AI for Social Media Summaries: An Encoder-Decoder Transformer System vs ChatGPT

Afrodite Papagiannopoulou[1], Chrissanthi Angeli[1] and Panagiotis Makrigiannis[2]
[1]Dept of Electrical and Electronic Engineering, University of West Attica, Ancient Olive Grove Campus, Athens, Greece
[2]Dept of Civil Engineering, University of West Attica, Ancient Olive Grove Campus, Athens, Greece

apapagiannop@uniwa.gr
afpapag@gmail.com
angeli@uniwa.gr
c_angeli@otenet.gr
pmakrig@uniwa.gr

Abstract: In recent years, automatic text summarization has become a vital area of research due to its role in improving access to and understanding of vast information across domains. The rise of social media has intensified the need for summarization tools capable of handling user-generated content such as posts, comments, and discussions. Unlike structured texts, social media content is often informal, fragmented, context-dependent, and noisy. It frequently includes slang, abbreviations, emojis, and diverse writing styles, posing unique challenges for traditional summarization methods. While conventional approaches perform well on formal text, they often struggle to capture the nuances of online discourse. This highlights the need for specialized models that can generate coherent and context-aware summaries tailored to the characteristics of social media language. Recent advances in neural architectures, particularly Transformer-based sequence-to-sequence models, have shown promise in overcoming these challenges. These models excel at capturing long-range dependencies and contextual relationships, making them well-suited for summarizing dynamic and unstructured inputs. Despite technical progress, evaluating the quality of summaries remains difficult. Standard metrics like ROUGE may not fully reflect subjective qualities such as fluency, coherence, and semantic fidelity, which are essential for human-like summarization. This paper introduces a Transformer-based summarization system designed specifically for social media comments related to topical posts. We benchmark its performance against models like ChatGPT, assessing outputs across multiple linguistic and semantic dimensions. By combining both traditional and advanced evaluation metrics, our work provides a more holistic view of summarization quality and identifies key areas for future improvement.

Keywords: Social media summarization, Transformers, Multidimensional evaluation, Natural language generation

1. Introduction

The exponential growth of social media has transformed how individuals communicate, engage with information, and influence public discourse on critical issues such as politics, economics, and social change. Vast volumes of user-generated content—ranging from posts and comments to reactions and threads—are shared across platforms daily, reflecting a wide spectrum of public opinion and societal trends. This unstructured and dynamic content offers a valuable resource for analyzing social behavior. However, its overwhelming scale and informal nature pose significant challenges in efficiently extracting meaningful insights. Without effective summarization techniques, essential information can be lost amidst redundancy, noise, and conflicting viewpoints, making it difficult for users, organizations, and policymakers to interpret online discourse.

Text summarization plays a pivotal role in addressing this challenge by condensing large volumes of information into concise, coherent, and relevant representations. During high-impact events or fast-moving discussions, summarization can reduce cognitive overload, enhance comprehension, and ensure timely dissemination of critical updates. For journalists, researchers, businesses, and governments, robust summarization systems offer strategic value by enabling rapid identification of trends, key themes, and public sentiment.

Historically, Automatic Text Summarization (ATS) (Gupta and Lehal, 2010) has focused on formal, structured content such as academic articles and reports. Early methods—rooted in linguistic, statistical, and heuristic techniques—used strategies like keyword extraction, sentence ranking, and clustering (Luhn, 1958). While effective for static documents, they struggle with the informal language, brevity, and contextual ambiguity that characterize social media communication. As digital platforms evolved, so did the need for more flexible and semantically aware summarization methods.

Recent advances in artificial intelligence have given rise to deep learning (DL) techniques, particularly transformer-based architectures, which have significantly advanced the state of summarization. Unlike traditional models, transformers leverage self-attention mechanisms to process entire text sequences in parallel, enabling greater contextual understanding and improved handling of long-range dependencies. These features make transformers well-suited for abstractive summarization—an approach that rephrases and synthesizes content rather than extracting verbatim sentences (Varma et al., 2017; Rachabathuni, 2017). By generating human-like summaries that capture underlying meaning, abstractive methods enhance both readability and informativeness.

Despite their impact, the application of transformer-based abstractive summarization to social media remains relatively underexplored. The informal grammar, fragmented syntax, and rapidly shifting topics common in social platforms introduce linguistic and contextual challenges that require domain-specific adaptations. Moreover, existing studies often focus on a narrow range of platforms, overlooking variability across different types of social media interactions.

In response, this thesis investigates the integration of transformer models with abstractive summarization techniques to improve the quality, relevance, and efficiency of summarizing social media content. The approach aims to generate summaries that are not only coherent and semantically rich but also tailored to the nuances of user-generated language. Specifically, the system addresses challenges of informal expression, thematic grouping, and the evolving nature of online conversations. Unlike prior work limited to datasets from Twitter or Reddit, this study incorporates a more diverse set of sources and employs innovative training strategies to enhance generalizability. Furthermore, the evaluation framework extends beyond conventional metrics to include a multidimensional assessment of summary quality, coherence, and utility.

By advancing summarization techniques for social media communication, this work contributes to tools that support faster decision-making, improved content comprehension, and more effective engagement with digital public discourse.

2. Related Work

Advancements in Artificial Intelligence (AI) and neural architectures have significantly propelled the capabilities of Natural Language Processing (NLP), particularly in tasks involving large-scale data and complex linguistic structures. Among these tasks, text summarization has emerged as a central challenge, evolving from early neural models to contemporary transformer-based systems.

Initial explorations in automatic text summarization employed Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs) (Gao et al., 2019). These architectures were among the first to show promise in abstractive summarization, especially within domains characterized by informal language such as social media. Notable contributions in this space include the Attentional Encoder-Decoder framework (Liang, 2020), as well as attention-augmented models (Wang and Ren, 2021), which improved the identification of salient information in noisy, user-generated text. The introduction of sequence-to-sequence models with integrated attention mechanisms (Bhandarkar and Thomas, 2023) marked a pivotal step, enabling more nuanced and context-aware summary generation through enhanced encoder-decoder coupling.

The emergence of Transformer architectures (Vaswani et al., 2017) revolutionized NLP by replacing recurrence with self-attention mechanisms, allowing more efficient modeling of long-range dependencies. Transformers rapidly achieved state-of-the-art performance across tasks (Gupta 2022), including summarization. For example, a multi-stage model for Chinese summarization (Su 2020) combined BERT encoders with bidirectional LSTMs and a document-level transformer, demonstrating the effectiveness of hybrid designs in long-form summarization.

Transformer models have also been adapted for dialogue summarization, capturing conversational dynamics with greater fidelity (Singhal et al., 2020). Comparative evaluations of leading transformer-based models—including BART, PEGASUS, and T5—have shown consistent superiority of T5 in ROUGE metrics (Blekanov 2022). Beyond traditional text corpora, summarization has expanded to include sources such as the WikiHow dataset (Pal 2022) and web content enhanced by social media signals (Nguyen et al., 2020), leveraging models like BERT and T5 for broader generalization across styles.

Despite these advances, the targeted summarization of user-generated comments remains underexplored. Prior work has focused predominantly on summarizing primary post content or event narratives, with less emphasis on downstream user interactions. Studies drawing from Twitter and Reddit (Blekanov 2022; Li 2020; Kerui 2020)—and, in some cases, Sina Weibo—have used hybrid architectures pairing pre-trained BERT encoders with Transformer-based decoders (Li 2020; Kerui 2020) or alternative summarization modules (Tampe 2021). While these models yield promising outputs, research on comment-level summarization, particularly in informal and fragmented threads, remains nascent. The T5 model, though strong in general-purpose tasks, has only recently been applied to this domain (Blekanov 2022).

Parallel efforts have explored sentence-level summarization in review datasets (Rawat et al., 2021), using the Universal Sentence Encoder with statistical heuristics and graph-based reduction to extract informative sentence fragments. These methods enhance content density while addressing constraints of brief textual inputs.

In recent years, prompt engineering has emerged as a novel strategy for adapting large language models (LLMs) without full retraining. By designing task-specific input prompts, researchers can steer model behavior and improve performance across NLP tasks. Frameworks like OpenPrompt (Ding et al., 2022) support this process with modular tools for prompt-based learning, spanning architectures like masked language models (MLM), autoregressive language models (LM), and sequence-to-sequence (Seq2Seq) frameworks.

Further refinements include entity chain representations (Narayan et al., 2021), which improve coherence by structuring input, and prefix-tuning (Li 2021), a parameter-efficient fine-tuning method introducing task-specific vectors while keeping base parameters frozen. Prompt-tuning (Lester 2021) extends this by learning tunable prompts for frozen models, offering a cost-effective, reusable alternative to full fine-tuning. These lightweight methods are useful in resource-constrained settings or domains requiring rapid adaptation.

In specialized fields like clinical NLP, where conventional methods often miss domain-specific subtleties, prompt engineering has shown notable improvements in both flexibility and performance (Wang et al., 2024). These developments support a broader shift toward prompt-based methodologies as an efficient alternative to traditional training (Liu et al., 2021).

Taken together, this body of work highlights growing interest in adaptable, context-sensitive summarization—particularly those leveraging transformers' expressive power and prompt engineering’s task-specific control.

3. Methodological Framework for Transformer-Based Summarization

The methodological foundation for the proposed abstractive summarization framework is grounded in the transformer architecture, selected for its proven ability to model complex linguistic dependencies and handle the semantic variability inherent in user-generated social media content. Transformers are especially well-suited for this domain, given their demonstrated efficacy in capturing subtle contextual cues across fragmented and noisy text on platforms like Twitter and Reddit.

Unlike traditional extractive techniques or earlier neural models such as RNNs—including LSTM and GRU—that suffer from sequential processing and limited memory, transformers use self-attention mechanisms to process entire sequences in parallel (Vaswani et al., 2017). This capability is critical for modeling the long-range dependencies prevalent in dispersed, context-sensitive online discourse.

Summarization is framed here as a multi-dimensional task defined by three core criteria: coherence, ensuring logical flow; informativeness, preserving core content; and conciseness, reducing redundancy while maintaining clarity. This principled approach underlies the system’s design and supports the use of abstractive strategies over extractive ones. While extractive models select and combine source fragments, transformer-based abstractive models generate fluent, semantically rich summaries better aligned with the intent of the original material.

The system is structured into four modules:

- Preprocessing: Cleans informal input, normalizes abbreviations/emojis, and removes noise.
- Representation Learning: Encodes text using transformer embeddings like BERT (Devlin et al., 2019) or T5 (Raffel et al., 2019).
- Summarization Engine: Uses fine-tuned transformer models with reinforcement learning and attention refinement.
- Evaluation Framework: Combines automatic metrics with human benchmarks for comprehensive validation.

The choice of transformers reflects not only performance but conceptual fit, supported by their success across NLP tasks (Gupta 2022; Blekanov 2022; Nguyen et al. 2020). Ultimately, this framework treats transformers as foundational to addressing the unique linguistic challenges of modern social media summarization.

3.1 System Description

3.1.1 Motivation for transformer-based social media summarization

The exponential growth of content across digital and social media platforms has led to a deluge of information that users must parse to derive meaningful insights. Social media platforms, especially during unfolding events, are flooded with posts, comments, and reactions, often resulting in redundancy and thematic drift. This volume impedes comprehension and masks the core information users seek.

Abstractive summarization serves as a vital solution for distilling essential content from noisy data. Social media content holds value far beyond ephemeral communication—it contributes to digital journalism, informs policy debates, and enables the study of societal behavior at scale. Effective summarization of user-generated comments supports applications in crisis response, social analytics, and public opinion monitoring.

However, summarizing social media data poses distinctive challenges:

- Informality and Non-standard Language: Social texts often forgo grammatical norms, using slang, abbreviations, emojis, and emoticons (e.g., “<3”, “:P”), mirroring spoken language and emotional spontaneity.
- Contextual Fragmentation and Topic Drift: Threads diverge from the original topic, producing disjointed, non-linear conversations that hinder coherent summarization.
- Lexical Sparsity and Multimodality: Posts are concise, embedding meaning in short text, symbols, and images—limiting the context available to traditional models.

Transformer-based architectures offer a robust solution. With multi-head self-attention, they model long-range dependencies and complex, non-linear context. Pre-trained on diverse corpora (Gupta 2022), they generalize across registers, interpreting irregular syntax, emotive symbols, and fragmented structures.

Transformers preserve factual content, sentiment, and communicative intent—essential for representing social discourse. Their scalability enables real-time, high-volume social media summarization. Building on these strengths, we present a modular, transformer-based system for comment-level abstractive summarization across social platforms.

3.1.2 Overview of the proposed system

The proposed system focuses on generating abstractive summaries from clusters of user comments associated with individual social media posts. Unlike prior models that center on tweet threads (Blekanov 2022; Li 2020) or Reddit discussions (Kerui 2020), this system although trained exclusively on data from a single platform (Facebook), the system is designed to generalize across platforms, demonstrating adaptability to diverse speech structures and linguistic characteristics, thereby ensuring applicability in a variety of communicative environments.

The system architecture integrates the foundational transformer model with a four-stage processing pipeline (Figure 1):

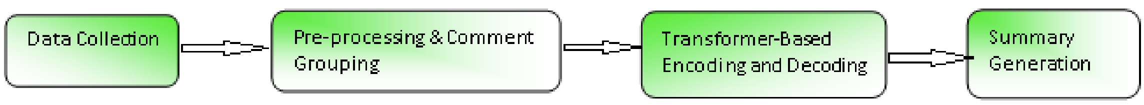

Figure 1: Social Media Summarization Pipeline

1. Data Collection: The dataset is sourced from Facebook1, a platform with short, unstructured, and variable comment threads. These threads often include repeated content, emojis, informal language, and user-specific references—making them ideal for testing robust summarization techniques. Unlike structured document collections, this dataset reflects the idiosyncrasies of organic online discourse.
2. Pre-processing and Comment Grouping: Pre-processing reduces linguistic noise and normalizes input through:

- Removing emojis, special characters, dashes, and punctuation
- Tokenizing text while preserving case-sensitive abbreviations
- Grouping comments by post ID or title to form semantically related clusters

Thematic grouping helps the model focus on coherent discourse units, reducing cross-topic interference. This strategy draws on dictionary-based clustering approaches used in prior pipelines (Bhandarkar 2023; Li 2020).

3. Transformer-Based Encoding and Decoding: Tokens are converted into IDs and passed through embeddings processed by a multi-head attention mechanism (eight heads). Following Vaswani et al. 2017, the encoder includes:

- Multi-head self-attention
- Feed-forward layers
- Residual connections and layer normalization

The decoder receives encoder outputs and a right-shifted target sequence. Masked self-attention prevents future token access. Cross-attention layers integrate contextual representations, and a softmax classifier outputs token distributions.

4. Summary Generation: The decoder auto-regressively generates one token at a time, conditioned on previous tokens and encoder context. This enables fluent, coherent, and human-like summaries with semantic integrity and pragmatic tone.

Unlike extractive or graph-based reductions (Rawat et al., 2021), our abstractive method captures paraphrasing, inferred meaning, and dialogic nuance—critical for informal, sentiment-rich environments like social media. This modular pipeline remains scalable and adaptable for evolving datasets and platform constraints.

4. Comparison With ChatGPT

To benchmark the performance of our summarization framework, we conducted a comparative evaluation against ChatGPT, a widely adopted large language model developed by OpenAI in 2022. Built on the GPT (Generative Pre-trained Transformer) architecture, ChatGPT leverages deep learning and natural language processing to generate human-like text. Its fine-tuning via Reinforcement Learning from Human Feedback (RLHF) enables alignment with user intent, making it highly effective in various language tasks, including summarization.

For this comparison, a random subset of user comment threads was selected. Each set of comments was prefixed with the prompt “summarize:” and input into ChatGPT to generate corresponding summaries. These outputs were evaluated using the same suite of metrics applied to our system—grouped into reference-based, reference-free, and human-aligned categories.

4.1 Comparison Using Reference-Based Metrics

Reference-based evaluation metrics quantify the quality of machine-generated text by comparing it against human-authored references. These metrics are widely adopted to assess dimensions such as lexical overlap, semantic similarity, and content coverage. This section provides a concise overview of the reference-based metrics employed in the current analysis, with emphasis on their roles in evaluating alignment between system-generated summaries and gold-standard outputs.

Evaluation was conducted using standard metrics from the ROUGE (Lin 2004) family, including ROUGE-1, ROUGE-L, ROUGE-Lsum, ROUGE-S, and ROUGE-WE. These metrics capture a range of textual similarities, from exact lexical matches to structurally and semantically informed correspondences.

[1]https://github.com/jbencina/facebook-news

As illustrated in Figure 2, the Proposed Model demonstrates superior performance across all ROUGE variants in comparison to ChatGPT. For instance, the ROUGE-1 F1-score achieved by the Proposed Model (0.3292) exceeds that of ChatGPT (0.2739), reflecting greater overlap with key content words in the reference summaries. While both models yield moderate recall, the higher precision and F1-scores associated with the Proposed Model indicate stronger alignment with salient information in the source texts.

Scores for ROUGE-L, ROUGE-Lsum, and ROUGE-S are comparatively lower, as expected given their more stringent criteria regarding sequence structure and skip-bigram matches. Nevertheless, the Proposed Model maintains a consistent advantage across these metrics, suggesting enhanced preservation of both lexical and structural aspects of the target summaries.

The ROUGE-WE metric, which incorporates semantic similarity through pre-trained word embeddings, further underscores this trend. The Proposed Model attains a ROUGE-WE score of 0.9568, slightly higher than ChatGPT's 0.9495. Although the numerical difference (Δ = 0.0073) is modest, it may reflect a meaningful improvement in semantic fidelity, particularly in contexts where conceptual accuracy is prioritized over surface-form similarity.

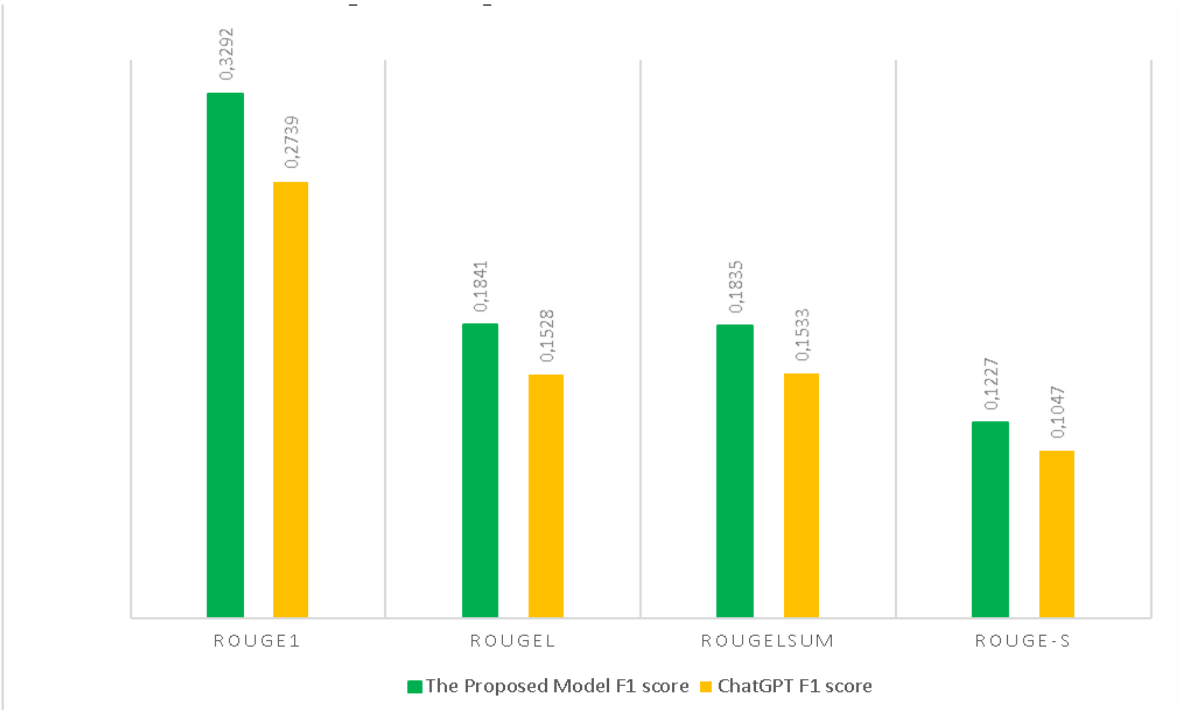

Figure 2: Comparison on reference-based metrics

4.2 Comparison Using Reference-Free Metrics

In contexts where human-authored reference summaries are unavailable, reference-free evaluation metrics provide an effective alternative for assessing the quality of automatically generated summaries. These metrics are particularly valuable in real-world applications where the construction of gold-standard references is infeasible due to resource constraints. Unlike reference-based methods, reference-free evaluation relies on intrinsic properties of the generated text—such as fluency, coherence, informativeness, and semantic consistency—enabling comprehensive quality assessment without dependence on external benchmarks.

This section presents a comparative analysis of generated summaries using a set of reference-free metrics, including Shannon Consistency, Shannon Coherency (Egan 2022), Semantic Distance-based Clustering (SDC) (Li 2022b), and SUPERT (Gao 2020). These metrics were selected to capture various dimensions of summary quality, particularly semantic fidelity and structural coherence. Figure 3 visualizes the average metric scores for the Proposed model and ChatGPT.

Across all evaluated dimensions, the Proposed model demonstrates superior performance. Higher scores in Shannon Consistency (0.9512 vs. 0.9447) and Shannon Coherency (0.9519 vs. 0.9471) suggest stronger logical structure and internal alignment within its outputs. These findings indicate that the Proposed model generates summaries with greater semantic stability across different input variations.

The most substantial performance gap is observed in the SDC metric, where the Proposed model achieves a score of 0.5746, substantially exceeding ChatGPT's 0.2646. This metric is designed to assess the preservation of semantic intent and minimize interpretive divergence. The lower score for ChatGPT implies greater variability and potential semantic drift, which may undermine reliability in applications requiring accurate content representation.

On the SUPERT metric, both models exhibit strong performance, with the Proposed model slightly outperforming ChatGPT (0.9990 vs. 0.9977). Although the margin is minimal, it further supports the Proposed model's ability to generate summaries that align closely with expected semantic clusters, even in the absence of reference summaries.

As a result, the Proposed model outperforms ChatGPT across all reference-free evaluation criteria, with particularly notable improvements in semantic consistency and coherence. These results underscore the effectiveness of reference-free metrics as a scalable and domain-agnostic evaluation framework, especially in settings where human-generated references are unavailable or impractical to obtain.

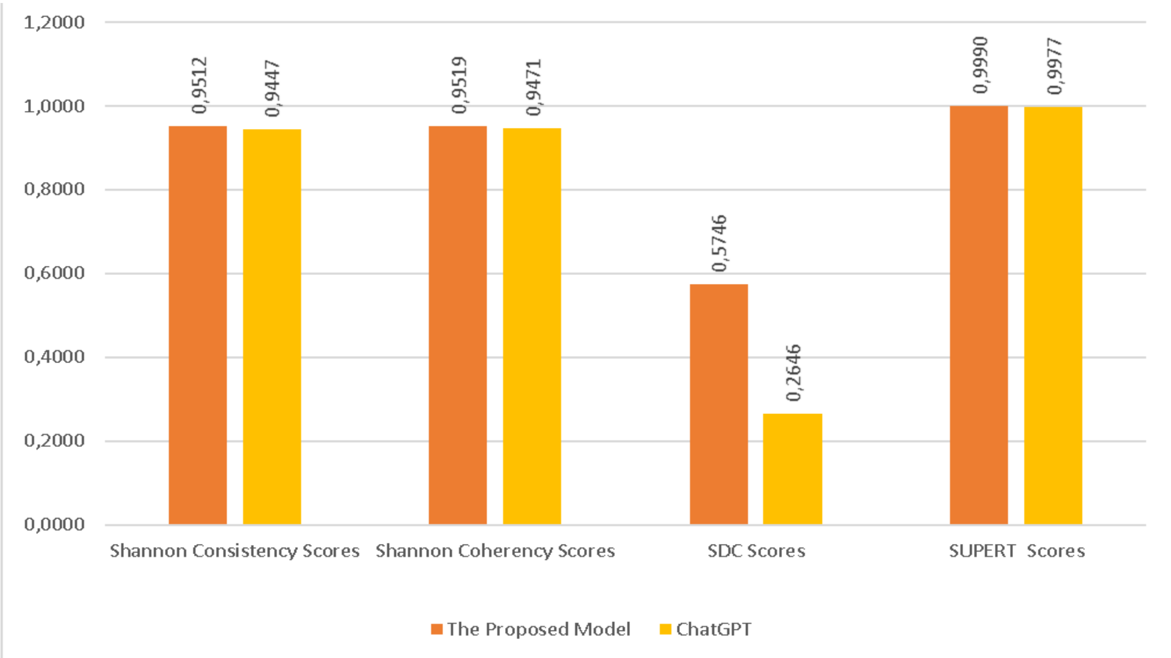

Figure 3: Comparison of reference free metrics

4.3 Comparison Using Human-Shaped Metrics

Evaluating abstractive summarization presents significant challenges, particularly within the domain of social media discourse, due to the subjective and multi-dimensional nature of the task. In contrast to extractive methods, effective abstractive summarization must accurately and succinctly convey the source text's essential meaning while employing novel and contextually appropriate phrasing, thereby emulating human summarization strategies (Ng, 2015). The presence of multiple valid summary interpretations for a given input further complicates the evaluation process, as no single reference summary can comprehensively capture all plausible variations.

These challenges are exacerbated in social media environments, where texts are often informal, grammatically inconsistent, and characterized by non-standard linguistic elements such as slang, emojis, and abbreviations. The dialogic and non-linear structure of discussions—frequently involving references to prior posts or thematic divergence—introduces additional contextual dependencies. Furthermore, the brevity and fragmented nature of social media content typically limit lexical richness and cohesion, complicating both the generation and evaluation of abstractive summaries. As such, a rigorous evaluation framework must account for linguistic variability, contextual coherence, and semantic fidelity across heterogeneous and dynamic textual inputs.

To address these concerns and evaluate qualitative aspects beyond traditional reference-based metrics, a comparative analysis was conducted using human-centered evaluation criteria across a Transformer-based summarization model (referred to here as the Proposed model) and ChatGPT. The assessment focused on six dimensions: consistency, lexical diversity, fluency, readability, semantic coherence, and overall coherence.

Results indicate complementary strengths across the models. The Proposed model demonstrated notable improvements in consistency, lexical diversity, readability, and semantic coherence. Specifically, consistency scores more than doubled those of ChatGPT (0.5627 vs. 0.2490), reflecting a stronger capacity to maintain internal logical structure and thematic stability across generated content. This attribute is especially pertinent in applications requiring structured outputs, such as automated reporting or multi-turn dialogue systems.

In terms of coherence, both models performed comparably, with a marginal advantage observed for the Proposed model (0.4054 vs. 0.3914). Lexical diversity was also higher in the Proposed model, indicating more varied vocabulary usage. While this may enhance richness, it introduces a potential trade-off with simplicity, which remains a key factor in user-facing applications.

Fluency was the one area in which ChatGPT outperformed the Proposed model (0.7925 vs. 0.7876), likely attributable to extensive pre-training on a broad and diverse corpus. Nevertheless, the Proposed model exhibited higher readability (0.4455 vs. 0.3626), suggesting it produces more accessible outputs, despite slightly lower fluency.

Semantic coherence scores were closely aligned, with a slight edge again favoring the Proposed model (0.4132 vs. 0.3948), indicating its greater aptitude for maintaining meaning across sentence boundaries. As illustrated in Figure 4, the Proposed model outperformed ChatGPT in five of the six human-shaped evaluation dimensions.

These findings highlight the differing strengths of the models under consideration. While ChatGPT maintains superior fluency, the Proposed model delivers a more balanced performance profile, excelling in dimensions critical for interpretive quality and structural consistency. These results underscore the Proposed model's suitability for applications requiring readability, consistency, and diversity, whereas ChatGPT may be more appropriate for tasks emphasizing conversational naturalness and fluid expression.

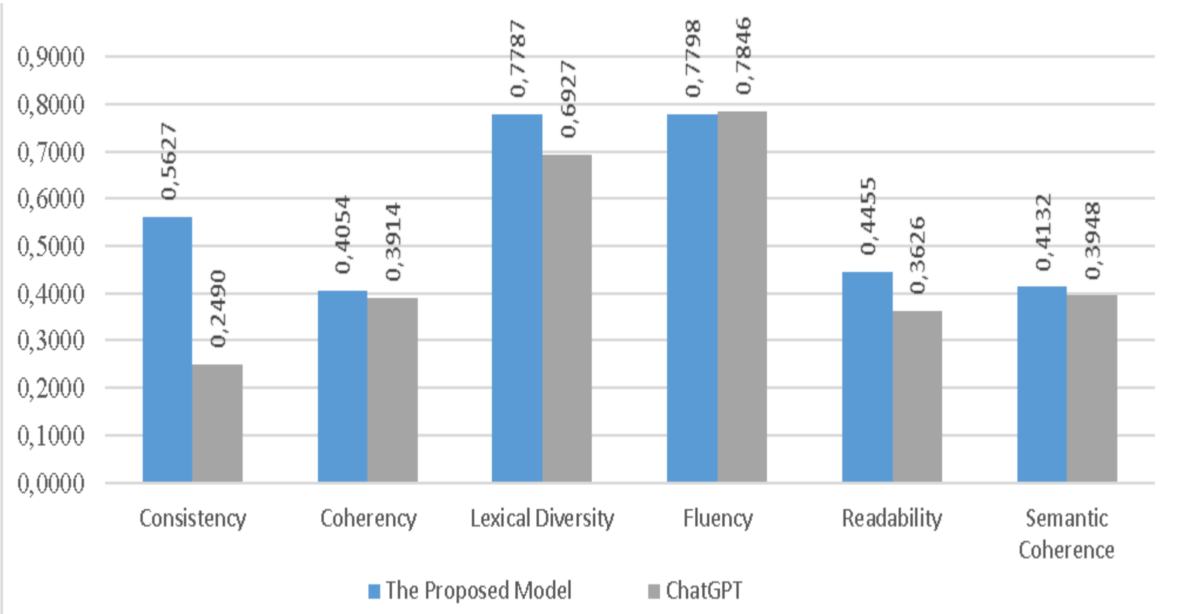

Figure 4: Comparison on human-shaped metrics

Future work will seek to enhance the fluency of the Proposed model without compromising its demonstrated strengths, aiming to achieve a more optimal integration of naturalness, consistency, and accessibility in abstractive summarization of social media content.

5. Conclusion and Future Work

In an era marked by the explosive growth of user-generated content, the ability to extract coherent, relevant, and timely insights from social media has never been more crucial. This research introduces a domain-adaptive abstractive summarization system that effectively addresses the inherent challenges of processing noisy, unstructured data typical of social networking platforms. By harnessing the power of transformer-based architectures and transfer learning, the system demonstrates significant advancements in fluency, contextual awareness, and semantic alignment, outperforming existing models—including ChatGPT 4.0—across a range of both quantitative and qualitative evaluation metrics. Its robust preprocessing pipeline and attention-driven encoder-decoder design enable it to capture not only the factual essence but also the emotional and cultural nuances embedded in digital conversations.

Beyond its technical contributions, this work underscores the practical relevance of intelligent summarization in real-world applications such as sentiment tracking, trend analysis, public communication, and customer feedback synthesis. Importantly, the system's flexibility, scalability, and domain adaptability highlight its potential for both academic inquiry and industry deployment. However, the study also illuminates several critical avenues for future research. Enhancing computational efficiency through lightweight architectures, extending capabilities to multilingual and multimodal contexts, and enabling cross-platform and real-time summarization are all promising directions. These next steps are essential to addressing the dynamic, diverse, and increasingly global nature of social media discourse. Ultimately, this work lays a strong foundation for developing summarization technologies that do more than condense content—they interpret and contextualize it in ways that are meaningful to human understanding and decision-making.

Acknowledgements

The publication fees were totally covered by the University of West Attica.

Ethical Declaration: I confirm that ethical clearance was not required for this research.

AI Declaration: I have not used any AI tool to create this paper

References

Bhandarkar, P., Thomas, K. T. (2023) "Text Summarization Using Combination of Sequence-To-Sequence Model with Attention Approach". *Springer Science and Business Media Deutschland GmbH*: 283–293, https://doi.org/10.1007/978-981-19-3035-5_22

Blekanov, I. S., Tarasov N., and Bodrunova S. S. (2022) "Transformer-Based Abstractive Summarization for Reddit and Twitter: Single Posts vs. Comment Pools in Three Languages". *Future Internet* 14, 69. https://doi.org/10.3390/fi14030069

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019) "BERT: Pre-training of deep bidirectional transformers for language understanding". In *Proceedings of NAACL-HLT* (pp. 4171–4186).

Ding N., Hu S., Zhao W., Chen Y., Liu Z., Zheng H., et al. (2022) "OpenPrompt: An Open-source Framework for Prompt-learning". *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*: System Demonstrations; p. 105-13. https://doi.org/10.18653/v1/2022.acl-demo.10

Egan, N., Vasilyev, O., Bohannon, J. (2022) "Play the Shannon game with language models: a human-free approach to summary evaluation". *In Proceedings of the AAAI conference on Artificial Intelligence*, 10599–10607. https://arxiv.org/abs/2103.10918

Gao, S., Chen, X., Li, P., Ren, Z., Bing, L., Zhao, D. and Yan, R. (2019) "Abstractive Text Summarization by Incorporating Reader Comments". *In The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*, 33(01): 6399-6406. DOI:10.1609/aaai.v33i01.33016399

Gao, Y., Zhao, W., Eger, S. (2020) "SUPERT: towards new frontiers in unsupervised evaluation metrics for multi-document summarization". In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1347–1354, Online. Association for Computational Linguistics. https://doi.org.10.18653/v1/2020.acl-main.124

Gupta, V., & Lehal, G. S. (2010) "A survey of text summarization extractive techniques". *Journal of Emerging Technologies in Web Intelligence, 2*(3), 258–268. https://doi.org/10.4304/jetwi.2.3.258-268

Gupta, A., Chugh, D. and Katarya, R. (2022) "Automated News Summarization Using Transformers". *In Sustainable Advanced Computing*, 2022, Volume 840. ISBN: 978-981-16-9011-2. https://arxiv.org/pdf/2108.01064

Kerui, Z., Haichao, H., Yuxia, L. (2020) "Automatic text summarization on social media". *ACM International Conference Proceeding Series, Association for Computing Machinery*, https://doi.org/10.1145/3440084.3441182

Lester, B., Al-Rafou, R., and Constant, N. (2021) "The Power of Scale for Parameter-Efficient Prompt Tuning". *In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 3045–3059, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.243

Li, Q. and Zhang, Q. (2020) "Abstractive Event Summarization on Twitter". *In The Web Conference 2020 - Companion of the World Wide Web Conference*, WWW 2020, Association for Computing Machinery: 22–23, https://doi.org/10.1145/3366424.3382678

Li, X.L., and Liang, P. (2021) "Prefix-tuning: Optimizing continuous prompts for generation". *In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* (Volume 1: Long Papers), 4582–4597, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.acl-long.353

Liang, Z., Du, J. and Li, C. (2020) "Abstractive social media text summarization using selective reinforced Seq2Seq attention model". *Neurocomputing*, 410, 432–440, https://doi.org/10.1016/j.neucom.2020.04.137

Lin, C. Y. (2004) "ROUGE: A package for automatic evaluation of summaries". In *Text Summarization Branches Out* (pp. 74–81). Barcelona, Spain: Association for Computational Linguistics.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., and Neubig, G. (2021) "Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing". *ACM Computing Surveys* 55, 1-35. https://doi.org/10.1145/3560815

Liu, Y., Jia, Q., Zhu, K. (2022b) "Reference-free summarization evaluation via semantic correlation and compression ratio". *In Proceedings of the 2022 conference of the North American Chapter of the Association for Computational Linguistics: human language technologies*, 2022, 2109–2115. https://doi.org/10.18653/v1/2022.naacl-main.153

Luhn, H. P. (1958) "The Automatic Creation of Literature Abstracts". *IBM Journal of Research and Development*, vol. 2, no. 2, Apr. 1958, pp. 159-165, https://doi: 10.1147/rd.22.0159

Pal, A., Fan, L., Igodifo, V. (2022) "Text Summarization using BERT and T5". https://anjali001.github.io/Project_Report.pdf

Narayan, S., Zhao, Y., Maynez, J., Simões, G., Nikolaev, V., and McDonald, R. (2021) "Planning with Learned Entity Prompts for Abstractive Summarization". *Transactions of the Association for Computational Linguistics*, 9: 1475–1492. https://doi.org/10.1162/tacl_a_00438

Ng, J-P., and Abrecht, V. (2015) "Better Summarization Evaluation with Word Embeddings for ROUGE". *In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal. Association for Computational Linguistics, 1925–1930. https://doi.org/10.18653/v1/D15-1222

Nguyen, M. T., Nguyen, V. C., Vu, H. T., Nguyen, V. H. (2020) "Transformer-based Summarization by Exploiting Social Information". *In Proceedings - 2020 12th International Conference on Knowledge and Systems Engineering*, KSE 2020, Institute of Electrical and Electronics Engineers Inc.: 25–30, https://doi.org/10.1109/KSE50997.2020.9287388

Rachabathuni, P. K. (2017) "A survey on abstractive summarization techniques". In *2017 International Conference on Inventive Computing and Informatics (ICICI)* (pp. 762–765). IEEE. https://doi.org/10.1109/ICICI.2017.8365239

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narag, S., Matena, M., Zhou, Y., Li, W., and Liu P. J. (2019) "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer". *The Journal of Machine Learning Research*, Volume 21, Issue 1, 140:1-140:67. https://dl.acm.org/doi/abs/10.5555/3455716.3455856

Rawat, R., Rawat, P., Elahi V. and Elahi, A. (2021) "Abstractive Summarization on Dynamically Changing Text". *In the 5th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, 2021, pp. 1158-1163, https://doi.org/10.1109/ICCMC51019.2021.9418438

Singhal, D., Khatter, K., Tejaswini, A., Jayashree, R. (2020) "Abstractive Summarization of Meeting Conversations". *In 2020 IEEE International Conference for Innovation in Technology*, INOCON 2020, Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/INOCON50539.2020.9298305

Su, M. H., Wu, C. H., Cheng, H. T. (2020) A Two-Stage Transformer-Based Approach for Variable-Length Abstractive Summarization. *IEEE/ACM Transactions on Audio Speech and Language Processing*, 28, 2061–2072. https://doi.org/10.1109/TASLP.2020.3006731

Tampe, I., Mendoza, M. and Milios, E. (2021) "Neural Abstractive Unsupervised Summarization of Online News Discussions". Arai, K. (Eds) *Intelligent Systems and Applications*. IntelliSys 2021. Lecture Notes in Networks and Systems, vol 295. Springer, Cham. https://arxiv.org/abs/2106.03953

Varma, V., Kurisinkel, L. J., & Radhakrishnan, P. (2017) "Social media summarization". In *A practical guide to sentiment analysis* (Chapter 7, pp. 135–153).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I. (2017) "Attention Is All You Need". *In 31st Conference on Neural Information Processing Systems* (NIPS 2017), Long Beach, CA, USA. https://arxiv.org/abs/1706.03762

Wang, J., Shi, E., Yu, S., Wu, Z., Ma, C., Dai, H., Yang, Q., Kang, Y., Wu, J., Hu, H., Yue, C., Zhang, H., Liu, Y., Li, X., Ge, B., Zhu, D., Yuan, Y., Shen, D., Liu, T., and Zhang, S. (2024) "Prompt Engineering for Healthcare: Methodologies and Applications". https://arxiv.org/html/2304.14670v2

Wang, Q. and Ren, J. (2021) "Summary-aware attention for social media short text abstractive summarization". *Neurocomputing*, 425, 290–299, https://doi.org/10.1016/j.neucom.2020.04.136