cinus, F., & Ferrara, E. (2024b). *Uncovering Coordinated Cross-Platform Information Operations Threatening the Integrity of the 2024 U.S. Presidential Election Online Discussion*. *arXiv* (Cornell University). <https://doi.org/10.48550/arxiv.2409.15402>
- Mishra, H., Singh, P., Kurien, M., & Quraishi, S. J. (2025). *Detecting Bots in Misinformation Campaigns: Strategies and Societal Implications*. 917–921. <https://doi.org/10.1109/icpct64145.2025.10940500>
- Mohamed, E. A., Ismail, W. N., Ibrahim, O. A. S., & Younis, E. M. G. (2024). A two-stage framework for Arabic social media text misinformation detection combining data augmentation and AraBERT. *Social Network Analysis and Mining*, 14(1). <https://doi.org/10.1007/s13278-024-01201-4>
- Mohawesh, R., Obaidat, I., Alqarni, A. A., Aljubailan, A. A., Al-Shannaq, M. A., Salameh, H. B., Al-Yousef, A., Saifan, A. A., Alkhushayni, S., & Maqsood, S. (2025). Truth be told: a multimodal ensemble approach for enhanced fake news detection in textual and visual media. *Journal Of Big Data*, 12(1). <https://doi.org/10.1186/s40537-025-01182-x>
- Montoro-Montarroso, A., Cantón-Correa, J., Rosso, P., Chulvi, B., Panizo-Lledot, Á., Huertas-Tato, J., Figueras, B. C., Rementería, M. J., & Gómez-Romero, J. (2023). Fighting disinformation with artificial intelligence: fundamentals, advances and challenges. *El Profesional de La Informacion*. <https://doi.org/10.3145/epi.2023.may.22>
- Mudavadi, K. C., & Madrid-Morales, D. (2024). *Countering Political Disinformation* (pp. 450–459). <https://doi.org/10.4324/9781003334774-52>
- Muñoz, P., Díez, F., & Bellogín, A. (2024). Modeling disinformation networks on Twitter: structure, behavior, and impact. *Applied Network Science*, 9(1). <https://doi.org/10.1007/s41109-024-00610-w>
- Nabila, M. H., Shaibu, R. A., Afeti, G. E., & Adza, E. A. (2025). Disinformation as a driver of political polarization: A strategic framework for rebuilding civic trust in the U.S. *World Journal of Advanced Research and Reviews*, 27(1), 916–925. <https://doi.org/10.30574/wjarr.2025.27.1.2564>

- Ng, L. H. X., Robertson, D. C., & Carley, K. M. (2024). Cyborgs for strategic communication on social media. *Big Data & Society*, 11(1). <https://doi.org/10.1177/20539517241231275>
- Olanipekun, S. O. (2025). Computational propaganda and misinformation: AI technologies as tools of media manipulation. *World Journal of Advanced Research and Reviews*, 25(1), 911–923. <https://doi.org/10.30574/wjarr.2025.25.1.0131>
- Osasona, F., Amoo, O. O., Atadoga, A., Abrahams, T. O., Farayola, O. A., & Ayinla, B. S. (2024). Reviewing The Ethical Implications of AI in Decision Making Processes. *International Journal of Management & Entrepreneurship Research*, 6(2), 322–335. <https://doi.org/10.51594/ijmer.v6i2.773>
- Padalko, H., Chomko, V., & Chumachenko, D. (2024). A novel approach to fake news classification using LSTM-based deep learning models. *Frontiers in Big Data*, 6. <https://doi.org/10.3389/fdata.2023.1320800>
- Padalko, H., Chomko, V., Yakovlev, S., & Chumachenko, D. (2023). Ensemble machine learning approaches for fake news classification. *Radioelectronic and Computer Systems*, 4, 5–19. <https://doi.org/10.32620/reks.2023.4.01>
- Parimanoharan, S., & Nawarathna, R. D. (2025). Assessing Classical Machine Learning and Transformer-based Approaches for Detecting AI-Generated Research Text. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2509.20375>
- Park, S., & Nan, X. (2025). Generative AI and misinformation: a scoping review of the role of generative AI in the generation, detection, mitigation, and impact of misinformation [Review of *Generative AI and misinformation: a scoping review of the role of generative AI in the generation, detection, mitigation, and impact of misinformation*]. *AI & Society*. Springer Nature. <https://doi.org/10.1007/s00146-025-02620-3>
- Pierson, J., Kerr, A., Robinson, S. C., Fanni, R., Steinkogler, V. E., Milan, S., & Zampedri, G. (2023). Governing artificial intelligence in the media and communications sector. *Internet Policy Review*, 12(1). <https://doi.org/10.14763/2023.1.1683>
- Rakowski, R., & Kowaliková, P. (2024). The political and social contradictions of the human and online environment in the context of artificial intelligence applications. *Humanities and Social Sciences Communications*, 11(1). <https://doi.org/10.1057/s41599-024-02725-y>
- Raza, S., Khan, T., Paulen-Patterson, D., Chatrath, V., Rahman, M., & Bamgbose, O. (2024). FakeWatch: A Framework for Detecting Fake News to Ensure Credible Elections. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2403.09858>
- Raza, S., Paulen-Patterson, D., & Chen, D. (2025). Fake news detection: comparative evaluation of BERT-like models and large language models with generative AI-annotated data. *Knowledge and Information Systems*. <https://doi.org/10.1007/s10115-024-02321-1>
- Rita, P., António, N., & Afonso, A. P. (2023). Social media discourse and voting decisions influence: sentiment analysis in tweets during an electoral period. *Social Network Analysis and Mining*, 13(1). <https://doi.org/10.1007/s13278-023-01048-1>
- Rodilosso, E. (2024). Filter Bubbles and the Unfeeling: How AI for Social Media Can Foster Extremism and Polarization. *Philosophy & Technology*, 37(2). <https://doi.org/10.1007/s13347-024-00758-4>
- Romanchuk, R., & Vysotska, V. (2024). Automatic Content Analysis Systems: Detecting Disinformation in Social Networks. *Qeios*. <https://doi.org/10.32388/tyb6ze>

- Romanishyn, A., Malytska, O., & Goncharuk, V. A. (2025). AI-driven disinformation: policy recommendations for democratic resilience. *Frontiers in Artificial Intelligence*, 8, 1569115–1569115. <https://doi.org/10.3389/frai.2025.1569115>
- Rosadi, I. T., Pravitasari, A. A., & Andriyana, Y. (2024). Data tweet clustering using bidirectional gated recurrent unit and k-prototype for the Indonesian political year. *International Journal of Data and Network Science*, 8(2), 907–920. <https://doi.org/10.5267/j.ijdns.2023.12.015>
- Saeed, M. H., Ali, S., Paudel, P., Blackburn, J., & Stringhini, G. (2024). Unraveling the Web of Disinformation: Exploring the Larger Context of State-Sponsored Influence Campaigns on Twitter. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2407.18098>
- Saini, M. (2025). Fake News Detection and Fact-Checking System. *International Journal for Research in Applied Science and Engineering Technology*, 13(4), 6003–6006. <https://doi.org/10.22214/ijraset.2025.69692>
- Sanaullah, A., Das, A., Das, A., Kabir, M. A., & Shu, K. (2022). Applications of machine learning for COVID-19 misinformation: a systematic review [Review of *Applications of machine learning for COVID-19 misinformation: a systematic review*]. *Social Network Analysis and Mining*, 12(1). Springer Science+Business Media. <https://doi.org/10.1007/s13278-022-00921-9>
- Sangiorgio, E., Marco, N. D., Etta, G., Cinelli, M., Cerqueti, R., & Quattrociocchi, W. (2024). Evaluating the effect of viral news on social media engagement. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2407.13549>
- Schmitt, M., & Fléchais, I. (2024). Digital deception: generative artificial intelligence in social engineering and phishing. *Artificial Intelligence Review*, 57(12). <https://doi.org/10.1007/s10462-024-10973-2>
- Shah, A. V., Shah, H., Bafna, V. H., Khandor, C., & Nair, S. (2024). VERITAS-NLI: Validation and Extraction of Reliable Information Through Automated Scraping and Natural Language Inference. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2410.09455>
- Shoaib, M. R., Wang, Z., Ahvanooey, M. T., & Zhao, J. (2023). Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models. <https://doi.org/10.1109/icca59364.2023.10401723>
- Shubham, Irfan, A., Singh, R., Joshi, C., Gehlot, A., & Mahala, R. (2026). Artificial Intelligence in Mitigating Political Disinformation on Social Media: A Comprehensive Review. *International Journal for Global Academic & Scientific Research*, 4(4). <https://doi.org/10.55938/ijgasr.v4i4.241>
- Sikosana, M., Maudsley-Barton, S., & Ajao, O. (2025). Linguistic patterns in pandemic-related content: a comparative analysis of COVID-19, Constraint, and Monkeypox datasets. *Frontiers in Artificial Intelligence*, 8. <https://doi.org/10.3389/frai.2025.1627522>
- Singh, P., Srivastava, R., Rana, K. P. S., & Kumar, V. (2022). SEMI-FND: Stacked ensemble based multimodal inferencing framework for faster fake news detection. *Expert Systems with Applications*, 215, 119302–119302. <https://doi.org/10.1016/j.eswa.2022.119302>
- Singh, R. K., & Bharati, R. (2023). Tackling the Fake News Epidemic using Machine Learning Algorithm. *International Journal for Research in Applied Science and Engineering Technology*, 11(4), 1857–1861. <https://doi.org/10.22214/ijraset.2023.50344>

- Sovia, R., Valkyrie, D. A., Zain, R. H., & Firdaus, F. (2025). Language Processing for Detecting Fake News on Twitter Using a Long Short-Term Memory Architecture. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 9(4), 935–943. <https://doi.org/10.29207/resti.v9i4.6570>
- Staender, A., Humprecht, E., Esser, F., Morosoli, S., & Aelst, P. V. (2025). Is Sensationalist Disinformation More Effective? Three Facilitating Factors at the National, Individual, and Situational Level. In *Routledge eBooks* (pp. 54–74). Informa. <https://doi.org/10.4324/9781032647807-4>
- Surjatmodjo, D., Unde, A. A., Cangara, H., & Sonni, A. F. (2024). Information Pandemic: A Critical Review of Disinformation Spread on Social Media and Its Implications for State Resilience. *Social Sciences*, 13(8), 418–418. <https://doi.org/10.3390/socsci13080418>
- Taha, M., Zayed, H. H., Azer, M., & Gadallah, M. (2022). Automated COVID-19 misinformation checking system using encoder representation with deep learning models. *IAES International Journal of Artificial Intelligence*, 12(1), 488–488. <https://doi.org/10.11591/ijai.v12.i1.pp488-495>
- Tahat, D., Alfaisal, R., Tahat, K., & Salloum, S. A. (2024). *Navigating the Newscape: Long Short-Term Memory Models for Misinformation Detection*. 112–114. <https://doi.org/10.1109/sds64317.2024.10883921>
- Trivedi, A., Suhm, A., Mahankal, P., Mukuntharaj, S., Parab, M. D., Mohan, M., Berger, M., Sethumadhavan, A., Jaiman, A., & Dodhia, R. (2021). Defending Democracy: Using Deep Learning to Identify and Prevent Misinformation. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2106.02607>
- Vieira, R., & Figueira, Á. (2025). Emotional Sequencing as a Marker of Manipulation in Social Media Disinformation. *Future Internet*, 17(12), 546–546. <https://doi.org/10.3390/fi17120546>
- Vysotska, V., & Nazarkevych, M. (2025). Development of an information technology for detecting the sources and networks of disinformation dissemination in cyberspace based on machine learning methods. *Eastern-European Journal of Enterprise Technologies*, 4, 35–51. <https://doi.org/10.15587/1729-4061.2025.335501>
- Wandwi, G., & Ngaiza, R. (2026). Harnessing Deep Learning to Detect Irony and Sarcasm in News Headlines for Combating Misinformation in Digital Media. *Open Journal of Applied Sciences*, 16(1), 16–31. <https://doi.org/10.4236/ojapps.2026.161003>
- Zhang, H., & Shafiq, M. O. (2024). Survey of transformers and towards ensemble learning using transformers for natural language processing. *Journal Of Big Data*, 11(1). <https://doi.org/10.1186/s40537-023-00842-0>
- Zhou, M., Lü, L., Hu, C., Li, Y., Wang, W., & Fang, J. (2025). *Hierarchical Sentiment-Syntax-Semantic Enhancement with Adaptive External Knowledge for Fake News Detection*. 1330–1336. <https://doi.org/10.1109/icflic68075.2025.11325043>