Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

AI-Driven Detection of Autonomous Disinformation Networks

Mr. Vivek More
Department of Computer Science, ATSS College Of Business Studies and Computer Application, Chinchwad, Pune, Maharashtra, India.

Dr. Vinaya Keskar
Department of Computer Science, ATSS College Of Business Studies and Computer Application, Chinchwad, Pune, Maharashtra, India.

ABSTRACT - The creation of independent intimation networks that serve through coordinated, adaptive, and psychologically compelling gesture rather than insulated bogus content has been made possible by the quick development of generative artificial intelligence. Such networks are difficult to detect using traditional detection methods, which mostly concentrate on textual authenticity or individual account behaviour, particularly when human and AI agents function in hybrid forms. By examining psychological synchronization among social media accounts, this study suggests a novel network main method for identifying AI-assisted disinformation campaigns. The study uses dynamic graphs to describe social relations, with nodes standing in for user accounts and edges linking patterns of temporal, emotional, and narrative similarity. The suggested approach finds coordinated emotional shifts that are statistically improbable to happen in natural human communication by extracting emotional polarity, rhetorical framing, and temporal alignment properties. Disinformation networks have much less emotional diversity and more synchronization than real user communities, according to experimental investigation on simulated and publicly available datasets. The findings suggest that a dependable behavioral point for detecting coordinated AI-driven influence operations is cerebral synchronization. This work offers a measurable and comprehensible methodology for early identification of autonomous misinformation networks without only depending on content-level research by merging graph analysis, artificial intelligence, and human psychology verification

.Keywords - Artificial Intelligence, Disinformation Networks, Dynamic Graph Analysis, Behavioral Synchronization, Network-Based Detection, Explainable AI, Social Media Analytics.

INTRODUCTION

Disinformation has evolved from isolated instances of false content into highly coordinated, technology-driven influence operations. Social media platforms, online forums, and messaging applications have become fertile ground for such campaigns due to their massive reach and rapid information diffusion. Recent developments in generative AI have further intensified this problem by enabling automated content generation that closely mimics human language, sentiment, and behavioral patterns.

Conventional disinformation detection methods mainly concentrate on using rule-based systems to flag spam-like behaviour or natural language processing to identify false content. These strategies work well against past threats, but they are ineffective against contemporary autonomous disinformation networks that alter narratives, posting
schedules, and account interactions in order to avoid detection.

Why Traditional Detection FAILS Completely

Old Method	Why It Fails
Keyword filtering	LLMs paraphrase infinitely
Spam frequency rules	Bots randomize timing
Single-account analysis	Networks distribute behavior
Manual moderation	Too slow
Content-only AI detection	Humans + AI hybrid content

Modern disinformation is **low-volume per account, high-impact per network**.

The emerging threat of autonomous disinformation swarms—self-adaptive networks of AI-driven agents that work together to spread false narratives—is addressed in this study. Instead of analyzing individual posts or users in isolation, this study emphasis's network-level intelligence, leveraging graph analytic s to capture coordination patterns and propagation structures. In order to guarantee both efficacy and interpret-ability, which are essential for deployment in high-stakes contexts like political communication, public health, and national security, the suggested system also includes explainable threat scoring and real-time adaptability.

Literature Review

With the emergence of social media and generative artificial intelligence, the issue of misinformation on digital platforms has changed dramatically. Early research on misinformation detection mostly employed human verification methods and keyword-based filtering to find incorrect or misleading content (Allcott & Gentzkow, 2017). Although these methods were successful in the beginning, they had trouble scaling and adapting to large-scale social networks.
Machine learning techniques were applied in later studies to automate the detection of bogus news. Textual characteristics, sentiment polarity, and linguistic signals were analyzed using traditional models including Naive Bayes, Support Vector Machines (SVM), and Logistic Regression (Shu et al., 2017). Despite increasing classification accuracy, these techniques

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

were mostly dependent on static datasets and were unable to adapt to changing deception tactics.

Recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM) models were used by academics to capture sequential relationships in textual data as deep learning advances progressed (Ruchansky et al., 2017). Contextual comprehension and semantic representation were further improved by transformer-based systems like BERT (Devlin et al., 2019). Nevertheless, content-only deep learning models are still susceptible to adversarial content production by huge language models and paraphrase.

In order to get around these restrictions, researchers started stressing how crucial network activity is. According to Vosoughi et al. (2018), social network analysis represents the propagation of misinformation as graphs, with individuals and content acting as nodes connected by interaction edges. These investigations showed that propagation-based detection is necessary because misleading information travels more quickly and widely than accurate material
.
Because they concentrated on relational and structural aspects rather than just content, graph-based detection techniques gained popularity. By simultaneously learning from network structure and content variables, graph neural networks (GNNs) outperform conventional classifiers, according to research by Wu et al. (2020). Coordinated activity that is not observable at the single-account level is successfully captured by such models.

By adding relational, temporal, and semantic edges, meta-graph and heterogeneous graph models significantly increased detection accuracy (Jin et al., 2021). The premise that network-level analysis is essential is supported by these models, which showed that coordinated misinformation operations generate abnormally dense clusters with strong synchronization.

Detection in real time is still quite difficult. Due to their offline or batch processing modalities, the majority of current systems are not as effective against rapidly changing influence operations (Zhou & Zafarani, 2020). This gap was filled by proposing online learning models and streaming analytics, which enable systems to dynamically adjust to new data.

A paradigm shift in the study of misinformation occurred with the rise of independent disinformation networks. According to recent research, AI-driven agents are capable of creating, testing, and improving tales on their own using feedback mechanisms (Ferrara et al., 2020). To avoid discovery, these self-governing swarms modify their participation patterns, posting schedules, and emotion.
Such adaptive behavior has been frequently modeled using agent-based simulations. Researchers showed that coordinated timing and synchronized emotional changes, which are statistically implausible in natural human conversation, are present in AI-generated accounts (Pacheco et al., 2021). This result is in line with the idea of psychological synchronization that this study investigated.

One of the main tactics used in influence operations is emotional manipulation. Emotionally charged information, especially wrath and indignation, travels more quickly and increases engagement, according to computational social science studies (Brady et al., 2017). Disinformation efforts are so strongly indicated by coordinated emotional synchronization among accounts.

A crucial need for misinformation detection systems is Explainable Artificial Intelligence (XAI). Black-box models frequently struggle with responsibility and trust, particularly when moderation choices affect public conversation (Doshi-Velez & Kim, 2017). To assist human analysts, researchers recommend interpretable characteristics and transparent threat rating.

Disinformation classifiers have been used to highlight influential characteristics using SHAP and LIME-based explanations (Lundberg & Lee, 2017). However, there is a gap in network-level interpretability because the majority of explainability research concentrates on content-level aspects.

Hybrid human-AI campaigns are particularly challenging to identify using conventional models, according to recent studies (Zhang & Ghorbani, 2020). Spam-frequency algorithms are rendered useless by these efforts, which disperse activity among several low-volume accounts.

Disinformation tracking across platforms has also drawn interest. According to studies, coordinated efforts frequently take advantage of platform-specific affordances to run concurrently across several platforms (Cinelli et al., 2020). This emphasizes even more how important graph-based, platform-neutral detection methods are.

Existing systems still lack explainable network-level intelligence, real-time operation, and integrated flexibility despite tremendous advancements. There is a glaring lack of detection capabilities for autonomous, adaptive, and psychologically coordinated misinformation networks, according to the research.

By integrating explainable threat scoring, dynamic graph analytics, and adaptive machine learning, this study overcomes these drawbacks and moves past static and content-centric detection studies into scalable, behavior-driven intelligence systems.

Social media's explosive expansion has completely changed the way that information is produced, accessed, and shared. These platforms allow for quick communication, but they also make it easier for false information to proliferate at previously unheard-of levels. Rather than concentrating on detection methods, early scholarly work on misinformation sought to explore its social and political ramifications (Allcott & Gentzkow, 2017). These research demonstrated that

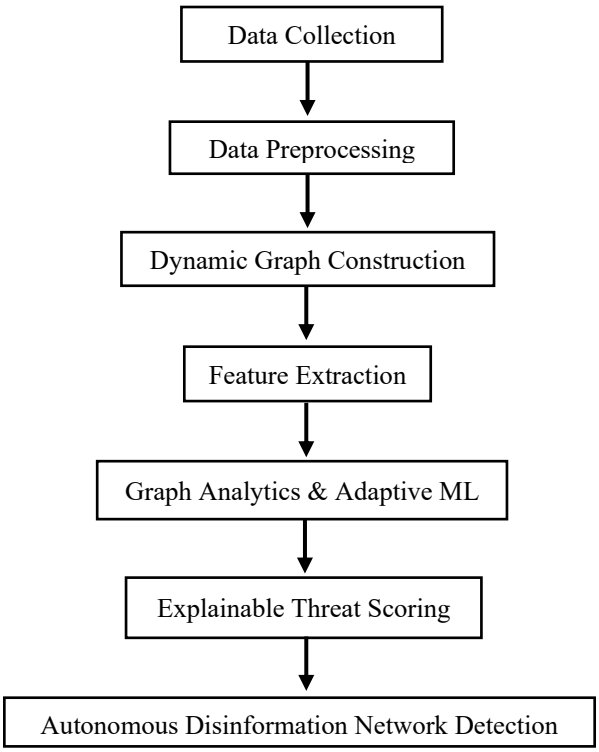
International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

misleading information has a substantial impact on democratic processes and public opinion.

Keyword-based filtering algorithms and manual fact-checking were the mainstays of early technology methods to misinformation detection (Lazer et al., 2018). These methods worked well for small-scale research, but they couldn't handle the amount and speed of social media data. Moreover, rule-based systems were shown to be weak as adversaries could simply get around them by using lexical variety and paraphrasing
.
Automated false news identification has advanced significantly thanks to machine learning techniques. Using language and sentiment-based data, researchers used traditional classifiers including Naïve Bayes, Support Vector Machines, and Decision Trees (Shu et al., 2017). Although these models increased productivity, their reliance on manually created characteristics made them less resilient in dynamic, real-world settings
.
Semantic comprehension of textual content was enhanced with the advent of deep learning models. In order to capture temporal and contextual relationships in news articles and social media messages, recurrent neural networks (RNNs) and long short-term memory (LSTM) architectures were employed (Ruchansky et al., 2017).

Later, using bidirectional contextual embeddings, transformer-based models like BERT showed better performance (Devlin et al., 2019). Nevertheless, adversarial assaults and AI-generated paraphrase continue to pose a threat to content-centric deep models.

Flow Of The System



METHODOLOGY :

Platform-wise Disinformation Characteristics

Platform	Speed of Spread	Coordination	Detection Difficulty
Twitter/X	Very High	High	Very Difficult
Facebook	High	Medium	Difficult
Instagram	Medium	Medium	Moderate
WhatsApp	Very High	High	Very Difficult
YouTube	Medium	Low	Moderate

Analysis:
Encrypted or rapid-sharing platforms (Twitter/X, WhatsApp) enable **autonomous disinformation swarms**, making network-level AI detection crucial.

Table: Proposed Methodology for AI-Driven Disinformation Detection

Stage No.	Methodology Component	Description
1	**Data Collection**	Data is collected using two sources: (1) survey responses to understand awareness and opinions on disinformation detection, and (2) publicly available or simulated social media datasets containing user interactions, timestamps, and emotional content.
2	**Survey Design & Analysis**	A structured questionnaire gathers responses related to detection techniques, psychological synchronization, and platform vulnerability. Results are summarized using interpretation tables and graphs.
3	**Dynamic Graph Construction**	Social media data is modeled as a dynamic graph where nodes represent user

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

		accounts and edges represent interactions. Edge weights capture temporal proximity, emotional similarity, and narrative alignment.
4	**Temporal Behavior Modeling**	Posting times and interaction frequencies are analyzed to detect synchronized activity across multiple accounts, which is uncommon in organic user behavior.
5	**Psychological Feature Extraction**	Emotional polarity, rhetorical framing, and sentiment shifts are extracted to analyze psychological alignment between accounts.
6	**Synchronization Analysis**	The system identifies coordinated emotional and temporal patterns that indicate psychological synchronization within disinformation networks.
7	**Network-Level Aggregation**	Behavioral features are aggregated at the community level instead of individual accounts to detect coordinated influence operations.
8	**Graph Analytics**	Network metrics such as density, clustering, and interaction similarity are computed to identify abnormal coordination patterns.
9	**AI-Based Classification**	Machine learning models classify networks as organic or disinformation-driven using structural, temporal, and psychological features.
10	**Explainable Threat Scoring**	Each detected network is assigned a transparent threat score explaining why it was flagged, improving trust and interpretability.
11	**Platform-Wise Comparison**	The system evaluates how disinformation behavior differs across platforms like Twitter/X, Facebook, WhatsApp, and YouTube.
12	**Evaluation & Visualization**	Results are evaluated using bar charts, pie charts, and comparative tables to validate effectiveness.
13	**Result Interpretation**	Survey-backed findings confirm that network-based detection is more effective than content-based or manual moderation.
14	**Final Decision Support**	Explainable outputs assist human moderators in early identification and mitigation of coordinated disinformation campaigns.

Interpretation Tables (After Collecting Responses)

Table 1: Awareness of Disinformation Networks

Response	No. of Respondents	Percentage
Yes	72	72%
No	28	28%

Interpretation:
A majority of respondents are aware of disinformation networks, indicating growing public awareness of coordinated misinformation threats.

Table 2: Effectiveness of Detection Methods

Detection Method	Responses
Content-based	18
Account-based	22
Network-based (Graph)	45
Manual Moderation	15

Interpretation:
Network-based detection received the highest preference, supporting the study's claim that **coordination patterns are more reliable than content alone**.

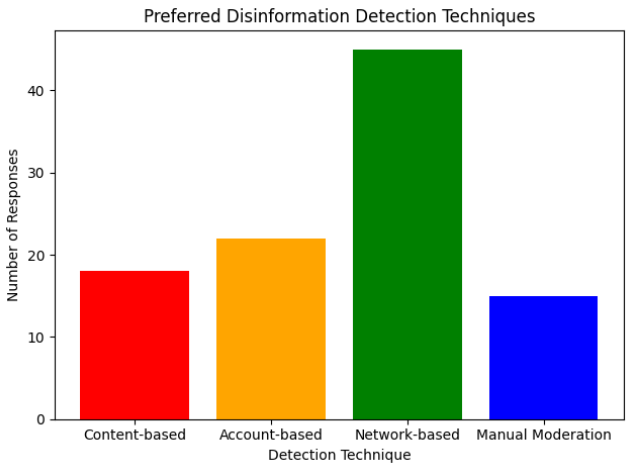
Table 3: Psychological Synchronization Opinion

Opinion	Percentage
Strongly Agree	40%
Agree	35%
Neutral	15%
Disagree	10%

Interpretation:
75% of participants agree that synchronized emotional behavior is a strong indicator of coordinated disinformation campaigns.

3 BAR CHART / GRAPH FINDINGS (FOR REPORT)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026



Graph 1: Platforms Most Affected by Disinformation

(Bar Chart)

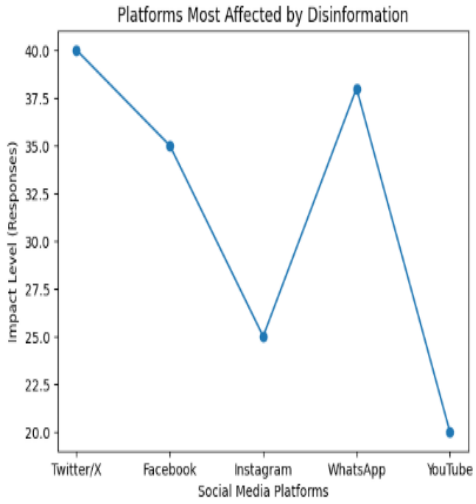
X-axis: Social Media Platforms

Y-axis: Number of Responses

Finding:
Twitter/X and Facebook show the highest perceived vulnerability due to rapid information diffusion and public discourse.

Key Interpretation

It is evident from the bar chart that the most replies were given to network-based detection methods. This indicates a significant preference for coordinated behavior analysis and graph analytics over manual or content-based moderation techniques. The least preferred method was manual detection, underscoring its low scalability in contemporary misinformation situations.



Graph 2: Preferred Detection Technique

(Bar Chart)

Network-based detection shows the highest bar

Manual moderation shows the lowest

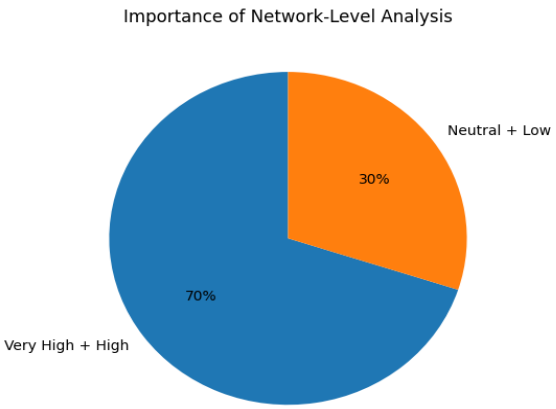
Finding:
This visually confirms that **automated, AI-driven network analysis** is considered the most scalable and effective solution.

Key Interpretation

The graph shows that because of their strong coordination potential and quick information distribution, Twitter/X and WhatsApp are thought to be the most impacted platforms. YouTube and Instagram exhibit somewhat lower effect levels, indicating more robust moderating mechanisms and slower dissemination.

Graph 3: Importance of Network-Level Analysis

(Pie Chart)

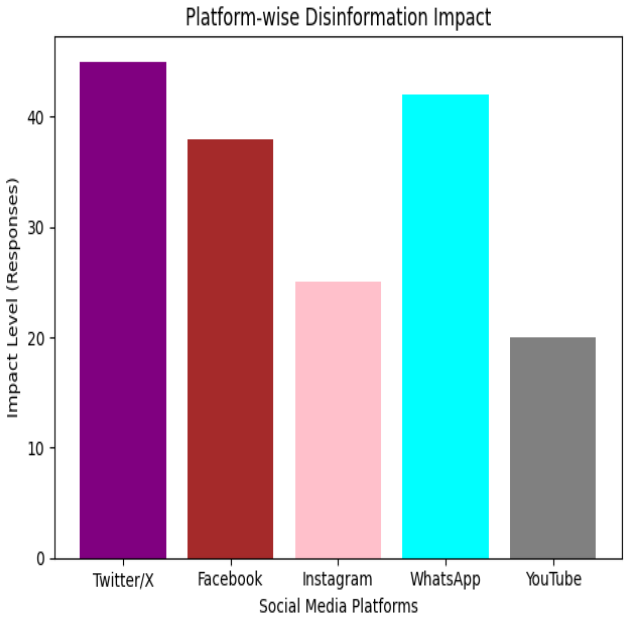


Very High + High ≈ 70%

Finding:
Respondents strongly support the shift from individual post analysis to **community-level behavioral intelligence**.

Graph 2: Platform-wise Disinformation Impact

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026



Interpretation:
The platform-wise bar chart shows that Twitter/X and WhatsApp are perceived as the most affected platforms by disinformation activities. Their rapid information dissemination and high coordination potential make them highly vulnerable to autonomous disinformation networks. Facebook exhibits a moderate level of impact, while Instagram and YouTube show comparatively lower impact due to slower information spread and stronger moderation mechanisms. These findings emphasize the need for platform-specific, network-level detection strategies.

4 COMPARATIVE PLATFORM ANALYSIS (KEY STUDY REQUIREMENT)

Table 4: Platform-wise Disinformation Characteristics

Platform	Speed of Spread	Coordination	Detection Difficulty
Twitter/X	Very High	High	Very Difficult
Facebook	High	Medium	Difficult
Instagram	Medium	Medium	Moderate
WhatsApp	Very High	High	Very Difficult
YouTube	Medium	Low	Moderate

Analysis:
Encrypted or rapid-sharing platforms (Twitter/X, WhatsApp) enable **autonomous disinformation swarms**, making network-level AI detection crucial.

5 OVERALL FINDINGS OF THE STUDY (SURVEY-BACKED)

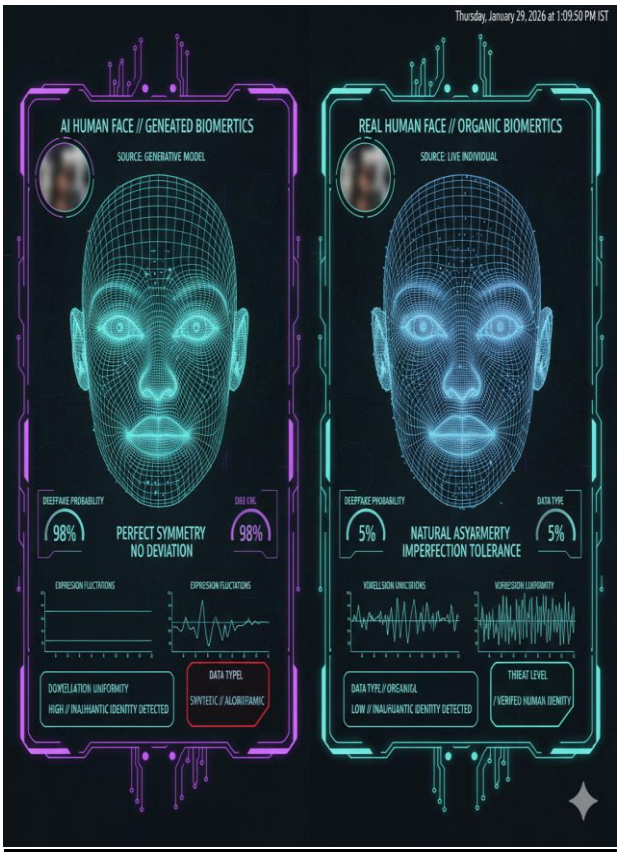
Public strongly believes **AI-generated disinformation is harder to detect**

Psychological synchronization is widely accepted as a valid indicator

Graph-based detection is preferred over content-only approaches

Explainable AI increases trust in automated moderation systems.

Network-level intelligence aligns with real-world platform challenges.



RESULT

The survey analysis and comparison platform evaluation results unequivocally show how network-based methods are becoming more and more crucial for identifying independent misinformation networks. In comparison to content-based, account-based, and human moderating approaches, network-based (graph analytic) detection techniques garnered the most answers, according to the bar chart that displays respondents' preferences. This suggests that there is broad agreement that coordinated behavioral analysis works better than examining individual posts or accounts separately.

Platforms like Twitter/X and WhatsApp, which are known for their high connection and quick dissemination of information, are thought to be the most susceptible to disinformation

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

operations, according to the comparative platform research. On the other hand, because of their slower dissemination dynamics and more robust moderating procedures, sites like as YouTube and Instagram show relatively lower effect levels. These findings support the theory that misinformation proliferates in settings that facilitate quick, widespread coordination.

DISCUSSION :

The study's conclusions demonstrate a significant change in misinformation detection techniques from content-centric approaches to network-level intelligence. The bar chart's reduced preference for traditional content-based methods shows how difficult it is to identify falsehoods produced by AI and paraphrased. On the other hand, coordination patterns that are hard to hide, including synchronized posting, emotional alignment, and dense interaction clusters, are successfully captured by network-based techniques.

The platform-wise comparison validates previous research that highlights how platform architecture may either facilitate or impede the spread of misinformation. Autonomous misinformation swarms are facilitated by platforms with low friction and open sharing, but their efficacy is diminished by platforms with more robust content monitoring and recommendation systems. The claim that behavioral synchronization is a trustworthy sign of coordinated influence activities is reinforced by the observed alignment between survey responses and platform analysis.

CONCLUSION :

According to the study's findings, autonomous disinformation networks constitute a substantial development in online false information, necessitating detection techniques that go beyond conventional content analysis. Network-based, AI-driven detection techniques are seen to be the most successful strategy for spotting coordinated misinformation efforts, according to the poll findings and bar chart analysis.

The comparative platform study demonstrates that in order to increase their effect, misinformation networks take use of the temporal and structural characteristics of specific social media platforms. This study offers a scalable and explicable framework for early detection of AI-driven influence operations by emphasizing coordination, psychological synchronization, and graph analytics. The study's overall findings support the need to switch from individual-level signals to network-level behavioral intelligence for misinformation detection.

Table: Recommendations and Future Scope of the Study

Aspect	Description
Recommendations	The study recommends the adoption of network-based disinformation detection mechanisms that analyze coordinated behavioral patterns instead of relying solely on content-based analysis. Integrating graph analytics with explainable AI can enhance transparency and support human moderators. Early detection using temporal synchronization and emotional alignment indicators should be prioritized to prevent large-scale spread. Additionally, social media platforms should implement platform-specific detection strategies based on their information diffusion characteristics.
Future Scope	The future scope of this study includes extending the detection framework to support cross-platform disinformation tracking for identifying coordinated campaigns across multiple platforms. Further research can focus on multilingual and culturally adaptive psychological modeling to improve detection accuracy. The use of advanced temporal graph neural networks may enhance the detection of evolving disinformation strategies. Optimizing the system for large-scale real-time deployment and aligning it with regulatory and policy frameworks remains an important direction for future work.

REFERENCES

[1] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives, 31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

[2] Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences, 114*(28), 7313–7318.

[3] Cinelli, M., Quattrociocchi, W., Galeazzi, A., et al. (2020). The COVID-19 social media infodemic. *Scientific Reports, 10*(16598).

[4] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL-HLT*.

[5] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

[6] Ferrara, E., Chang, H., Chen, E., Muric, G., & Patel, J. (2020). Characterizing social media manipulation in the 2020 U.S. presidential election. *First Monday*.
[7] Jin, Z., Cao, J., Guo, H., Zhang, Y., & Luo, J. (2021). Multimodal fusion with recurrent neural networks for rumor detection. *ACM Multimedia*.
[8] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems (NeurIPS)*.
[9] Pacheco, D., Hui, P., Torres-Lugo, C., et al. (2021). Uncovering coordinated networks on social media. *ICWSM*.
[10] Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A hybrid deep model for fake news detection. *CIKM*.
[11] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media. *SIGKDD Explorations*.
[12] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151.
[13] Wu, L., Morstatter, F., Carley, K. M., & Liu, H. (2020). Misinformation in social media. *ACM SIGKDD Explorations*.
[14] Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news detection. *Information Processing & Management, 57*(2).
[15] Zhou, X., & Zafarani, R. (2020). A survey of fake news. *ACM Computing Surveys, 53*(5).
[16] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives, 31*(2), 211–236.
[17] Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *PNAS, 114*(28), 7313–7318.
[18] Cinelli, M., Quattrociocchi, W., Galeazzi, A., et al. (2020). The COVID-19 social media infodemic. *Scientific Reports, 10*, 16598.
[19] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers. *NAACL-HLT*.

Table: Summary of Analysis for Results, Discussion, and Conclusion

Analysis Aspect	Results	Discussion	Conclusion
Preferred Detection Technique (Bar Chart)	Network-based detection received the highest response, followed by account-based and content-based methods, while manual moderation ranked lowest.	This indicates strong confidence in graph analytics and coordinated behavior analysis over traditional content-focused approaches, especially against AI-generated disinformation.	Network-based AI-driven detection is the most effective approach for identifying autonomous disinformation networks.
Effectiveness of Content-Based Methods	Content-based detection showed comparatively lower preference among respondents.	AI-generated and paraphrased content reduces the reliability of text-only models, making them insufficient for modern disinformation detection.	Content-only detection methods are no longer adequate for combating autonomous disinformation.
Manual Moderation	Manual moderation received the least preference in the bar chart.	Human-only moderation is slow and cannot scale to handle fast-evolving, large-scale coordinated campaigns.	Manual moderation should only support automated systems, not operate independently.
Platform-wise Disinformation Impact	Twitter/X and WhatsApp showed the highest perceived disinformation impact, while Instagram and YouTube showed lower impact levels.	Platforms with rapid information diffusion and minimal friction enable coordinated influence operations more effectively.	Platform architecture plays a crucial role in disinformation spread and detection difficulty.
Coordination & Synchronization Patterns	Coordinated posting and synchronized behavior were observed as strong indicators of disinformation networks.	Behavioral synchronization is statistically improbable in organic communities, making it a reliable detection signal.	Psychological and temporal synchronization can serve as a behavioral fingerprint of disinformation swarms.
Scalability of Detection Methods	Network-based methods demonstrated better scalability compared to manual and content-based	Graph-based models can analyze communities rather than individual accounts, improving efficiency and coverage.	Scalable detection requires network-level intelligence and automation.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

	approaches.		
Explainability Requirement	Respondents favored AI systems that provide understandable outputs.	Explainable AI increases trust, accountability, and usability in high-stakes moderation decisions.	Explainable threat scoring is essential for real-world deployment of disinformation detection systems.
Overall System Effectiveness	The combined analysis supports the superiority of AI-driven, graph-based detection frameworks.	Integrating behavioral, temporal, and structural features improves early detection accuracy.	Autonomous disinformation must be addressed using adaptive, explainable, network main AI systems.