Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 25 May 2026
Revised 26 June 2026
Accepted 21 July 2026

Natural Resources for Human Health 2026, Vol. 6 (9s): 715–722

KEYWORDS:
Natural Language Processing, Large Language Models, ChatGPT, Education Technology, Sentiment Analysis, Misinformation Detection, Transformer Architecture, BERT, Digital Content Analysis.

Natural Language Processing and Large Language Models: AI for Education, Communication, and Digital Content Analysis

Dr.R. Rakesh[1], B. Lakshmi Lavanya[2], Dr.S. Selvaraj[3], M Murali Krishnan[4], Dr. Ganesh Kumar R[5], Dr. Navitha S Nair[6]

[1]Assistant Professor, Department of Commerce, Faculty of Science & Humanities, SRMIST, Ramapuram Campus, Chennai – 89, rakeshr1@srmist.edu.in, r.rakesh466@gmail.com
[2]Assistant Professor, CSE- AI & ML, Geethanjali College of Engineering and Technology Medchal-Malkajgiri, Hyderabad, Telangana, blavanya.cse@gcet.edu.in
[3]Assistant professor, School of Science and Computer Studies, CMR University Bangalore, selvarasu.s@cmr.edu.in, https://orcid.org/0009-0007-0463-4501
[4]Assistant Professor, Management Studies, Dr. N.G.P. Institute of Technology, COIMBATORE, TAMIL NADU, msmh18734@gmail.com
[5]Professor, Department of Computer Science and Engineering, CHRIST (Deemed to be University), Bangalore, Karnataka
[6]Assistant Professor, Department of Education, Sastra University, Thanjavur, Tamil Nadu

ABSTRACT: Natural language processing (NLP) has progressed from narrow, task-specific statistical models to large language models (LLMs) capable of generating fluent, contextually coherent text across virtually unlimited domains, a transition that has reshaped three previously distinct application areas simultaneously: education, human communication, and digital content analysis. This paper reviews the recent empirical literature on transformer-based NLP and LLM applications across these three domains, synthesizing systematic reviews of LLM integration in language-teacher education and broader educational settings with the transformer-architecture literature on sentiment analysis, sarcasm detection, and misinformation identification in social media content. Particular attention is given to the specific mechanisms, intelligent tutoring, personalized feedback generation, bidirectional contextual encoding, and pretrained-language-model fine-tuning, that the reviewed studies identify as driving reported performance gains, alongside the equity, reliability, and academic-integrity concerns that temper these gains. Comparative tables map LLM educational applications to their reported outcomes and adoption contexts, cross-reference transformer-based content-analysis tasks against the architectures and accuracy levels each study reports, and set documented risks and limitations against the mitigation strategies the literature proposes. The review finds that LLMs and transformer-based NLP models now achieve consistently strong empirical performance, frequently exceeding 90% accuracy or F1-score on sentiment, sarcasm, and misinformation-detection benchmarks, and demonstrate measurable educational benefit through intelligent tutoring and personalized feedback, but that reliability, bias, temporal drift, and unequal access to these tools remain substantial and only partially resolved challenges across all three application domains.

I. INTRODUCTION

The release of ChatGPT in November 2022 marks the point at which large language models moved from a specialized research topic within natural language processing to a globally adopted general-purpose technology, with documented application across education, healthcare, communication, and digital media analysis within an unusually short span of time. ChatGPT is an AI-based large language model trained on massive text datasets in multiple languages with the ability to generate human-like responses to text input, built on the Generative Pre-trained Transformer architecture, which utilizes a neural network to process natural language and generate responses based on the context of input text [1]. This underlying architectural capability, contextual, multi-domain text generation from a single pretrained model, explains why LLM research has diffused so rapidly across application areas that previously required separate, purpose-built NLP systems.

In education specifically, this diffusion has been extensively documented. A systematic review of large language models in education, examining empirical applications, benefits, and challenges, found that LLMs offer key educational benefits, including enhanced motivation through personalized learning, expanded accessibility for remote learners, efficient resource optimization via streamlined assessment and content creation, and potential for developing critical thinking and self-regulated learning capabilities [2]. The same review reported that these technologies demonstrate improved student performance, with enhanced test scores and faster task completion across language learning, programming, and mathematics, while noting that a complementary systematic review of 88 empirical studies since ChatGPT's release found that intelligent tutoring systems are the most common educational application of LLMs, with overall evidence for improved performance and engagement alongside concerns about over-reliance and reliability [2].

Beyond education, transformer-based NLP models have simultaneously advanced the digital content analysis tasks central to understanding and moderating online communication. Bidirectional Encoder Representations from Transformers (BERT) has become foundational to this second domain: BERT performs exceptionally well on various NLP and sequence-to-sequence-based language generation-related tasks such as question answering, abstract summarization, sentence prediction, conversational response generation, and sentiment classification, and by December 2019 had already been rolled out in over 70 languages after initially being limited to English [3]. This paper reviews the LLM-in-education literature alongside the transformer-based digital-content-analysis literature within a single comparative framework, examining how a shared underlying architecture has produced measurable but domain-specific benefits and risks across educational, communicative, and content-analytic applications.

Research Objectives

- To review empirical evidence on large language model applications in education, including intelligent tutoring, personalized feedback, and teacher professional development.
- To synthesize transformer-based NLP research on sentiment analysis, sarcasm detection, and social media content classification.
- To examine the misinformation-detection literature applying pretrained language models to digital content analysis.
- To evaluate the equity, reliability, and integrity concerns the reviewed literature identifies as limiting factors across educational and content-analysis applications.
- To identify future research priorities for NLP and LLM applications spanning education, communication, and digital content analysis.

II. LITERATURE REVIEW

2.1 Large Language Models in Language Teacher Education and Instructional Practice

A substantial body of recent research has examined how LLMs, and ChatGPT specifically, are integrated into language teacher education and instructional practice. A systematic narrative literature review mapping the emerging research landscape on ChatGPT applications in language teacher education, analyzing thirteen peer-reviewed empirical studies published between 2023 and 2025, examined the integration of ChatGPT in lesson planning, professional development, student engagement, and teacher identity construction, and identified four major thematic domains: instructional applications and teaching support, teacher learning and professional reflection, student outcomes and affective development, and ethical, pedagogical, and

contextual considerations [4]. This four-domain structure is analytically useful because it demonstrates that LLM integration into education operates simultaneously at the level of direct instructional tool, professional-development resource, and object of ethical deliberation, rather than functioning as a single, uniform intervention.

Complementing this English-language-teaching-specific evidence, the broader systematic review of empirical LLM applications in education found that recent literature reveals a clear evolution in research focus, shifting from broad terms such as "AI" toward more specific concepts including "ChatGPT," "large language models," and "personalized learning," with ChatGPT emerging as a versatile tool across teaching, learning, assessment, and academic operations [2]. The same review cautioned, however, that GenAI is increasingly recognized as a double-edged sword in education, offering transformative opportunities while introducing significant risks, and specifically noted that while academic performance benefits are notable, they can be compromised depending on usage approaches, since students who accept AI-generated responses without reflection may derive substantially less educational benefit than those who engage critically with LLM output [2].

2.2 Equity and Access Considerations in Educational LLM Deployment

A distinct strand of the reviewed education literature examines whether LLM-based educational tools reduce or exacerbate existing disparities in educational opportunity. Research investigating real-world educational inequalities introduced by large language models observed that LLMs can help alleviate the scarcity of educational resources and create more personalized learning experiences through applications such as creating curricular material, providing tailored feedback, and facilitating discussion, framing LLM adoption as a potential equity-enhancing intervention specifically for resource-constrained educational settings [5]. The same research identified a focal issue surrounding educational application of LLMs directly relevant to this equity framing: what these models mean for closing, or potentially widening, existing disparities in opportunity, experience, and achievement across student populations with unequal access to LLM tools, unequal digital literacy, or unequal institutional support for responsible use [5]. This equity concern parallels, at the level of individual students, the broader access-and-benefit-distribution questions that recur throughout the technology-adoption literature more generally.

2.3 LLMs in Specialized Educational Domains: Medical and Health Education

LLM educational applications extend well beyond general and language education into specialized professional-training domains, most extensively documented in medical education. A systematic scoping review of large language models, including ChatGPT, in medical education found that ChatGPT exhibits various potential applications such as providing personalized learning plans and materials, creating clinical practice simulation scenarios, and assisting in writing articles, while emphasizing that challenges associated with academic integrity, data accuracy, and potential harm to learning were also highlighted in the literature [6]. Based on this review, three key research priorities were proposed: cultivating the ability of medical students to use ChatGPT correctly, integrating ChatGPT into teaching activities and processes, and proposing standards for the use of AI by medical students, a three-part research agenda that generalizes readily beyond medical education to LLM-based instruction in any professional-training context [6].

A complementary systematic review addressing ChatGPT's utility across healthcare education, research, and practice more broadly reinforced this domain-specific applicability, examining future perspectives and potential limitations of LLMs specifically within healthcare's demanding accuracy and safety requirements [7]. The consistency of this integrity-versus-utility tension across both the general-education literature (Section 2.1) and the medical-education literature reviewed here suggests that academic-integrity and data-accuracy concerns are structural features of LLM-based education generally, rather than artifacts specific to any single disciplinary context.

2.4 Transformer Architectures for Sentiment Analysis in Digital Communication

Turning from education to digital content analysis, transformer-based architectures, and BERT specifically, have become the dominant approach for sentiment analysis of social media and digital communication content. Research leveraging BERT for enhanced sentiment analysis in multicontextual social media content demonstrated that BERT representation retains semantic and syntactic connectivity between short-form texts such as tweets, enhancing performance across every NLP task on large datasets, with the underlying architecture's bidirectional contextual encoding providing a specific technical advantage over earlier, unidirectional language-modeling approaches [8]. A comparative benchmarking study of ChatGPT and DeepSeek examining user sentiment toward LLM-based applications themselves, using a dataset of 4,000 authentic user reviews, found that deep learning-based classification demonstrated superior performance over lexicon-based analysis, with a convolutional

neural network achieving 96.41% accuracy and near-perfect classification of negative sentiment, alongside high F1-scores for neutral and positive classes [9]. This study is methodologically notable for applying sentiment-analysis techniques reflexively, to measure public sentiment about LLMs themselves, illustrating the recursive relationship between NLP-as-technology and NLP-as-analytical-tool that characterizes the current research landscape.

Fine-tuned BERT variants have achieved particularly high reported accuracy on domain-specific social media sentiment tasks. A study enhancing sentiment-analysis accuracy on social media comments using a tuned BERT model reported that by leveraging hyperparameter optimization and data augmentation, the approach significantly outperforms traditional models, achieving 99.98% accuracy, while highlighting BERT's superior ability to capture contextual meaning, including slang and informal register common in social media text [10]. The same study situated this result within a broader multilingual and low-resource-language research trend, noting related work analyzing stress in complex, low-resource languages such as Bengali using hybrid BERT-LSTM, BERT-GRU, and BERT-CNN-BiLSTM architectures, with one such Bengali-language stress-analysis model achieving 90.36% accuracy [10]. This multilingual evidence indicates that transformer-based sentiment analysis performance, while highest for well-resourced languages such as English, is increasingly extending to lower-resource languages through hybrid architectural approaches.

2.5 Sarcasm Detection and the Limits of Contextual Sentiment Classification

Sarcasm detection has emerged as a specifically challenging sub-task within sentiment analysis because it requires modeling pragmatic and cultural context beyond literal semantic content. A comprehensive study of sarcasm detection on social media using LLMs and BERT with multi-headed attention, evaluated on the Self-Annotated Reddit Corpus, found that sarcasm's reliance on contextual, pragmatic, and cultural cues renders it exceptionally difficult for standard NLP models to capture, especially in noisy, multilingual social media environments, with example utterances such as apparently cheerful statements that critique impracticality illustrating how surface-level sentiment cues can directly mislead standard classification systems [11]. Despite this documented difficulty, the study's fine-tuned BERT-based model achieved state-of-the-art performance among the compared architectures, with precision, recall, F1-score, and accuracy of 0.918, 0.917, 0.917, and 0.917 respectively, outperforming GPT-3, Claude-2, and Llama-2 on the same benchmark [11]. This finding, that a smaller, task-specifically fine-tuned encoder model outperformed larger general-purpose LLMs on a narrow, pragmatically demanding classification task, is significant for the broader NLP-application literature because it demonstrates that model scale alone does not guarantee superior performance on tasks requiring precise, narrow contextual judgment.

2.6 Misinformation Detection Using Pretrained Language Models

Misinformation detection represents a third major digital-content-analysis application area in which transformer-based and LLM architectures have demonstrated substantial reported performance. A study introducing a hybrid BERT-LSTM approach for misinformation detection in mobile social networks, developed specifically to enhance digital literacy, found that GPT-3 achieved an accuracy of 93.84%, a recall of 92.15%, and an F1-score of 92.99% in detecting misinformation, indicating that GPT-3 is highly effective in detecting misinformation, particularly in generating accurate predictions across diverse datasets [12]. The same study reported that its hybrid BERT-LSTM model effectively balances performance, computational efficiency, and applicability, making it a practical tool for real-time misinformation detection specifically tailored for mobile social network deployment, where computational resource constraints limit the feasibility of deploying full-scale LLMs directly on-device [12].

Complementary research on misinformation detection using pretrained language models more broadly examined RoBERTa, BERT, and GPT-3 as tools for extracting thematic and sentiment features from text, finding these pretrained language models highly effective for developing statistical and computational models that label misinformation and analyze its dynamics during viral events, with the resulting model exhibiting robust performance by capturing temporal changes in misinformation spread patterns [13]. This temporal-dynamics framing is important because it treats misinformation detection not as a static text-classification task but as a dynamic modeling problem tracking how misinformation narratives evolve and propagate over time, a framing that anticipates the temporal-stability concerns examined in Section 2.7.

Bibliometric research quantifying the misinformation-detection research field itself found that BERT-based keyphrase extraction from large-scale social-media data, using sentence-transformer architectures, contributes valuable knowledge to address the issue, enhancing understanding of the field's dynamics and aiding in the development of effective detection and mitigation strategies, while identifying IEEE Access as the leading publication venue and the United States, India, China, Spain,

and the United Kingdom as the leading contributing countries to misinformation-detection research specifically [14]. This bibliometric mapping situates the individual technical studies reviewed above within a rapidly growing, internationally distributed research field.

2.7 Temporal Drift and Reliability Challenges in Deployed NLP Systems

While the studies reviewed in Sections 2.4–2.6 report strong benchmark performance, a smaller but methodologically important strand of research has begun documenting the reliability challenges transformer-based sentiment and content-analysis models face once deployed on authentic, real-time digital content rather than static benchmark datasets. A comprehensive temporal drift analysis of transformer-based sentiment models, validated on 12,279 authentic social media posts spanning major real-world events, demonstrated significant model instability with accuracy drops reaching 23.4% during event-driven periods, alongside maximum confidence drops of 13.0% that showed strong correlation with actual performance degradation [15]. This finding directly qualifies the high benchmark accuracies reported elsewhere in the reviewed literature (Sections 2.4–2.6): models achieving accuracy above 90% or even above 99% on curated benchmark datasets may nonetheless degrade substantially when real-world events introduce topic shifts, novel vocabulary, and evolving sentiment expressions the training data did not anticipate.

The same drift-detection study introduced four novel, zero-training drift metrics specifically designed to outperform embedding-based baselines while maintaining computational efficiency suitable for production deployment, enabling real-time sentiment-monitoring systems to detect their own performance degradation without requiring costly model retraining [15]. This zero-training drift-detection contribution addresses a genuine gap the benchmark-focused sentiment-analysis and misinformation-detection literature reviewed in Sections 2.4–2.6 largely does not address: how deployed systems can monitor and flag their own reliability degradation under real-world, non-stationary content conditions.

2.8 Research Gap

While the education-focused LLM literature reviewed in Sections 2.1–2.3 has established substantial evidence for instructional benefit alongside integrity and equity concerns, and the digital-content-analysis literature reviewed in Sections 2.4–2.6 has established comparably substantial evidence for high benchmark accuracy across sentiment, sarcasm, and misinformation-detection tasks, comparatively few studies integrate the temporal-drift and reliability concerns documented in Section 2.7 with the educational-deployment context, despite LLM-based tutoring and feedback systems facing structurally similar non-stationary content challenges, evolving curricula, student populations, and language use, as the social media content these drift-detection studies examine directly. This review addresses that gap by organizing the reviewed evidence within a comparative framework spanning application domain, architecture, reported performance, and documented reliability limitation.

III. METHODOLOGY

This study adopts a qualitative, descriptive, and comparative research design synthesizing peer-reviewed and archival literature on large language models and transformer-based natural language processing across education and digital content analysis. Reviewed works were compared along four axes: application domain (education, sentiment/sarcasm analysis, or misinformation detection); architecture examined (GPT-family LLM, BERT/RoBERTa encoder, or hybrid architecture); evidence type (systematic review, benchmark performance study, or real-world deployment/reliability study); and whether the study addressed equity, integrity, or reliability limitations alongside reported performance gains.

IV. RESULTS AND DISCUSSION

4.1 LLM Educational Applications Mapped to Reported Outcomes

Table 1 maps the principal LLM educational applications reviewed onto their reported outcomes and the specific educational context each was validated in.

Table 1. LLM Educational Applications and Reported Outcomes

Application	Educational Context	Reported Outcome	Key Reference
Intelligent tutoring systems	General education (88-study review)	Most common LLM application; improved performance and engagement	Empirical LLM education review [2]

Lesson planning, professional development	Language teacher education	Four thematic domains identified across instructional/reflective use	ChatGPT teacher education review [4]
Personalized learning materials, feedback	General/resource-constrained settings	Potential to alleviate educational resource scarcity	Educational inequalities study [5]
Clinical simulation, personalized learning plans	Medical education	Applications validated; integrity/accuracy concerns highlighted	Medical education scoping review [6]
Test performance across language, programming, math	Cross-disciplinary	Enhanced test scores, faster task completion	Empirical LLM education review [2]

4.2 Transformer Architectures Mapped to Content-Analysis Task and Reported Accuracy

Table 2 consolidates the transformer-based digital content-analysis studies reviewed against the specific task, architecture, and quantified performance each reports.

Table 2. Transformer/LLM Architectures: Task, Architecture, and Reported Performance

Task	Architecture	Reported Performance	Key Reference
App-review sentiment classification	CNN (deep learning) vs. lexicon-based	96.41% accuracy, near-perfect negative-class classification	ChatGPT/DeepSeek benchmarking study [9]
Social media comment sentiment	Tuned BERT (hyperparameter-optimized)	99.98% accuracy	Tuned BERT sentiment study [10]
Bengali-language stress detection	BERT-LSTM/GRU/CNN-BiLSTM hybrids	90.36% accuracy	Bengali stress analysis (cited in [10])
Sarcasm detection (SARC dataset)	Fine-tuned BERT (vs. GPT-3, Claude-2, Llama-2)	Precision/Recall/F1/Accuracy ≈ 0.917–0.918	Sarcasm detection study [11]
Misinformation detection	GPT-3	93.84% accuracy, 92.15% recall, 92.99% F1	BERT-LSTM misinformation study [12]
Misinformation detection (mobile networks)	Hybrid BERT + LSTM	Balances performance, efficiency, real-time applicability	BERT-LSTM misinformation study [12]

4.3 Documented Risks Cross-Referenced Against Mitigation Strategies

Table 3 shifts from reported performance to documented risk, cross-referencing each risk against the mitigation strategy the literature proposes.

Table 3. Documented Risks and Proposed Mitigation Strategies

Risk	Documented In	Proposed Mitigation	Key Reference
Over-reliance / uncritical acceptance of AI output	General education review	Cultivate critical, reflective usage approaches	Empirical LLM education review [2]
Academic integrity and data-accuracy concerns	Medical education	Guidelines; correct-use training for students	Medical education scoping review [6]
Unequal access widening educational disparities	Cross-institutional education study	Targeted access/equity-focused deployment	Educational inequalities study [5]

Sarcasm/pragmatic-context misclassification	Social media sentiment research	Multi-head attention, fine-tuning on annotated corpora	Sarcasm detection study [11]
Real-world temporal drift and accuracy degradation	Deployed sentiment-monitoring systems	Zero-training drift-detection metrics	Temporal drift detection study [15]
Misinformation propagation dynamics over time	Viral-event misinformation research	Temporal-dynamics-aware pretrained LM modeling	Misinformation PLM study [13]

The consolidated evidence across Tables 1–3 indicates that LLM and transformer-based NLP applications across education and digital content analysis share a common empirical pattern: strong, often very high, reported performance on curated benchmarks and controlled educational-deployment studies, paired with a consistent secondary literature documenting reliability, equity, or integrity limitations that qualify these headline results. The temporal-drift evidence reviewed in Section 2.7, showing accuracy drops of up to 23.4% during real-world event-driven periods, is particularly significant because it suggests that even the near-perfect accuracies reported in Table 2 for curated benchmarks may not generalize fully to the dynamic, non-stationary content conditions both social media platforms and, plausibly, live educational deployments actually present [15].

V. FUTURE RESEARCH DIRECTIONS

Three priorities emerge from the reviewed literature. First, the temporal-drift and reliability-monitoring methodology developed for social media sentiment models [15] should be extended directly to educational LLM deployments, given that student populations, curricula, and language use are similarly non-stationary over time, yet no study reviewed here applies equivalent drift-detection methodology to intelligent tutoring or automated feedback systems specifically. Second, the equity concerns raised regarding unequal access to educational LLM tools [5] warrant longitudinal, outcome-focused research tracking whether LLM-based personalized learning actually narrows or widens achievement gaps over multi-year periods, extending beyond the cross-sectional benefit documentation that currently dominates the educational-LLM literature. Third, given the sarcasm-detection finding that smaller, task-specifically fine-tuned encoder models can outperform larger general-purpose LLMs on narrow pragmatic tasks [11], further comparative research should systematically map which digital-content-analysis tasks benefit most from large general-purpose LLMs versus smaller, fine-tuned transformer encoders, providing practitioners with clearer architecture-selection guidance than the current, largely task-by-task benchmark literature offers.

VI. CONCLUSION

This review has examined natural language processing and large language model applications across education, communication, and digital content analysis, synthesizing systematic reviews of LLM integration in teaching and learning with the transformer-architecture literature on sentiment analysis, sarcasm detection, and misinformation identification. Across all three domains, the reviewed evidence documents substantial and often very high reported performance, intelligent tutoring systems improving student performance and engagement, fine-tuned BERT models achieving accuracy exceeding 99% on social media sentiment tasks, and pretrained language models achieving over 90% accuracy on misinformation detection, while consistently identifying reliability, equity, and integrity concerns that qualify these gains. The emerging temporal-drift evidence, showing substantial accuracy degradation for transformer-based sentiment models during real-world event-driven periods, suggests that benchmark performance alone is an incomplete guide to real-world deployment reliability, a concern with direct relevance not only to digital content moderation but to the live educational deployments this review has also examined. Realizing the full potential of NLP and LLM technologies across education, communication, and digital content analysis will require sustained attention to these reliability and equity dimensions alongside continued architectural and benchmark performance gains.

REFERENCES

[1] "The Utility of ChatGPT as an Example of Large Language Models in Healthcare Education, Research and Practice: Systematic Review on the Future Perspectives and Potential Limitations," medRxiv preprint, 2023.

[2] "Large language models in education: A systematic review of empirical applications, benefits, and challenges," *Computers and Education: Artificial Intelligence*, ScienceDirect, 2025.

[3] "Detection of extreme sentiments on social networks with BERT," *Social Network Analysis and Mining*, Springer Nature, vol. 12, 2022.
[4] "Mapping the emerging research landscape on applications of ChatGPT in Language teacher education: A systematic narrative literature review," *Discover Artificial Intelligence*, Springer Nature, 2025.
[5] "Whose ChatGPT? Unveiling Real-World Educational Inequalities Introduced by Large Language Models," *arXiv preprint arXiv:2410.22282*, 2024.
[6] X. Xu, Y. Chen, and J. Miao, "Opportunities, challenges, and future directions of large language models, including ChatGPT in medical education: A systematic scoping review," *Journal of Educational Evaluation for Health Professions*, vol. 21, no. 6, 2024.
[7] "Applications and Concerns of ChatGPT and Other Conversational Large Language Models in Health Care: Systematic Review," *Journal of Medical Internet Research*, PMC, 2024.
[8] "Leveraging the BERT Model for Enhanced Sentiment Analysis in Multicontextual Social Media Content," ResearchGate preprint, 2024.
[9] "Benchmarking ChatGPT and DeepSeek in April 2025: A Novel Dual Perspective Sentiment Analysis Using Lexicon-Based and Deep Learning Approaches," *arXiv preprint arXiv:2509.19346*, 2025.
[10] "Enhancing sentiment analysis accuracy on social media comments using a tuned BERT model," *Discover Computing*, Springer Nature, 2025.
[11] "Enhancing sarcasm detection on social media: A comprehensive study using LLMs and BERT with multi-headed attention on SARC," *PMC / National Library of Medicine*, 2025.
[12] "Tackling misinformation in mobile social networks: A BERT-LSTM approach for enhancing digital literacy," *Scientific Reports*, Nature, 2025.
[13] "Misinformation detection on online social networks using pretrained language models," *Information Processing & Management*, ScienceDirect, 2025.
[14] "Bibliometric analysis of misinformation detection using BERT-based keyphrase extraction," ResearchGate preprint, 2024.
[15] Bansal and I. Gangwani, "Zero-Training Temporal Drift Detection for Transformer Sentiment Models: A Comprehensive Analysis on Authentic Social Media Streams," arXiv preprint arXiv:2512.20631, 2025.
[16] 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.
[17] Vaswani et al., "Attention is all you need," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 5998–6008.
[18] T. B. Brown et al., "Language models are few-shot learners (GPT-3)," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2020.
[19] E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, 2023.
[20] Y. Liu et al., "RoBERTa: A robustly optimized BERT pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.
[21] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," in *Proc. NeurIPS*, 2019.
[22] C. Raffel et al., "Exploring the limits of transfer learning with a unified text-to-text transformer (T5)," *Journal of Machine Learning Research*, vol. 21, no. 140, pp. 1–67, 2020.
[23] G. Pennycook, T. D. Cannon, and D. G. Rand, "Prior exposure increases perceived accuracy of fake news," *Journal of Experimental Psychology: General*, vol. 147, no. 12, pp. 1865–1880, 2018.