https://doi.org/10.54361/ajmas.258438

Original article

From Tweets to Insights: Enhancing Summarization through Social Interaction Context

Anwar Alhenshiri*, **Hanan Suliman**, **Hoda Badesh**

Department of Computer Science, Faculty of Information Technology, Misurata University, Misurata, Libya
Corresponding Email. alhenshiri@it.misuratau.edu.ly

Abstract
Social media platforms, particularly Twitter (now X), generate massive volumes of short and fragmented posts that pose significant challenges for information retrieval and knowledge extraction. Traditional tweet summarization methods rely primarily on the content of individual tweets, often producing incomplete and shallow summaries that fail to capture the broader discourse context. This study introduces a comment-aware summarization framework that integrates user replies and interactions into the summarization process, aiming to enhance informativeness, coherence, and contextual richness. A dataset of tweets and associated comments was manually collected and preprocessed. Two summarization interfaces were developed: (1) a baseline tweet-only system and (2) an experimental comment-aware system that combines tweets with semantically relevant, engagement-weighted comments. Both systems employed TF–IDF vectorization, cosine similarity, and the TextRank algorithm for extractive summarization. A within-subjects user study with 32 participants compared the two systems across quantitative metrics (task completion time, number of queries, correctness of results) and qualitative dimensions (satisfaction, credibility, ease of use, diversity of perspectives). Results showed that the comment-aware system required slightly longer task times but significantly reduced the number of queries ($p < 0.001$) and improved accuracy ($p < 0.01$). Participants consistently rated the comment-aware summaries as more informative and reflective of diverse perspectives. These findings highlight the value of socially enriched summarization in overcoming the limitations of short-form content and provide design recommendations for next-generation information retrieval systems.
Keywords. Human Computer Interaction, Information Retrieval, User Tasks, Social Networks.

Introduction

The exponential growth of social media platforms has transformed the ways in which individuals communicate, exchange information, and engage in public discourse. Among these platforms, Twitter (rebranded as X) occupies a central role as a microblogging service characterized by brevity, immediacy, and virality. Its defining features—short-form content, real-time updates, and rapid diffusion—have made it an influential medium for news dissemination, political debate, crisis response, marketing, and cultural trends. However, these same features exacerbate the problem of information overload, as millions of tweets are generated hourly, producing a fragmented and overwhelming stream of User-Generated Content (UGC). This large volume often hinders efficient access to relevant insights and poses significant challenges for information retrieval [1].

To address this challenge, automatic text summarization has emerged as a critical research domain. Summarization techniques aim to condense large volumes of text into concise and coherent summaries without losing essential meaning, thereby facilitating faster consumption and supporting downstream tasks such as sentiment analysis, event detection, and misinformation tracking. Existing approaches generally fall into two categories: extractive summarization, which selects salient words, phrases, or sentences from the original text, and abstractive summarization, which generates new sentences that paraphrase and reframe the source material. While extractive methods have historically dominated tweet summarization due to the short and fragmented nature of tweets, recent advances in transformer-based architectures and large language models (LLMs)—including BERT, GPT, and T5—have enabled more fluent and context-aware abstractive techniques [2, 3].

Despite these advancements, most existing approaches treat tweets as isolated units, ignoring the broader conversational context in which they are embedded. Tweets are rarely self-contained; they are typically accompanied by replies, retweets with comments, and reactions that provide elaborations, clarifications, counterarguments, or emotional responses. The omission of this interactional context leads to shallow summaries that fail to capture the richness of online discourse. Recent work in news and forum summarization has demonstrated the benefits of integrating comments into summarization processes, improving informativeness, coherence, and diversity of perspectives [4, 3]. However, systematic application of comment-aware summarization to Twitter remains limited, with most studies either disregarding replies or treating them as noise.

The present study addresses this gap by proposing and empirically evaluating a comment-aware tweet summarization framework. The framework integrates user interactions—specifically replies and comments—into the summarization process to generate outputs that are not only concise and coherent but also contextually enriched and socially grounded. To this end, two prototype systems were implemented: a baseline tweet-only summarization interface and an experimental comment-aware interface. Both systems

Received*: 17-08-2025 -* ***Accepted****: 14-10-2025 -* ***Published****: 22-10-2025*

employed TF–IDF vectorization, cosine similarity, and the TextRank algorithm for extractive summarization, but the latter additionally incorporated semantically relevant and engagement-weighted comments.
The study was evaluated through a within-subjects user experiment with 32 participants. Results demonstrated that while the comment-aware system required slightly longer task completion times, it significantly reduced the number of queries and improved retrieval accuracy. Participants also reported higher satisfaction, credibility, and perceived diversity of perspectives with the enriched summaries. These findings underscore the potential of socially enriched summarization systems to overcome the limitations of short-form content and inform the design of next-generation information retrieval (IR) platforms.
The following section situates this research within the broader evolution of information retrieval and text summarization**.** It first reviews foundational models in IR and their transition toward neural and transformer-based approaches. It then surveys the progression of summarization techniques—from statistical and extractive methods to deep learning and hybrid models—before focusing on comment-aware and socially enriched frameworks. This contextualization clarifies how the present study extends prior work by operationalizing social interaction signals within tweet summarization and empirically assessing their impact on retrieval effectiveness.
The remainder of this paper is structured as follows. Section 2 reviews related work on text summarization and comment-aware approaches. Section 3 outlines the methodology, including dataset construction, preprocessing, and the proposed framework. Section 4 describes the empirical study, including participants, design, and procedures. Section 5 presents results, while Section 6 discusses their implications. Finally, Section 7 concludes with key contributions, limitations, and directions for future research.
Building upon the challenges outlined in the introduction, this section reviews the foundational and contemporary research that informs the present study. It traces the evolution of information retrieval and text summarization methods, with a particular focus on their adaptation to social media contexts and the emerging role of user interactions in enhancing summary quality.

Information Retrieval Foundations

Information Retrieval (IR) has undergone a profound transformation from keyword-based indexing systems to advanced, context-aware models capable of handling heterogeneous and large-scale data. Classical approaches such as the Boolean model and the Vector Space Model (VSM) established early foundations but were limited in their ability to capture semantic relationships. Probabilistic models such as BM25 [5] improved retrieval precision by incorporating term frequency and document length normalization, and continue to serve as baselines in modern IR systems.
The advent of neural architectures introduced more sophisticated approaches. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) enabled representation learning for textual relevance tasks [6]. More recently, transformer-based models [7], exemplified by BERT and GPT, have revolutionized IR by capturing contextual semantics and enabling fine-tuned retrieval and ranking [8]. In the social media domain, where brevity and informality complicate retrieval, transformer-based architectures have proven particularly effective in bridging lexical gaps and modeling semantic similarity [9].

Evolution of Text Summarization

Early approaches to text summarization (TS) relied on statistical and feature-based methods. Sentence scoring based on term frequency, positional features, and linguistic cues was widely used [10, 11]. With the rise of machine learning, extractive techniques increasingly incorporate supervised classifiers trained on sentence-level features [12].
The introduction of deep learning and attention mechanisms brought a major shift toward abstractive summarization. Transformer-based architectures such as T5, BART, and GPT demonstrated state-of-the-art performance across summarization benchmarks [13, 14]. Hybrid models that combine extractive precision with abstractive fluency emerged as promising solutions, particularly in noisy, short-text domains such as social media.

Summarization in Social Media Contexts.

Summarizing microblogging platforms such as Twitter presents unique challenges: brevity, informality, multilingualism, and context fragmentation. Early microblog summarization focused on clustering and ranking tweets around events [15, 16]. More recent approaches leveraged embedding-based semantic similarity [17] and contextual models to improve coherence in summaries.
Domain-specific advancements include ATSumm, an abstractive summarizer incorporating auxiliary information to handle sparse disaster-related data [18], and LaMSUM, a multi-level extractive framework using LLMs [19]. Real-time methods, such as Dirichlet score classification for streaming tweets [18], highlight the importance of timeliness in summarization. Additionally, fairness-aware approaches such as FairExtract and FairGPT [19] have sought to mitigate representational bias in summaries.

Received*: 17-08-2025 -* ***Accepted****: 14-10-2025 -* ***Published****: 22-10-2025*

Comment-Aware and Socially Enriched Summarization
A growing body of work emphasizes the importance of user interactions in improving summarization quality. Social search studies have shown that incorporating social signals [20] improves relevance and personalization. In forum and news contexts, integrating user comments enhances coverage and informativeness [21, 22]. On Twitter, contextual enrichment through replies, retweets, and linked articles has demonstrated benefits for misinformation tracking and event detection [23].
Recent advances also include knowledge graph–augmented summarization [20] and retrieval-augmented generation (RAG) with LLMs [13], both of which demonstrate the potential of external and contextual information in improving summary quality. However, systematic efforts to incorporate comment-aware strategies in tweet summarization remain limited, with most approaches treating replies as noise rather than as valuable sources of elaboration and perspective.

Research Gap
The literature demonstrates clear progress from feature-based extractive methods toward context-aware, semantically grounded, and socially enriched approaches. Yet, in the specific domain of Twitter summarization, most techniques remain narrowly focused on tweets themselves, neglecting the broader conversational ecosystem. This omission results in summaries that are factually concise but contextually shallow. The present research addresses this gap by systematically integrating user comments into tweet summarization, empirically evaluating their impact on retrieval effectiveness, and providing design recommendations for next-generation IR systems.

Methodology
Research Design
This study adopted a comparative experimental design to evaluate the effectiveness of a comment-aware tweet summarization framework against a baseline tweet-only system. Both systems were developed using extractive summarization techniques, specifically TF–IDF vectorization, cosine similarity, and the TextRank algorithm. The baseline system summarized tweets in isolation, whereas the experimental system integrated tweets with semantically relevant and engagement-weighted comments. A within-subjects user study was conducted to assess the performance of the two systems across quantitative and qualitative measures.

Dataset Collection
Due to restrictions associated with the Twitter API, a manual data collection approach was employed. Tweets and their associated user comments were harvested and stored in structured CSV format. The dataset focused on highly topical and widely discussed events, ensuring that the retrieved material was both meaningful and contextually rich.
Although smaller in scale compared to automated API-based datasets, this collection method produced a realistic, interaction-rich corpus suitable for evaluating the role of user comments in summarization.

Preprocessing
Given the noisy and informal nature of social media text, a structured preprocessing pipeline was applied to both tweets and comments. Metadata such as retweet markers, user mentions, and hyperlinks were removed, and compound hashtags were normalized into plain words. Text was then lowercased, tokenized with NLTK, and stripped of common stop-words. Finally, stemming was applied to reduce lexical variation. This sequence produced clean, semantically enriched tokens suitable for retrieval and summarization.

Summarization Pipelines
Baseline: Tweet-Only Summarization
Each tweet was treated as a separate document. Using TF–IDF vectorization, tweets were represented numerically, and cosine similarity was used to retrieve tweets relevant to user queries. Extractive summarization was performed via TextRank, which ranks sentences within the retrieved set and assembles the top-ranked sentences into a coherent summary.

Experimental: Comment-Aware Summarization
In the experimental system, each tweet was enriched with a subset of top-ranked comments. Selection was based on three factors: semantic similarity to the original tweet (computed using Sentence-BERT embeddings), engagement signals (normalized likes and replies), and a weighted scoring function:

$$S = \alpha \cdot \text{Semantic Similarity} + (1-\alpha) \cdot (\beta \cdot \text{Likes}_{norm} + \gamma \cdot \text{Replies}_{norm})$$

with α=0.5\, β=0.6, and γ=0.4. This ensured that the chosen comments were both contextually relevant and socially significant. The top three comments were selected and merged with the corresponding tweet, yielding a richer input text. This enriched content was then processed through the same TextRank summarization pipeline as the baseline, enabling direct comparability between the two systems. The summarization process is depicted in (Figure 1).

***Received**: 17-08-2025 - **Accepted**: 14-10-2025 - **Published**: 22-10-2025*

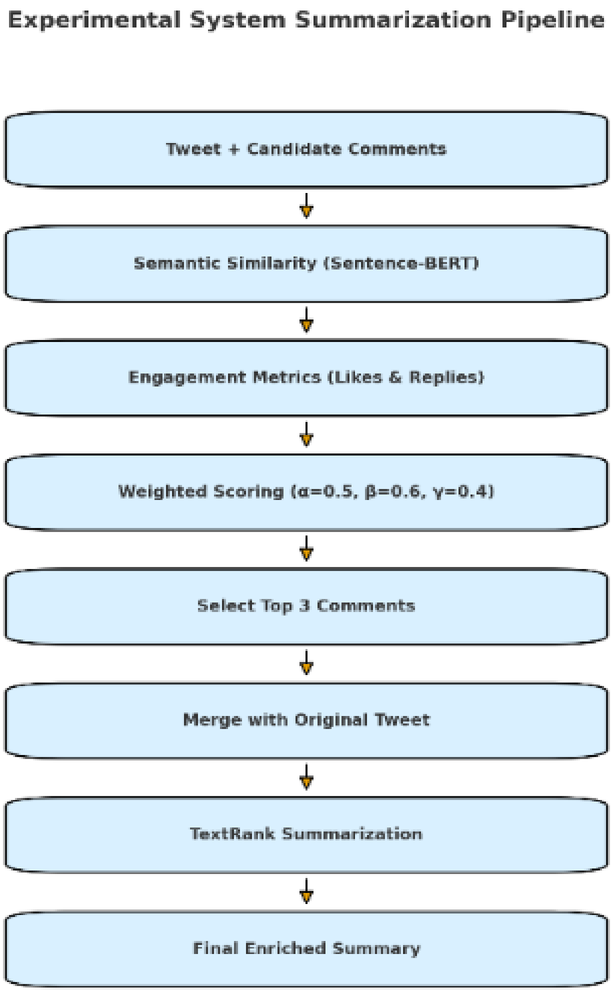

Figure 1. The Summarization Process

Evaluation Metrics

System performance was evaluated through both quantitative and qualitative measures. Quantitative metrics included task completion time, number of queries, accuracy of retrieved results, and standard summarization scores (ROUGE, BLEU, precision, recall, F1). Qualitative evaluation was based on a Likert-scale questionnaire covering usability and credibility, supplemented by open-ended feedback on system strengths and weaknesses.

Summary of Methodology

The methodological framework established a controlled comparison between tweet-only and comment-aware summarization systems. By standardizing preprocessing, retrieval, and summarization pipelines, and differing only in the integration of comments, the study was able to isolate the effect of user interactions on summarization performance. This dual evaluation—quantitative and qualitative—ensured a comprehensive assessment of both technical effectiveness and human-centered usability.

Results

The comparative evaluation of the tweet-only baseline and the comment-aware system is presented in two parts: quantitative system performance and qualitative user feedback. Together, these analyses provide a comprehensive understanding of the trade-offs associated with integrating comments into tweet summarization.

Quantitative Results

Task Completion Time

Descriptive statistics revealed that the baseline system yielded a mean task completion time of 168.16 seconds (SD = 32.85), whereas the comment-aware system averaged 174.72 seconds (SD = 30.53). Although the difference was modest, statistical analysis indicated significance. Both a paired t-test ($t(31) = -2.18$, $p = 0.037$) and a Wilcoxon signed-rank test ($p = 0.045$) confirmed that the comment-aware condition generally required longer task execution.

This outcome reflects the additional cognitive load imposed by processing enriched content. While tweets alone provide brevity, tweets plus comments introduce more text, which users must evaluate before identifying relevant material.

***Received**: 17-08-2025 - **Accepted**: 14-10-2025 - **Published**: 22-10-2025*

Number of Queries Submitted
The average number of queries submitted per task decreased significantly with the experimental system, from 3.16 queries (baseline) to 2.19 queries (comment-aware). Statistical analysis confirmed this reduction was highly significant (paired t-test: $t(31) = 4.65$, $p < 0.001$; Wilcoxon signed-rank: $p < 0.001$).
This result suggests that the integration of comments reduced user effort by providing additional lexical and semantic cues, thereby decreasing the need for query reformulation.

Number of Correct Results Retrieved
Participants retrieved an average of 6.53 correct results with the baseline system compared to 7.19 correct results with the comment-aware system. The improvement was statistically significant (paired t-test: $t(31) = -3.48$, $p = 0.0015$; Wilcoxon signed-rank: $p = 0.002$).
The consistent gain in accuracy reflects the role of comments in providing clarifications, elaborations, and alternative perspectives, which enhanced the relevance of retrieved summaries.

Summary of Quantitative Findings
The quantitative analysis highlights a clear trade-off: the comment-aware system improved accuracy and reduced query effort, but at the cost of slightly longer completion times. The quantitative results are summarized in (Table 1).

Table 1. Quantitative Results

Measure	Baseline System	Comment-Aware System	Statistical Significance
Task Completion Time (sec)	168.16	174.72	$p = 0.037$ (t-test), $p = 0.045$ (Wilcoxon)
Number of Queries Submitted	3.16	2.19	$p < 0.001$ (t-test & Wilcoxon)
Correct Results Retrieved	6.53	7.19	$p = 0.0015$ (t-test), $p = 0.002$ (Wilcoxon)

Qualitative Results
Participants' subjective evaluations reinforced the quantitative findings. Across all six survey dimensions, the comment-aware system received higher levels of agreement:
*Speed of reaching relevant results***:** 63% agreement (baseline) vs. 72% (comment-aware).
*Satisfaction with results***:** 68.8% vs. 81.2%.
*Credibility of content***:** 65.6% vs. 78.1%.
Alignment with information needs: 71.9% vs. 78.1%.
Ease of completion: 62.5% vs. 68.8%.
Diversity of perspectives: 59.4% vs. 78.1%.
While none of the differences reached statistical significance at the 5% threshold, the consistent upward trend favored the comment-aware system. The strongest improvement was observed in diversity of perspectives, suggesting that participants valued the contextual richness provided by comments. The results are shown in (Table 2) for further clarification.

Table 2. Qualitative Results

Survey Dimension	Baseline System (%)	Comment-Aware System (%)
Speed of reaching relevant results	63.0	72.0
Satisfaction with results	68.8	81.2
Credibility of content	65.6	78.1
Alignment with information needs	71.9	78.1
Ease of completion	62.5	68.8
Diversity of perspectives	59.4	78.1

Discussion
The findings of this study demonstrate the dual impact of integrating user comments into tweet summarization: while accuracy and user satisfaction improved, this was counterbalanced by a moderate increase in task completion time. The consistent improvement in correctness ($6.53 \rightarrow 7.19$) and reduction in query effort ($3.16 \rightarrow 2.19$) confirm that user comments provide valuable contextual resources, supplying elaborations, fact-checks, and emotional signals that tweets alone cannot convey. At the same time, the statistically significant increase in completion time indicates that richer input increases cognitive load, reflecting a fundamental tension in information retrieval between brevity and contextual depth.
These results suggest that comment-aware summarization can substantially enhance effectiveness in retrieval tasks where accuracy and comprehensiveness are prioritized over speed, such as crisis response,

***Received**: 17-08-2025 - **Accepted**: 14-10-2025 - **Published**: 22-10-2025*

misinformation tracking, or social research. In contrast, for time-sensitive contexts like live event monitoring, the efficiency costs may outweigh accuracy gains unless optimized filtering mechanisms are introduced. Participants' qualitative feedback further revealed that the perceived diversity of perspectives was the most valued contribution of comments, aligning with broader social computing research showing that users often seek multiple viewpoints, not just factual information, when navigating online discourse. However, some participants also noted frustration with the increased reading burden, emphasizing the need for interfaces that present comment-derived insights selectively or visually.
These findings complement prior studies demonstrating the benefits of context-enriched summarization in forums and news domains [1, 3]. Unlike earlier work, this study provides a systematic, controlled comparison between tweet-only and comment-aware summarization in a real-world setting, reinforcing calls for socially enriched information retrieval systems while quantifying the trade-offs associated with such enrichment.

Limitations
Several limitations must be acknowledged. First, the dataset was manually collected and relatively small, limiting generalizability. Second, the summarization approach was extractive; abstractive methods leveraging large language models (LLMs) may produce more fluent and efficient summaries. Third, the study population, though relevant, was not demographically representative.

Conclusion
This study proposed a comment-aware tweet summarization framework that enriches traditional tweet-only approaches by incorporating user replies and interactions. A study with 32 participants showed that while including comments slightly increased task completion time, it reduced the number of queries and improved summary accuracy, credibility, and informativeness. Participants highlighted those comments often added missing details, clarified confusing tweets, and provided diverse perspectives. The work contributes a novel framework for integrating comments, empirical evidence of trade-offs between efficiency and accuracy, and design guidance for balancing brevity with contextual richness. Future research should validate the framework on larger and more diverse datasets, explore abstractive summarization with LLMs for fluent and semantically nuanced outputs, implement adaptive comment filtering to optimize efficiency, and extend the approach to multilingual and multimodal contexts. In sum, comment-aware summarization bridges tweet-level brevity with community-level meaning, enhancing microblog information retrieval and supporting the development of next-generation, socially responsive systems.

Conflict of interest. Nil

References
1. Garg S, Chakraborty T, Dandapat S. ATSumm: An abstractive summarizer for disaster-related tweets using auxiliary information. Inf Process Manag. 2024;61(2):103280. doi:10.1016/j.ipm.2023.103280.
2. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT; 2019; Minneapolis, MN. p. 4171–4186. doi:10.48550/arXiv.1810.04805.
3. Zhang Y, Li H, Wang J. Multi-perspective comment integration for news summarization. Inf Retr J. 2024;27(2):135–158. doi:10.1007/s10791-023-09456-7.
4. Bhatia P, Dhingra B, Singh M, Joshi A. Comment-aware summarization for news and online forums: Improving informativeness through social context. J Comput Linguist. 2023;49(2):287–305. doi:10.1162/coli_a_00456.
5. Robertson S, Zaragoza H. The probabilistic relevance framework: BM25 and beyond. Found Trends Inf Retr. 2009;3(4):333–389. doi:10.1561/1500000019.
6. Mitra B, Craswell N. An introduction to neural information retrieval. Found Trends Inf Retr. 2018;13(1):1–126. doi:10.1561/1500000061.
7. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Polosukhin I. Attention is all you need. In: Advances in Neural Information Processing Systems (NeurIPS); 2017. p. 5998–6008.
8. Chen Z. Abstractive summarization with transformer-based architectures: A comprehensive survey. ACM Comput Surv. 2022;55(3):1–35. doi:10.1145/3512785.
9. Hightower S. Transformer-based models in information retrieval: Trends, challenges, and opportunities. Found Trends Inf Retr. 2025;19(1):1–64. doi:10.1561/1500000099.
10. Delort JY, Bouchon-Meunier B, Yvon F. Description and evaluation of the LIA text summarization system at DUC-2003. In: Proceedings of HLT-NAACL Text Summarization Workshop; 2003; Edmonton, Canada. p. 1–8.
11. Nagao K, Hasegawa T, Imai K. Text summarization based on rhetorical structure extraction. In: Proceedings of the International Conference on Computational Linguistics; 1998; ACL. p. 703–707.
12. Nagy A, Novák A, Prószéky G. Automatic sentence-level extractive summarization using supervised machine learning. Acta Cybern. 2009;19(2):315–329.

Received: *17-08-2025* - ***Accepted***: *14-10-2025* - ***Published***: *22-10-2025*

13. Li J, Zhang Y, Chen X. Retrieval-augmented generation for contextualized summarization: A case study in social media. Trans Assoc Comput Linguist. 2025;13:22–37. doi:10.1162/tacl_a_00630.
14. Karanja J, Matheka D. Advances in abstractive summarization: From recurrent neural networks to transformers. Int J Artif Intell. 2022;20(4):87–106.
15. Inouye D. Summarizing microblogs for event detection: A statistical approach. In: Proceedings of the International Conference on Social Computing; 2010; IEEE. p. 120–127.
16. Ramanujam E, Krishnan R, Kumar S. Event summarization in Twitter using clustering techniques. Int J Inf Technol. 2016;8(4):235–244.
17. Reimers N, Gurevych I. Sentence-BERT: Sentence embeddings using Siamese BERT networks. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing; 2019; ACL. p. 3980–3990. doi:10.48550/arXiv.1908.10084.
18. Jin X, Li Y, Sun L. Real-time tweet summarization using Dirichlet score classification. In: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval; 2024. p. 1602–1611. doi:10.1145/3626772.3654390.
19. Nezhad MH, Tjoa E, Feng Y. FairExtract and FairGPT: Addressing fairness in summarization. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2024;38(5):4321–4328. doi:10.1609/aaai.v38i5.29471.
20. Morris MR, Teevan J. Exploring the value of social search. In: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems; 2012. p. 1825–1834. doi:10.1145/2207676.2208327.
21. Ramanujam E, Krishnan R, Kumar S. Event summarization in Twitter using clustering techniques. Int J Inf Technol. 2016;8(4):235–244.
22. Chhikara M, Chhikara P, Gupta R. LaMSUM: Multi-level extractive summarization of long-form text using large language models. Inf Process Manag. 2024;61(1):103265. doi:10.1016/j.ipm.2023.103265.
23. Dhokar M, Jain R, Kumar A. Leveraging social context for misinformation detection in microblogs. Soc Netw Anal Min. 2023;13(24):1–17. doi:10.1007/s13278-023-00987-3.

***Received**: 17-08-2025 - **Accepted**: 14-10-2025 - **Published**: 22-10-2025*