International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 5, Sep-Oct-2026, ISSN (Online): 2395-566X

Deep Learning And Social-Context Approaches For Fake News Detection: A Review Of Neural, Propagation-Based And Explainable Methods

Durafe Sejal Dattatray; Prof. N. G. Rokade; Prof. M. C. Kolse; Prof. S. K. Said
Department Of Artificial Intelligence And Data Science Jaihind College Of Engineering Pune, Maharashtra, India;

Abstract—Fake news and misinformation on digital platforms threaten public trust and information integrity, and manual verification cannot keep pace with the volume of content created daily. This paper reviews AI-based fake news detection beyond simple text classification, focusing on deep learning and social-context methods. It examines early credibility features on social media, recurrent networks that model the time series of user responses to rumours, hybrid models that combine article text with user behaviour, explainable models that use user comments to justify decisions, and graph-based geometric deep learning that analyses how news spreads. Machine-generated fake news and multimodal datasets combining text and images are also discussed, together with support vector machines and random forests as baselines. A five-layer real-time detection architecture is proposed that fuses content, user and propagation signals, followed by research gaps, objectives, a comparison of approaches and a discussion of challenges including early detection before content spreads, cross-platform generalization, adversarial manipulation and the need for explainability. The review concludes that combining what is said, who says it and how it spreads produces more accurate and trustworthy fake news detection than content analysis alone.

Index Terms—Fake news detection, misinformation, deep learning, social context, propagation graphs, explainable AI, rumour detection, NLP

I. INTRODUCTION

1.1 Motivation

A false story about a public figure, a health remedy or a communal incident can reach millions of people within hours through social media and messaging apps. Vosoughi *et al.* [1] showed that false news spreads more widely and quickly than true news. The consequences range from confusion and panic to violence and damage to democratic processes.

Text classifiers that look only at the words of an article are useful but limited: skilled writers can make fake news look stylistically normal, and Potthast *et al.* [2] found that style alone separates hyperpartisan from mainstream news better than fake from real news. Additional evidence comes from the social context, meaning who shares the story, how users react and how it propagates.

This paper reviews deep learning and social-context approaches to fake news detection and proposes a multi-signal architecture.

1.2 Objectives Of The Review

- To review deep learning text models for fake news detection.
- To examine user-based and propagation-based social-context methods.
- To propose a multi-signal real-time detection architecture.
- To compare content-only, social-context and hybrid approaches.
- To discuss challenges and future directions, including explainability.

II. OVERVIEW OF KEY CONCEPTS

2.1 Signals For Detection

Fig. 1 organizes detection signals into content (text, images), source (publisher and author credibility), users (accounts that share and comment) and propagation (the structure and timing of sharing cascades).

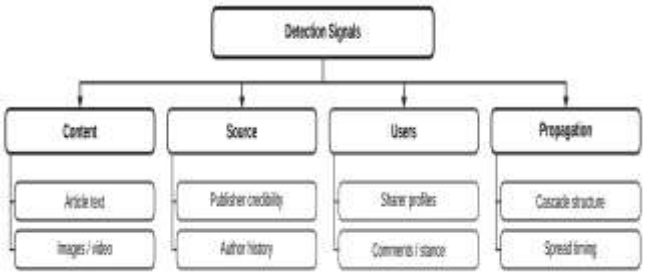


Fig. 1. Signals used for fake news detection.

2.2 Deep Text Models

Convolutional networks for sentence classification [3] capture local phrase patterns, recurrent networks [4] capture sequence order, and transformers such as BERT [5] capture context across a whole article.

© 2026 IJSRET

III. REVIEW OF EXISTING WORK

Castillo *et al.* [6] studied information credibility on Twitter and showed that features of the message, the user and the propagation pattern can automatically assess whether trending news is credible. Ma *et al.* [7] modelled the stream of posts responding to an event with recurrent neural networks and detected rumours earlier and more accurately than feature-based methods.

Ruchansky *et al.* [8] proposed CSI, a hybrid model that captures the text of articles, the temporal pattern of user engagement and the characteristics of users who promote them. Shu *et al.* [9] introduced dEFEND, which co-attends to article sentences and user comments, both improving accuracy and producing explanations by highlighting check-worthy sentences and informative comments.

Monti *et al.* [10] applied geometric deep learning to propagation graphs and found that fake and real news spread in distinguishable ways, allowing detection within hours of publication. Rashkin *et al.* [11] analysed linguistic characteristics of untrustworthy news, and Zellers *et al.* [12] showed with Grover that neural generators can produce convincing fake articles, but that the same models are good at detecting them. Nakamura *et al.* [13] released Fakeddit, a large multimodal dataset of text and images. Table I summarizes these works.

TABLE I Summary of Literature Survey

Ref.	Method	Key Findings	Limitations
[6]	Message, user and propagation features	Automatic credibility assessment	Hand-crafted features
[7]	RNN over response streams	Earlier rumour detection	Needs response posts
[8]	CSI hybrid model	Text + engagement + users	Requires user data
[9]	dEFEND co-attention	Accurate and explainable	Needs user comments
[10]	Geometric deep learning on cascades	Early propagation-based detection	Platform data access
[11]	Linguistic analysis of fake news	Identified deceptive language cues	Style can be imitated
[12]	Grover neural fake news	Generator-based detection	Arms race with generators
[13]	Fakeddit multimodal dataset	Large text + image benchmark	Reddit-specific

IV. IDENTIFIED RESEARCH GAPS

- Propagation-based methods need time for news to spread, delaying detection.
- Social context data is increasingly hard to access due to platform restrictions.
- Models trained on one platform or language often fail on others.
- Adversaries can manipulate engagement with bots to mislead social-context detectors.
- Explanations understandable to ordinary users remain rare.

V. SYSTEM ARCHITECTURE

The proposed architecture appears in Fig. 2.

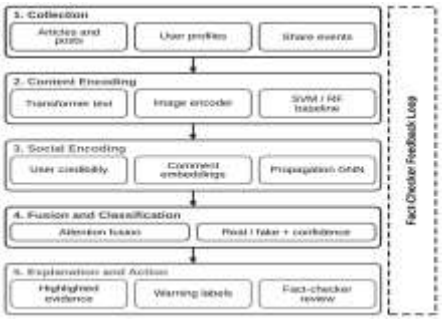


Fig. 2. Proposed multi-signal fake news detection architecture.

5.1 Collection Layer

Articles, posts, images, user profiles, comments and share events are collected from news sites and social platforms through permitted APIs.

5.2 Content Encoding Layer

A transformer encodes article text; an image encoder processes attached images; baseline SVM and random forest models provide fast initial scores.

5.3 Social Encoding Layer

User credibility features and comment embeddings are computed, and the share cascade is represented as a graph encoded with a graph neural network.

5.4 Fusion And Classification Layer

Content, user and propagation representations are combined with attention and classified as real or fake, with confidence.

© 2026 IJSRET

5.5 Explanation And Action Layer

Influential sentences and comments are highlighted; fact-checkers review flagged items, and users see warning labels with links to verified sources.

VI. IMPLEMENTATION METHODOLOGY

Fig. 3 shows the processing flow. The fused representation h is formed from content c, users u and propagation g with attention weights α obtained by a softmax over learned scores:

$$h = \alpha_c \cdot c + \alpha_u \cdot u + \alpha_g \cdot g \quad (1)$$

and the final prediction is $\hat{y} = \sigma(Wh + b)$. Models are trained with cross-entropy and evaluated with F1-score and early-detection accuracy at fixed time points after publication.

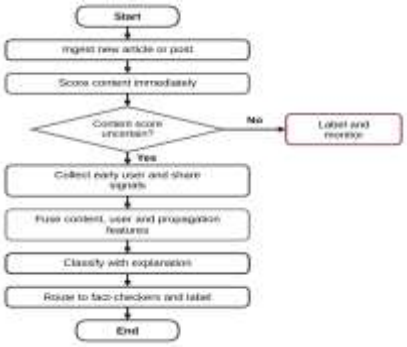


Fig. 3. Multi-signal detection workflow.

VII. PERFORMANCE COMPARISON

Table II compares approaches. Content-only models enable instant detection; social-context and propagation models improve accuracy as data accumulates; hybrid models give the best overall performance and support explanation.

TABLE II Comparison of Detection Approaches

Approach	Timeliness	Accuracy	Data Need
Content-only (SVM / RF)	Instant	Medium	Text only
Content-only (BERT)	Instant	High	Text only
User / comment based	Minutes–hours	High	Engagement data
Propagation graphs	Hours	High	Share cascades
Hybrid fusion	Progressive	Very high	All signals

VIII. OPEN ISSUES AND FUTURE DIRECTIONS

Table III summarizes challenges. Future work includes early detection using the first few shares, cross-lingual models for Indian languages, detection of AI-generated text and images, and collaborative systems connecting platforms with fact-checking organizations.

TABLE III Challenges and Mitigation

Challenge	Mitigation
Late detection	Content-first, refine with early signals
Data access limits	Partnerships; privacy-preserving APIs
Cross-platform shift	Domain adaptation; multilingual models
Bot manipulation	Bot detection; robust aggregation
Explainability	Evidence highlighting; fact-check links

IX. CONCLUSION

Fake news detection is most effective when it considers not only what is said but also who spreads it and how. The reviewed literature shows that deep text models, user engagement modelling, propagation graphs and explainable attention mechanisms each add useful signals. A hybrid, explainable system that fuses these signals and supports human fact-checkers offers the strongest defence against digital misinformation.

9.1 Acknowledgment

The authors express their sincere gratitude to the Department of Artificial Intelligence and Data Science, Jaihind College of Engineering, Pune, for providing the facilities, guidance and encouragement needed to carry out this work. The authors also thank the faculty members and colleagues whose valuable suggestions helped improve the quality of this review.

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. M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, “A stylometric inquiry into hyperpartisan and fake news,” in Proc. 56th Annu. Meeting ACL, 2018, pp. 231–240.
3. Y. Kim, “Convolutional neural networks for sentence classification,” in Proc. EMNLP, 2014, pp. 1746–1751.

© 2026 IJSRET

4. S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.
5. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. NAACL-HLT, 2019, pp. 4171–4186.
6. C. Castillo, M. Mendoza, and B. Poblete, “Information credibility on Twitter,” in Proc. 20th Int. Conf. World Wide Web (WWW), 2011, pp. 675–684.
7. J. Ma et al., “Detecting rumors from microblogs with recurrent neural networks,” in Proc. 25th Int. Joint Conf. Artif. Intell. (IJCAI), 2016, pp. 3818–3824.
8. N. Ruchansky, S. Seo, and Y. Liu, “CSI: A hybrid deep model for fake news detection,” in Proc. ACM Conf. Inf. Knowl. Manage. (CIKM), 2017, pp. 797–806.
9. K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, “dEFEND: Explainable fake news detection,” in Proc. 25th ACM SIGKDD, 2019, pp. 395–405.
10. F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, “Fake news detection on social media using geometric deep learning,” arXiv:1902.06673, 2019.
11. 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. EMNLP, 2017, pp. 2931–2937.
12. R. Zellers et al., “Defending against neural fake news,” in Proc. NeurIPS, vol. 32, 2019.
13. K. Nakamura, S. Levy, and W. Y. Wang, “r/Fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection,” in Proc. 12th Lang. Resour. Eval. Conf. (LREC), 2020, pp. 6149–6157.

© 2026 IJSRET