Journal of the Brazilian Computer Society, 2025, 31:1, doi: 10.5753/jbcs.2025.5786
This work is licensed under a Creative Commons Attribution 4.0 International License.

Long-Text Abstractive Summarization using Transformer Models: A Systematic Review

Abubakar Salisu Bashir [**Northwest University, Kano State | Aliko Dangote Universities of Science and Technology, Wudil, Nigeria |** ***bsalisu2016@gmail.com***]
Abdulkadir Abubakar Bichi [**Northwest University, Kano State |** ***engraabubakar@gmail.com***]
Usman Mahmud [**Northwest University, Kano State |** ***usmanmahmud@auk.edu.ng***]
Abdulrahman Mohammed Bello [**Aliko Dangote Universities of Science and Technology, Wudil, Nigeria|** ***dzabdul@gmail.com***]

Department of Computer Science, Northwest University, Kano State, Nigeria

Received: 30 March 2025 • **Accepted:** 08 August 2025 • **Published:** 29 October 2025

Abstract Transformer models have significantly advanced abstractive summarization, achieving near-human performance. However, while effective for short texts, long-text summarization remains a challenge. This systematic review analyzes 56 studies on transformer-based long-text abstractive summarization published between 2017 and 2024, following predefined inclusion criteria. Findings indicate that 69.64% of studies adopt a hybrid approach while 30.36% focus on improving transformer attention mechanisms. News articles and scientific papers are the most studied domains, with widely used datasets including CNN/Daily Mail, PubMed, arXiv, GovReport, QMSum, and XSum. ROUGE is the dominant evaluation metric (61%), followed by BERTScore (20%), with others such as BARTScore, human evaluation, METEOR, and BLEU-4 also used. Despite progress, challenges persist, including contextual information loss, high computational costs, implementation complexity, lack of standardized evaluation metrics, and limited model generalization. These findings highlight the need for more robust hybrid approaches, efficient attention mechanisms, and standardized evaluation frameworks to enhance long-text abstractive summarization. This review provides a comprehensive analysis of existing methods, datasets, and evaluation techniques, identifying research gaps and offering insights for future advancements in transformer-based long-text abstractive summarization.

Keywords: Long text summarization, Transformer models, Text summarization, Long documents, Systematic literature review

1 Introduction

The advent of transformer models [Vaswani *et al.*, 2017] has revolutionized natural language processing (NLP), driving significant advancements in machine translation [He, 2024], question answering [Madan *et al.*, 2024], and text summarization [Wang *et al.*, 2024b]. Among these, abstractive summarization of long texts presents unique challenges due to its requirement for semantic understanding, contextual coherence, and fluency [Fikri *et al.*, 2024; Gokhan *et al.*, 2024]. Unlike extractive summarization methods that merely select key sentences, abstractive summarization paraphrases and synthesizes information, demanding models that can generate human-like summaries [Dai and He, 2024; Sun *et al.*, 2024]. This review focuses on abstractive summarization because it produces more coherent and informative summaries; particularly valuable for complex or technical documents. While extractive methods are computationally efficient, they often lead to redundancy and fragmented narratives. We also focus on long-text summarization due to its critical relevance in domains such as legal, scientific, and financial writing, where maintaining discourse structure and contextual integrity across extended input is essential.

Transformer architectures, particularly self-attention mechanisms, have proven highly effective in handling long-range dependencies and processing vast amounts of text [Bettayeb *et al.*, 2024; Chen *et al.*, 2024; Chu *et al.*, 2024]. However, summarizing long-text documents introduces challenges such as handling extensive text length, preserving contextual integrity, and mitigating redundancy [Gokhan *et al.*, 2024; Wang *et al.*, 2024b]. To address these, researchers have explored techniques like modifying attention mechanisms [Liu *et al.*, 2024; Pang *et al.*, 2022], and employing hybrid methods that combine extractive and abstractive approaches [Hardy *et al.*, 2022; Rahman *et al.*, 2024]. While these strategies improve summarization effectiveness, they also introduce trade-offs, such as increased computational cost during modification of the attention mechanism [Zhao *et al.*, 2020] and potential information loss during segmentation or extraction [Ulker and Ozer, 2024].

Several researchers have reviewed long-text abstractive summarization within specific domains. Akter *et al.* [2025] conducted a survey on legal document summarization, identifying dataset limitations, domain-specific language challenges, and jurisdictional variations. Jain *et al.* [2021] also analyzed legal text summarization techniques, highlighting the dominance of extractive approaches and advocating for reinforcement learning and hierarchical models to improve abstractive methods.

Similarly, Rennard *et al*. [2023] and Kirstein *et al*. [2025] explore dialogue systems. Rennard et al. highlight challenges in meeting summarization, including speech disfluencies, overlapping dialogues, and dataset limitations, with AMI and ICSI being the most used. Kirstein et al. analyze 1,262 papers, identifying key obstacles in abstractive dialogue summarization, such as long-range dependency handling, entity resolution, and the lack of diverse, high-quality datasets. Other studies were on scientific documents. Chaves *et al*. [2022] highlight the growing use of transformer-based models for biomedical literature summarization but note that multi-document summarization and EHRs remain underexplored. Altmami and Menai [2022] discuss the dominance of extractive techniques, citation-based summarization, and the challenge of incorporating figures and tables.

Despite advancements in transformer-based long-text abstractive summarization, existing reviews predominantly focus on domain-specific applications such as legal, biomedical, and dialogue summarization. While these studies provide valuable insights, their findings are often not generalizable to broader long-text abstractive summarization tasks due to differences in text structure, linguistic complexity, and summarization objectives. Additionally, many prior surveys rely on empirical evaluations rather than systematic methodologies, making it difficult to reproduce results and identify overarching trends across domains.

Moreover, while hybrid techniques and attention modifications have been explored, a systematic comparison of these approaches remains scarce, particularly regarding their trade-offs in computational efficiency, coherence, and factual consistency. This study addresses these gaps by conducting a systematic literature review (SLR) on transformer-based long-text abstractive summarization, encompassing diverse domains, methodologies, datasets, and evaluation metrics. By adopting a structured review approach, this study aims to provide a comprehensive synthesis of existing research, highlighting key challenges, trends, and future directions.

To bridge these gaps, this study systematically reviews transformer-based long-text abstractive summarization with the following key contributions:

- Mapping the Evolution of Techniques: Categorizing key approaches, identifying research gaps, and formulating research question.
- Evaluating Datasets and Metrics: Assessing the effectiveness of commonly used datasets and evaluation metrics.
- Identifying Underexplored Areas: Highlighting less-examined applications and research gaps.
- Addressing Challenges and Future Directions: Analyzing scalability and contextual understanding issues while proposing solutions.

The paper is structured as follows: Section 2 details the methodology, Section 3 presents the review findings and discussion, and Section 4 concludes the study.

2 Methodology

This research employs a Systematic Literature Review (SLR) [Angioi and Hiller, 2023] as its foundational methodology to investigate the use of transformer-based models in long-text abstractive summarization. SLR is recognized for its rigorous process of identifying, analyzing, and synthesizing scholarly evidence to address specific research questions [Angioi and Hiller, 2023]. As illustrated in Figure 1, the SLR framework encompasses three core phases: conceptualization, implementation, and dissemination [Saleh *et al*., 2024]. These phases are iterative rather than strictly sequential, allowing for revisiting earlier steps to incorporate new insights and improve the review's comprehensiveness [Saleh *et al*., 2024]. This adaptability ensures the SLR remains responsive to the evolving body of knowledge in the field.

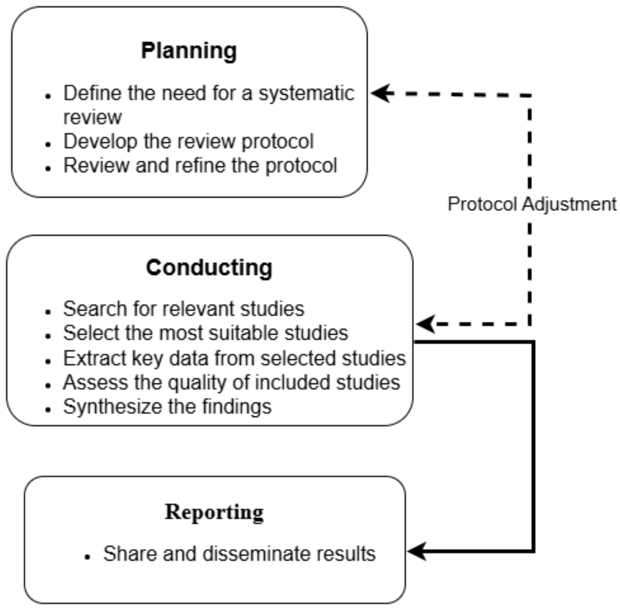

Figure 1. Systematic Literature Review Process Flowchart.

2.1 Research questions

Formulating research questions is a crucial step in conducting a SLR [Angioi and Hiller, 2023], ensuring clarity and focus [Saleh *et al*., 2024]. This study applies the PICOC framework—Population, Intervention, Comparison, Outcome, and Context [Petticrew and Roberts, 2008]—to structure its research questions (Table 1), with Table 2 detailing the questions and their motivations. The questions explore advancements in transformer-based long-text abstractive summarization, including model types, trends, datasets, evaluation metrics, application domains, limitations, and future research directions. This approach aims to provide a comprehensive overview and identify opportunities for innovation.

2.2 Search strategy and study selection

The search strategy in this study follows a systematic, multi-step approach to ensure comprehensive and relevant data collection. It involves identifying data sources, defining precise

Table 1. PICOC Structure

PICOC Element	Description
Population	Transformer Model-based long-text Abstractive Summarization.
Intervention	Techniques/Approaches, trends in long-text abstractive summarization, domains of application, datasets, and evaluation metrics.
Comparison	Comparing the performance of approaches used in long-text abstractive summarization.
Outcome	Serves as a theoretical basis for researchers interested in improving the performance of long-text abstractive summarization using transformer models. Also details the domains of application, datasets used, and evaluation metrics employed.
Context	Research conducted by researchers in the field of long-text abstractive summarization.

Table 2. Research Questions and Motivations

ID	Research Question	Motivation
RQ1	Is long-text abstractive summarization using transformer models an active research area?	To determine the relevance and importance of research in long-text abstractive summarization using transformer models.
RQ2	What techniques are used to address the problem of long-text in abstractive summarization?	To identify techniques specifically designed to handle the challenges of summarizing long texts.
RQ3	What are the most frequently used datasets for long-text abstractive summarization using transformer models?	To identify widely used datasets that support research in long-text abstractive summarization using transformer models.
RQ4	Which domains have benefited from transformer-based long-text abstractive summarization?	To explore the application of transformer-based summarization across various domains and their impact.
RQ5	What evaluation methods are commonly used in transformer-based long-text abstractive summarization, and what are their limitations?	To identify commonly used evaluation methods and understand their limitations in assessing transformer-based long-text abstractive summarization.
RQ6	What are the limitations of the present studies and future directions?	To uncover the gaps in existing research and propose potential future directions for advancing long-text abstractive summarization.

search strings, conducting an initial search, and iteratively refining search terms for optimal results. Three leading digital libraries were selected for their extensive coverage of artificial intelligence, natural language processing, and related fields (see Table 3). These platforms were chosen for their reliability, accessibility, and high-quality indexed publications.

The search strings were initially developed using the term ”long-text abstractive summarization” for exploratory searches. Additional keywords, such as ”long document”, “evaluation metric” and ”transformer models” were incorporated to refine the focus. Boolean operators (”OR” and ”AND”) were applied strategically to retrieve studies specifically addressing transformer models in long-text summarization.

Through iterative refinements, the search terms were adjusted based on retrieved articles to eliminate gaps or redundancies. The finalized search strings, tailored for each digital library, are presented in Table 3. The study focused on publications from 2017 to 2024 to capture recent advancements, starting from the introduction of the transformer models, a pivotal moment in natural language processing.

The study selection process followed PRISMA-S guidelines [Rethlefsen *et al*., 2021] and is summarized in Figure 2. The initial search across three digital libraries—Science Direct, Google Scholar, and IEEEXplore—retrieved 147 records from Science Direct, 642 from Google Scholar, and 22 from IEEEXplore. For Science Direct, filtering by the 2017–2024 year range reduced the records to 132, and further restricting the search to ”research articles” resulted in 112 records. Applying the same year range filter to Google Scholar reduced the records to 621, while the records from IEEEXplore remained unchanged at 22. The retrieved records were consolidated into a single Excel file, yielding 755 entries. After removing duplicates, 699 unique records remained. An eligibility assessment, based on the inclusion/exclusion criteria in Table 4, further excluded 313 records, leaving a final set of 386 records for analysis.

Table 3. Search String from the Three Data Sources

Data Source	Search String
ScienceDirect	(”Long text” OR ”long-text” OR ”long document”) AND (”abstractive summarization” OR ”abstractive text summarization”)
Google Scholar	(”long text” OR ”long-text” OR ”long document”) AND (”abstractive summarization” OR ”abstractive text summarization”) AND (”transformer” OR ”BERT” OR ”GPT” OR ”T5” OR ”Pegasus” OR ”Longformer” OR ”BigBird” OR ”Bart”) -survey -review
IEEE Xplore	(”long text” OR ”long-text” OR ”long document”) AND (”abstractive text summarization” OR ”abstractive summarization”)

The titles and abstracts of the remaining 386 records underwent detailed screening, resulting in the exclusion of 276 records due to irrelevance. This left 110 records for further evaluation. The full texts of these 110 papers were then thoroughly reviewed, and 54 papers were excluded for insufficient

Table 4. Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Research articles written in English.	Studies on multimodal abstractive summarization.
Studies published between 2017 and 2024.	Studies written in languages other than English.
Complete journal articles, book chapters, or conference proceedings.	Documents without full-text access.
Focused on long-text abstractive summarization.	Studies on short-text abstractive summarization.
Techniques involve transformer-based models.	Studies that do not use transformer models.

relevance, lack of methodological detail, or misalignment with the study's objectives. 56 papers were selected for the final analysis.

2.3 Data extraction

During the data extraction phase, 56 articles were critically appraised and analyzed, forming the basis for the thematic discussion. The discussion was systematically organized around key aspects, including publication year, techniques for long-text abstractive summarization, datasets, domain-specific applications, evaluation methods, and study limitations. This structured approach ensured a thorough and coherent synthesis of the reviewed literature.

2.4 Study quality assessment and data synthesis

To ensure reliable findings, 56 selected articles underwent a quality assessment using predefined criteria, focusing on research clarity, appropriateness of methods, dataset suitability, evaluation validity, and result transparency. The PRISMA checklist [Page *et al*., 2021] was used for a systematic review, and a double-review approach resolved any discrepancies to reduce bias. Data synthesis was organized thematically, covering publication year, transformer advancements, datasets, domain-specific applications (e.g., legal, scientific, or news), evaluation metrics, and study limitations. Quantitative data, like metric frequencies (e.g., ROUGE, BERTScore), were tabulated, while qualitative analysis focused on interpreting methodologies and findings. This thematic approach highlighted research progress, trends, and gaps, such as scalability, and generalizability. The findings provide a foundation for discussing advancements and future opportunities in long-text abstractive summarization.

3 Results and Discussion

This section presents the key findings of the systematic review on long-text abstractive summarization using transformer models, structured around the predefined research questions. An initial search yielded 755 articles related to long-text abstractive summarization with transformer models. Following a rigorous screening and eligibility assessment, 56 articles (see Table 5) published between 2017 and 2024 were identified as relevant for this study. These selected articles provide insights into the advancements, challenges, and methodologies employed in transformer-based abstractive summarization for long-text documents.

RQ1 : Is long-text abstractive summarization using transformer models an active research area?

The findings confirm that long-text abstractive summarization using transformer models has been an active and expanding research area since its emergence in 2019 (Figure 3). The increasing publication trend, rising from six studies in 2020 to 15 in 2024, underscores the growing scholarly interest in this domain. This growth aligns with prior studies [Fu, 2024; Steblianko *et al*., 2024], which identify transformer-based summarization as a rapidly evolving field with increasing applications in scientific document processing. This study presents a quantitative assessment of methodological evolution, revealing a paradigm shift in modeling approaches over time.

A key contribution of this study is the identification of a clear transition from Improved Transformer (I.T.) models to Hybrid models, signifying a fundamental shift in methodological preferences. Initially, in 2019, I.T. models—characterized by architectural enhancements, fine-tuning strategies, and optimized attention mechanisms—were the sole approach. Hybrid models, which integrate a preprocessing stage (e.g., sentence ranking, content selection) with a transformer-based summarizer, emerged in 2020 but remained marginal, with only two occurrences compared to four for I.T. models. However, by 2021, Hybrid models surpassed I.T. models for the first time, indicating a growing preference for hybridization.

This trend became more pronounced in subsequent years, with Hybrid model usage peaking at 12 instances in 2022, while I.T. models declined to only two occurrences. The growing preference for Hybrid models can be attributed to their ability to address key limitations of improved transformers, including:

- Factual consistency: By incorporating content selection mechanisms, Hybrid models reduce hallucinations in generated summaries.
- Computational efficiency: Hybrid approaches mitigate the quadratic complexity of self-attention, improving scalability for longer texts.
- Content relevance: Preprocessing techniques enhance information retention, reducing the risk of critical content being truncated.

An unexpected finding of this study is the resurgence of I.T. models in 2024, increasing from two instances in 2023 to five in 2024. This resurgence challenges the assumption that Hybrid models represent the ultimate optimization strategy for long-text summarization. Instead, it suggests a re-evaluation of direct transformer enhancements as a viable alternative to hybrid approaches.

This shift is likely driven by recent advancements in transformer architectures, including:

- Extended context windows (e.g., LongT5, Memorizing Transformers) that significantly increase the input length capacity of transformer models.
- Memory-efficient attention mechanisms (e.g., sparse attention, recurrent memory augmentation) that reduce computational overhead, allowing direct I.T. models to handle longer texts without truncation.

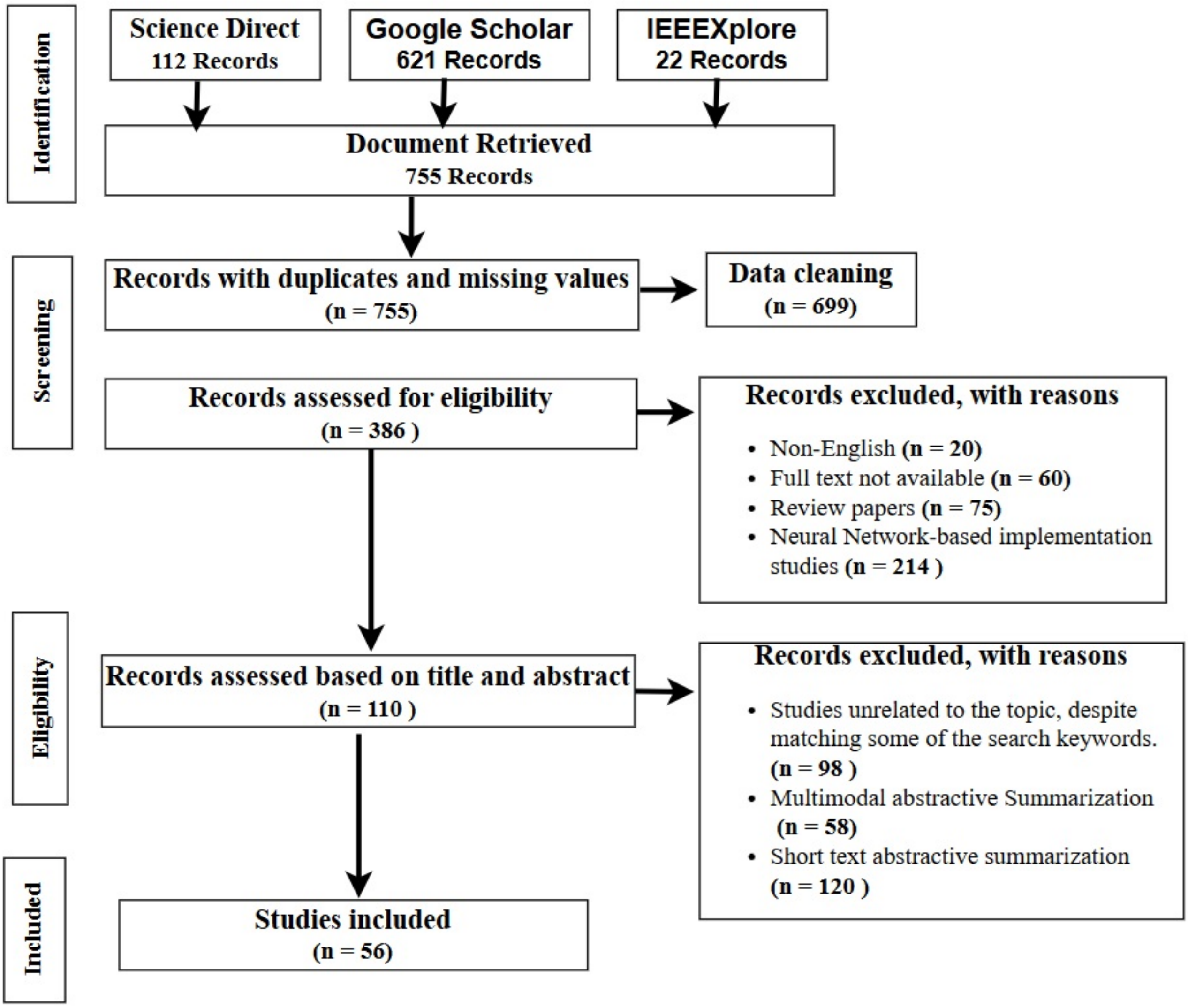

Figure 2. Systematic Literature Review Process Flowchart.

Unlike previous studies that primarily advocate for hybridization [Fu, 2024], this study identifies a countertrend where architectural refinements in transformers are revitalizing the standalone I.T. approach. This presents a crucial research question:

Under what conditions do direct architectural improvements (I.T.) outperform hybrid approaches in long-text abstractive summarization?

This question has significant implications for scalability, factual consistency, and domain-specific applications. While Hybrid models have proven effective in mitigating truncation-based information loss, recent transformer improvements may offer a more direct and efficient solution without requiring additional preprocessing steps.

RQ2 : What techniques are used to address the problem of long-text in abstractive summarization?

An analysis of 56 papers reveals that two primary approaches dominate long-text abstractive summarization: hybrid methods and improved Transformer architectures (Figure 4). Hybrid methods, adopted in 69.6% of studies, involve extractive or segmentation-based preprocessing before abstractive summarization, while 30.4% focus on enhancing Transformer architectures for better memory efficiency and extended context processing. Within hybrid methods, extraction-based preprocessing (28 studies) selects salient sentences to form an intermediate summary, whereas segmentation-based preprocessing (11 studies) divides text into smaller chunks before feeding it into the Transformer model (Figure 5).

The evolution of hybrid approaches indicates a transition in long-text abstractive summarization techniques. Early studies (2020) employed neural networks, such as LSTMs, for extractive preprocessing, but limitations in handling long-range dependencies led to a shift towards heuristic rule-based approaches. Six studies applied rule-based segmentation, and two implemented rule-based extraction, but these methods suffered from fixed-rule constraints that negatively impacted summary quality. To address these limitations, Transformer-based preprocessing emerged, with 12 studies utilizing Transformer models for extraction and 3 studies for segmentation.

Table 5. Selected Articles for the Review

Abbreviation:
I.T.: Improve Transformer

S/N	Author	Year	Technique	Dataset used	Evaluation Metric
1	Liao *et al*. [2020]	2020	I.T.	CNN/Daily Mail	ROUGE
2	Wang *et al*. [2024a]	2024	Hybrid	Webis-TLDR-17,XSum	ROUGE, BERTScore
3	Pang *et al*. [2022]	2022	I.T.	PubMed, arXiv, TVMegaSite, BookSum, ForeverDreaming, CNN/Daily Mail	ROUGE
4	Kiruluta *et al*. [2021]	2021	I.T.	PubMed	ROUGE
5	Pilault *et al*. [2020]	2020	Hybrid	arXiv, PubMed, Newsroom, BigPatent	ROUGE
6	Jeeson-Daniel *et al*. [2021]	2021	I.T.	PubMed	ROUGE
7	Karlbom and Clifton [2020]	2020	I.T.	Podcast	ROUGE
8	Pu *et al*. [2024]	2024	I.T.	BookSum, Multi-LexSum, eLife	ROUGE, BERTScore, Meteor
9	Cao and Wang [2023]	2023	I.T.	GovReport, QMSum, SummScreen, arXiv, BookSum	ROUGE
10	Rahman *et al*. [2024]	2024	Hybrid	XLsum, SWAS	ROUGE, BERTScore, BARTScore
11	Aksenov *et al*. [2020]	2020	I.T.	CNN/Daily Mail, SwissText	ROUGE
12	Cao and Wang [2022]	2022	I.T.	GOVREPORT-QS, GovReport, WIKIBIOSUM	ROUGE, BLEU-4
13	Dat *et al*. [2024]	2024	I.T.	Gigaword, CNN/Daily Mail, arXiv	ROUGE
14	Wu *et al*. [2024]	2024	I.T.	CNN/Daily Mail, Reddit, PubMed	ROUGE, BARTScore
15	Ying *et al*. [2021]	2021	Hybrid	Blog posts	ROUGE
16	Lim and Song [2023]	2023	Hybrid	AMI, ICSI, QMSum, SummScreen-FD, GovReport, SummScreen-TMS	ROUGE
17	Calizzano *et al*. [2022]	2022	Hybrid	WikinewsSum	ROUGE
18	Liu *et al*. [2022b]	2022	Hybrid	arXiv, PubMed, GovReport, Multi-News	ROUGE
19	Pant and Chopra [2022]	2022	Hybrid	Financial Report	ROUGE
20	Wu *et al*. [2023]	2023	I.T.	eLife, PLOS	ROUGE, BERTScore, BARTScore
21	Yuan *et al*. [2023]	2023	Hybrid	arXiv, QMSum	ROUGE
22	Zhang *et al*. [2022]	2022	Hybrid	AMI, ICSI, QMSum, SummScreen, GovReport	ROUGE, Human Evaluation
23	Zhang *et al*. [2019]	2019	I.T.	CNN/Daily Mail	ROUGE
24	Hardy *et al*. [2022]	2022	Hybrid	BOOKSUM	ROUGE, BERTScore, Human Evaluation
25	Zhang *et al*. [2021]	2021	Hybrid	CNN/Daily Mail, Gigaword	ROUGE
26	Mei *et al*. [2022]	2022	Hybrid	WikiHow, Reddit$_TIFU$	ROUGE
27	Liu and Chen [2022]	2022	Hybrid	AMI	ROUGE
28	Ma and Zong [2020]	2020	Hybrid	WikiSum	ROUGE
29	Kumar *et al*. [2022]	2022	Hybrid	MuP-2021, CNN/Daily Mail	ROUGE
30	Kashyap [2022]	2022	Hybrid	BOOKSUM	ROUGE
31	Yu *et al*. [2023]	2023	Hybrid	QMSum	ROUGE
32	You *et al*. [2024]	2024	Hybrid	eLife, PLOS	ROUGE, BERTScore
33	Liu and Xu [2023]	2023	Hybrid	QMSum	ROUGE
34	Huang *et al*. [2024]	2024	I.T.	CNN/Daily Mail, XSum, PubMed	ROUGE, BERTScore
35	Benedetto *et al*. [2024]	2024	Hybrid	Facebook$_{post}$	ROUGE, BERTScore
36	Huang *et al*. [2021b]	2021	Hybrid	LPO-News, CNN/Daily Mail	ROUGE, BERTScore
37	Wilman *et al*. [2024]	2024	Hybrid	CNN/Daily Mail, XSum, PubMed	ROUGE
38	Ezzat *et al*. [2024]	2024	I.T.	Mukhtasar	ROUGE
39	Moro and Ragazzi [2023]	2023	Hybrid	BillSum, PubMed, GovReport	ROUGE
40	Jain *et al*. [2024]	2024	Hybrid	BillSum, Fire$_legal$	ROUGE
41	Zhang *et al*. [2024]	2024	Hybrid	Multi-News, WCEP-10	ROUGE
42	Bajaj *et al*. [2021]	2021	Hybrid	Amicus	ROUGE
43	Chen and Iwaihara [2023]	2023	Hybrid	arXiv, PubMed, GovReport	ROUGE, BERTScore
44	Ranggianto *et al*. [2023]	2023	Hybrid	TripAdvisor, Amazon, DUC2004	ROUGE, BERTScore
45	Huang *et al*. [2023]	2023	Hybrid	CAIL2020, LCRD	ROUGE
46	Liu *et al*. [2022a]	2022	Hybrid	Scriptbase	ROUGE, BERTScore
47	Huang *et al*. [2021a]	2021	I.T.	GovReport, PubMed, arXiv	ROUGE
48	Tretyak and Stepanov [2020]	2020	Hybrid	arXiv	ROUGE
49	Saxena and Keller [2024]	2024	Hybrid	MENSA	ROUGE, BERTScore
50	Lu *et al*. [2023]	2023	Hybrid	Articles, Books	ROUGE, BERTScore
51	Obonyo *et al*. [2022]	2022	Hybrid	Cochrane, MS^2	ROUGE, BERTScore
52	Zhao *et al*. [2020]	2020	I.T.	arXiv, PubMed, Search2Wiki	ROUGE
53	Ulker and Ozer [2024]	2024	Hybrid	ArXivComp, SciTLDR, SciSummNet	ROUGE
54	Nguyen and Ding [2023]	2023	Hybrid	LexisNexis	ROUGE
55	Wu *et al*. [2021]	2021	Hybrid	CNN/Daily Mail, XSum, WikiHow	ROUGE
56	Liu *et al*. [2024]	2023	Hybrid	CLSum	ROUGE

However, the computational cost of using multiple Transformers posed challenges, especially in low-resource settings, necessitating alternative solutions.

To mitigate computational overhead, recent studies have explored graph-based techniques for both extraction and segmentation. However, our analysis identifies a major limitation: all graph-based approaches rely solely on a single-objective function (similarity score) for sentence ranking or segmentation. This fails to address redundancy and coverage, two critical factors for generating high-quality summaries. A contribution of this study is the identification of this limitation and the proposal of a multi-objective optimization approach that integrates coverage, coherence, relevance, and sentence length as additional ranking criteria with particle swarm optimization. This framework is expected to enhance summary quality while reducing the computational burden associated with Transformer-based preprocessing.

Unlike previous studies Fu [2024]; Koh *et al*. [2023], which primarily list existing approaches, our study formulates a research question:

How does a multi-objective approach (considering coverage, coherence, relevance, and sentence length) integrated with particle swarm optimization compare to single-objective methods (similarity score) in improving the quality of long-text abstractive summarization?

Addressing this question will provide deeper insights into the effectiveness of multi-objective ranking functions and their potential to optimize long-text summarization performance.

RQ3 : Which Domains Have Benefited from Transformer-Based Long-Text Abstractive Summarization?

Transformer-based long-text abstractive summarization has been applied across diverse domains, leveraging domain-

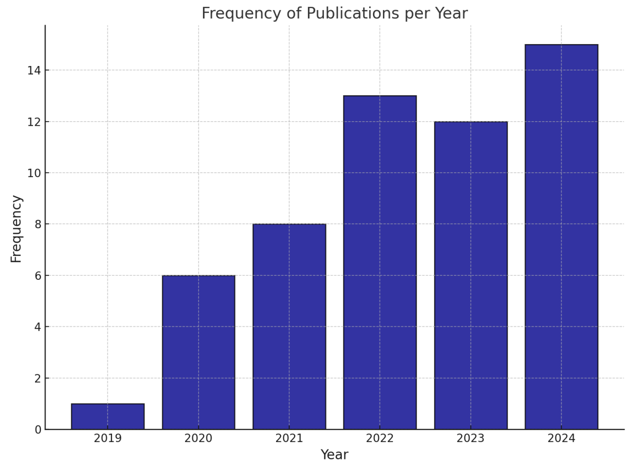

Figure 3. Yearly Publications on Transformer-Based Long-Text Summarization.

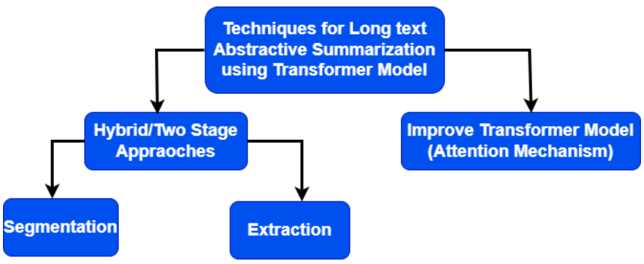

Figure 4. Techniques in Long-text Abstractive Summarization

specific datasets to enhance summarization quality (Table 6). Expanding on the classification framework by Wibawa *et al.* [2024], this study identifies eight primary application areas: news, scientific literature, legal documents, dialogues, books/literature, social media, product reviews, and reports, with biography-based summarization emerging as a distinct yet underexplored category. The news domain remains the most dominant, supported by well-established datasets such as CNN/Daily Mail, XSum, and Multi-News, which facilitate structured summarization. Scientific literature follows, with datasets like PubMed, arXiv, and GovReport presenting challenges in summarizing multi-section documents without compromising critical information. Legal and conversational domains are also gaining attention, utilizing datasets such as BillSum and QMSum, which require precision and coherence in summarization.

Despite advancements, key research gaps remain. Social media summarization, for instance, is constrained by the brevity, noise, and informal nature of data from platforms such as Reddit, Facebook Post, and Webis TLDR 17. Addressing these challenges requires context-aware filtering and fine-tuned models optimized for short-form, high-variability content. Similarly, biography summarization, though relevant for digital identity generation and historical documentation, has seen limited research, with WikiBioSum being one of the few available datasets. Future studies should explore methods for ensuring factual accuracy and temporal consistency in summarizing biographical data. Another underexplored area is product review summarization, which primarily relies on Amazon and TripAdvisor datasets. A promising direction involves leveraging vision-language models to enhance summarization by incorporating textual sentiment alongside image-based cues.

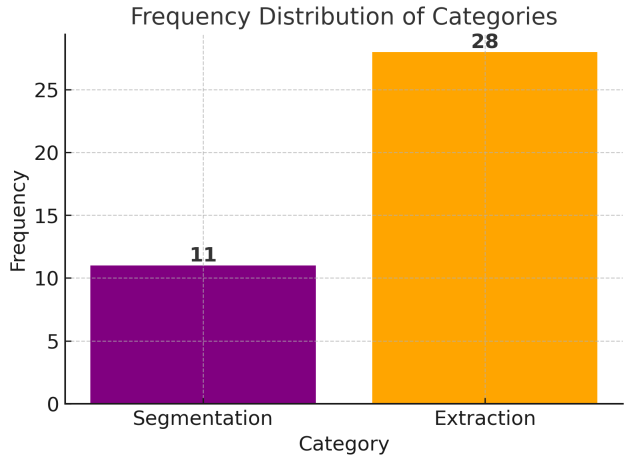

Figure 5. Frequency Distribution of Segmentation and Extraction

This study refines domain classification, highlights emerging areas, and identifies research gaps. The findings suggest that further investigation into social media, and biography summarization could enhance the adaptability and real-world applicability of transformer-based models. Addressing these gaps will contribute to the development of more efficient and context-aware summarization techniques.

RQ4 : What are the most frequently used datasets for long-text abstractive summarization?

Our analysis identifies a diverse range of datasets employed in long text abstractive summarization, reflecting the need for models to generalize across multiple domains. A total of 61 unique datasets were examined, highlighting the extensive benchmarking efforts in this field (Figure 6). Over the years, the CNN/Daily Mail and PubMed datasets have remained dominant, primarily due to their structured nature and accessibility, making them standard benchmarks for summarization models. The CNN/Daily Mail dataset is widely used for news summarization, while PubMed serves as a critical resource for summarizing scientific literature. However, recent trends indicate a shift toward domain specific datasets, particularly in legal (e.g., BillSum, Fire Legal, Amicus), scientific (e.g., arXiv, SciSummNet, eLife), and financial domains (e.g., Financial Report, LexisNexis), driven by the increasing need for specialized summarization models.

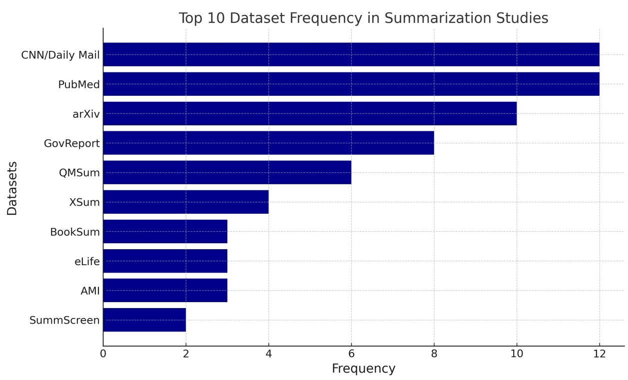

Figure 6. Top 10 Datasets for Long-Text Summarization

The timeline analysis further reveals an increasing adoption

Table 6. Domain of Application

Domain	Datasets
News	CNN/Daily Mail, XSum, Newsroom, Multi-News, SwissText, WCEP-10, LPO-News, WikinewsSum, DUC2004, Gigaword, WikiSum
Scientific/Research	PubMed, arXiv, GovReport, Cochrane, SciSummNet, SciTLDR, ArXiv-Comp, MS², Search2Wiki, eLife, PLOS, Multi-LexSum
Legal	BillSum, Fire_Legal, CAIL2020, Amicus, GOVREPORT-QS
Dialogue/Conversations	AMI, ICSI, SummScreen, QMSum, Scriptbase
Books/Literature	BookSum, BOOKSUM, WikiHow, ForeverDreaming, TVMegaSite, MENSA
Social Media	Reddit, Facebook_Post, Webis-TLDR-17
Product Reviews	TripAdvisor, Amazon
Reports	GovReport, Financial_Report
Biography/Profiles	WIKIBIOSUM

of hybrid approaches that integrate extractive and abstractive techniques, particularly in recent years (2022 to 2024). Hybrid methods are frequently applied to datasets such as Webis TLDR 17, Facebook Post, and Reddit, reflecting the growing interest in social media and user-generated content summarization. This shift highlights the evolving nature of summarization tasks, which now encompass not only traditional news and research articles but also conversational and informal text. Despite their frequent usage, some widely adopted datasets pose challenges. For example, CNN/Daily Mail exhibits extractive bias, which limits its effectiveness in truly abstractive summarization. Similarly, while PubMed and arXiv provide valuable domain-specific data, they present challenges related to technical jargon and specialized discourse, necessitating more sophisticated approaches to handle scientific text effectively.

Despite these advancements, several critical gaps remain. Low resource and multilingual datasets are significantly underrepresented, with Mukhtasar and XLSum among the few that cater to non English languages. This underrepresentation underscores the need for research efforts focused on expanding dataset diversity and improving cross-lingual summarization techniques. As the field progresses, future research should prioritize the development of benchmarks for conversational and user-generated content summarization, explore cross-domain datasets, and enhance domain adaptation techniques to improve performance across specialized fields. Addressing these gaps will be crucial in advancing the robustness and applicability of long-text abstractive summarization models.

RQ5 : What evaluation methods are commonly used in transformer-based long-text abstractive summarization, and what are their limitations?

Our analysis reveals that ROUGE remains the predominant evaluation metric, used in 61% of studies (Figure 7), consistent with prior findings [Sanchan, 2024]. Its widespread adoption stems from its ability to measure content overlap between generated and reference summaries [Lin, 2004]. However, its reliance on lexical matching limits its applicability to abstractive summarization, where paraphrasing and novel sentence construction occurs.

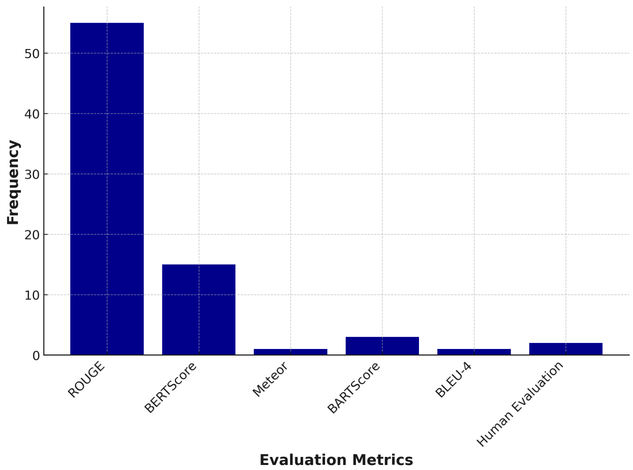

Figure 7. Distribution of Evaluation Metrics in Long-Text Abstractive Summarization

To address these limitations, BERTScore, used in 20% of studies, leverages contextual embeddings to assess semantic similarity, offering a more effective measure for abstractive summaries [Liao *et al.*, 2020]. BARTScore, found in 5% of studies, integrates generative modeling to evaluate fluency and coherence, though its adoption is constrained by computational costs. Human evaluation, despite being the most reliable for assessing coherence, relevance, and readability, appears in only 2% of studies due to time and resource constraints. Other traditional lexical-based metrics, such as BLEU-4 and METEOR, are increasingly being replaced by semantic-aware methods.

Emerging approaches, including reinforcement learning-based evaluation and learned reward models, seek to align evaluation with human judgment more effectively. These methods train models on human-annotated quality scores, moving beyond conventional lexical and embedding-based comparisons. However, their application remains limited, highlighting a research opportunity.

Current metrics also struggle with assessing discourse coherence and long-range dependencies in long-text summarization. Recent studies propose graph-based coherence metrics (e.g., Entity Graph Coherence) and knowledge-graph-augmented BERTScore to capture logical flow and topic continuity in lengthy documents. These novel approaches require further validation.

Given the limitations of fully automated evaluation, hybrid frameworks integrating human-in-the-loop feedback with automated metrics are gaining interest. Such approaches leverage active learning and limited human annotations to refine automated evaluation models, balancing efficiency with reliability.

While ROUGE and BERTScore continue to dominate summarization evaluation, emerging research highlights the need for reinforcement learning-based evaluation, discourse-aware metrics, and hybrid evaluation models. Addressing these gaps will be crucial for advancing transformer-based long-text abstractive summarization.

RQ6 : What are the limitations of the present studies and future direction

Limitations

Despite significant progress in long-text abstractive summarization, several critical challenges remain:

- Loss of Context and Semantic Coherence: Hybrid approaches, such as segmentation and extractive preprocessing, often disrupt the semantic flow of text, leading to incomplete or fragmented summaries. Current methods lack mechanisms to preserve contextual dependencies across segmented chunks, resulting in factual inconsistencies.
- Computational Inefficiency of Transformer Models: Advanced transformer models like Longformer and BigBird, despite their effectiveness, demand excessive memory and processing power, making them impractical for large-scale applications and inaccessible to researchers with constrained computational resources. The lack of efficient low-resource implementations remains a significant barrier.
- Dataset Limitations and Generalizability: Most summarization models are trained on a limited set of benchmark datasets (e.g., CNN/Daily Mail, PubMed), which do not represent domain-specific requirements (e.g., legal, medical, financial texts). This limits the models' ability to generalize across diverse real-world applications.
- Lack of Empirical Comparisons between Hybrid and Transformer-Only Models: While hybrid approaches enhance efficiency, recent advancements in transformer optimizations (e.g., flash attention, token pruning, and retrieval-augmented generation) suggest that pure transformer models may achieve comparable efficiency without external preprocessing. However, comprehensive empirical studies comparing both approaches remain scarce.

Future Research Directions

To overcome these challenges, future research should focus on the following key areas:

- Context-Preserving Hybrid Approaches: Investigate adaptive segmentation techniques that dynamically adjust chunk sizes based on discourse structure. Graph-based models leveraging BERT-enhanced content selection and reinforcement learning-driven preprocessing could significantly improve coherence in segmented text summarization.
- Development of Low-Resource and Efficient Transformer Models: Future studies should prioritize lightweight transformer architectures that integrate sparse attention mechanisms and memory-efficient training strategies to make long-text summarization more accessible for resource-constrained environments.
- Expansion of Domain-Specific and Multilingual Datasets: To improve generalization, future research should focus on developing datasets that reflect real-world document structures across legal, scientific, and financial domains. Additionally, expanding low-resource language datasets will enhance inclusivity in summarization research.
- Comparative Studies on Hybrid vs. Transformer-Only Models: Future research should conduct large-scale comparative analyses to determine the optimal trade-offs between hybrid approaches and improved transformer architectures. Understanding when preprocessing-based hybrid methods outperform end-to-end transformer models remains an open challenge.

In addition to these directions, we propose a theoretical framework for multi-objective sentence selection, detailed in Appendix A. This framework addresses challenges such as redundancy, lack of coherence, and poor topic coverage in current graph-based preprocessing techniques by incorporating multiple objectives (semantic relevance, discourse coherence, topical coverage, and brevity) optimized via Particle Swarm Optimization. Though not implemented in this study, it represents a structured roadmap for future research aimed at improving the quality and efficiency of preprocessing in long-text abstractive summarization.

4 Limitation of the study and Conclusion

While this review focuses on transformer-based architectures specifically designed or adapted for long-text processing, it is important to acknowledge the growing influence of Large Language Models (LLMs) such as GPT-4, Claude, and Gemini. These models have demonstrated impressive performance in zero-shot and few-shot summarization, including for long documents. However, their proprietary nature, lack of reproducibility, and substantial computational requirements limit their integration in this study. Consequently, we prioritize open-source models and frameworks that support controlled comparisons and are feasible for resource-constrained research environments. Future work may incorporate LLMs into empirical evaluations or benchmark them against more efficient long-context transformers.

This systematic review provides a comprehensive analysis of long-text abstractive summarization using transformer models, highlighting key trends, methodologies, and challenges. The findings indicate that hybrid approaches—combining extractive preprocessing with abstractive techniques—are the most widely explored strategy for long-text summarization. While transformer models have significantly advanced the field, challenges remain in terms of computational cost, information loss, and scalability. Addressing these limitations requires improved segmentation and extraction techniques, more efficient transformer architectures, and standardized evaluation frameworks. Future research should focus on enhancing model generalization across diverse domains, integrating adaptive hybrid techniques, and expanding applications to low-resource languages. By addressing these gaps, the field can move toward more scalable and effective long-text summarization models.

Declarations

Acknowledgements

The authors gratefully acknowledge Aliko Dangote University, Wudil, Kano, Nigeria, for providing the opportunity and institutional support to conduct this research.

Funding

This research did not receive any funding.

Authors' Contributions

Bashir: Conceptualization, data curation, investigation, formal analysis, and writing—original draft. Usman and Abdulrahman: Visualization. Dr. Bichi: Proofreading and manuscript editing.

Competing Interests

The authors declare that they have no competing interests.

Availability of Data and Materials

The data used in this study are available upon request.

Appendix A

A Theoretical Framework for Multi-Objective Preprocessing in Long-Text Abstractive Summarization

Graph-based extractive preprocessing is a critical component in hybrid summarization pipelines. Traditional methods like TextRank and LexRank rely on single-objective similarity measures (e.g., cosine similarity, ROUGE), which often result in redundancy, poor coherence, and limited topic diversity—especially in long documents.

To overcome these limitations, we propose a multi-objective graph-based preprocessing framework that optimizes sentence selection across multiple criteria: semantic relevance, coherence, topical coverage, and brevity. Our framework employs Particle Swarm Optimization (PSO), selected for its fast convergence, low computational cost, and proven effectiveness in multi-objective optimization tasks [Patel *et al.*, 2025; Xue *et al.*, 2015]. Compared to Genetic Algorithms and Ant Colony Optimization, PSO avoids complex operations such as crossover or path modeling and has demonstrated superior performance in text-based applications [Chandrashekar and Sahin, 2014].

A.1 Multi-Objective Graph-Based Preprocessing Technique

The proposed framework enhances traditional graph-based approaches by integrating a meta-heuristic optimizer to dynamically balance the competing objectives of redundancy reduction, discourse coherence, and topic diversity. This preprocessing step improves the quality of inputs to downstream abstractive summarizers, especially transformer-based models.

- Relevance (R): Selecting sentences that contain core information aligned with the document's primary content.
- Coherence (C): Prioritizing logically structured sentences to enhance fluency and readability.
- Coverage (V): Maximizing the inclusion of diverse topics to ensure comprehensive summarization.
- Summary Length (L): Controlling the number of extracted sentences to maintain an optimal balance between brevity and informativeness.

A.2 Problem Definition

Let D be a document containing a set of n sentences:

$$D = \{S_1, S_2, \dots, S_n\}$$

The goal is to select a subset of sentences $X \subseteq D$ that optimally preserves key information while minimizing redundancy and ensuring coherence. This can be formulated as an optimization problem where the objective function is defined as:

$$\max \sum_{i=1}^{|X|} \left(w_1 R(S_i) + w_2 C(S_i) + w_3 V(S_i) + w_4 L(S_i)\right)$$

Where:

- $R(S_i)$: represents the relevance score of sentence S_i to the document.
- $C(S_i)$: represents the coherence score, ensuring logical sentence transitions.
- $V(S_i)$: represents the coverage score, maximizing topic diversity.
- $L(S_i)$: represents the summary length constraint, ensuring brevity.
- w_1, w_2, w_3, w_4: are weight parameters controlling the contribution of each factor.

This optimization problem is multi-objective and requires an efficient search algorithm to identify the optimal subset X.

A.3 Graph-Based Sentence Representation

We represent the document as a weighted undirected graph:

$$G = (V, E, W)$$

Where:

- V : is the set of nodes representing sentences.
- E : is the set of edges, where an edge exists between two sentences if their similarity exceeds a given threshold.
- W : is the weight matrix, where w_{ij} represents the similarity score between sentences S_i and S_j.
- The edge weights w_{ij} are computed using semantic similarity metrics, such as:

$$w_{ij} = \text{cosine}(\text{BERT}(S_i), \text{BERT}(S_j))$$

where BERT embeddings are used to capture semantic context rather than relying on traditional TF-IDF-based approaches.

A.4 Multi-Objective Optimization Formulation

- Objective 1: Relevance Maximization Relevance is computed based on sentence-document similarity:

$$R(S_i) = \frac{1}{|D|} \sum_{j=1}^{|D|} \text{cosine}(\text{BERT}(S_i), \text{BERT}(D_j))$$

Where: - D_j represents document-level sentence embeddings.
- Objective 2: Coherence Score We ensure coherence by maximizing adjacent sentence semantic alignment:

$$C(S_i) = \frac{1}{|X| - 1} \sum_{j \in X, j \neq i} w_{ij}$$

where higher $C(S_i)$ values indicate better local sentence transitions.
- Objective 3: Coverage Optimization Coverage is computed by ensuring that selected sentences span multiple topics using Latent Dirichlet Allocation (LDA) topic distributions:

$$V(S_i) = \sum_{k=1}^{K} p_k(S_i) \log p_k(D)$$

where $p_k(S_i)$ represents the probability of sentence S_i belonging to topic k, and $p_k(D)$ represents the document-wide topic distribution.
- Objective 4: Length Constraint Summary length $L(X)$ is controlled by penalizing excessive sentence selection:

$$L(X) = \exp\left(-\alpha \left|X - L_{\text{target}}\right|\right)$$

where L_{target} is the desired summary length, and α controls the penalty strength.

A.5 Particle Swarm Optimization for Sentence Selection

Given the multi-objective nature of the problem, we employ Particle Swarm Optimization (PSO) for sentence selection. PSO is an iterative meta-heuristic algorithm inspired by swarm intelligence. Each candidate solution (particle) represents a ranked list of sentences.

A.6 PSO Algorithm for Sentence Selection

1. **Initialization:** Generate an initial population of sentence subsets $X(0)$.
2. **Fitness Evaluation:** Compute the fitness of each subset using:

$$f(X) = w_1 R(X) + w_2 C(X) + w_3 V(X) + w_4 L(X)$$

3. **Velocity Update:** Each particle updates its velocity v_i as:

$$v_i^{(t+1)} = \omega v_i^{(t)} + c_1 r_1 (p_{\text{best},i} - X_i^{(t)}) + c_2 r_2 (g_{\text{best}} - X_i^{(t)})$$

where:
   - ω is the inertia weight controlling exploration.
   - c_1, c_2 are cognitive and social learning coefficients.
   - r_1, r_2 are random values in $[0, 1]$.
   - $p_{\text{best},i}$ is the best solution found by particle i.
   - g_{best} is the best solution found by the swarm.
4. **Position Update:**

$$X_i^{(t+1)} = X_i^{(t)} + v_i^{(t+1)}$$

5. **Convergence Check:** The algorithm stops when a convergence criterion is met (e.g., no significant improvement in fitness).

A.7 Implementation Strategy of the Proposed Framework

The Multi-Objective Graph-Based Preprocessing Technique is designed to improve sentence selection before feeding long-text data into a transformer-based summarization model. This framework enhances context retention, coherence, and diversity while minimizing redundancy and computational overhead.

To achieve this, the framework consists of the following phases:

1. **Graph-Based Sentence Representation** – Each sentence is represented as a node in a weighted graph, with edge weights computed using semantic similarity metrics derived from BERT embeddings.
2. **Multi-Objective Optimization with PSO** – The sentence selection process is optimized using Particle Swarm Optimization (PSO), balancing four key objectives: relevance, coherence, coverage, and length constraints.
3. **Preprocessed Output for Summarization** – The optimized subset of sentences is passed to a transformer-based abstractive summarization model to generate the final summary.

Table 7. The following pseudo code outlines the structured steps of the proposed technique

Step	Description
Step 1: Preprocessing	Tokenize document D into sentences $\{S_1, S_2, ..., S_n\}$ Generate sentence embeddings E using a pre-trained BERT model Normalize embeddings for consistency
Step 2: Graph Construction	Initialize an empty graph $G = (V, E)$ For each sentence S_i in D: - Add node V_i to graph G - Compute cosine similarity between V_i and all other nodes - If similarity > threshold, add edge $E_{i,j}$ with weight W_{ij}
Step 3: Multi-Objective Optimization (PSO)	Initialize a swarm of particles (candidate summaries) Define the multi-objective function: - Relevance $R(X) \rightarrow$ Selects sentences with high document similarity - Coherence $C(X) \rightarrow$ Ensures logical sentence transitions - Coverage $V(X) \rightarrow$ Maximizes inclusion of diverse topics - Length Constraint $L(X) \rightarrow$ Penalizes overly long summaries While termination criteria not met: - Evaluate fitness of each particle based on the objective function - Update particle velocities and positions using: $$v_i^{(t+1)} = \omega v_i^{(t)} + c_1 r_1 (p_{\text{best},i} - X_i^{(t)}) + c_2 r_2 (g_{\text{best}} - X_i^{(t)})$$ - Adjust particle positions (selected sentences) accordingly - Store the best candidate summary found
Step 4: Abstractive Summarization	Select optimized set of sentences X_{best} from the best particle Concatenate selected sentences into an input sequence Feed the sequence into a transformer-based summarization model (e.g., BART, Pegasus) Generate final abstractive summary S
Return	Optimized Summary S

This pseudo code (Table 7) provides a structured implementation roadmap, ensuring reproducibility and clarity for future research. While this study presents the framework as a theoretical contribution, future work should focus on empirical validation, benchmarking against existing summarization techniques, and optimizing computational efficiency.

While this framework is theoretical, it directly addresses the shortcomings of existing graph-based preprocessing methods by introducing a balanced, multi-faceted optimization process. Future work will involve implementing and benchmarking this approach against current baseline models to assess gains in coherence, informativeness, and computational efficiency.

References

Aksenov, D., Moreno-Schneider, J., Bourgonje, P., Schwarzenberg, R., Hennig, L., and Rehm, G. (2020). Abstractive text summarization based on language model conditioning and locality modeling. *arXiv preprint arXiv:2003.13027*. DOI: 10.48550/arxiv.2003.13027.

Akter, M., Çano, E., Weber, E., Dobler, D., and Habernal, I. (2025). A comprehensive survey on legal summarization: Challenges and future directions. *arXiv preprint arXiv:2501.17830*. DOI: 10.48550/arxiv.2501.17830.

Altmami, N. I. and Menai, M. E. B. (2022). Automatic summarization of scientific articles: A survey. *Journal of King Saud University-Computer and Information Sciences*, 34(4):1011–1028. DOI: 10.1016/j.jksuci.2020.04.020.

Angioi, M. and Hiller, C. E. (2023). Systematic literature reviews. *Research Methods in the Dance Sciences, M. Angioi and CE Hiller, Eds. University Press of Florida*, pages 265–280. DOI: 10.2307/j.ctv33jb41z.24.

Bajaj, A., Dangati, P., Krishna, K., Kumar, P. A., Uppaal, R., Windsor, B., Brenner, E., Dotterrer, D., Das, R., and McCallum, A. (2021). Long document summarization in a low resource setting using pretrained language models. *arXiv preprint arXiv:2103.00751*. DOI: 10.18653/v1/2021.acl-srw.7.

Benedetto, I., La Quatra, M., Cagliero, L., Vassio, L., and Trevisan, M. (2024). Tasp: Topic-based abstractive summarization of facebook text posts. *Expert Systems with Applications*, 255:124567. DOI: 10.1016/j.eswa.2024.124567.

Bettayeb, M., Halawani, Y., Khan, M. U., Saleh, H., and Mohammad, B. (2024). Efficient memristor accelerator for transformer self-attention functionality. *Scientific Reports*, 14(1):24173. DOI: 10.1038/s41598-024-75021-z.

Calizzano, R., Ostendorff, M., Ruan, Q., and Rehm, G. (2022). Generating extended and multilingual summaries with pretrained transformers. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1640–1650. Available at:`https://aclanthology.org/`

2022.lrec-1.175/.

Cao, S. and Wang, L. (2022). Hibrids: Attention with hierarchical biases for structure-aware long document summarization. *arXiv preprint arXiv:2203.10741*. DOI: 10.18653/v1/2022.acl-long.58.

Cao, S. and Wang, L. (2023). Awesome: Gpu memory-constrained long document summarization using memory mechanism and global salient content. *arXiv preprint arXiv:2305.14806*. DOI: 10.18653/v1/2024.naacl-long.330.

Chandrashekar, G. and Sahin, F. (2014). A survey on feature selection methods. *Computers & electrical engineering*, 40(1):16–28. DOI: 10.1016/j.compeleceng.2013.11.024.

Chaves, A., Kesiku, C., and García-Zapirain, B. (2022). Automatic text summarization of biomedical text data: A systematic review.

Chen, W. and Iwaihara, M. (2023). Efficient summarization of long documents using hybrid extractive-abstractive method. Available at:https://proceedings-of-deim.github.io/DEIM2023/1b-5-2.pdf.

Chen, Y., Wan, Z., Li, Y., He, X., Wei, X., and Han, J. (2024). Graph curvature flow-based masked attention. *Journal of Chemical Information and Modeling*, 64(21):8153–8163. DOI: 10.1021/acs.jcim.4c01616.

Chu, C.-L., Chen, Y.-C., Cheng, W., Lin, C., and Chang, Y.-H. (2024). Attentionrc: A novel approach to improve locality sensitive hashing attention on dual-addressing memory. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 43(11):3925–3936. DOI: 10.1109/tcad.2024.3447217.

Dai, W. and He, Q. (2024). Automatic summarization model based on clustering algorithm. *Scientific Reports*, 14(1):15302. DOI: 10.21203/rs.3.rs-3992927/v1.

Dat, D. H., Anh, D. D., Luu, A. T., and Buntine, W. (2024). Discrete diffusion language model for long text summarization. *arXiv e-prints*, pages arXiv–2407. DOI: 10.48550/arxiv.2407.10998.

Ezzat, Z., Khalfallah, A., and Khoriba, G. (2024). Fused transformers: Fused information of arabic long article for summarization. *Procedia Computer Science*, 244:96–104. DOI: 10.1016/j.procs.2024.10.182.

Fikri, F. B., Oflazer, K., and Yanıkoğlu, B. (2024). Abstractive summarization with deep reinforcement learning using semantic similarity rewards. *Natural Language Engineering*, 30(3):554–576. DOI: 10.1017/s1351324923000505.

Fu, X. (2024). Transformer models in text summarization. *Applied and Computational Engineering*, 101(1):35–41. DOI: 10.54254/2755-2721/101/20240946.

Gokhan, T., Price, M. J., and Lee, M. (2024). Graphs in clusters: a hybrid approach to unsupervised extractive long document summarization using language models. *Artificial Intelligence Review*, 57(7):189. DOI: 10.1007/s10462-024-10828-w.

Hardy, H., Ballesteros, M., Ladhak, F., Khalifa, M., Castelli, V., and McKeown, K. (2022). Novel chapter abstractive summarization using spinal tree aware sub-sentential content selection. *arXiv preprint arXiv:2211.04903*. DOI: 10.48550/arxiv.2211.04903.

He, Y. (2024). Research on the optimization model of semantic coherence and fluency in language translation. *Applied Mathematics and Nonlinear Sciences*, 9(1). DOI: 10.2478/amns-2024-2769.

Huang, L., Cao, S., Parulian, N., Ji, H., and Wang, L. (2021a). Efficient attentions for long document summarization. *arXiv preprint arXiv:2104.02112*. DOI: 10.18653/v1/2021.naacl-main.112.

Huang, Y., Li, Z., Chen, Z., Zhang, C., and Ma, H. (2024). Sentence salience contrastive learning for abstractive text summarization. *Neurocomputing*, 593:127808. DOI: 10.1016/j.neucom.2024.127808.

Huang, Y., Sun, L., Han, C., and Guo, J. (2023). A high-precision two-stage legal judgment summarization. *Mathematics*, 11(6):1320. DOI: 10.3390/math11061320.

Huang, Y., Yu, Z., Guo, J., Xiang, Y., and Xian, Y. (2021b). Element graph-augmented abstractive summarization for legal public opinion news with graph transformer. *Neurocomputing*, 460:166–180. DOI: 10.1016/j.neucom.2021.07.013.

Jain, D., Borah, M. D., and Biswas, A. (2021). Summarization of legal documents: Where are we now and the way forward. *Computer Science Review*, 40:100388. DOI: 10.1016/j.cosrev.2021.100388.

Jain, D., Borah, M. D., and Biswas, A. (2024). Summarization of lengthy legal documents via abstractive dataset building: An extract-then-assign approach. *Expert Systems with Applications*, 237:121571. DOI: 10.1016/j.eswa.2023.121571.

Jeeson-Daniel, A., Lin, C., Mila-Quebec AI Institute, and Lau, E. (2021). Abstractive summarization using longformer-pegasus. Available at:https://yunglau.github.io/assets/pdf/Long_PEGASUS.pdf.

Karlbom, H. and Clifton, A. (2020). Abstractive podcast summarization using bart with longformer attention. In *The 29th Text Retrieval Conference (TREC) notebook. NIST*. Available at:https://trec.nist.gov/pubs/trec29/papers/hk_uu_podcast.P.pdf.

Kashyap, P. (2022). Coling 2022 shared task: Led finteuning and recursive summary generation for automatic summarization of chapters from novels. In *Proceedings of The Workshop on Automatic Summarization for Creative Writing*, pages 19–23. Available at:https://aclanthology.org/2022.creativesumm-1.3/.

Kirstein, F., Wahle, J. P., Gipp, B., and Ruas, T. (2025). Cads: A systematic literature review on the challenges of abstractive dialogue summarization. *Journal of Artificial Intelligence Research*, 82:313–365. DOI: 10.1613/jair.1.16674.

Kiruluta, A., Lemos, A., and Lundy, E. (2021). New approaches to long document summarization: Fourier transform based attention in a transformer model. *arXiv preprint arXiv:2111.15473*. DOI: 10.48550/arXiv.2111.15473.

Koh, H. Y., Ju, J., Liu, M., and Pan, S. (2023). An empirical survey on long document summarization: Datasets, models, and metrics. *ACM computing surveys*, 55(8):1–35. DOI: 10.1145/3545176.

Kumar, S., Kohli, G. S., Shinde, K., and Ekbal, A. (2022). Team ainlpml@ mup in sdp 2021: scientific document summarization by end-to-end extractive and abstractive approach. In *Proceedings of the Third Workshop on Schol-*

arly Document Processing, pages 285–290. Available at:https://aclanthology.org/2022.sdp-1.36/.

Liao, P., Zhang, C., Chen, X., and Zhou, X. (2020). Improving abstractive text summarization with history aggregation. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9. IEEE. DOI: 10.1109/ijcnn48605.2020.9207502.

Lim, J. and Song, H.-J. (2023). Improving multi-stage long document summarization with enhanced coarse summarizer. In *Proceedings of the 4th New Frontiers in Summarization Workshop*, pages 135–144. DOI: 10.18653/v1/2023.newsum-1.13.

Lin, C.-Y. (2004). Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81. Available at:https://aclanthology.org/W04-1013.pdf.

Liu, D., Hong, X., Lin, P.-J., Chang, E., and Demberg, V. (2022a). Two-stage movie script summarization: An efficient method for low-resource long document summarization. In *Proceedings of The Workshop on Automatic Summarization for Creative Writing*, pages 57–66. Available at:https://aclanthology.org/2022.creativesumm-1.9/.

Liu, S., Cao, J., Li, Y., Yang, R., and Wen, Z. (2024). Low-resource court judgment summarization for common law systems. *Information Processing & Management*, 61(5):103796. DOI: 10.1016/j.ipm.2024.103796.

Liu, X. and Xu, Y. (2023). Learning to rank utterances for query-focused meeting summarization. *arXiv preprint arXiv:2305.12753*. DOI: 10.18653/v1/2023.findings-acl.538.

Liu, Y., Ni, A., Nan, L., Deb, B., Zhu, C., Awadallah, A. H., and Radev, D. (2022b). Leveraging locality in abstractive text summarization. *arXiv preprint arXiv:2205.12476*. DOI: 10.18653/v1/2022.emnlp-main.408.

Liu, Z. and Chen, N. F. (2022). Dynamic sliding window modeling for abstractive meeting summarization. In *INTERSPEECH*, pages 5150–5154. Available at:https://www.isca-archive.org/interspeech_2022/liu22b_interspeech.pdf.

Lu, G., Larcher, S. B., and Tran, T. (2023). Hybrid long document summarization using c2f-far and chatgpt: A practical study. *arXiv preprint arXiv:2306.01169*. DOI: 10.48550/arxiv.2306.01169.

Ma, Y. and Zong, L. (2020). Neural abstractive multi-document summarization: Hierarchical or flat structure? In *Proceedings of the Second International Workshop of Discourse Processing*, pages 29–37. DOI: 10.18653/v1/2020.iwdp-1.6.

Madan, S., Lentzen, M., Brandt, J., Rueckert, D., Hofmann-Apitius, M., and Fröhlich, H. (2024). Transformer models in biomedicine. *BMC medical informatics and decision making*, 24(1):214. DOI: 10.1186/s12911-024-02600-5.

Mei, A., Kabir, A., Bapat, R., Judge, J., Sun, T., and Wang, W. Y. (2022). Learning to prioritize: Precision-driven sentence filtering for long text summarization. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 313–318. Available at:https://aclanthology.org/2022.lrec-1.33/.

Moro, G. and Ragazzi, L. (2023). Align-then-abstract representation learning for low-resource summarization. *Neurocomputing*, 548:126356. DOI: 10.1016/j.neucom.2023.126356.

Nguyen, H. and Ding, J. (2023). Keyword-based augmentation method to enhance abstractive summarization for legal documents. In *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law*, pages 437–441. DOI: 10.1145/3594536.3595120.

Obonyo, I., Casola, S., and Saggion, H. (2022). Exploring the limits of a base bart for multi-document summarization in the medical domain. In *Proceedings of the Third Workshop on Scholarly Document Processing*, pages 193–198. Available at:https://aclanthology.org/2022.sdp-1.23/.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., and Moher, D. (2021). Updating guidance for reporting systematic reviews: development of the prisma 2020 statement. *Journal of clinical epidemiology*, 134:103–112. DOI: 10.31222/osf.io/jb4dx.

Pang, B., Nijkamp, E., Kryściński, W., Savarese, S., Zhou, Y., and Xiong, C. (2022). Long document summarization with top-down and bottom-up inference. *arXiv preprint arXiv:2203.07586*. DOI: 10.18653/v1/2023.findings-eacl.94.

Pant, M. and Chopra, A. (2022). Multilingual financial documentation summarization by team_tredence for fns2022. In *Proceedings of the 4th Financial Narrative Processing Workshop@ LREC2022*, pages 112–115. Available at:https://aclanthology.org/2022.fnp-1.18/.

Patel, D., Saxena, A. K., and Dubey, A. (2025). A particle swarm optimization based model for feature selection. In *Innovative and Intelligent Digital Technologies; Towards an Increased Efficiency: Volume 2*, pages 329–340. Springer. DOI: 10.1007/978-3-031-71649-$2_2$8.

Petticrew, M. and Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons. Book.

Pilault, J., Li, R., Subramanian, S., and Pal, C. (2020). On extractive and abstractive neural document summarization with transformer language models. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, pages 9308–9319. DOI: 10.18653/v1/2020.emnlp-main.748.

Pu, D., Wang, Y., and Demberg, V. (2024). Incorporating distributions of discourse structure for long document abstractive summarization. *arXiv preprint arXiv:2305.16784*. DOI: 10.18653/v1/2023.acl-long.306.

Rahman, S., Labib, M. A., Murad, H., and Das, U. (2024). Cuet_sstm at the gem'24 summarization task: Integration of extractive and abstractive method for long text summarization in swahili language. In *Proceedings of the 17th International Natural Language Generation Conference: Generation Challenges*, pages 112–117. DOI: 10.18653/v1/2024.inlg-genchal.12.

Ranggianto, N. A., Purwitasari, D., Fatichah, C., and Sholikah, R. W. (2023). Abstractive and extractive approaches for summarizing multi-document travel reviews. *Jurnal RESTI*, 7(6):1464–1475. DOI:

10.29207/resti.v7i6.5170.

Rennard, V., Shang, G., Hunter, J., and Vazirgiannis, M. (2023). Abstractive meeting summarization: A survey. *Transactions of the Association for Computational Linguistics*, 11:861–884. DOI: 10.1162/tacl$_{a0}$0578.

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., and Koffel, J. B. (2021). Prisma-s: an extension to the prisma statement for reporting literature searches in systematic reviews. *Systematic reviews*, 10(1):39. DOI: 10.31219/osf.io/sfc38.

Saleh, M. E., Wazery, Y. M., and Ali, A. A. (2024). A systematic literature review of deep learning-based text summarization: Techniques, input representation, training strategies, mechanisms, datasets, evaluation, and challenges. *Expert Systems with Applications*, 252:124153. DOI: 10.1016/j.eswa.2024.124153.

Sanchan, N. (2024). Comparative study on automated reference summary generation using bert models and rouge score assessment. *Journal of Current Science and Technology*, 14(2):26–26. DOI: 10.59796/jcst.v14n2.2024.26.

Saxena, R. and Keller, F. (2024). Select and summarize: Scene saliency for movie script summarization. *arXiv preprint arXiv:2404.03561*. DOI: 10.18653/v1/2024.findings-naacl.218.

Steblianko, O., Shymkovych, V., Kravets, P., Novatskyi, A., and Shymkovych, L. (2024). Scientific article summarization model with unbounded input length. *Information, Computing and Intelligent systems*, (5):150–158. DOI: 10.20535/2786-8729.5.2024.314724.

Sun, W., Fang, C., Chen, Y., Zhang, Q., Tao, G., You, Y., Han, T., Ge, Y., Hu, Y., Luo, B., *et al*. (2024). An extractive-and-abstractive framework for source code summarization. *ACM Transactions on Software Engineering and Methodology*, 33(3):1–39. DOI: 10.1145/3632742.

Tretyak, V. and Stepanov, D. (2020). Combination of abstractive and extractive approaches for summarization of long scientific texts. *arXiv preprint arXiv:2006.05354*. DOI: 10.48550/arxiv.2006.05354.

Ulker, M. and Ozer, A. B. (2024). Abstractive summarization model for summarizing scientific article. *IEEE Access*, 12:91252–91262. DOI: 10.1109/access.2024.3420163.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30. DOI: 10.48550/arxiv.1706.03762.

Wang, G., Garg, P., and Wu, W. (2024a). Segmented summarization and refinement: A pipeline for long-document analysis on social media. *Journal of Social Computing*, 5(2):132–144. DOI: 10.23919/jsc.2024.0010.

Wang, T., Yang, C., Zou, M., Liang, J., Xiang, D., Yang, W., Wang, H., and Li, J. (2024b). A study of extractive summarization of long documents incorporating local topic and hierarchical information. *Scientific Reports*, 14(1):10140. DOI: 10.1038/s41598-024-60779-z.

Wibawa, A. P., Kurniawan, F., *et al*. (2024). A survey of text summarization: Techniques, evaluation and challenges. *Natural Language Processing Journal*, 7:100070. DOI: 10.1016/j.nlp.2024.100070.

Wilman, P., Atara, T., and Suhartono, D. (2024). Abstractive english document summarization using bart model with chunk method. *Procedia Computer Science*, 245:1010–1019. DOI: 10.1016/j.procs.2024.10.329.

Wu, W., Li, W., Xiao, X., Liu, J., Cao, Z., Li, S., Wu, H., and Wang, H. (2021). Bass: Boosting abstractive summarization with unified semantic graph. *arXiv preprint arXiv:2105.12041*. DOI: 10.18653/v1/2021.acl-long.472.

Wu, Y., Li, H., Nenadic, G., and Zeng, X.-J. (2024). Extract-and-abstract: Unifying extractive and abstractive summarization within single encoder-decoder framework. *arXiv preprint arXiv:2409.11827*. DOI: 10.48550/arxiv.2409.11827.

Wu, Y.-H., Lin, Y.-J., and Kao, H.-Y. (2023). Ikm_lab at biolaysumm task 1: Longformer-based prompt tuning for biomedical lay summary generation. In *The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks*, pages 602–610. DOI: 10.18653/v1/2023.bionlp-1.64.

Xue, B., Zhang, M., Browne, W. N., and Yao, X. (2015). A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on evolutionary computation*, 20(4):606–626. DOI: 10.26686/wgtn.14214497.v1.

Ying, S., Zhao, Z. Y., and Zou, W. (2021). Longsumm 2021: Session based automatic summarization model for scientific document. In *Proceedings of the Second Workshop on Scholarly Document Processing*, pages 97–102. DOI: 10.18653/v1/2021.sdp-1.12.

You, Z., Radhakrishna, S., Ming, S., and Kilicoglu, H. (2024). Uiuc_bionlp at biolaysumm: an extract-then-summarize approach augmented with wikipedia knowledge for biomedical lay summarization. In *Proceedings of the 23rd Workshop on Biomedical Natural Language Processing*, pages 132–143. DOI: 10.18653/v1/2024.bionlp-1.11.

Yu, T., Ji, Z., and Fung, P. (2023). Improving query-focused meeting summarization with query-relevant knowledge. *arXiv preprint arXiv:2309.02105*. DOI: 10.18653/v1/2023.findings-ijcnlp.5.

Yuan, R., Wang, Z., Cao, Z., and Li, W. (2023). Preserve context information for extract-generate long-input summarization framework. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 13932–13939. DOI: 10.1609/aaai.v37i11.26631.

Zhang, M., Lu, J., Yang, J., Zhou, J., Wan, M., and Zhang, X. (2024). From coarse to fine: Enhancing multi-document summarization with multi-granularity relationship-based extractor. *Information Processing & Management*, 61(3):103696. DOI: 10.1016/j.ipm.2024.103696.

Zhang, M., Zhou, G., Yu, W., and Liu, W. (2021). Ki-habs: Key information guided hierarchical abstractive summarization. *KSII Transactions on Internet & Information Systems*, 15(12). Available at:`https://koreascience.kr/article/JAKO202104554882565.page`.

Zhang, X., Meng, K., and Liu, G. (2019). Hie-transformer: a hierarchical hybrid transformer for abstractive article summarization. In *International Conference on Neural Information Processing*, pages 248–258. Springer. DOI: 10.1007/978-3-030-36718-3$_2$1.

Zhang, Y., Ni, A., Mao, Z., Wu, C. H., Zhu, C., Deb, B., Awadallah, A. H., Radev, D., and Zhang, R. (2022).

Summ^ n: A multi-stage summarization framework for long input dialogues and documents. *arXiv preprint arXiv:2110.10150*. DOI: 10.48550/arXiv.2110.10150.

Zhao, Y., Saleh, M., and Liu, P. J. (2020). Seal: Segment-wise extractive-abstractive long-form text summarization. *arXiv preprint arXiv:2006.10213*. DOI: 10.48550/arxiv.2006.10213.