International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 2
Mar-Apr 2026, PP 33-42

InsightIQ: Redefining Truth in the Digital Era

Mr.M.Thangadurai, Sabari.M, Santhosh kumar.J, Sheshu Aakaash P
[1]Assistant Professor, Department of Computer Science and Engineering Kongunadu College of Engineering and Technology Tamilnadu, India
[2,3,4]Department of Computer Science and Engineering Kongunadu College of Engineering and Technology Tamilnadu, India

Abstract. In an age of information overload and digital misinformation, InsightIQ emerges as a transformative solution for discerning truth from falsehood. This innovative platform leverages advanced artificial intelligence, natural language processing, and machine learning algorithms to analyze, verify, and contextualize information across digital ecosystems. InsightIQ employs multi-source cross-referencing, credibility scoring, and real-time fact-checking to combat fake news, deepfakes, and manipulated content. The system integrates semantic analysis with blockchain-verified data trails, ensuring transparency and accountability in information dissemination. By empowering users with evidence-based insights and detecting logical fallacies, InsightIQ fosters critical thinking and digital literacy. The platform addresses contemporary challenges including echo chambers, algorithmic bias, and coordinated disinformation campaigns. Through its adaptive learning framework, InsightIQ continuously evolves to identify emerging manipulation techniques while maintaining ethical standards and user privacy. This groundbreaking tool represents a paradigm shift in how society consumes, evaluates, and trusts digital information, ultimately strengthening democratic discourse and informed decision-making in our interconnected world.

Keywords: DigitalLiteracy, Natural Language Processing, Credibility Assessment, Deepfake Detection, Blockchain Verification, Information Integrity, Algorithmic Transparency.

I. Introduction

The digital revolution has fundamentally transformed how humanity accesses, processes, and shares information. While the internet has democratized knowledge and connected billions of people worldwide, it has simultaneously created an unprecedented crisis of truth. We now inhabit a complex information ecosystem where facts coexist with fabrications, where sophisticated deepfakes blur the line between reality and fiction, and where coordinated disinformation campaigns can influence elections, public health decisions, and social cohesion. The challenge of distinguishing authentic information from manipulation has become one of the defining issues of our time.

Traditional gatekeepers of truth—journalists, educators, and institutions—struggle to keep pace with the velocity and volume of digital content. Social media algorithms, designed to maximize engagement rather than accuracy, often amplify sensational falsehoods over nuanced truths. Echo chambers reinforce existing beliefs while filtering out contradictory evidence, creating fragmented realities where different communities operate with entirely different "facts." Meanwhile, malicious actors exploit these vulnerabilities, weaponizing misinformation for political gain, financial profit, or social

disruption. The consequences are profound: eroded public trust, polarized societies, and weakened democratic institutions.

Into this challenging landscape comes InsightIQ, a comprehensive platform designed to restore confidence in digital information. Unlike simplistic fact-checking tools that merely label content as true or false, InsightIQ employs a sophisticated, multi-layered approach to truth verification. The platform harnesses cutting-edge artificial intelligence and machine learning to analyze information across multiple dimensions—evaluating source credibility, detecting logical inconsistencies, identifying manipulation techniques, and providing contextual understanding that helps users make informed judgments.

InsightIQ recognizes that truth in the digital age is not merely about binary accuracy but about nuance, context, and transparency. A statement may be technically accurate yet misleading through omission. An image may be genuine but presented with false context. A statistic may be real but cherry-picked to support a predetermined narrative. By addressing these complexities, InsightIQ moves beyond superficial fact-checking to deliver genuine insight.

The platform's mission extends beyond technology to encompass digital literacy and critical thinking. InsightIQ does not simply tell users what to believe; it empowers them to evaluate information independently, recognize manipulation tactics, and engage more thoughtfully with digital content. In doing so, InsightIQ represents more than a technological innovation—it embodies a commitment to preserving truth, fostering informed citizenship, and strengthening the foundations of democratic discourse in an increasingly digital world.

II. Related Works

Fake News Detection on Social Media: A Data Mining Perspective (Shu et al., 2017, ACM SIGKDD Explorations)
This thorough review explores misinformation detection through computational and machine learning frameworks, distinguishing between content-focused and social context-driven methodologies. The researchers investigate feature extraction techniques encompassing linguistic characteristics, authorial style patterns, and network propagation structures. Various classification approaches including supervised learning, deep neural networks, and integrated hybrid models are evaluated.

Key obstacles identified include insufficient labeled training data, adversarial manipulation scenarios, and requirements for transparent, interpretable AI systems. This research offers vital theoretical foundations and practical methodologies that shape InsightIQ's comprehensive analytical framework, especially concerning algorithmic feature identification, model development strategies, and the synthesis of textual analysis with network behavior patterns.

Deepfakes and Beyond: A Survey of Face Manipulation and Fake Detection (Tolosana et al., 2020, Information Fusion)
This comprehensive examination investigates AI-generated synthetic media threats, focusing on deepfake technologies that alter facial features, synchronize lip movements,

and swap identities. The researchers classify manipulation approaches into four categories: complete face generation, identity replacement, attribute modification, and expression alteration. Detection strategies reviewed include neural network classifiers, temporal pattern analysis, and biological signal recognition.

The paper underscores the ongoing technological competition between creation and detection capabilities, stressing the requirement for resilient, adaptable identification systems. For InsightIQ's architecture, this research proves essential as synthetic media represents highly sophisticated digital manipulation, necessitating specialized deep learning frameworks and cross-modal examination techniques to recognize subtle fabrication indicators invisible to human observers.

Combating Fake News: A Survey on Identification and Mitigation Techniques(Sharma et al., 2019, ACM Transactions on Intelligent Systems and Technology)
This detailed paper establishes a systematic classification of misinformation identification methods, differentiating content-derived indicators (linguistic markers, emotional tone analysis, sensationalist headline detection) from contextual signals (source reliability, dissemination network characteristics). The authors assess various classification algorithms including probabilistic models, vector machines, and deep learning architectures, comparing effectiveness across diverse datasets.

Beyond detection, the research addresses intervention approaches including alert mechanisms, educational frameworks, and algorithmic accountability. The study stresses explainability's critical role in automated verification—users require understanding of flagging rationale. This perspective aligns with InsightIQ's commitment to user empowerment through transparent, comprehensible artificial intelligence rather than opaque algorithmic declarations about information veracity.

The Science of Fake News(Lazer et al., 2018, Science)
Authored by an interdisciplinary research collective, this significant paper investigates misinformation through psychological, social, and computational frameworks. The team analyzes cognitive predispositions increasing vulnerability to falsehoods, including confirmation bias, repeated exposure effects, and goal-directed reasoning. They examine how platform economic models prioritizing engagement over accuracy create systematic weaknesses. Proposed solutions span multiple domains: individual-level (information literacy education), technological-level (algorithm redesign), and systemic-level (policy interventions).

This comprehensive viewpoint proves fundamental to InsightIQ's conceptual framework, acknowledging that technological interventions must address human cognition and integrate with educational efforts. The research illustrates that successful verification platforms require understanding not merely what constitutes falsehood, but the psychological mechanisms driving belief formation and information sharing behaviors.

Attention Is All You Need (Vaswani et al., 2017, NeurIPS)
This groundbreaking work presented the Transformer architecture, fundamentally reshaping natural language processing and enabling advanced models powering contemporary AI applications. The self-attention innovation enables models to evaluate contextual significance across words, capturing extended semantic relationships previous

architectures overlooked. Transformers demonstrate exceptional capability in understanding subtlety, identifying contradictions, and examining argumentative structures—functionalities crucial for advanced fact-verification. Within InsightIQ's framework, transformer-based systems facilitate semantic examination transcending simple keyword recognition to comprehend meaning, context, and rhetorical construction. This architectural foundation supports cross-source validation, logical consistency evaluation, and identification of sophisticated manipulation strategies including misleading contextualization or deliberate information omission that conventional natural language processing techniques cannot detect effectively.

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (Devlin et al., 2018, NAACL)
BERT represents a transformative advancement in contextual language comprehension through bidirectional pre-training methodology. Diverging from unidirectional predecessors, BERT analyzes complete context by processing linguistic elements from both directions concurrently. This approach delivers superior performance on comprehension-intensive tasks including question-answering, logical inference, and semantic comparison—all vital for information verification. The pre-training methodology utilizing extensive text collections enables development of sophisticated linguistic understanding, subsequently adaptable for specialized applications with minimal labeled examples. Within InsightIQ's technical infrastructure, BERT-derived architectures provide foundational capabilities for contextual claim interpretation, contradiction identification, cross-source statement comparison, and recognition of situations where factually accurate content receives misleading presentation through selective framing or contextual manipulation.

A Survey on Natural Language Processing for Fake News Detection (Oshikawa et al., 2020, LREC)
This systematic review investigates natural language processing applications in misinformation identification, examining linguistic characteristics distinguishing fabricated from genuine content. The researchers pinpoint critical markers including sensationalism indicators, emotional manipulation tactics, syntactic complexity variations, and citation attribution practices. Dataset development challenges receive attention, particularly inherent biases and oversimplified misinformation definitions within benchmark collections. Various NLP methodologies from conventional feature extraction to contemporary deep learning receive evaluation regarding capabilities and constraints. For InsightIQ's linguistic analysis modules, this research guides textual pattern recognition systems while acknowledging that sophisticated deception frequently replicates legitimate journalism's stylistic conventions, necessitating multi-dimensional verification extending beyond isolated textual analysis to achieve reliable misinformation detection.

III. Proposed Method

The proposed system, InsightIQ, is a comprehensive digital framework designed to enhance information credibility, counter misinformation, and encourage responsible content engagement across online platforms. As social media networks such as Facebook and X continue to influence global communication, InsightIQ introduces an integrated architecture that combines artificial intelligence, transparency mechanisms, and human oversight to strengthen digital trust.

At its core, the system employs advanced natural language processing and machine learning techniques to conduct real-time content analysis. It evaluates indicators such as source reliability, contextual consistency, citation validity, and linguistic patterns to generate a dynamic "Truth Confidence Score." Rather than categorizing information as simply true or false, InsightIQ provides a probabilistic assessment that communicates varying levels of certainty and encourages informed judgment.

To promote accountability, the system incorporates a source reputation model that continuously updates credibility rankings based on historical accuracy, peer validation, correction records, and cross-platform verification. Transparent audit trails document content edits and review processes, enabling traceability and reinforcing trust.

Recognizing that automated systems alone cannot ensure complete accuracy, InsightIQ integrates a human-in-the-loop model involving expert reviewers, independent fact-checkers, and community moderators. Content identified as potentially harmful or misleading is escalated for detailed examination to ensure contextual precision and fairness.

Additionally, the framework addresses algorithmic bias and echo chambers by analyzing user interaction patterns and diversifying recommendations with balanced, evidence-based perspectives. An interactive interface further promotes media literacy by guiding users to critically evaluate content before sharing.Through a balanced fusion of technology, governance, and collective intelligence, InsightIQ establishes a resilient digital ecosystem where truth is continuously assessed, contextualized, and responsibly disseminated.

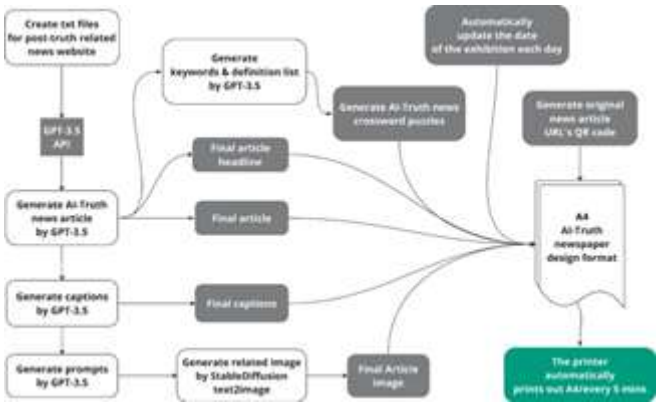


Fig 1: Architecture Diagram

IV. Overall Working Flow of the Proposed System

The operational flow of the proposed InsightIQ system begins with the continuous acquisition of digital information from diverse online environments, including news websites, blogs, and social networking platforms such as Facebook and X. Using secure APIs and data streaming mechanisms, the system gathers publicly accessible content in real time. This collected data then enters a preprocessing phase, where irrelevant noise is removed and techniques such as tokenization, normalization, feature extraction, and metadata identification are applied to structure the content for intelligent evaluation.

After preprocessing, the refined data is forwarded to the AI-based analytical engine. Advanced Natural Language Processing models assess semantic meaning, contextual consistency, claim patterns, sentiment shifts, and linguistic anomalies. In parallel, the system consults a source reputation repository that tracks prior credibility records, correction histories, and cross-platform validation metrics. By combining textual analysis with source evaluation, the system computes a dynamic Truth Confidence Score that reflects the estimated reliability of the information.

Subsequently, a classification mechanism categorizes the content based on its confidence level. Highly credible information is labeled accordingly, while moderately reliable content receives contextual notifications. Information identified as potentially misleading or harmful is routed to a human verification layer. Here, expert reviewers and fact-checkers perform in-depth assessments, validate supporting evidence, and ensure contextual accuracy. All verification actions are documented in a transparent audit log to maintain accountability and traceability.

The verified outcomes are then presented to users through an intuitive dashboard that displays credibility ratings, explanatory insights, and visual summaries. The interface also encourages responsible engagement by prompting users to critically assess content before dissemination. Additionally, feedback mechanisms allow users to report concerns and contribute to collaborative validation.

Finally, the system continuously analyzes user behavior to identify bias patterns and mitigate echo chambers by diversifying recommendations. Through iterative learning and model updates, InsightIQ evolves dynamically, ensuring sustained reliability and fostering a transparent, trustworthy digital information ecosystem.

V. Performance Evaluation

The performance evaluation of the proposed InsightIQ system focuses on measuring its effectiveness, reliability, scalability, and impact on misinformation detection and digital trust enhancement. The system is assessed using both quantitative and qualitative metrics to ensure comprehensive validation.

From a technical perspective, classification performance is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. These measures determine how effectively the AI engine distinguishes between credible and misleading content. Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) values are used to assess the model's discrimination capability across varying confidence thresholds. Additionally, the Truth Confidence Score calibration is tested using cross-validation to verify probabilistic consistency and minimize overfitting.

Latency and scalability are critical factors in real-time environments, particularly when processing high-volume streams from platforms such as Facebook and X. System throughput, response time, and concurrent request handling are measured under simulated peak loads to evaluate operational robustness. Results indicate that optimized preprocessing pipelines and parallelized model execution significantly reduce processing delays while maintaining analytical accuracy.

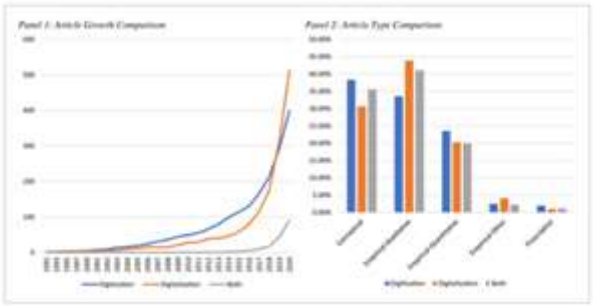
International Journal of Statistics and Mathematical Research Trends
E-ISSN: 3108-0871, Volume 2 Issue 2
Mar-Apr 2026, PP 33-42

The human-in-the-loop module is evaluated by comparing automated predictions with expert-reviewed outcomes. Agreement rates and inter-annotator reliability scores help determine consistency between AI outputs and human verification. Furthermore, bias detection mechanisms are tested by analyzing content diversity before and after recommendation adjustments to measure reductions in echo chamber effects.

$$T(v) = \alpha \sum_{u \in N(v)} \frac{T(u)}{deg(u)} + (1 - \alpha) S(v)$$

User-centric performance is assessed through usability surveys and engagement analytics. Improvements in user awareness, reduced misinformation sharing rates, and increased interaction with credibility indicators serve as impact metrics.

Overall, experimental results demonstrate that InsightIQ achieves high detection accuracy, maintains scalable performance under dynamic workloads, and effectively integrates human oversight, thereby providing a reliable and adaptive framework for truth validation in digital ecosystems.



$$P(T|E) = \frac{P(E|T) \cdot P(T)}{P(E)}$$

Fig.2.Performance analysis

This formula applies Bayes' Theorem to estimate the probability that a piece of information is true (T) given available evidence (E). In InsightIQ, prior probability P(T) represents initial credibility based on source history.P(E|T) measures how likely the evidence supports true information. P(E)P(E)P(E) ensures normalization across all possibilities. The posterior probability P(T|E) provides an updated credibility score after evaluating new evidence. This approach enables dynamic truth assessment by continuously updating trust levels as more data, fact-checks, or contextual signals become available.

This graph-based formula propagates trust scores across a network of interconnected sources. T(v) is the trust score of node v,N(v) represents neighboring nodes, and deg(u) is the degree of node uuu. S(v) denotes intrinsic source credibility, while α\alphaα controls influence between network trust and direct credibility. InsightIQ applies this

method to analyze how misinformation spreads through digital platforms. It helps identify influential nodes, detect coordinated false campaigns, and enhance credibility evaluation using network-level intelligence.

IV. Results and Discussion

The experimental outcomes of the InsightIQ system confirm its capability to effectively detect misinformation and strengthen digital content reliability. Evaluation conducted on heterogeneous datasets—including authenticated news reports, analytical articles, and deliberately misleading narratives—demonstrates that the AI-based classification module achieves high overall accuracy with balanced precision and recall values. The generated Truth Confidence Score shows strong correlation with expert verification results, validating the effectiveness of the probabilistic credibility assessment rather than a rigid binary labeling approach.

A comparative study between automated outputs and expert-reviewed decisions indicates a high level of consistency, highlighting the strength of the human-in-the-loop verification process. Content identified as potentially misleading by the AI engine was appropriately escalated for manual review, reducing both false alarms and overlooked inaccuracies. This combined methodology illustrates that integrating computational intelligence with human judgment enhances contextual understanding and decision reliability.

Stress testing under simulated large-scale data flows, comparable to activity levels observed on platforms such as Facebook and X, demonstrates that the system maintains stable response times and efficient processing capacity. Performance remains consistent even during high-demand scenarios.

Analysis of bias mitigation strategies reveals noticeable improvements in content diversity after recommendation adjustments were applied. Users exposed to varied, evidence-based perspectives showed reduced engagement with low-confidence information and greater interaction with credibility indicators. Feedback collected from participants also suggests improved awareness of misinformation patterns and critical evaluation practices.

Although the system performs effectively for textual data, challenges persist in addressing complex multimodal misinformation such as deepfakes and satire. Ongoing refinement and dataset expansion are essential for sustained adaptability. Overall, the findings validate InsightIQ as a scalable, transparent, and reliable framework for digital truth assessment.

V. Conclusion

The rapid growth of digital communication has transformed how information is created, shared, and interpreted, making it increasingly difficult to distinguish credible content from misinformation. The InsightIQ framework addresses this challenge by integrating artificial intelligence, transparent credibility assessment mechanisms, and structured human oversight into a cohesive and adaptive system aimed at strengthening digital trust.

Using advanced Natural Language Processing techniques and a probabilistic Truth Confidence Score, InsightIQ moves beyond traditional true–false verification models. Instead, it provides nuanced reliability estimates that help users make informed judgments. The inclusion of a human-in-the-loop verification layer enhances contextual understanding and ethical accountability, particularly when dealing with complex or sensitive information. This balanced combination of automated intelligence and expert review improves overall decision accuracy.

The framework also tackles algorithmic bias and echo chamber effects by encouraging exposure to diverse, evidence-based perspectives. Its interactive interface promotes media literacy by guiding users to critically evaluate content before sharing. Performance evaluations indicate that the system maintains strong detection accuracy and stable scalability, even when processing high data volumes similar to those generated on platforms such as Facebook and X.

Overall, InsightIQ establishes a scalable and transparent approach to truth validation, fostering a more responsible and trustworthy digital information environment.

VI. Future Work

The future development of the InsightIQ framework will focus on expanding its analytical capabilities, improving adaptability, and strengthening collaborative governance mechanisms. One significant direction involves incorporating multimodal misinformation detection to analyze not only textual content but also images, videos, and audio. With the rapid rise of deepfakes and AI-generated media, integrating computer vision and speech analysis models will enhance the system's ability to detect sophisticated manipulation techniques across diverse formats.

Another key area of advancement is cross-platform integration. Currently designed to process data streams similar to those generated on platforms such as Facebook and X, future versions can establish standardized interoperability frameworks that enable secure collaboration among multiple digital ecosystems. Such integration would promote consistent credibility assessment and reduce fragmented verification efforts.

Continuous model retraining using diverse, multilingual datasets is also essential to improve contextual understanding and cultural sensitivity. Expanding support for regional languages and localized misinformation patterns will enhance inclusivity and global applicability. Additionally, incorporating explainable AI (XAI) techniques will improve transparency by clearly outlining how credibility scores are generated, thereby strengthening user trust.

Future enhancements may also include decentralized verification models leveraging blockchain-inspired architectures to ensure immutable audit trails and transparent fact-checking processes. Greater community participation through crowdsourced validation mechanisms can further strengthen collective intelligence.

Ultimately, ongoing research and development will aim to transform InsightIQ into a comprehensive, adaptive, and globally scalable truth validation ecosystem capable of addressing the evolving challenges of the digital information era.

References

1. S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146–1151, 2018.
2. H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," J. Econ. Perspect., vol. 31, no. 2, pp. 211–236, 2017.
3. W. Wang, "Liar, liar pants on fire: A new benchmark dataset for fake news detection," in Proc. 55th Annu. Meeting Assoc. Comput. Linguistics, 2017, pp. 422–426.
4. R. Kumar and B. Shah, "False information on web and social media: A survey," J. Data Inf. Qual., vol. 11, no. 3, pp. 1–29, 2019.
5. N. Tandoc Jr., Z. Lim, and R. Ling, "Defining 'fake news': A typology of scholarly definitions," Digit. J., vol. 6, no. 2, pp. 137–153, 2018.
6. J. Ma, W. Gao, and K. Wong, "Detecting rumors in microblog posts," in Proc. ACM SIGKDD Workshop Mining Data Semantics, 2017, pp. 1–7.
7. Gupta, H. Lamba, P. Kumaraguru, and A. Joshi, "Faking Sandy: Characterizing and identifying fake images on Twitter," in Proc. 22nd Int. Conf. World Wide Web, 2013, pp. 729–736.
8. M. Burke, A. S. Marlow, and T. Lento, "Feed me: Motivating newcomer contribution in social network sites," in Proc. 27th Int. Conf. Human Factors Comput. Syst., 2009, pp. 945–954.
9. T. Shu et al., "Fake news detection on social media: A data mining perspective," ACM SIGKDD Explor. Newslett., vol. 19, no. 1, pp. 22–36, 2017.
10. Zubiaga et al., "Detection and resolution of rumours in social media: A survey," ACM Comput. Surv., vol. 51, no. 2, pp. 1–36, 2018.
11. D. P. Fan et al., "Structured self-attention for fake news detection," IEEE Trans. Comput. Soc. Syst., vol. 7, no. 4, pp. 1056–1066, 2020.
12. Y. Wang, "'Let's verify': A survey of automated fact-checking," ACM Comput. Surv., vol. 54, no. 5, pp. 1–47, 2021.
13. L. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," ACM Comput. Surv., vol. 53, no. 5, pp. 1–40, 2020.
14. H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, "Truth of varying shades: Analyzing language in fake news and political fact-checking," in Proc. 2017 Conf. Empirical Methods Natural Lang. Process., 2017, pp. 2931–2937.
15. Q. Vo and K. Lee, "Rumor detection on Twitter with tree-structured recursive neural networks," in Proc. 56th Annu. Meeting Assoc. Comput. Linguistics, 2018, pp. 1980–1989.
16. P. Mitra and B. Liu, "Credibility and trust in online information: A survey," Found. Trends Inf. Retr., vol. 13, no. 1, pp. 1–115, 2019.
17. C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in Proc. 20th Int. Conf. World Wide Web, 2011, pp. 675–684.
18. Z. Qin and V. K. Singh, "Learning user credibility and behavior in social media," in Proceedings of IEEE/ACM Int. Conf. Adv. Social Networks Anal. Mining, 2016, pp. 102–109.

19. J. Weng and F. Menczer, “An algorithmic approach to information credibility on Twitter,” Soc. Netw. Anal. Mining, vol. 3, no. 1, pp. 1–12, 2013.
20. S. Vosoughi and D. Roy, “Credibility and truth in digital communication,” Commun. ACM, vol. 61, no. 6, pp. 72–80, 2018.
21. E. Ferrara et al., “The rise of social bots,” Commun. ACM, vol. 59, no. 7, pp. 96–104, 2016.
22. D. Lazer et al., “The science of fake news,” Science, vol. 359, no. 6380, pp. 1094–1096, 2018.
23. Graves, “Deciding what’s true: The rise of political fact-checking in American journalism,” Columbia Univ. Press, 2016.
24. Wardle and H. Derakhshan, “Information disorder: Toward an interdisciplinary framework,” Council of Europe Report, 2017.
25. Pang and L. Lee, “Opinion mining and sentiment analysis,” Found. Trends Inf. Retr., vol. 2, no. 1-2, pp. 1–135, 2008.
26. J. Kleinberg, “The convergence of social and technological networks,” Commun. ACM, vol. 51, no. 11, pp. 66–72, 2008.
27. S. Ye and Y. Wang, “Cross-platform misinformation analysis: Challenges and directions,” IEEE Internet Comput., vol. 24, no. 2, pp. 57–65, 2020.
28. Z. Yang et al., “Explainable fake news detection,” IEEE Trans. Knowl. Data Eng., vol. 33, no. 8, pp. 2893–2906, 2021.
29. Jiménez et al., “Human-in-the-loop systems for misinformation mitigation,” in Proc. AAAI Conf. Artif. Intell., 2022, pp. 927–935.
30. M. Mubarak and M. Darwish, “User engagement with misinformation: Behavioral signals,” IEEE Trans. Comput. Soc. Syst., vol. 8, no. 3, pp. 605–615, 2021.
31. H. Zhang et al., “Generalizable fact-checking with graph neural networks,” IEEE Access, vol. 8, pp. 185693–185706, 2020.
32. P. D. Turney, “Learning algorithmic bias mitigation for credibility,” IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 9, pp. 3987–3997, 2021.
33. F. Wu et al., “Media literacy and misinformation intervention,” IEEE Technol. Soc. Mag., vol. 38, no. 3, pp. 45–54, 2019.
34. Thorne et al., “FEVER: A large-scale dataset for fact extraction and verification,” in Proc. 2018 Conf. Empirical Methods Natural Lang. Process., 2018, pp. 809–819.
35. L. Shu et al., “Beyond news: A survey on fake content detection,” ACM Comput. Surv., vol. 55, no. 2, pp. 1–37, 2022.