International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 11 | Nov 2026 www.irjet.net p-ISSN: 2395-0072

AI-Driven Misinformation Analysis System with Propagation Visualization, Influence Analysis and Audience-Exposure Estimation

[1]Bhagya Lakshmi T, [2]Keerthan, [3]Nagaraj Shenoy S, [4]Madhan Gowda M B, [5]Mahesha

[12345]Department of Computer Science & Engineering, Sir M. Visvesvaraya Institute of Technology, Bengaluru, India

---

Abstract - *The rapid dissemination of information through social media has increased the need for systems that can assess potentially misleading content and examine its dissemination patterns. This paper presents an AI-driven misinformation analysis system integrating AI-assisted content analysis, evidence-based fact verification, optical character recognition (OCR), engagement analysis, simulated propagation-network visualization, and influence analysis. The system processes user-submitted text and text extracted from social-media screenshots, while allowing users to review extracted engagement metrics before further analysis. Its spread-analysis component distinguishes observed engagement changes from future audience-exposure estimation. For X, an experimental forecasting method combines the median of clipped historical daily log-growth rates with a growth rate estimated by an engagement-informed Ridge model. For Instagram and Facebook, provisional engagement-based estimates are generated using projected engagement and benchmark-based methods, but they require further empirical validation. The system represents potential dissemination relationships through a simulated graph and applies graph-based analysis to examine influential nodes. Further evaluation using larger time-stamped datasets is required to assess the reliability of these estimates across platforms and forecast horizons.*

Index Terms - Misinformation Detection, Artificial Intelligence, Natural Language Processing, AI-Assisted Content Analysis, Optical Character Recognition, Propagation Tracking, Influence Detection, Spread Prediction, Graph Analysis.

I. INTRODUCTION

Social media platforms have transformed how information is produced, shared, and consumed. Although these platforms support rapid communication and access to information, they also enable misleading claims to circulate before they can be adequately verified. The volume and speed of online content make credibility assessment challenging when relying on manual review alone. These challenges motivate integrated systems that combine content analysis, evidence verification, and examination of potential dissemination patterns. Such systems can help users assess suspicious content and explore how observed engagement changes and potential network relationships may relate to its dissemination.

Existing misinformation-detection methods commonly examine textual content and linguistic patterns to identify potentially misleading information. Machine-learning and natural language processing techniques have been widely applied to fake-news detection, while deep-learning models such as BERT can capture contextual relationships within news content [3]. However, content analysis alone provides limited information about how misleading content circulates through online networks, which sources contribute to its dissemination, and how its audience exposure may change. Analyses based only on text, visual content, or individual social-context signals may therefore provide an incomplete view of misinformation dissemination. These limitations motivate approaches that examine content characteristics alongside engagement patterns and potential propagation relationships.

To address these limitations, an integrated misinformation analysis system can combine content analysis, social-context information, and network properties. Multimodal approaches can incorporate textual and visual information to support the identification of potentially misleading content [4], [5]. Social-context features, such as user interactions, temporal activity, and source-related characteristics, can provide additional signals for analysis [1], [6]. Furthermore, network information can represent potential dissemination relationships as a graph, enabling the examination of propagation patterns and connections between nodes beyond content analysis [8], [9]. Combining these complementary perspectives can support a broader investigation of misinformation and its potential dissemination across social-media platforms.

The integration of content and network-based analysis offers several advantages. Textual content can be analyzed using machine learning and natural language processing techniques to generate classification and confidence-based results [2], [3]. In addition, Optical Character Recognition (OCR) enables text contained in images, screenshots, and similar visual content to be extracted and processed within the same analysis framework. Network graphs can represent potential dissemination relationships between nodes, supporting graph-based analysis for identifying influential sources. Furthermore, engagement patterns and

temporal observations can provide signals for exploratory estimation of future audience exposure [7]. Combining these capabilities can therefore support a deeper understanding of misinformation beyond content analysis alone.

Recent studies have addressed different aspects of misinformation analysis with promising results. Deep-learning approaches have improved contextual content analysis [3], while multimodal methods have demonstrated the value of combining textual and visual information [4], [5]. Social-context models have also shown that user activity and source-related information can provide useful signals for fake-news detection [1], [6]. Research on information cascades highlights the relevance of network structure to understanding how information may spread over time [7]. In addition, Bi-GCN and FANG demonstrate how graph-based representations can capture propagation structures and social-context information for misinformation analysis [8], [9]. These studies highlight the potential of combining complementary content, social, and network-level information.

The motivation for this work is to develop an **AI-Driven Misinformation Analysis System with Propagation Tracking, Influence Detection and Spread Prediction** that integrates these complementary capabilities within a unified framework. The specific objectives of this work are (i) to analyze textual and image-based content for misinformation using AI-assisted content analysis, natural language processing, and OCR, (ii) to generate potential propagation networks and analyze influential nodes using graph-based network analysis, and (iii) to examine observed engagement changes and estimate future audience exposure using the implemented engagement-based methods. The contributions of this paper are as follows:

- An integrated framework combining misinformation analysis, OCR-based text extraction, potential propagation-network visualization, influence analysis, and spread estimation within a single system.
- A content analysis module that applies AI-assisted content analysis and natural language processing to generate analysis results and confidence-based outputs.
- A graph-based mechanism for representing potential dissemination relationships and examining influential nodes within the generated network.
- An engagement-based spread-analysis module that examines observed engagement changes and provides exploratory estimates of future audience exposure where supported by the implemented methods.

II. LITERATURE SURVEY

The rapid growth of social media has made misinformation a significant challenge. Online content can be created and shared rapidly without proper verification, allowing misleading information to reach a large audience. Existing studies have explored misinformation detection using textual, visual, social, and network-based features. This section reviews major research approaches and their limitations in detecting and analyzing misinformation.

Early research on misinformation detection focused on news content and social-context information. Shu et al. [2] highlighted the importance of auxiliary social information for fake-news detection, while Ma et al. [1] examined temporal social-context information for early rumor detection. These studies demonstrate that combining textual content with changing social signals can support misinformation detection.

Deep-learning approaches have further improved misinformation detection by learning contextual representations from news content. Jwa et al. [3] presented exBAKE, a BERT-based model for fake-news detection that explores the relationship between news headlines and article bodies. However, such approaches primarily focus on textual information and do not incorporate visual and broader social-context cues.

Multimodal approaches combine textual and visual information to enhance misinformation detection. Wang et al.[4] proposed EANN, which extracts textual and visual features and learns event-invariant representations for fake-news detection on newly emerging events. Similarly, Khattar et al. [5] proposed MVAE, a bimodal variational autoencoder for learning multimodal representations from textual and visual features. These studies demonstrate the significance of combining multiple forms of content for the analysis of online misinformation.

Social-context information provides additional signals beyond the textual content of a news item. Shu et al. [2] discussed the importance of auxiliary social information, including user engagements, for fake-news detection. Ruchansky et al. [6] proposed CSI, a hybrid deep model that integrates article text, temporal user activity, and source-user information through its Capture, Score, and Integrate modules. These studies indicate that social-context features can complement content-based evidence in misinformation analysis.

The proliferation of misinformation on social networks has also made the prediction of information spread an important research direction. Li et al. [7] proposed DeepCas, an end-to-end deep-learning model that represents information cascades using graph structures and sampled propagation paths for cascade-size prediction. The study shows that cascade structure can

provide useful signals for predicting future growth, thereby complementing misinformation detection by addressing the potential extent of information dissemination.

Graph-based approaches focus on the structural properties of misinformation propagation. Bian et al. [8] introduced Bi-GCN, which models top-down and bottom-up structures during rumor propagation and dispersion. Nguyen et al. [9] proposed FANG, a graph-based framework that encodes social context for fake-news detection. These studies demonstrate how graph and network information can complement textual features and provide additional insight into the propagation of misinformation.

Despite these advances, existing approaches often focus on individual aspects of misinformation, such as content, social context, or propagation structure. This motivates an integrated framework combining content analysis, potential propagation-network visualization, influence analysis, and engagement-based spread estimation. Further evaluation is needed to assess the reliability of future audience-exposure estimates.

III. DESIGN AND METHODOLOGY OF PROPOSED WORK

The proposed system provides an integrated workflow for analyzing misinformation from textual content and image-based social-media screenshots. The system accepts either directly entered text or uploaded images as input. For image-based content, OCR is used to extract the embedded text, which is then passed through the same content-analysis and fact-verification process. The system further extracts relevant engagement information and allows the user to verify or correct the detected values. The finalized social information is then used to assess spreading behaviour and estimate future audience exposure where supported by the implemented method. The system also generates a potential propagation network for analyzing influence, community structure, and potential information-spread patterns.

A. Input Processing and Content Acquisition

The proposed system supports two primary input paths:

direct text input and image-based social-media content are the two supported input paths. When the user enters text directly, the submitted content is passed to the content-analysis stage. When an image or social-media screenshot is uploaded, Optical Character Recognition (OCR) is first used to extract the text present in the image. The extracted text is then passed to the same content-analysis and fact-verification workflow used for direct text input.

For uploaded social-media content, the system also extracts available engagement information, such as likes, comments, shares, reposts, bookmarks, and follower counts, depending on the platform and the information available. The detected values are presented to the user for verification. The user can review, edit, or correct the extracted engagement values before finalizing them. This verification step helps ensure that the social information used in subsequent analysis reflects the user's confirmed input.

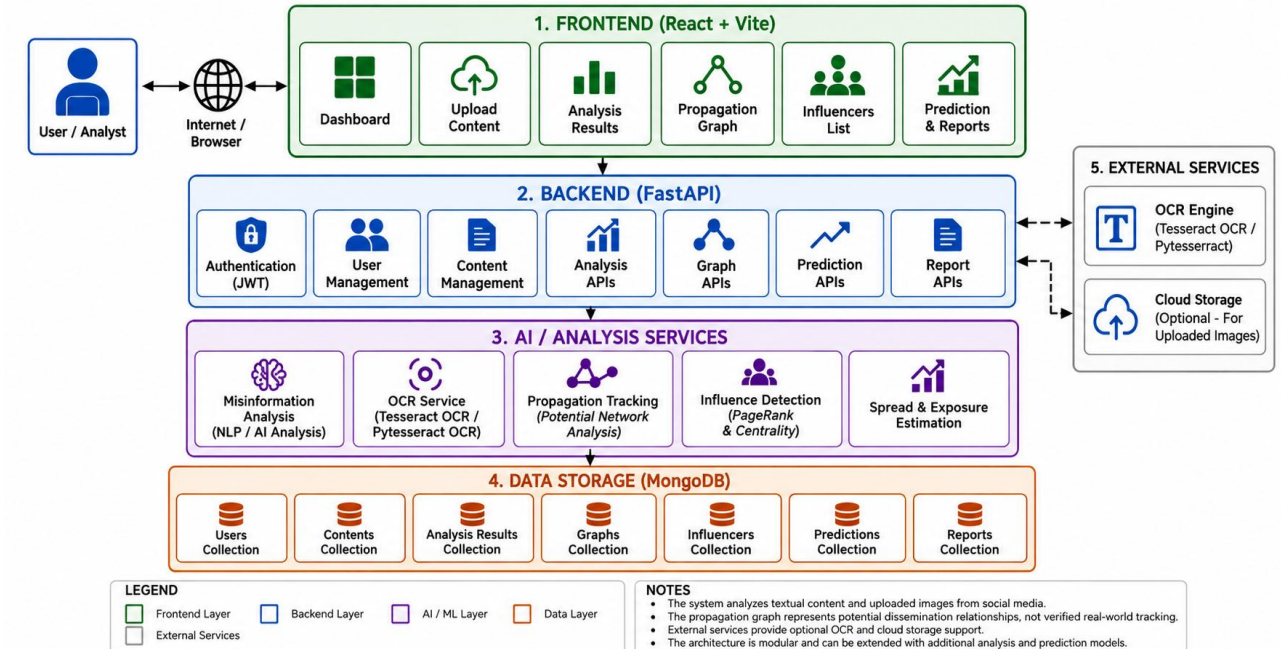


Fig. 1. Overall architecture of the proposed misinformation analysis framework.

B. Content, Engagement and Social Information Analysis

The system analyzes submitted content for misinformation using NLP-based and AI-based content analysis. For image or social-media screenshot inputs, OCR first extracts the embedded text, which is then passed through the same content-analysis and fact-verification workflow. The system also extracts available engagement and social

information, such as likes, comments, shares, reposts, bookmarks, and followers. These values are presented to the user for verification, allowing the user to edit or correct them before finalization. The verified engagement information is used to assess spreading behaviour and estimate future audience exposure where supported by the implemented method. Potential propagation networks are generated separately to examine dissemination patterns and influential nodes.

This user-verification step helps improve the reliability of the social information used in subsequent analysis.

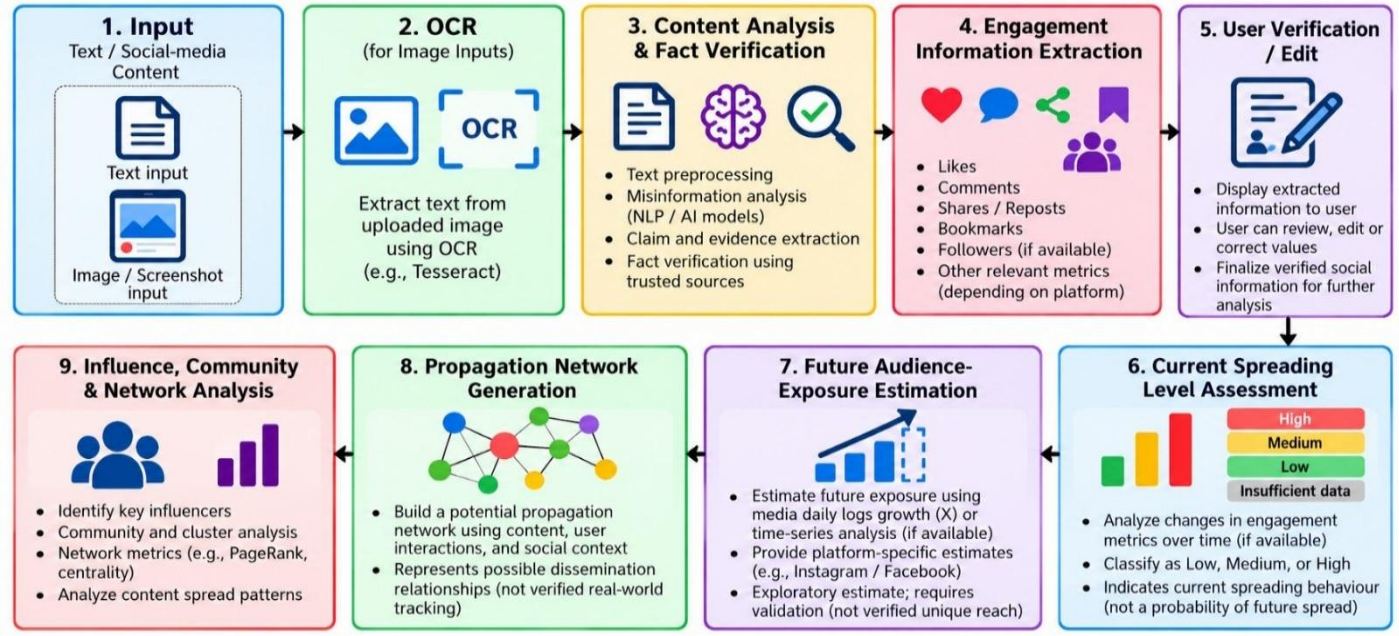


Fig. 2. Functional Workflow of the Proposed Misinformation Analysis and Spread Estimation System.

C. Misinformation Analysis and Spread Prediction

The analyzed content is evaluated using AI-assisted content analysis and fact-verification information to assess its credibility and misinformation risk. After the user verifies the extracted engagement and follower information, the finalized social metrics are used to assess spreading behaviour and estimate future audience exposure where supported by the implemented method. The system distinguishes observed spreading behaviour from experimental future audience-exposure estimates, which should not be interpreted as verified unique reach. The potential propagation network is generated separately to visualize simulated dissemination relationships and support network-structure analysis.

Current Spread Assessment and Future Audience-Exposure Estimation

The spread-analysis component distinguishes current spreading behaviour from future audience-exposure estimation. The system assesses spreading behaviour using the available engagement information. Where time-stamped observations are available, changes in engagement metrics can help characterize spreading behaviour over time. This assessment describes observed engagement patterns and should not be interpreted as a calibrated probability of future spread.

For X, future audience exposure is estimated experimentally by combining the median of clipped historical daily logarithmic growth rates with a growth rate predicted by an engagement-informed Ridge regression model. The historical growth rates are calculated from eligible pairs of time-stamped view observations, clipped to the interval [-0.03, 0.08], and summarized using their median. The two growth-rate estimates are blended using a configured weight of 0.25 for the Ridge-model component. For Instagram and Facebook, provisional exposure estimates are calculated using projected engagement counts and benchmark-based engagement-to-view methods. These estimates are exploratory and should not

be interpreted as verified unique audience reach or validated predictions. Further evaluation using larger, time-stamped datasets is required to assess reliability across platforms and forecast horizons.

D. Propagation Network Generation

The propagation-network generation module scales the size of a simulated graph according to the estimated future audience exposure of the analyzed post. A linear min–max mapping converts the exposure estimate into a target node count between 10 and 100 nodes, using 152 and 8,959,565 as the lower and upper reference exposure boundaries, respectively. Values outside this interval are clamped to the corresponding node-count boundary. These reference values are based on the selected dataset and forecasting configuration and do not represent guaranteed limits on real-world exposure. The generated nodes and edges represent simulated potential propagation relationships rather than verified sharing events or actual audience reach.

The target node count is calculated as:

$$N(E) = 10 + \text{round}(90 \times \text{clip}((E - 152) / (8{,}959{,}565 - 152), 0, 1))$$

where E is the predicted future audience-exposure estimate and N(E) is the target number of simulated nodes. The function clip(x, 0, 1) restricts the normalized exposure to the interval [0, 1]. The boundary conditions are $N(E) = 10$ for $E \leq 152$ and $N(E) = 100$ for $E \geq 8{,}959{,}565$. Intermediate values are mapped linearly to the node-count range.

IV. SYSTEM IMPLEMENTATION AND ANALYSIS

A. System Implementation Environment

The proposed system is implemented as an integrated web-based application for misinformation analysis, engagement verification, spread assessment, and propagation-network analysis. The frontend is developed using React and Vite with Tailwind CSS, while FastAPI provides backend services and API integration. MongoDB is used for data storage. Groq provides AI/LLM-assisted content analysis, while Tesseract OCR, accessed through Pytesseract, supports text extraction from uploaded images and screenshots. React Flow is used for propagation-network visualization. The analysis services support content assessment, influence analysis, and exploratory future audience-exposure estimation based on available engagement data.

The experimental dataset consists of 43 unique social-media posts: 18 from Instagram, 7 from Facebook, and 18 from X. These posts were used to examine the system's analysis and audience-exposure estimation functionality. Because the dataset is limited in size, the results should be interpreted as preliminary rather than as evidence of generalizable predictive performance. The combined hybrid forecasting method has not yet been separately evaluated; therefore, its predictive performance relative to the individual baseline models remains unverified.

TABLE I IMPLEMENTATION TECHNOLOGIES

Component	Technology
Frontend	React, Vite, Tailwind CSS
Backend	FastAPI
Database	MongoDB
AI/NLP Analysis	Groq API / LLM-based analysis
OCR	Tesseract OCR (via Pytesseract)
Graph Visualization	React Flow
Graph Analysis	NetworkX
Deployment	Vercel and Render

B. Content and System Processing

The implemented system accepts directly entered text and image-based social-media content for misinformation analysis. For image inputs, OCR is performed using Tesseract OCR through Pytesseract to extract textual content, which is then processed by the AI-assisted analysis pipeline. The system supports content analysis and fact verification for supported social-media content, while platform-specific handling is used to extract available engagement metrics. Since OCR may misread text or engagement values due to image quality and layout differences, users can review and correct the extracted metrics before they are used for subsequent spread assessment.

C. Spread Prediction and Risk Interpretation

After the user verifies the extracted engagement metrics, the system analyzes the available information to assess spreading behaviour and estimate future audience exposure where supported by the implemented method. Current spreading behaviour, future audience-exposure estimation, and misinformation risk are treated as separate outputs. For X, daily logarithmic growth rates are calculated from eligible time-stamped view observations, clipped to [-0.03, 0.08], and summarized using their median. This estimate is blended with an engagement-informed Ridge regression growth rate using a weight of 0.25. The X-view experiment used 18 eligible view-observation pairs and leave-one-out validation to compare persistence, Ridge regression, and the median daily log-growth baseline. Ridge did not outperform persistence in

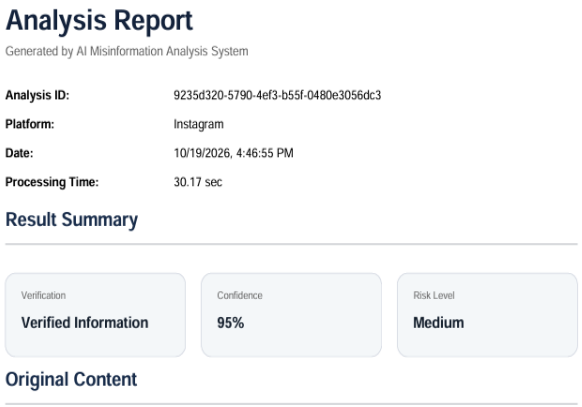
MAE. For Instagram and Facebook, provisional estimates use projected engagement and benchmark-based methods. The forecasting horizon ranges from 0 to 88 days. All estimates are experimental and require further validation; predicted views should not be interpreted as verified unique audience reach.

TABLE II EXAMPLE OUTPUT OF THE MISINFORMATION ANALYSIS SYSTEM

Parameter	Result
Platform	Instagram
Fact Verification Verdict	Verified Information
Verification Confidence	95%
AI Detection Prediction	Needs Verification
Risk Level	Medium
Future Audience-Exposure Estimate	30,346 predicted views (exposure proxy)
Prediction Horizon	7 days

Parameter	Result
Platform	Instagram
Propagation Nodes	10
Propagation Edges	9
Observed Engagement Growth Tier	High

Table II presents an example of the system's misinformation analysis and future audience-exposure estimation for Instagram. The system estimates 30,346 predicted views over the selected 7-day horizon using the available engagement information. This value is an exposure proxy, not verified unique audience reach. The estimate is exploratory and requires further validation using larger, time-stamped datasets.



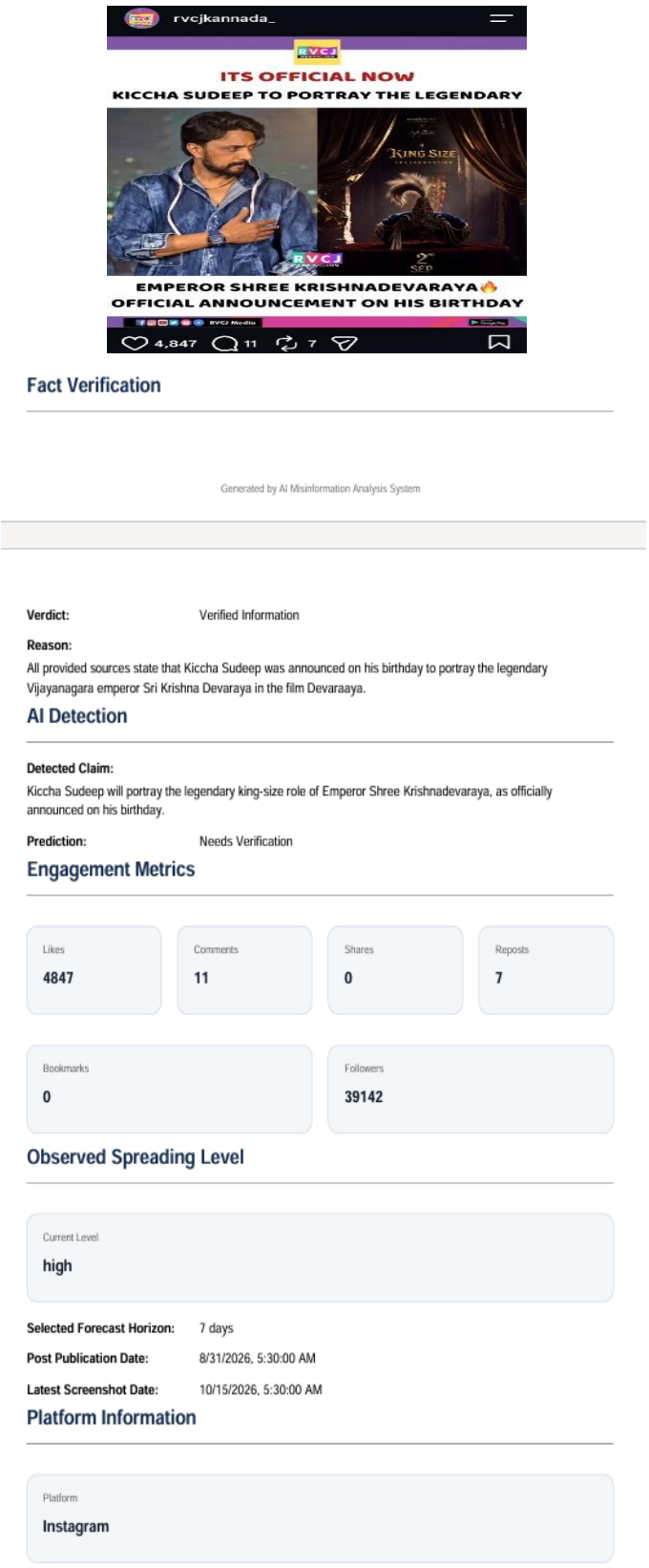


Fig. 3. System-generated report showing misinformation analysis, fact verification, and spread assessment.

D. Misinformation Analysis and Verification

The implemented system performs misinformation analysis using AI-assisted content analysis and fact-verification information. For image-based inputs, the text extracted through OCR is passed through the same analysis pipeline as directly entered text. The system generates a verification verdict and confidence value based on the information processed by its verification component, while AI-based analysis provides an additional prediction that can be reviewed by the user. These outputs should be interpreted as assessments of the submitted content rather than absolute determinations of truth or falsehood. The reliability of the verification results depends on the evidence and methods used by the implemented system.

E. Propagation and Influence Analysis

The implemented system generates a potential propagation network based on the analyzed content and its associated node relationships. The example shown in Table III contains 10 nodes and 9 edges. These values describe the structure of the generated network and do not represent verified real-world information diffusion. The graph visualizes potential dissemination relationships and supports network-structure analysis. Therefore, the results should be interpreted as simulated potential propagation rather than actual tracking of social-media users.

TABLE III POTENTIAL PROPAGATION NETWORK

Test Case	Input / Function	Result
Direct-Text Analysis	User-entered textual content	Successful
Image OCR	English/Hindi/Kannada screenshot	Successful
Facebook Screenshot	Engagement extraction and analysis	Successful
Twitter/X Screenshot	Platform-specific engagement extraction	Successful
Instagram Post	Engagement extraction from screenshot	Successful
Instagram Reel	Engagement extraction from screenshot	Successful
Fact Verification	Evidence-based claim verification	Successful

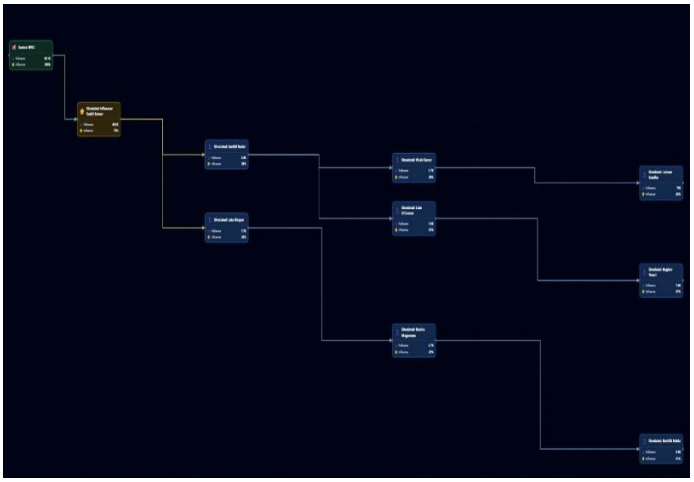


Fig. 4. Simulated propagation network generated using exposure-based linear node scaling.

TABLE IV FUNCTIONAL EVALUATION OF THE IMPLEMENTED SYSTEM

Spread Analysis	Spreading-level assessment and exploratory exposure estimation	Successful
Propagation Analysis	Potential propagation-network generation and visualization	Successful
User Verification	User review of extracted engagement metrics	Successful

The functional evaluation summarizes the tested system modules, including text analysis, image-based text extraction, engagement extraction, fact verification, spread assessment, and propagation-network generation. These tests assess the operation of the implemented functions and do not establish the predictive accuracy of future audience-exposure estimates. Further evaluation using larger, time-stamped datasets is required to assess the reliability of these estimates across platforms and forecast horizons.

F. Discussion

The proposed system integrates content analysis, evidence-based fact verification, engagement verification, and potential propagation-network analysis within a unified workflow. The representative Instagram case illustrates how the system combines content-related findings with available social-media metrics to support misinformation analysis. The system provides verification outputs and application-generated spread indicators, while the propagation graph enables exploration of potential dissemination patterns. However, the reliability of these outputs requires further evaluation. Future work will focus on testing larger time-stamped datasets, improving

platform-specific engagement extraction, expanding multilingual OCR support, and evaluating the robustness of propagation and influence analysis.

V. CONCLUSION

The proposed AI-driven misinformation analysis system integrates content analysis, evidence-based fact verification, OCR-based text extraction, engagement verification, and simulated propagation-network visualization within a unified framework. The system supports text and social-media screenshot inputs, allowing extracted content and available engagement metrics to be analyzed through a common workflow. The representative Instagram case demonstrates the system's analysis capabilities, including verification outputs and experimental audience-exposure estimates. The generated network represents simulated potential propagation relationships rather than verified real-world diffusion. Future work will focus on evaluating output reliability using larger time-stamped datasets, improving platform-specific engagement extraction, expanding multilingual OCR support, and assessing the robustness of propagation and influence analysis.

REFERENCES

[1] J. Ma, W. Gao, Z. Wei, Y. Lu, and K.-F. Wong, "Detect Rumors Using Time Series of Social Context Information on Microblogging Websites," in Proc. 24th ACM Int. Conf. Inf. Knowl. Manage., 2015, pp. 1751–1754, doi: 10.1145/2806416.2806607.

[2] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media," Big Data, vol. 8, no. 3, pp. 171–188, 2020, doi: 10.1089/big.2020.0062.

[3] H. Jwa, D. Oh, K. Park, J. M. Kang, and H. Lim, "exBAKE: Automatic Fake News Detection Model Based on Bidirectional Encoder Representations from Transformers (BERT)," Applied Sciences, vol. 9, no. 19, Art. no. 4062, 2019, doi: 10.3390/app9194062.

[4] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, and J. Gao, "EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection," in Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2018, pp. 849–857, doi: 10.1145/3219819.3219903.

[5] D. Khattar, J. S. Goud, M. Gupta, and V. Varma, "MVAE: Multimodal Variational Autoencoder for Fake News Detection," in Proc. 2019 World Wide Web Conf., 2019, pp. 2915–2921, doi: 10.1145/3308558.3313552.

[6] N. Ruchansky, S. Seo, and Y. Liu, "CSI: A Hybrid Deep Model for Fake News Detection," in Proc. 2017 ACM Conf. Inf. Knowl. Manage., 2017, pp. 797–806, doi: 10.1145/3132847.3132877.

[7] C. Li, J. Ma, X. Guo, and Q. Mei, "DeepCas: An End-to-End Predictor of Information Cascades," in Proc. 26th Int. Conf. World Wide Web, 2017, pp. 577–586, doi: 10.1145/3038912.3052643.

[8] T. Bian, X. Xiao, T. Xu, P. Zhao, W. Huang, Y. Rong, and J. Huang, "Rumor Detection on Social Media with Bi-Directional Graph Convolutional Networks," in Proc. AAAI Conf. Artif. Intell., vol. 34, no. 1, pp. 549–556, 2020, doi: 10.1609/aaai.v34i01.5393.

[9] V.-H. Nguyen, K. Sugiyama, P. Nakov, and M.-Y. Kan, "FANG: Leveraging Social Context for Fake News Detection Using Graph Representation," in Proc. 29th ACM Int. Conf. Inf. Knowl. Manage., 2020, pp. 1165–1174, doi: 10.1145/3340531.3412046.