International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VII **Month of publication:** July 2025

DOI: https://doi.org/10.22214/ijraset.2025.73052

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

Enhancing Technical Documentation through Intelligent Text Summarization Techniques

Rajpal Kaur
Technical Writer, Ivedha Inc., 18 Wynford Drive, Suite 306, Toronto ON M3C 3S2.

Abstract: In an era of fast digital transformation, technical documentation is more important than ever in aiding user knowledge, upkeep of systems, and operational efficiency across a variety of organizations. However, the ever-growing complexity of software platforms, enterprise applications, and IT infrastructures has resulted in a massive amount of technical content that is challenging to navigate and time-consuming to comprehend. Users, including developers, executives, end users, and support engineers, deserve accurate and easily accessible documentation. This study investigates the use of text summarizing techniques in technical documentation workflows to address the issues and improve the overall quality, usability, and efficacy of such content. Text Summarization (TS) entails condensing extensive text into brief forms while preserving its basic meaning. In technical documentation, this feature promotes faster information extraction, comprehension, and user engagement. The study defines two main summary techniques—extractive and abstractive—and assesses their efficacy in a documentation setting. Extractive summarization extracts essential lines or phrases straight from the source material while keeping the underlying structure and vocabulary, which is especially useful in circumstances that need technical precision. In contrast, abstractive summarization paraphrases and rewrites the text in a more reduced manner, resulting in greater fluidity and readability. This study proposes a hybrid model that combines these approaches to achieve a balance of clarity and accuracy. The process involves integrating traditional and transformer-based models like BERT, T5, and PEGASUS to technical documentation datasets. Using supervised fine-tuning and domain-specific corpora, the models are trained to provide summaries that are suited to different user needs. Finally, using text summarizing algorithms in technical documentation is a significant step toward more efficient, user-friendly, and intelligent content delivery. This study establishes the groundwork for creating adaptable documentation systems that match the changing needs of current users.
Keywords: BERT, Deep learning, PEGASUS, T5, Technical documentation, Text summarization.

I. INTRODUCTION

Enterprise IT systems are naturally complicated due to their size, integration of numerous technologies, diversified user base, and the importance of the commercial operations they support, which range from finance and human resources to supply chain management and customer interaction systems. These systems frequently have custom variations, older components, and ongoing changes, making them difficult to use and manage without clear instructions. From a research standpoint, this degree of complexity emphasizes the critical necessity for excellent documentation that not only transmits accurate technical information but also improves readability by meeting users' diverse cognitive and operational needs [1]. Poorly written documentation can result in higher support costs, decreased system effectiveness, and user irritation, emphasizing the necessity of a human-centered approach to technological communication in bridging the gap between system functionality and user comprehension [2].

Conventional documentation in enterprise IT systems frequently fails to assist end users effectively because it focuses on system performance and technical correctness over user satisfaction and its context of use. According to research, such documentation tends to come from a developer-centric or system-oriented viewpoint, with jargon-heavy vocabulary, linear structures, and static layouts that do not reflect how users obtain information in real-world tasks [3]. This gap causes difficulties in navigation, comprehension, and task execution, especially for non-technical users or those working under time limitations. As a result, users may forgo documentation entirely in favor of casual support avenues such as assistance from others or trial-and-error, increasing the likelihood of errors and inefficiencies. These issues highlight the important need for reconsidering documentation via a human-centered, usability-driven lens that aligns information with user intent, behaviour, and learning styles [4].

To make IT documentation more readable, readily available, and human-centered, research advises implementing Human-Centered Design (HCD) concepts that focus on recognizing user needs, behaviors, and settings. This process begins with conducting user research through interviews, questionnaires, and usability testing to understand how various user groups—such as novice users, administrators, and support staff—interact with documentation.

The content should then be adapted to these personas, utilizing clear language, visual aids, and modular architectures that facilitate efficient navigation and task fulfilment [5]. The structure of information should be instinctive, with concise headings, searchable material, and progressive exposure strategies that allow users to discover details as needed without feeling overburdened. Using responsive design improves compatibility across devices, while standards such as WCAG (Web Content Accessibility Guidelines) promote inclusivity for persons with disabilities.

Making documentation human-centered requires the integration of feedback chains and continuous content upgrades. Tools that let users review, comment, or flag unclear stuff can help encourage continuous development. Furthermore, including documentation into user interfaces—as contextual help, tooltips, or chat-based assistants—can significantly minimize cognitive strain and improve real-time issue solving. Research also recommends the use of analytics to measure how documentation is used, which material is often accessed or disregarded, and where users drop off, allowing for data-driven modification [6]. Finally, a human-centered documentation strategy improves usability while also lowering support costs and increasing general system adoption and contentment.

This research paper explores various techniques employed in technical documentation, with a primary emphasis on the role and impact of TS. It examines how summarization contributes to enhancing the clarity, accessibility, and efficiency of documentation by condensing complex technical content into concise, user-friendly formats [7]. TS is valuable in technical documentation because it allows for the synthesis of difficult, brittle, and often extremely specialized information into compact and effortlessly digestible formats, increasing accessibility and ease of use for a wide variety of users. Summaries are useful in contexts where time and transparency are crucial, such as software manuals, user guides, release notes, or API documentation [8]. They allow users to quickly understand the core ideas or methodological overviews without navigating through detailed technical information. This is especially useful for stakeholders like as decision-makers, new users, and non-technical audiences who demand a high-level knowledge without the need for detailed details. Furthermore, a summary enables layered material presentation, with concise overviews accompanied by extensive explanations for both novice and expert users [9]. It also plays an important part in digital documentation ecosystems by enhancing searchability and allowing AI-powered applications like chatbots, auto-suggest systems, and content indexing engines to respond faster and more relevantly [10]. As corporate systems become more complex, including summarizing techniques—whether manual or automated—into technical documentation streamlines information delivery, decreases cognitive burden, and greatly enhances user experience, comprehension, and productivity.

II. HANDLING GLITCHES IN TECHNICAL DOCUMENTATION

Technical documentation flaws, such as discrepancies, contradictions, broken links, or usability concerns, must be addressed by an amalgam of editorial rigor, user feedback, and technology-driven quality assurance. Table 1 depicts essential strategies from a research and best-practices perspective.

Table 1. Techniques of handling Technical Documentation

Category	Technique	Purpose
User-Centered Testing	Usability Testing	Identifies real-world pain points and unclear instructions
	Cognitive Walkthroughs	Detects logical gaps from the user's perspective
	Feedback Mechanisms (comments, ratings)	Captures user suggestions and confusion directly
Consistency and Standards	Use of Style Guides (e.g., Microsoft, Google)	Maintains uniform tone, terminology, and structure
	Terminology Management	Prevents inconsistent use of technical terms
	Content Templates	Ensures structural uniformity across documentation
Automation and Tools	Spell/Grammar Checkers (e.g., Grammarly, LanguageTool)	Catch language errors
	Text Summarization	Summarizes the lengthy texts

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

Category	Technique	Purpose
	Link Validators	Detects broken or incorrect hyperlinks
	Code Snippet Validators	Verifies code samples run as intended
Version Control	Git-Based Documentation Repositories	Tracks changes and supports collaborative editing
	Change Logs	Helps align documentation with product/system updates
Structured Authoring	Topic-Based Authoring (TBA)	Promotes modular and reusable content
	DITA, Markdown, or XML Documentation Frameworks	Enables consistent formatting and reuse
Review Process	Peer Review by SMEs	Ensures technical accuracy and relevance
	Documentation Review Sprints	Syncs documentation updates with agile development cycles
Accessibility & Readability	Readability Tools (e.g., Hemingway, Readable.io)	Improves clarity and reading ease for diverse users
	WCAG Compliance Checkers	Ensures accessibility for users with disabilities

III. SIGNIFICANCE OF TEXT SUMMARIZATION IN TECHNICAL DOCUMENTATION

Text summary has grown into an essential technique in technical documentation, particularly as the sheer amount and complexity of digital content increases across corporate systems, software platforms, and user manuals. In such cases, end customers frequently demand quick, precise access to information without having to read large publications in their entirety. Text summary helps to close this gap by compressing complicated technical knowledge into simple, short summaries that preserve key context. This not only minimizes cognitive stress for consumers but also improves the overall usability of documentation. Summarization, when properly performed, allows users to quickly comprehend core functionality, procedures, or troubleshooting steps, increasing productivity, minimizing support questions, and encouraging a more effective overall user experience. Summarization also works well with human-centered design principles since it caters to varied user preferences, whether a rapid overview is required or a greater engagement with complete content is preferable. Furthermore, in the age of automation and AI, TS is becoming a growing aspect of intelligent documentation systems. Technical content can be proactively synthesized for various use cases, such as tooltips, onboarding procedures, or chatbot responses, using extractive or abstractive techniques, which are frequently powered by advanced models such as BERT, T5, or PEGASUS. This adaptability guarantees that documentation is not only accessible but also relevant to the situation and scalable across platforms. Summarization helps technical authors and developers maintain consistency and relevancy by emphasizing major modifications or additions in huge amounts of text. Finally, text summary helps to make technical documentation more accessible, responsive, and relevant to user tasks, transforming static material into a dynamic knowledge asset that effectively meets both technical and non-technical audiences. Text summary can dramatically improve the readability of technical documentation by reducing dense, complicated information to short easy to easy-to-comprehend formats. Long narratives, complex methods, and domain-specific terminology are common features of technical papers, which may overwhelm or confound users, particularly non-experts. Summing up key points, instructions, or notions makes documentation more digestible and user-friendly, permitting readers to immediately absorb important information without having to trawl through vast blocks of text. This is especially useful in user manuals, product guides, and API documentation, where users constantly search for specific answers or procedures. Furthermore, a summary enables layered content design, with brief overviews serving as entry points to more extensive interpretations. This accommodates readers with different degrees of experience and time constraints, increasing engagement and comprehension. It also adheres to minimalist documentation standards by reducing verbosity and emphasizing task-oriented content. When combined with AI technologies and participatory platforms, real-time summarization can potentially provide dynamic, context-aware assistance to users. Overall, text summarizing enhances readability by making content easier to navigate, scannable, and relevant to the user's goals, thus increasing the productivity and efficacy of technical communication.

IV. LITERATURE REVIEW

The literature review provides an overview of existing research and scholarly work related to the research topic. It helps identify gaps, trends, and key findings that inform and support the current study. This section establishes the theoretical foundation and context for the research.

Fahd A. Ghanem et al., 2025 [11] proposed a Deep Learning (DL) system for automatic brief TS on Twitter. The suggested solution combines Bidirectional Encoder Representations from Transformers (BERT) with a Transformer-based Encoder-Decoder Architecture (TEDA), which includes a mechanism for attention to improve contextual awareness. Long Short-Term Memory (LSTM) networks are also included in BERT to efficiently capture dependencies that are long in tweets and summaries. The proposed framework's performance was evaluated against three benchmark Twitter datasets—Hagupit, SHShoot, and Hyderabad Blast—using ROUGE scores as the assessment metric. Results from experiments show that the model outperforms previous techniques for reliably extracting crucial information from tweets. These results demonstrate the framework's efficiency in automated short text summaries, providing a solid method for efficiently processing and summarizing massive amounts of social media content.

Mohmmadali M. Saiyyad et al., 2023 [12] employed Deep Learning (DL), which caused a paradigm shift in the way Natural Language Processing (NLP) is performed. The use of DL approaches has enabled considerable development in the disciplines of sentiment analysis, text translation, and TS. The authors investigated various DL algorithms for TS using experimental results presented by other researchers. The authors found that all DL methods have advantages and disadvantages. After comparing various strategies, the authors discovered that the pre-trained transformer produces the best outcomes for TS. In the future, combinations of traditional approaches and pre-trained transformers will be examined for improved performance. Multi-modal reiterating, which contains not just text but also visuals and maybe audio data, has an opportunity to significantly enhance the overall production process.

Yang Yang, Zhilei Wu et al., 2022 [13] described the fundamental ideas of IE and DL, focusing on the research advances and accomplishments of DL technologies within the field of IE. At the level of IE tasks, it elaborates on three aspects: entity relationship, the extraction procedure, event extraction, and multi-modal information extraction, and conducts an in-depth comparison of various extraction approaches. The authors also describe the opportunities for growth and trends in DL in the realm of IE, as well as the challenges that require additional investigation. At the approach level, research is expected to focus on multi-model and multi-task joint extraction, information extraction based on knowledge improvements, and information fusion based on multi-modal. At the model level, more research should be conducted to strengthen theoretical research, make models lighter, and improve model generalization capacity.

K. K. Mamidala et al., 2021 [14] presented an experimental method to extract an abstract from e-news articles in the Telugu language. An innovative lexical parameter-based method for information extraction has been introduced, intended for scoring sentences. Depending on the occurrence of the events or named entities in the document, the sentences are designated for the summary. The performance metrics like recall, precision, and F1 score have been calculated to measure the performance of the anticipated method.

W. S. El-Kassas et al., 2021 [15] explained the different categorizations and applications of the ATS. The authors provided a systematic review of the different methods of ATS and performed a categorization of different building blocks and techniques used for designing and implementing the ATS, which comprises ATS operations, the statistical and linguistic features, and the building blocks for the TS.

K. K. C. Reddy et al., 2021 [16] researched TS using ML technology and NLP with NLTK. A logical TS tool has been constructed using an extractive approach to produce a precise and flowing summary. The motive of the tool is to generate a brief and intelligible form of the summary.

P. Bhattacharya et al. 2021 [17] anticipated an unsupervised summarization algorithm, DELSumm, capable of methodically integrating the strategies from legal specialists into an optimization setup. The authors considered the case documents gathered from the Indian Supreme Court. The conducted tests demonstrate that the anticipated unsupervised method outperforms some strong standards in terms of ROUGE scores.

A. Qaroush et al., 2019 [18] proposed a generic extractive method intended to generate an informative summary from an Arabic document. Each sentence is evaluated considering the coverage, significance, and variety using a combination of semantic and statistical features. The efficiency of the anticipated technique is declared using a set of trials under the EASC corpus using the ROUGE measure.

Ayham Alomari et al., 2022 [19] conducted studies that reveal that approaches like Sequence-to-Sequence (Seq2Seq), Transfer Learning (TL), Reinforcement Learning (RL), and Pre-Trained Language Models (PTLMs) perform effectively for ATM. The authors complete the research paper by equating the finest models and conversing with upcoming research guidelines.

Tian Shi et al., 2021 [20] stated that several fascinating methods have been anticipated to mend seq2seq models, building them to manage diverse challenges. The authors developed an open-source library, Neural Abstractive Text Summarizer (NATS) toolkit for conducting abstractive TS. Multiple trials have been conducted using the CNN/DM dataset to evaluate the efficiency of multiple NN components.

Shashank Bhargav et al., 2021 [21] focused on the topics of documentation, explanation, summary group, and assessment of the outlines shaped. Extractive models such as KNN, TextRank, and BERT, and Abstractive models such as the Seq2Seq decoder were constructed for TS on the dataset for reviews on Amazon fine food.

In continuation of the above discussion, Table 2 provides a detailed elaboration of the literature review.

Table 2. Literature review

Category	Research Papers	Research conducted / Techniques used	Results / Outcomes
Rule-based	P. Verma and A. Verma, 2020 [22]	Performed methodical investigation of diverse TS techniques.	Declared the limitations of Graph-based, clustering-based, and MMR approaches.
Rule-based	T. Vodolazova and E. Lloret, 2019 [23]	Discussed the set of rules for transforming text into a semantic representation, and subject-verb-object concept frequency scoring.	The anticipated method outperformed the conventional abstractive methods while preserving the redundancy rate and linguistic quality.
Rule-based	M. E. Moussa, E. H. Mohamed et al., 2018 [24]	Discussed diverse approaches to opinion summarization.	Work has been conducted to assess opinion summarization in diverse ways.
DL	Y. Kumar, K. Kaur, et al., 2021 [25]	Surveyed the developments made in the area of ATS in diverse languages.	Research has been carried out on Indian Languages and foreign languages.
DL	S. Kadry, H. Yong, et al., 2021 [26]	GA-HC and PSO-HC for performing TS using Hierarchical Clustering.	Conducted the simulations and compared the performance of the anticipated models with the existing algorithms.
DL	D. Qiu and B. Yang, et al., 2021 [27]	Two attention mechanisms, MSAPN and MDAPN, have been anticipated.	The MSAPN and MDAPN model performs better with the ROUGE Recall score.
DL	A.A. Syed, F. L. Gaol, et al., 2021 [28]	Anticipated framework comprising encoder-decoder mechanisms, architecture, optimization and training strategies, dataset, and performance metrics.	Recommended BART and MASS for abstractive summarization.
DL	N. Lin, J. Li, et al., 2021 [29]	Proposed an efficient extractive method grounded on the LightGBM regression model for Indonesian text.	Outlined a formulation for computing the score of sentences as the objective function of the linear regression.
DL	D. Suleiman and A. Awajan, 2020 [30]	The Gigaword dataset for single sentences and the CNN/DM are used for multi-sentence summary techniques.	The pre-trained encoder model attained the maximum values of 43.85 for ROUGE1, 20.34 for ROUGE2, and 39.9 for ROUGE3.
DL	N. Bansal, A. Sharma, et al., 2020 [31]	A Seq2Seq encoder-decoder LSTM model with an attention mechanism has been employed to produce an abstractive summary of articles.	ROUGE is used as a performance metric to evaluate the similarity between the anticipated model and the existing model.
DL	W. Xu, C. Li, et al., 2020 [32]	Key information guide network grounded on a multi-task framework.	Multi-view attention guide network acquired the vibrant illustrations of the source text and

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

			the key information.
DL	A.P. Widyassari et al., 2020 [33]	Conducted a wide and systematic review of research in the area of TS available from 2008 to 2019.	Extractive summaries are easier in comparison to abstractive summaries.
DL	Y. Chen, Y. Ma, et al., 2019 [34]	Proposed a structure consisting of a collective encoder, decoder, and extractor.	Constrain the attention learned in the abstractive task with extractive task labels.
DL	Y. Zhang, D. Li, et al., 2019 [35]	Proposed a novel reproductive model grounded on a convolutional Seq2Seq architecture.	The hierarchical CNN model outperforms the orthodox RNN Seq2Seq model.
DL	M. M. Rahman and F. H. Siddiqui, 2019 [36]	Proposed an abstractive text model, MAPCoL, for the generation of an abstract.	MAPCoL outperformed the conventional LSTM-based models.
DL	Wang Q, Liu P, et al., 2019 [37]	A proposed hybrid model combining BERT word embedding with reinforcement learning.	CNN/Daily Mail and ROUGE have been used.
DL	S. Gupta and S. K. Gupta, 2019 [38]	Conducted a detailed literature review of different jobs performed.	The authors highlighted the pros and cons of different methods.

V. GENERALIZED RESEARCH METHODOLOGY FOR TECHNICAL DOCUMENTATION

Research technique is indispensable in a research article since it describes the methodical strategy utilized to conduct the investigation. It ensures the research's trustworthiness, reproducibility, and validity. A detailed technique enables others to assess the dependability of the consequences. Presented below in Fig. 1 is a flowchart outlining the investigation approach for improving the superiority and usability of technical documentation.

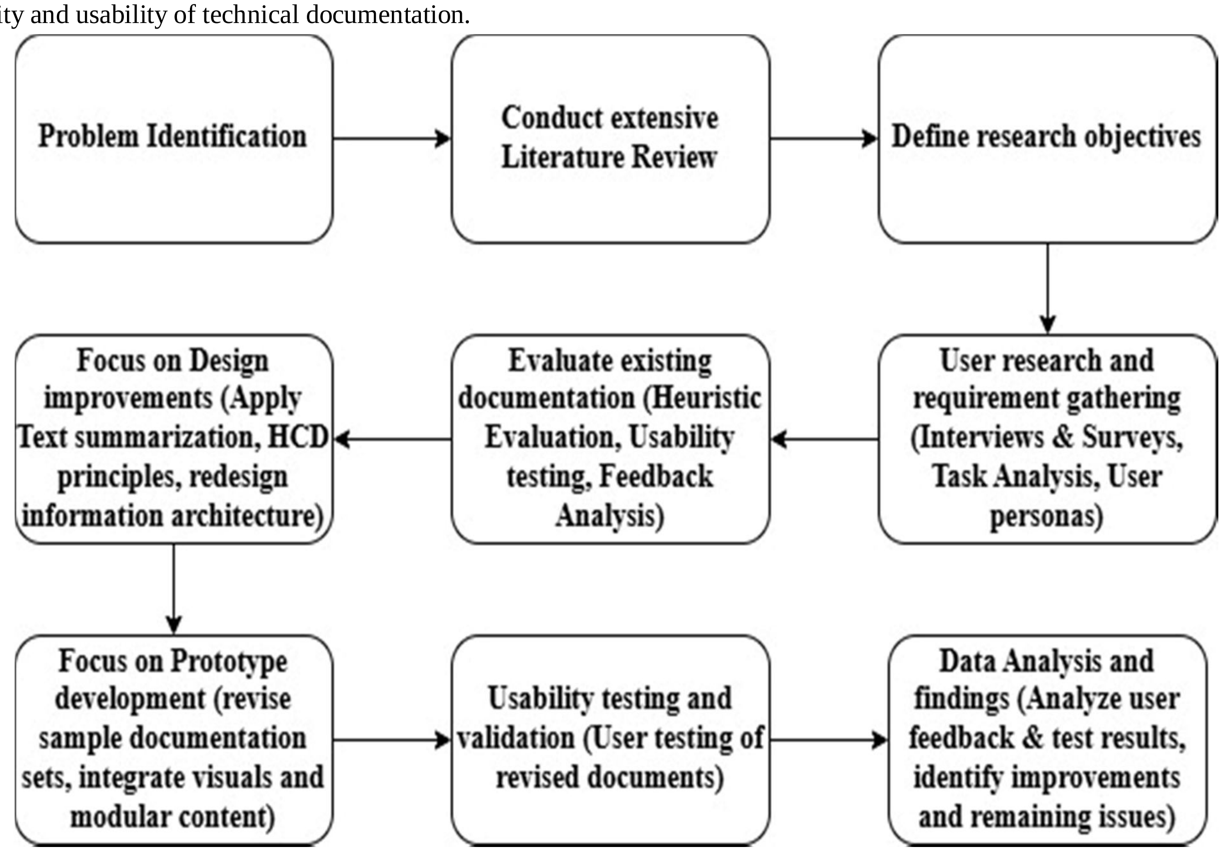


Fig. 1 Research Methodology

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

The flowchart in Fig. 1 effectively sketches a comprehensive, iterative research methodology tailored to improving technical documentation through a human-centered approach. It begins with problem identification, which lays the groundwork by recognizing inefficiencies or usability issues in existing documentation. This leads logically into an extensive literature review, grounding the research in existing theories and best practices. The flow then progresses to defining research objectives, a critical step to ensure that all subsequent activities are aligned with clear goals. From here, the process branches into user research and requirement gathering, emphasizing procedures like consultations, examinations, and task analysis to capture user potentials and context of use. In parallel, it includes evaluation of existing documentation through heuristic evaluation, usability challenges, and feedback analysis, which provides baseline insights into current weaknesses. These dual inputs feed into the design upgrading phase**,** where HCD principles and information architecture are applied to reshape content more intuitively. The redesigned concepts are advanced through prototype development, involving the revision of sample documentation sets with modular content and enhanced visuals. Following this, the methodology incorporates usability testing and validation of the revised prototypes with real users, ensuring that improvements are measurable and user-driven. Finally, data analysis and findings help close the loop by synthesizing feedback and test results to identify both improvements and residual issues, setting the stage for further refinement or implementation. The structured, user-centric, and cyclical nature of the flowchart reflects best practices in technical communication research, ensuring that documentation enhancements are not only technically sound but also genuinely responsive to user needs.

VI. TYPES OF TEXT SUMMARIZATION IN TECHNICAL DOCUMENTATION

Fig. 2 depicts a comprehensive classification of text summarizing procedures, dividing them into three categories.

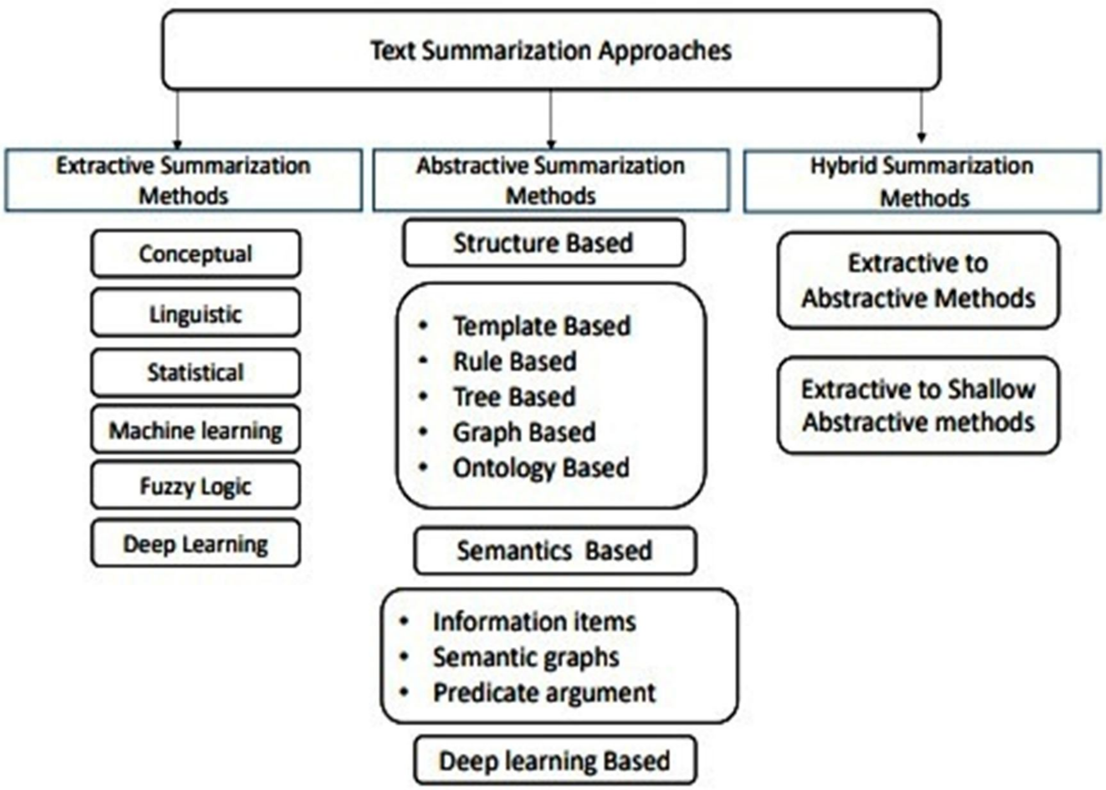


Fig. 2 Types of TS in Technical Documentation

- Extractive summarization employs a variety of techniques, including conceptual, linguistic, statistical, machine learning, fuzzy logic, and DL methods as mentioned in Table 3 below. These techniques select existing sentences or phrases from the original text based on their relevancy.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

Table 3. Extractive summarization techniques

Technique	Description
Frequency-Based (TF-IDF)	Ranks words and phrases according to term frequency and inverse document frequency.
TextRank	To score sentence importance, use a graph-based ranking model (such as PageRank).
LexRank	Similar to TextRank, but with cosine similarity and eigenvector centrality.
Linguistic Feature-Based	Uses syntactic, lexical, and discourse aspects (for example, POS tagging and sentence position).
Machine Learning-Based	Supervised models trained to classify sentence relevance.
DL-Based	Uses models like LSTMs, CNNs for extractive sentence scoring.

- Abstractive summarization is further classified into structure-based and semantics-based techniques. Template, rule, tree, graph, and ontology-based models all focus on understanding and rebuilding text structure. Semantics-based approaches use information items, semantic graphs, and predicate argument structures to get more insight. DL also plays an important part in structure and semantics-based abstractive approaches, allowing for more precise and human-like descriptions. Table 4 depicts different techniques of abstractive summarization.

Table 4. Abstractive summarization techniques

Technique	Description
Sequence-to-Sequence (Seq2Seq)	Summaries are generated using encoder-decoder models (often RNNs or LSTMs).
Transformer-Based (e.g., BERT, GPT,	Pretrained models that grasp context and produce coherent summaries.
Template-Based	Uses pre-defined templates to create summaries.
Semantic Graph-Based	Generates semantic linkages and summaries based on meaning rather than
Ontology-Based	Uses domain-specific knowledge bases to generate context-aware summaries.

- Hybrid summarizing approaches combine the advantages of extractive and abstractive techniques, as shown in Table 5. It focuses on two paths: shifting from extractive to abstractive methods and from extractive to shallow abstractive approaches, which imply layered or step-by-step summarizing processes.

Table 5. Hybrid summarization techniques

Technique	Description
Extractive-to-Abstractive Pipeline	The key content is first extracted and then rephrased using generative models.
Attention-Based Hybrid Models	Use attention processes to combine sentence extraction with context generation.
Reinforcement Learning Models	Improve the summary by training agents to maximize content quality and brevity.

Overall, the diagram effectively outlines the depth and diversity of summarization techniques and provides a clear framework for understanding how various methodologies contribute to different summarization goals, particularly in complex applications like technical documentation.

VII. JUSTIFYING THE NEED FOR HYBRID TEXT SUMMARIZATION

The most effective method for summarizing technical documentation is frequently a hybrid approach that combines extractive and abstractive summaries, with domain-specific fine-tuning as depicted in Table 6.

Table 6. Supremacy of Hybrid TS techniques

Factor	Why It Matters	Technique Advantage
Accuracy & Precision	Technical docs need to retain exact information	Extractive methods preserve original wording
Clarity & Brevity	Readers prefer clear and concise guidance	Abstractive methods improve readability
Terminology Sensitivity	Domain-specific terms must remain intact	Hybrid allows control over what gets rewritten
Context Awareness	Instructions often rely on steps or dependencies	DL models (e.g., BART, T5) handle context better
User Adaptability	Readers vary from beginners to experts	Hybrid allows multiple summary layers

VIII. DEEP LEARNING MODELS FOR TEXT SUMMARIZATION

In recent years, powerful DL models have greatly altered the field of TS, with techniques like as BERT, T5, GPT, and BART playing critical roles. These models use transformer topologies and perform exceptionally well in both extractive and abstractive summarization tasks.

1) BERT (Bidirectional Encoder Representations from Transformers) is commonly used for extractive summarization. Unlike typical models that read text in a single direction, BERT reads in either direction, allowing for better context capture. Models such as BERTSUM improve on BERT by using classification layers to identify and remove the most pertinent sentences from a document. Fig. 3 depicts the working principle of BERT.

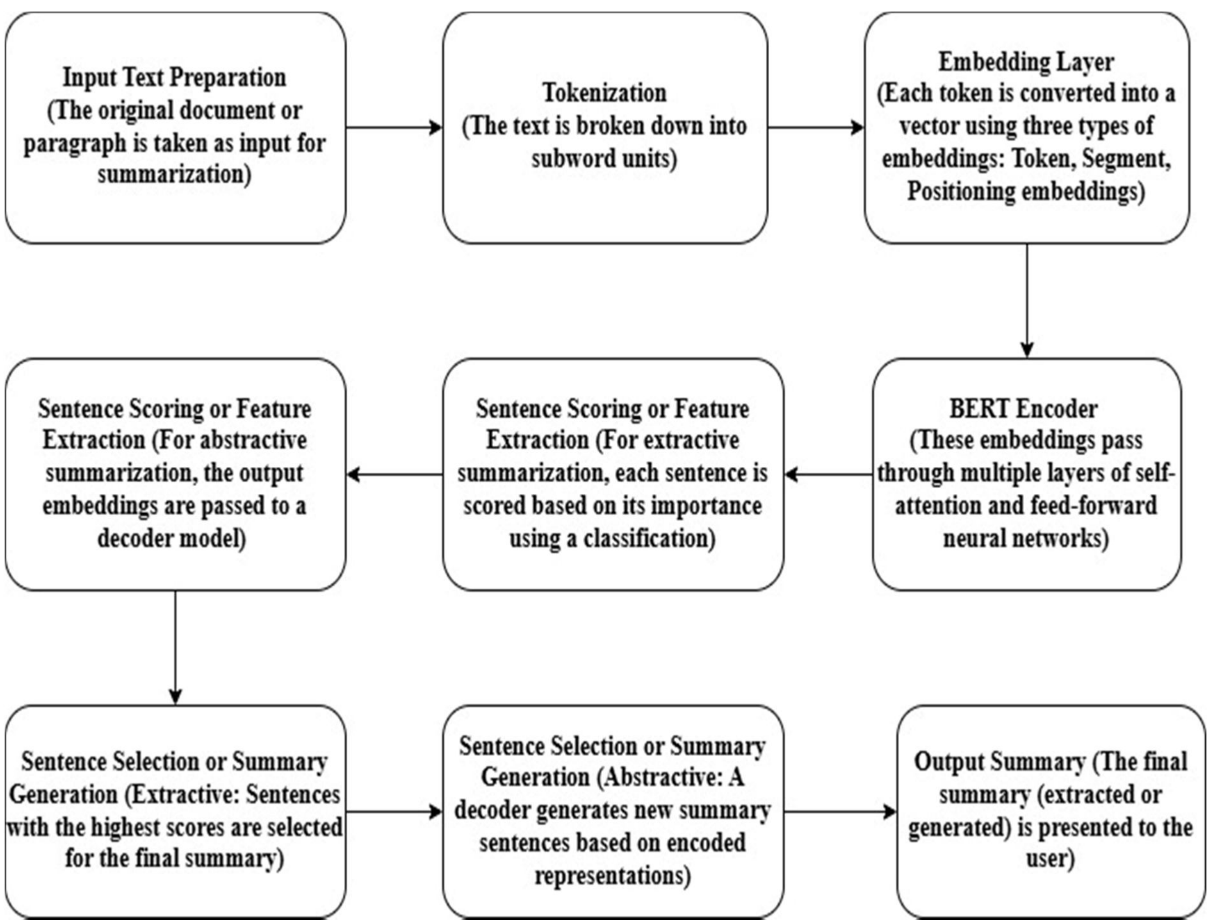


Fig. 3 Working of BERT

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

2) In contrast, T5 (Text-to-Text Transfer Transformer) is more widely utilized for abstractive summarization. T5 views all NLP tasks, including summary, as text generation problems, allowing it to produce comprehensible and grammatically correct summaries by learning to rephrase full sections. Fig. 4 shows the working principle of T5.

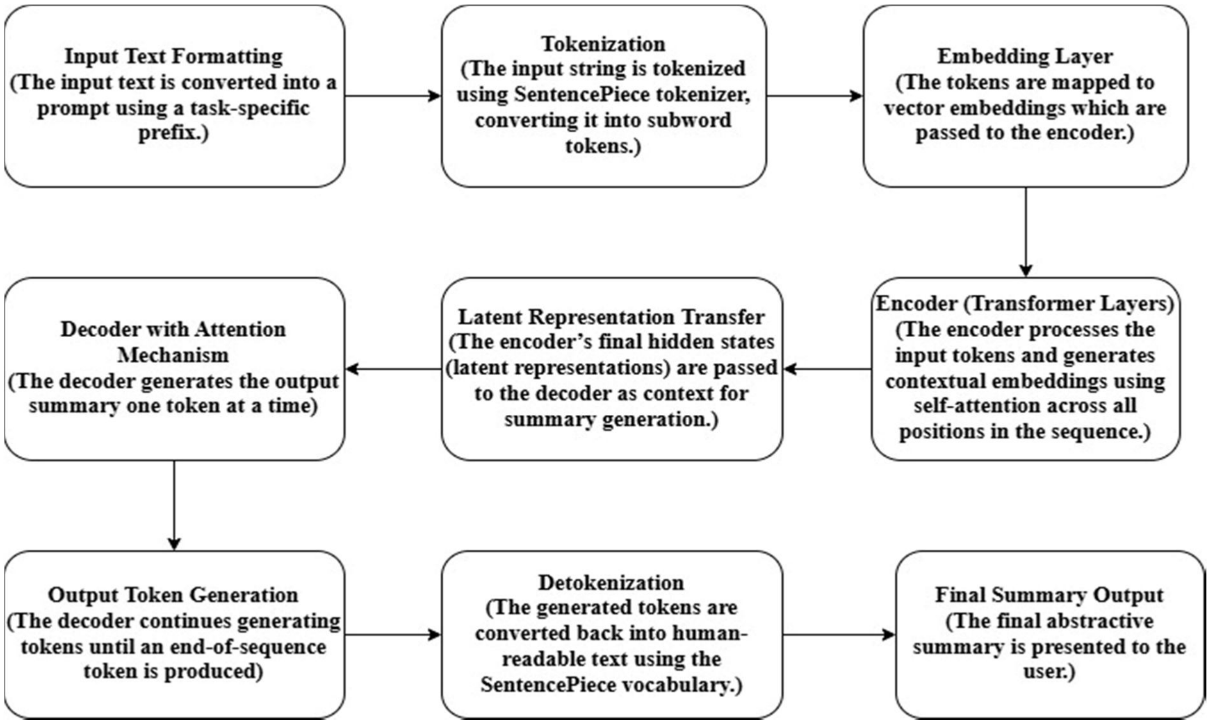


Fig. 4 Working of T5

3) Another important model is PEGASUS, which was designed expressly for abstractive summarization. It employs a novel pre-training objective in which crucial lines are obscured and the model is trained to anticipate them, closely resembling the summarization job itself. Fig. 5 shows the working principle of T5.

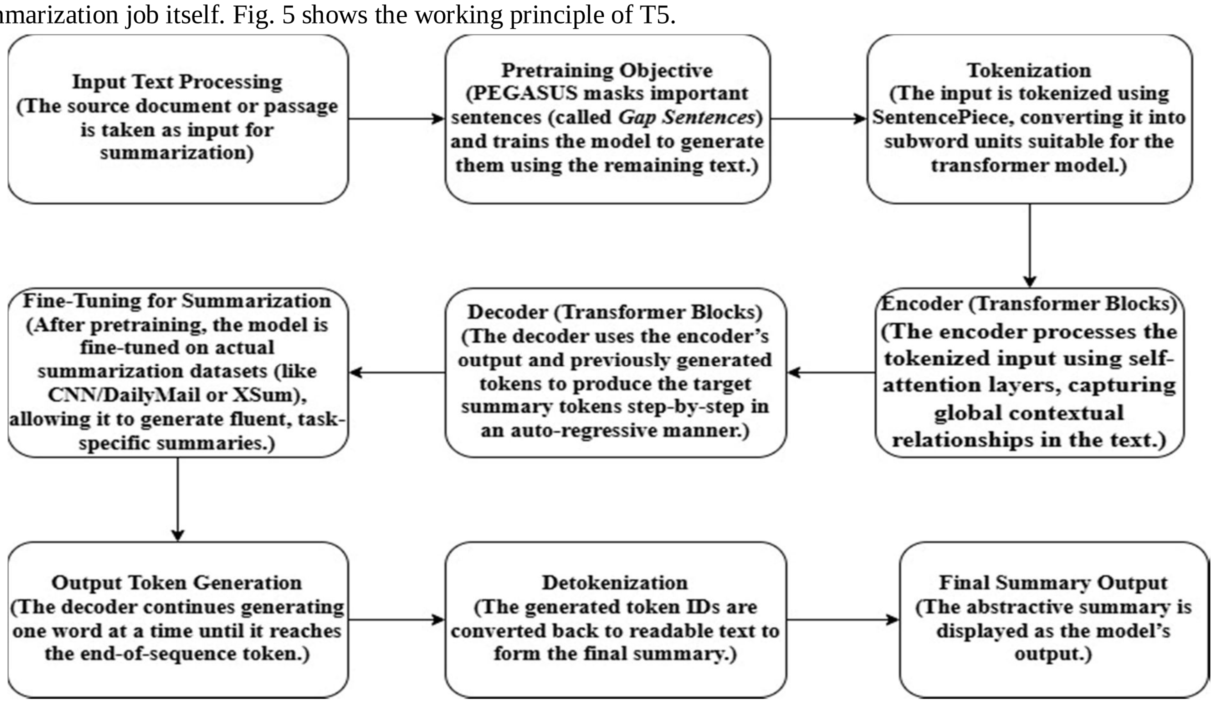


Fig. 5 Working of PEGASUS

These models have raised the bar for TS by recognizing deeper context, learning from massive amounts of data, and fine-tuning to specific topics. Their use in technical documentation, news summaries, and academic writing has greatly increased the efficiency and readability of complex textual content.

IX. FUTURE WORK ON TECHNICAL DOCUMENTATION USING HYBRID TEXT SUMMARIZATION TECHNIQUES

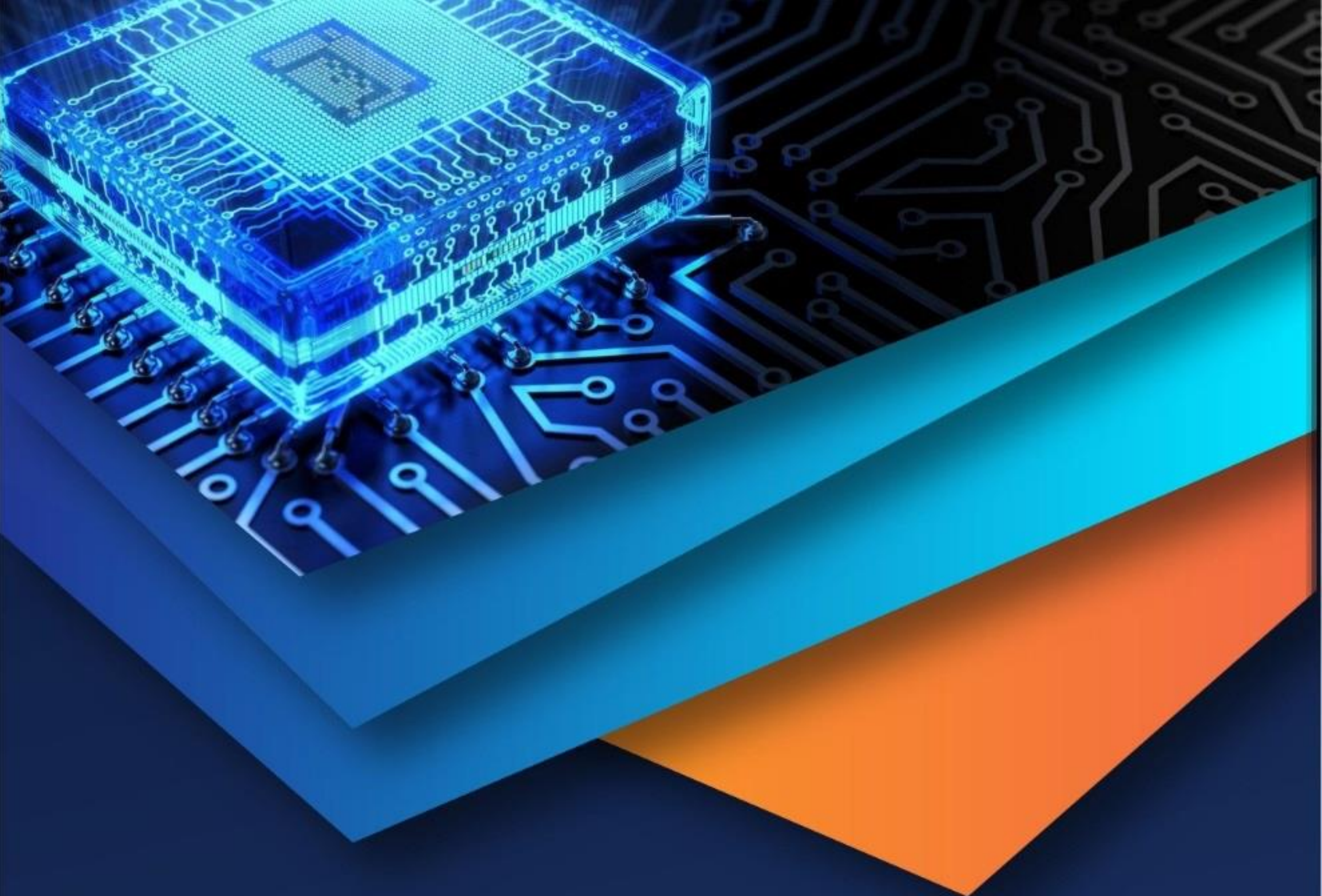
As the requirement for intelligent, user-centered documentation grows, future studies in technical documentation may concentrate on hybrid text summarizing algorithms that incorporate the benefits of both extractive and abstractive strategies. Extractive models maintain factual accuracy by picking essential lines from the source material, but abstractive models improve readability and natural language fluidity by producing compact paraphrased information. A hybrid approach may successfully balance these factors, resulting in high-quality summaries that are both dependable and simple to utilize. Future research could look into task-specific fine-tuning of transformer-based models (such as BERT for extraction and T5 or BART for abstraction) to produce multi-layered summaries customized to numerous user roles—such as developers, administrators, or end-users—based on their contextualized needs and technical knowledge.

Further developments might also concentrate on creating domain-adaptive hybrid workflows that combine semantic comprehension, user behavior statistical analysis, and document structure analysis to generate real-time, tailored summaries incorporated within technological interfaces. Integrating interactive summation tools into content management systems or documentation platforms enables dynamic summary production and customisation. Furthermore, merging summarizing with multimodal content processing (such as code fragments, schematics, and tables) is a largely untapped opportunity for improving technical documentation usability. Future studies should address appraisal issues by introducing new measures that consider not just linguistic quality but also task success, understanding efficiency, and user pleasure. Finally, hybrid summarization in technical documentation has the ability to transform static publications into versatile, responsive knowledge systems that meet the various and changing demands of users.

REFERENCES

[1] A. Chaves, C. Kesiku, and B. Garcia-zapirain, "Automatic Text Summarization of Biomedical Text Data : A Systematic Review," 2022.
[2] J. P. Verma et al., "Graph-Based Extractive Text Summarization Sentence Scoring Scheme for Big Data Applications," pp. 1–28, 2023.
[3] A. Pramita et al., "Review of automatic text summarization techniques & methods," J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 4, pp. 1029–1046, 2022, doi: 10.1016/j.jksuci.2020.05.006.
[4] A. Prasetya and F. Kurniawan, "A survey of text summarization : Techniques , evaluation and challenges," Nat. Lang. Process. J., vol. 7, no. April, p. 100070, 2024, doi: 10.1016/j.nlp.2024.100070.
[5] M. M. Saiyyad, "Text Summarization Using Deep Learning Techniques : A Review †," pp. 4–9, 2024.
[6] I. Mobin, M. H. Mahadi, and A. K. Pathan, "A Review of the State-of-the-Art Techniques and Analysis of Transformers for Bengali Text Summarization," vol. 1, no. Dl, pp. 1–29, 2025.
[7] B. Models and S. Abdel-salam, "Performance Study on Extractive Text Summarization Using," 2022.
[8] G. Padmapriya and K. Duraiswamy, "Multi-document-based text summarisation through deep learning algorithm," International Journal of Business Intelligence and Data Mining , vol. 16, no. 4, pp. 459–479, 2020, doi: 10.1504/IJBIDM.2020.107546.
[9] F. B. Goularte, S. M. Nassar, R. Fileto, and H. Saggion, "A text summarization method based on fuzzy rules and applicable to automated assessment," Expert Systems with Applications, vol. 115, pp. 264–275, 2019, doi: 10.1016/j.eswa.2018.07.047.
[10] N. Alami, M. Meknassi, and N. En-nahnahi, "Enhancing unsupervised neural networks based text summarization with word embedding and ensemble learning," Expert Systems with Applications, vol. 123, pp. 195–211, 2019, doi: 10.1016/j.eswa.2019.01.037.
[11] F. A. Ghanem, M. C. Padma, and H. M. Abdulwahab, "Deep Learning-Based Short Text Summarization : An Integrated BERT and Transformer Encoder – Decoder Approach," 2025.
[12] M. M. Saiyyad, "Text Summarization Using Deep Learning Techniques : A Review †," pp. 4–9, 2024.
[13] Y. Yang, Z. Wu, Y. Yang, S. Lian, F. Guo, and Z. Wang, "applied sciences A Survey of Information Extraction Based on Deep Learning," 2022.
[14] K. K. M. Et.al, "A Heuristic Approach for Telugu Text Summarization with Improved Sentence Ranking," Turkish Journal of Computer and Mathematics Education, vol. 12, no. 3, pp. 4238–4243, 2021, doi: 10.17762/turcomat.v12i3.1714.
[15] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic text summarization: A comprehensive survey," Expert Systems with Applications, vol. 165, no. November 2021, 2021, doi: 10.1016/j.eswa.2020.113679.
[16] K. K. C. Reddy, P. R. Anisha, N. G. Nguyen, and G. Sreelatha, "A Text Mining using Web Scraping for Meaningful Insights," Journal of Physics: Conference Series, vol. 2089, no. 1, 2021, doi: 10.1088/1742-6596/2089/1/012048.
[17] P. Bhattacharya, S. Poddar, K. Rudra, K. Ghosh, and S. Ghosh, Incorporating domain knowledge for extractive summarization of legal case documents, vol. 1, no. 1. Association for Computing Machinery, 2021.
[18] A. Qaroush, I. Abu Farha, W. Ghanem, M. Washaha, and E. Maali, "An efficient single document Arabic text summarization using a combination of statistical and semantic features," Journal of King Saud University - Computer and Information Sciences, vol. 33, no. 6, pp. 677–692, 2019, doi: 10.1016/j.jksuci.2019.03.010.
[19] A.Alomari, N. Idris, A. Q. M. Sabri, and I. Alsmadi, "Deep reinforcement and transfer learning for abstractive text summarization: A review," Computer

Speech & Language, vol. 71, no. August 2021, p. 101276, Jan. 2022, doi: 10.1016/j.csl.2021.101276.
[20] T. Shi, Y. Keneshloo, N. Ramakrishnan, and C. K. Reddy, "Neural Abstractive Text Summarization with Sequence-to-Sequence Models," ACM/IMS Transactions on Data Science vol. 2, no. 1, pp. 1–37, 2021, doi: 10.1145/3419106.
[21] S. Bhargav, A. Choudhury, S. Kaushik, R. Shukla, and V. Dutt, "A comparison study of abstractive and extractive methods for text summarization," Advances in Intelligent Systems and Computing, vol. In press, no. April, 2021.
[22] P. Verma and A. Verma, "A Review on Text Summarization Techniques," Journal of Scientific Research, vol. 64, no. 01, pp. 251–257, 2020, doi: 10.37398/jsr.2020.640148.
[23] T. Vodolazova and E. Lloret, "The Impact of Rule-Based Text Generation on the Quality of Abstractive Summaries," in Proceedings - Natural Language Processing in a Deep Learning World, Oct. 2019, vol. 2019-Septe, pp. 1275–1284, doi: 10.26615/978-954-452-056-4_146.
[24] M. E. Moussa, E. H. Mohamed, and M. H. Haggag, "A survey on opinion summarization techniques for social media," Future Computing and Informatics Journal, vol. 3, no. 1, pp. 82–109, 2018, doi: 10.1016/j.fcij.2017.12.002.
[25] Y. Kumar, K. Kaur, and S. Kaur, Study of automatic text summarization approaches in different languages, vol. 54, no. 8. Springer Netherlands, 2021.
[26] S. Kadry, H. Yong, and J. Choi, "Applied sciences Improved Text Summarization of News Articles Using GA-HC," 2021.
[27] D. Qiu and B. Yang, "Text summarization based on multi-head self-attention mechanism and pointer network," Complex & Intelligent Systems, 2021, doi: 10.1007/s40747-021-00527-2.
[28] A.A. Syed, F. L. Gaol, and T. Matsuo, "A survey of the state-of-the-art models in neural abstractive text summarization," IEEE Access, vol. 9, pp. 13248–13265, 2021, doi: 10.1109/ACCESS.2021.3052783.
[29] N. Lin, J. Li, and S. Jiang, "A simple but effective method for Indonesian automatic text summarisation," Connection Science, 2021, doi: 10.1080/09540091.2021.1937942.
[30] D. Suleiman and A. Awajan, "Deep Learning Based Abstractive Text Summarization: Approaches, Datasets, Evaluation Measures, and Challenges," Mathematical Problems in Engineering, vol. 2020, pp. 1–29, Aug. 2020, doi: 10.1155/2020/9365340.
[31] N. Bansal, A. Sharma, and R. K. Singh, "Recurrent neural network for abstractive summarization of documents," Journal of Discrete Mathematical Sciences and Cryptography, vol. 23, no. 1, pp. 65–72, Jan. 2020, doi: 10.1080/09720529.2020.1721873.
[32] W. Xu, C. Li, M. Lee, and C. Zhang, "Multi-task learning for abstractive text summarization with key information guide network," EURASIP Journal on Advances in Signal Processing, vol. 2020, no. 1, 2020, doi: 10.1186/s13634-020-00674-7.
[33] A.P. Widyassari et al., "Review of automatic text summarization techniques & methods," J. King Saud Univ. - Comput. Inf. Sci., no. xxxx, 2020, doi: 10.1016/j.jksuci.2020.05.006.
[34] Y. Chen, Y. Ma, X. Mao, and Q. Li, "Multi-Task Learning for Abstractive and Extractive Summarization," Data Science and Engineering, vol. 4, no. 1, pp. 14–23, 2019, doi: 10.1007/s41019-019-0087-7.
[35] Y. Zhang, D. Li, Y. Wang, Y. Fang, and W. Xiao, "Abstract text summarization with a convolutional seq2seq model," Applied Sciences, vol. 9, no. 8, 2019, doi: 10.3390/app9081665.
[36] M. M. Rahman and F. H. Siddiqui, "An optimized abstractive text summarization model using peephole convolutional LSTM," Symmetry (Basel). vol. 11, no. 10, 2019, doi: 10.3390/sym11101290.
[37] Wang Q, Liu P, Zhu Z, Yin H, Zhang Q, Zhang L. A text abstraction summary model based on BERT word embedding and reinforcement learning. Applied Sciences 2019;9(21). doi:10.3390/app9214701.
[38] S. Gupta and S. K. Gupta, "Abstractive summarization: An overview of the state of the art," Expert Systems With Applications, vol. 121, no. 2018, pp. 49–65, 2019, doi: 10.1016/j.eswa.2018.12.011.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY