iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: XII Month of publication: December 2025

DOI: https://doi.org/10.22214/ijraset.2025.76628

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

A Review on Fake News Detection and Personalized Recommendation on Social Media using BERT

Adithya AA[1], Hana Habeeb Rahman[2], Goutham Krishnan[3], Sylasree VA[4], Asst.Prof. Reshma P D[5]
Department of Computer Science and Engineering Universal Engineering College Thrissur,Kerala,India

Abstract: Therapidspreadofmisinformationonsocialmedia platforms poses a serious threat to public trust, social stability, and digital well-being, making fake news detection a critical researchproblem.Toaddressthischallenge,theintegration of intelligent detection models with privacy-preserving mecha- nisms has become increasingly important. This paper presents a comprehensive review of transformer-based fake news detection approaches, particularly those utilizing Bidirectional Encoder Representations from Transformers (BERT), along with Feder- ated Learning techniques for secure and decentralized person- alization. The study analyzes state-of-the-art methods, bench- markdatasets,andkeyperformancemetrics,demonstratinghow BERT-basedmodelseffectivelycapturedeepcontextualsemantics to accurately distinguish fake content from factual information. Inparallel,federatedlearningenablesdistributedmodeltraining while preserving user privacy, making it suitable for deployment in recommendation systems on social media platforms. Further- more,thisreviewdiscussesahybridarchitecturalperspectivethat combines a centralized BERT-based classifier with a lightweight, device-side federated recommender to deliver trustworthy and personalized news feeds. Finally, major challenges such as data imbalance, model explainability, and adversarial manipulation areexamined,andfutureresearchdirectionsareoutlinedtoward robust, interpretable, and ethically aligned artificial intelligence systems for combating misinformation.

Index Terms: Fake News Detection, BERT, Federated Learn- ing, Privacy, Recommendation Systems, Social Media

I. INTRODUCTION

Socialmediaplatformshavefundamentallytransformedthe way people consume and share information. News, opinions, and trending stories can spread across networks within min- utes, reaching millions of users with unprecedented speed. While this rapid dissemination enhances connectivity and awareness,ithasalsocreatedfertilegroundformisinformation and fake news. Inaccurate or intentionally deceptive content can manipulate public opinion, distort social discourse, and erode trust in legitimate information sources [5], [13], [14]. Traditional moderation strategies and manual fact-checking mechanisms are often too slow and resource-intensive to cope with the scale and velocity of modern social media streams[9], [16].

To address these challenges, researchers have increasingly turnedtoadvancedmachinelearningtechniquesforautomated fake news detection. Early approaches relied on handcrafted lexical,syntactic,andsemanticfeaturesprocessedbyclassical machinelearningalgorithmssuchasSupportVectorMachines, Decision Trees, and Random Forests [5], [13], [14]. Although these methods demonstrated reasonable performance under controlled conditions, they struggled to generalize across do- mains and adapt to evolving misinformation patterns.

The introduction of deep learning, particularly transformer- based architectures, has significantly enhanced detection ca- pabilities. Bidirectional Encoder Representations from Trans- formers (BERT) and its variants provide rich contextual em- beddings that capture subtle linguistic cues and long-range semantic dependencies [1], [6], [7]. Ayyub et al. [6] proposed a blended BERT-based approach that combines contextual embeddings with task-specific fine-tuning to improve classi- ficationstability, whileFarokhianandRafe[7]employeddual BERT networks to capture complementary contextual signals. Comparative studies by Chiang et al. [1] further emphasize BERT's superiority over traditional neural network models in source credibility recognition.

Beyond text-based models, ensemble learning and graph- based techniques have been introduced to enhance robust-ness by incorporating social context. Al-shaqi et al. [2] demonstrated that ensembles of diverse classifiers improve performance across heterogeneous datasets, particularly un- der noisy or adversarial conditions. Han et al. [3] applied Graph Neural Networks (GNNs) with continual learning to exploit relational structures in social media, highlighting the importance of user–post interactions and propagation patterns for accurate fake news detection.

Similarly, geometric deep learning frameworks [8] leverage graph structures to model complexrelationships betweenusers,content,andinformation sources.

Multimodal approaches have gained prominence as misin- formation increasingly combines text with images and meta- data. Frameworks such as CroMe [10] and contrastive multi- modalmodels[11]integratevisualandtextualfeaturesusing metric or contrastive learning strategies, achieving improved detection performance. Fu et al. [17] incorporate external knowledge and user interaction features to ground predic- tions in factual evidence and engagement signals. Likewise, Zhu et al. [20] propose shallow–deep multitask learning to jointly leverage unimodal and cross-modal representations, whileKGAlign[18]introducessemantic–structuralknowledge encoding to align textual and visual information with object- level knowledge graphs.

Knowledge integration and interpretability have also emergedascriticalaspectsofreliablefakenewsdetection.Hy- brid frameworks that combine machine learning with knowl- edge engineering enable models to validate claims against external sources and provide explainable decisions [15]. Liand Zhao [12] further augment deep learning models with vaguenessdetectiontechniquestoidentifyambiguousorinten- tionally misleading language, thereby improving transparency and trustworthiness in real-world deployments.

User privacy and personalization represent additional chal- lenges in modern fake news detection systems. Federated Learning (FL) enables local model training on user devices while sharing only aggregated updates with a central server, thus preserving user privacy and supporting personalized rec- ommendations [16], [17]. Although many existing systems focusprimarilyonbinaryfake–realclassification,theliterature demonstrates that federated learning is effective for scalable, privacy-aware news delivery [16].

Finally, human-centered factors such as critical thinking abilityandmedialiteracysignificantlyinfluencesusceptibility to misinformation. Alves and Costa [9] show that individual differences in analytical reasoning and media literacy affect the effectiveness of detection systems. Fernandez et al. [16] emphasize the importance of integrating fake news detection with recommendation systems to maintain content integrity whiledeliveringpersonalizeduserexperiences.Thesefindings support the development of hybrid frameworks that combine automated detection with user-adaptive mechanisms.

Overall, the surveyed literature reveals a clear evolution fromclassicalmachinelearningapproaches[5],[13],[14]
to transformer-based architectures [1], [6], [7], graph and geometric learning models [3], [8], multimodal fusion tech- niques[10],[11],[17],[18],[20],knowledge-drivenandinter- pretablesystems[12],[15],andprivacy-awarepersonalization strategies [16], [17]. This progression motivates the proposed hybrid framework that integrates high-capacity BERT-based fake news detection with privacy-conscious, user-centric per- sonalization to deliver trustworthy, intelligent, and adaptive news feeds.

II. LITERATURE REVIEW

Automaticfakenewsdetectiononsocialmediahasevolved into a multidisciplinary research domain encompassing nat- ural language processing (NLP), graph learning, multimodal representation, knowledge integration, and human-centered analysis. Early approaches primarily relied on traditional ma- chinelearningclassifiersusinghandcraftedlexical,syntactic, and stylistic features to distinguish deceptive content from legitimate information. Techniques such as Support Vector Machines, Decision Trees, and Random Forests demonstrated reasonable performance under controlled settings; however, their dependence on manual feature engineering limited ro- bustness across domains and rapidly evolving discourse pat- terns [5], [13], [14].

These limitations motivated the adoption of representa-tion learning methods capable of capturing deeper semantic and contextual signals. Deep neural networks, particularly transformer-based language models, have significantly ad- vanced text-based fake news detection. BERT-based architec- tures are widely used due to their bidirectional contextual embeddings and adaptability through fine-tuning. Ayyub etal. [6] propose a blended BERT framework that augments contextual embeddings with task-specific classification heads, improving robustness on noisy social media text. Farokhian and Rafe [7] further enhance contextual modeling by employ- ing dual BERT encoders to capture complementary semantic perspectives,yieldingimprovedgeneralization.Chiangetal.
[1] empirically compare traditional artificial neural networks with BERT-based models for source credibility detection, demonstrating the superior capability of transformer modelsin capturing subtle linguistic cues associated with deceptive content.

While transformer-based classifiers excel in semantic mod- eling, ensemble and hybrid architectures are frequently em- ployed to mitigate overfitting and enhance generalization. Al- shaqietal.[2]investigateensemblestrategiesthatcombinedi- verse base learners, showing improved detection performance across heterogeneous and adversarial datasets. Comparative surveys further advocate hybrid designs that integrate inter- pretable handcrafted features with learned representations to balance performance and explainability [5], [13].

Beyond textual analysis, several studies exploit social and structural signals by modeling user interactions and informa- tion propagation as graphs.

Graph Neural Networks (GNNs) leverage diffusion patterns, relational dependencies, and net- work topology to identify coordinated misinformation cam- paigns and anomalous spreading behaviors [3], [8]. These approaches provide complementary evidence to textual fea- tures and are particularly effective in distinguishing organic information diffusion from orchestrated manipulation.

Multimodal misinformation detection has gained promi- nence due to the increasing use of images, videos, and meta-dataalongsidetext.Cross-modaltransformerframeworkssuch asCroMealignvisualandtextualrepresentationsusingmetric or contrastive learning, improving detection accuracy when visualcontentsupportsorcontradictstextualclaims[10],[11]. Fuetal.[17] incorporate externalknowledgesourcesanduser interaction features into multimodal systems, demonstrating improvedreli abilitythroughfactualgrounding.Zhuetal.[20] introduce a shallow–deep multitask learning paradigm that jointly learns unimodal and cross-modal representations, further enhancing robustness.

Knowledgeintegrationandinterpretabilityareincreasingly recognized as critical components of reliable fake news detec-tionsystems.Knowledgegraph–basedmethodsvalidateclaims againstexternalfactsandprovideexplanatoryinsights,thereby reducing false positives where statistical correlations aloneare insufficient [15], [18]. Additionally, vagueness detection techniques identify hedging and ambiguous language patterns often used to mislead without making verifiable claims; inte- grating such cues with deep models enhances interpretability and assists human moderation [12].

Recent research also explores unsupervised and self- supervised learning to address the scarcity of labeled misin-formationdata.Structuralcontrastivelearningframeworksex- ploit propagation patterns and interaction graphs to learn dis- criminative representations from unlabeled streams, enabling early detection of emerging misinformation topics [19]. These methods complement supervised transformer-based pipelines and support semi-supervised or continual learning scenarios.

Human-centric studies provide insights into misinformation susceptibility and mitigation strategies. Alves and Costa [9] showthatcriticalthinkingskillsandnew-medialiteracysignif- icantly influence vulnerability to misinformation, highlighting the importance of combining automated detection with user- facing educational interventions. Fernandez et al. [16] advo- cate misinformation-aware recommender systems that adjust ranking and presentation strategies to reduce exposure to harmful content while preserving user engagement.

Finally, practical deployment considerations such as scal- ability, efficiency, and privacy are essential for real-world adoption. Recent works emphasize lightweight architectures and privacy-preserving or federated frameworks for personal- ization and on-device adaptation [16], [17]. Although explicit federatedlearningimplementationsarelimitedinthisliterature set,relatedresearchinformsthedesignofsystemsthatbalance personalization with user confidentiality.

Overall, the literature reveals a clear progression: (1) ashift from feature-engineered classifiers to deep contextual modelsfortextualunderstanding[1],[6],[7];(2)integra- tion of structural and multimodal signals through GNNs and cross-modal fusion [3], [8], [10], [11]; (3) incorporation of knowledge-based and vagueness-aware modules to improve factualgroundingand interpretability[12],[15],[18];and(4) emerging emphasis on privacy-aware and personalized detection strategies [16], [17]. Despite significant progress, most existing systems address only a subset of these dimen- sions, leaving a research gap for unified frameworks that simultaneouslyoffermultimodaldetection,explainability,per- sonalization, and privacy guarantees. The present work aimsto address this gap through a hybrid architecture that com- binescentralizedhigh-capacityclassificationwithlightweight, privacy-conscious personalization mechanisms.

III. COMPARISON BETWEEN MODELS

To provide a structured overview of existing approaches in fakenewsdetection,TableIsummarizesrepresentativemodels reported in prior research, highlighting their feature extrac- tionstrategiesandprivacyconsiderations.Thiscomparison facilitates a clearer understanding of prevailing trends and limitations in current systems and helps identify the research gap addressed by the proposed framework, which emphasizes accurate detection while preserving user privacy.

TableIpresentsacomparativeanalysisofcommonly used fake news detection models, including Artificial Neural Networks (ANNs), BERT-based architectures, Graph Neural Networks (GNNs), and multimodal frameworks. Traditional ANN-based models primarily rely on handcrafted textual features and demonstrate limited adaptability across domains. Transformer-based models such as BERT significantly im- prove detection accuracy by leveraging deep contextual em- beddings; however, they are typically trained in centralized settings that require access to large volumes of user-generated data, raising privacy concerns. Similarly, GNN-based ap-proaches effectively capture relational and propagation pat- terns within social networks but often depend on centralized aggregation of interaction data, which can expose sensitive user information.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

Recent studies have begun exploring federated deep learn- ing frameworks to address privacy challenges by enabling collaborative model training without sharing raw user data. While these approaches improve privacy preservation, they often focus on model training efficiency and may not fully exploithigh-capacitylanguagemodelsforfine-grainedcontent understanding.

The proposed hybrid system advances the state of the artby integrating a robust BERT-based fake news classifier with federated, on-device personalization mechanisms. This design combines the strong semantic modeling capability of trans-formerarchitectureswithprivacy-preservingdistributedlearn- ing,ensuringeffectivefakenewsdetectionwhilesafeguarding sensitive user data. As a result, the proposed approach repre- sentsacomprehensivesteptowardprivacy-aware,trustworthy, and personalized content recommendation systems.

TABLEI
COMPARISONOFDIFFERENTFAKENEWSDETECTIONMODELS

Model	FeatureExtraction
ANN&BERTHybrid	Contextualandsourcecredibilityembeddings
BlendedBERTModel	Linguisticandsemanticcontextualembeddings
DualBERTNetwork	Parallelcontextualrepresentations
GNN	User-postrelationalgraphembeddings
DNN	Deepsemanticandlexicalfeatures

IV. CONCLUSION

This survey highlights the rapid evolution of fake news detection research, tracing its progression from traditional feature-engineered approaches to advanced deep learning, multimodal, and graph-based architectures. Early studies pri- marily employed classical machine learning algorithms suchas Support Vector Machines, Decision Trees, and Random Forests, relying heavily on handcrafted linguistic and stylistic features.Althoughthesemethodsachievedinitialsuccess,their limitedcontextualunderstandingand poorcross-domaingen-eralization motivated the shift toward representation learning techniques.

The introduction of transformer-based models, particularly BERT and its variants, marked a significant advancement in fakenewsdetectionbyenablingdeepercontextualandseman- ticmodeling.WorkssuchasAyyubetal.[6]andFarokhianand Rafe[7] demonstratethat contextualembeddingssubstantially improve robustness and adaptability when dealing with noisy and informal social media text. Ensemble learning strategies and hybrid pipelines further enhance generalization and re- silience against domain shifts.

Recent research also emphasizes the importance of incor- porating multiple data modalities and relational structures. Multimodalframeworks[10],[11],[17],[20]integratetextual, visual, and contextual information to capture cross-modal inconsistencies and correlations, while graph neural networks and geometric deep learning approaches [3], [8] exploit social interactionandpropagationpatternstodetectcoordinatedmis- information campaigns. These findings indicate that effective misinformation detection requires understanding both content and its social dissemination dynamics.

In parallel, interpretability and factual grounding have emergedaskeypriorities.Knowledge-basedmethodsleverag- ing external knowledge graphs [15], [18], along with vague- ness detection techniques [12], enhance transparency and reliabilitybyaligningmodeldecisionswithverifiablefactsand explainable linguistic cues. Additionally, self-supervised and contrastive learning approaches [19] address label scarcity by enabling models to learn discriminative representations from unlabeled data streams.

Human-centered and ethical considerations further enrich this research landscape. Studies on media literacy and user behavior [9] underline that misinformation mitigation is not solely a technical challenge but also a socio-cognitive one. Misinformation-aware recommender systems and privacy- preserving personalization frameworks [16], [17] reflect a growingefforttobalancedetectionaccuracy,userengagement, and ethical data usage.

Overall, the surveyed literature reveals a clear trajectory toward integrated solutions. Future fake news detection sys-temsmustunifydeepcontextualmodeling,multimodalfusion, graph-based reasoning, and privacy-aware personalization to deliver scalable, interpretable, and user-centric solutions. The field is steadily advancing toward comprehensive frameworks that balance accuracy, transparency, and ethical responsibility in combating misinformation.

REFERENCES

[1] T. H. C. Chiang, C. S. Liao, and W. C. Wang, "Investigating thedifference of fake news source credibility recognition between ANNand BERT," *Journal of Information Technology and Media Studies*,vol. 9, no. 2, pp. 45–52, 2022.

[2] M. Al-shaqi, D. B. Rawat, and C. Liu, "Ensemble techniques for robustfake news detection," in *Proceedings of the International Conferenceon Machine Learning and Applications (ICMLA)*, pp. 1–8, 2024.

[3] Y. Han, S. Karunasekeran, and C. Leckie, "Graph neural networks withcontinual learning for fake news detection from social media," *IEEETransactions on Knowledge and Data Engineering*, vol. 32, no. 10, pp.2046–2059, 2020.

[4] G. Guler and S. Demirci, "Deep learning based fake news detection onsocialmedia,"InternationalJournalofComputerApplications,vol.180,no. 34, pp. 1–7, 2023.

[5] D.Singh,H.Panwar,andS.Chaturvedi,"Designingfakenewsdetectionsystem using machine learning techniques," International Journal ofAdvanced Research in Computer Science, vol. 16, no. 5, pp. 102–108,2025.

[6] S. Ayyub, A. Khan, and R. Malik, "BERT-based blended approach forfake news detection," Journal of Computational Linguistics Research,vol. 15, no. 3, pp. 215–223, 2021.

[7] M.FarokhianandV.Rafe,"FakenewsdetectionusingdualBERTdeepneuralnetworks,"NeuralComputingandApplications,vol.35,no.9,

[8] pp.7511–7520,2023.

[9] L. Zhang, Y. Liu, and P. Chen, "Fake news detection on social mediausing geometric deep learning," Expert Systems with Applications, vol.237, pp. 121–130, 2024.

[10] N. Alves and M. Costa, "Fake news detection on social media: Thepredictive role of university students' critical thinking dispositions andnew media literacy," Computers in Human Behavior, vol. 136, 2022.

[11] S. Patel, A. Nair, and J. Lee, "CroMe: Multimodal fake news detectionusing cross-modal tri-transformer and metric learning," IEEE Access,vol. 11, pp. 134056–134069, 2023.

[12] R. Kumar and V. Menon, "Multimodal fake news detection with con-trastivelearningandoptimaltransport,"InformationFusion,vol.103,

[13] pp.18–29,2024.

[14] H. Li and X. Zhao, "Combining vagueness detection with deep learningto identify fake news," Journal of Artificial Intelligence Research, vol.76, pp. 310–326, 2023.

[15] P.ReddyandM.Thomas,"Acomparativestudyofmachinelearninganddeep learning techniques for fake news detection," Procedia ComputerScience, vol. 197, pp. 512–520, 2022.

[16] D. Bhattacharya and S. Kumar, "Detection of fake news using machinelearning and natural language processing algorithms," Applied Intelli-gence, vol. 54, pp. 4205–4219, 2024.

[17] E.GarzaandM.Rodr´ıguez,"Combiningmachinelearningwithknowl-edge engineering to detect fake news in social networks—A survey,"Knowledge-Based Systems, vol. 233, pp. 107–125, 2021.

[18] A. Fernandez, S. Iqbal, and R. Green, "Advancing MisinformationAwareness in Recommender Systems for Social Media InformationIntegrity," Information Processing and Management, vol. 62, no. 4,pp.101–115, 2025.

[19] L. Fu, J. Zhang, H. Li, and X. Wang, "Multimodal FakeNews DetectionIncorporating External Knowledge and User Interaction Feature,"WileyOnline Library, 2023.

[20] T.-V. La, M.-H. Nguyen, and M.-S. Dao, l fake news detection," arXivpreprint,2025.

[21] X. Chen, L. Zhou, and Y. Li, "An unsupervised fake news detectionframeworkbasedonstructuralcontrastivelearning,"Cybersecurity,vol.8, no. 1, 2025.

[22] Y. Zhu, Y. Wang, and Z. Yu, "Multimodal fake news detection: MFNDdataset and shallow-deep multitask learning," in Proceedings of theInternational Joint Conference on Artificial Intelligence (IJCAI), pp.891–899, 2025.