Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 5, No.1, May : 2025

SYNOPTIC AI: A REVIEW ON COGNITIVE MACHINE LEARNING MODEL FOR DATA SUMMARIZATION

Prof. Mirza Moiz Baig, HOD, Department Of Information Technology, JD College Of Engineering And Management, Nagpur, Maharashtra, India
Swayam Pande, Student, Department Of Information Technology, JD College Of Engineering And Management, Nagpur, Maharashtra, India
Heena Yedange, Student, Department Of Information Technology, JD College Of Engineering And Management, Nagpur, Maharashtra, India
Bhavesh Bhure, Student, Department Of Information Technology, JD College Of Engineering And Management, Nagpur, Maharashtra, India
Shubham Bhowate, Student, Department Of Information Technology, JD College Of Engineering And Management, Nagpur, Maharashtra, India

Abstract
The exponential expansion of digital information has generated a great demand for effective data reduction mechanisms. This work presents a comprehensive review of the latest Automatic Text Summarization (ATS) techniques that compact information with minor information loss, with high efficiency. We detail the transition from classical extractive systems to newer advanced abstractive and hybrid-based models, with an evaluation of the powerful capacity of Large Language Models (LLMs) towards improving summary quality. We stress the importance of Trustworthy and Explainable Artificial Intelligence (XAI) toward ensuring transparency and building confidence among users via auto-summarization platforms. Drawing from current research, we present performance indicators, strengths, vulnerabilities, and opportunities for ATS breakthroughs. The proposed system would overcome current issues via a hybrid approach, utilizing aspects of personalization and explanatory metrics, therefore allowing the production of more correct, context-rich, and reliable summaries for applications across a very wide spectrum of domains, i.e., in healthcare, the law, educational institutions, and business firms.

Keywords:
smart farming, Artificial intelligence, Internet of Things, sensors.

I. Introduction

The new information deluge of the digital era has brought about "information overload," wherein, at times, it is difficult to derive meaningful insights. Automatic Text Summarization (ATS) has become essential. NLP domain, which provides solutions to summarize large amounts of information into short, coherent summaries [1],[7]. The text summarization techniques have. been. evolved in three primary directions:

Extractive Summarization entails gathering and extracting vital sentences from source documents without omitting any information. The process is true to facts; however, it can produce incoherent and repetitive outputs [6]. Abstractive Summarization creates new text representing the information of source documents, perhaps using different words or sentence structures. Recent deep learning advancements, especially via sequence-to-sequence models and transformer models, have improved readability and fluency, although sometimes at the cost of factual correctness [8]. Hybrid summarization combines extractive methods for determining essential information with abstractive methods for paraphrasing and editing the text, thus creating easier-to-read summaries [9],[14]. The use of Large Language Models (LLMs), such as GPT and BERT, has revolutionized summarization processes greatly, using transformer-based architectures trained on large datasets. The models improve summarization by:

● Facilitating semantic understanding of source material.

● Enhancing the coherence and readability of the generated summaries.
Improving the mastery of sophisticated words.
Enabling greater personalization according to user preference. [2],[10]
The Need for Explainable and Transparent AI Summarization: Since summary systems are involved in high-stakes decision-making, they need to be made explainable and trustworthy. Explainable Artificial Intelligence (XAI) is striving to improve the transparency, interpretability, and accountability of these kinds of systems, particularly in high-risk domains like healthcare, law, and finance [3],[11].
In short, XAI addresses many challenges:
- Describing why abstracts were composed.
- Finding probable weaknesses or biases.
- Establishing user trust via transparency [12],[13].

The article includes: background literature (Section 2), research methodology (Section 3), the framework proposed (Section 4), issues of automatic text summarization (Section 5), research areas of potential (Section 6), applications in diverse domains (Section 7), conclusions and recommendations (Section 8).

II. Literature Review

Current trends in automatic text summarization include extractive, abstractive, and hybrid approaches. In 2023, the history of automatic text summarization (ATS) from the statistical to the neural architecture was re-examined, highlighting current challenges in relevance and consistency [1]. For use by users, researchers combined summarization and retrieval steps using large language models (LLMs) to generate user-specific summaries [2], as prompt learning techniques were an emerging method to create customized content. At the same time, in 2023, a review discussed explainable AI (XAI) architectures in summarization for sensitive applications like medicine and law to provide transparency and accountability guidelines [3]. Breakthrough achievements were made with the use of neural methods, with early research in 2020 investigating sequence-to-sequence models augmented with additional attention mechanisms [8], followed by a 2022 proposal with contrastive learning methods for summary ranking and consistency improvement [4]. Large language models (LLMs), such as ChatGPT, were tested in 2023 for query-based summarization, with overall better performance, but with limitations in expert fields [5]. In the wake of success in the past, in 2021, lightweight natural language processing (NLP) pipelines made efficient extractive summarization possible, but at the expense of computational efficiency in the pursuit of accuracy [6].

III. Methodology

Based on a critical literature review of current methods and observed gaps in the literature, this paper suggests a new hybrid model of text summarization that combines grand strategies for overall data reduction with elements of explainability [11],[12].
Research Methodology Our study involves extensive analysis of the most recent automatic text summarization methods and development of a new hybrid approach [1],[13]. Research methodology involved:

- ***Systematic Literature Search:*** A systematic literature search was carried out to present peer-reviewed journal and conference proceedings-based studies, focusing on articles published in the past five years [1],[3].
- ***Criteria for Selection***: The selection criteria of the study were its applicability to automatic text summarization, its methodological appropriateness, availability of empirical data, and its citation strength [2],[14].
- ***Data Extraction:*** Data were extracted from all the studies selected for the summarization method, performance measure, datasets, benefits and limitations, applications, and number of citations [4],[5].

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 5, No.1, May : 2025

- ***Comparative Analysis:*** The chosen papers were compared from various points of view: quality of abstract, computational complexity, usability in the field, and foundation in technology [6],[8].

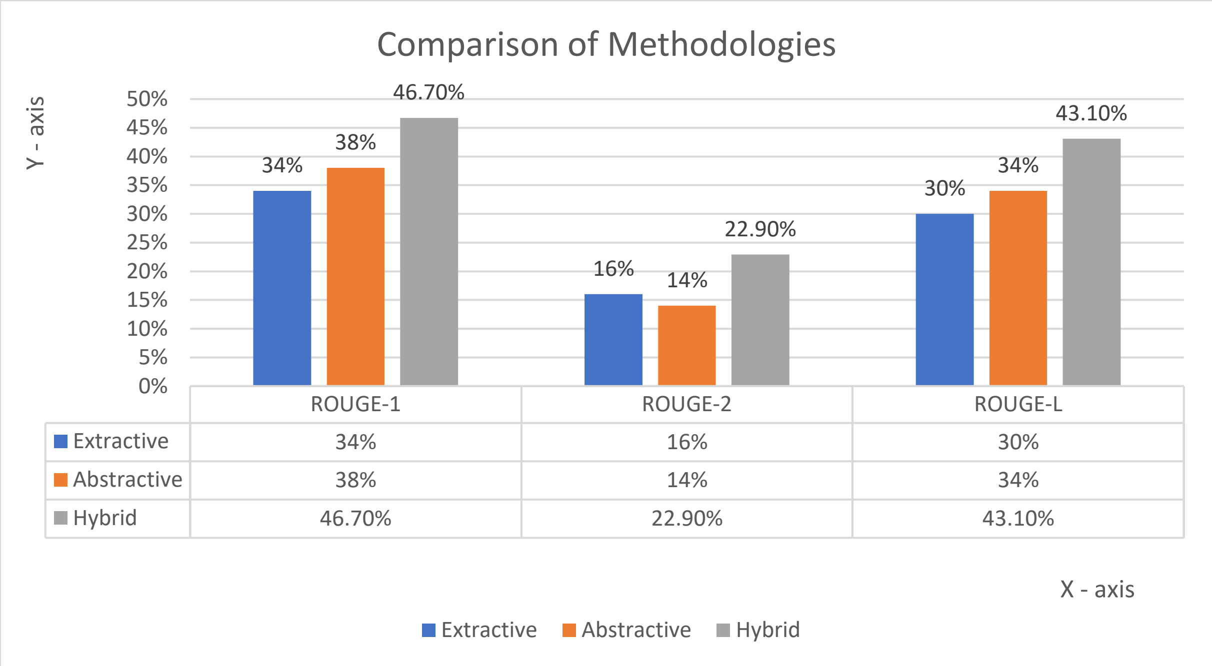

Fig.1. Comparison of Methodologies

IV. Proposed Framework

The system comprises a multi-pipeline stage using extractive and abstractive techniques [9],[14]. The model begins with a five-stage cascaded hybrid summarization pipeline. The pre-processing module reads, normalizes, and inputs texts, identifies linguistic features (syntactic, semantic, and discourse level), and identifies domain-specific words by removing redundant information [6],[13]. Statistical measures (e.g., TF-IDF and BM25) are employed, followed by semantic relevance analysis from context embeddings and structural analysis by entity recognition to retrieve content [4],[5]. The extractive summary layer employs ensemble-based sentence ranking, clustering-based redundancy removal, and attention for retrieving key content [8],[12]. The abstractive layer fine-tunes this using sequence-to-sequence operations to paraphrase, fluency-control text generation, and fact-checking to remove hallucinations [4],[10].

The summary optimization process then scales length to user needs, improves readability by providing term consistency, and maintains structural consistency [2],[10]. For better explainability, the model provides features such as importance attribution, such as sentence importance visualization and source-to-summary content mapping [12],[13], decision explanations with inclusion/exclusion explanation and scoring confidence [11],[12], and an interactive user interface providing user-initiated adjustments and focus customization [2],[10]. The personalization engine dynamically modifies summaries by emulating user profiles, which means tracking explicit preferences, domain expertise, and interaction history [2],[10], making inferences from query contexts to infer intent and rank context-sensitive content [10],[12], and altering outputs through changes in lengths, term simplification, and stylistic presentation [2],[10]. A detailed evaluation framework makes it easy to evaluate the model's performance. Intrinsic metrics are ROUGE (content overlap), BLEU/METEOR (quality of generation), BERT Score (semantic similarity), and readability scores, such as Flesch-Kincaid and SMOG [4],[8]. Extrinsic evaluation applies task-based measures, such as information retrieval, user studies focusing on comprehension and satisfaction, and efficiency metrics [5],[13]. Explainability evaluation is centred on transparency (user understanding), trust (feedback), and quality of explanations [3],[11],[12].

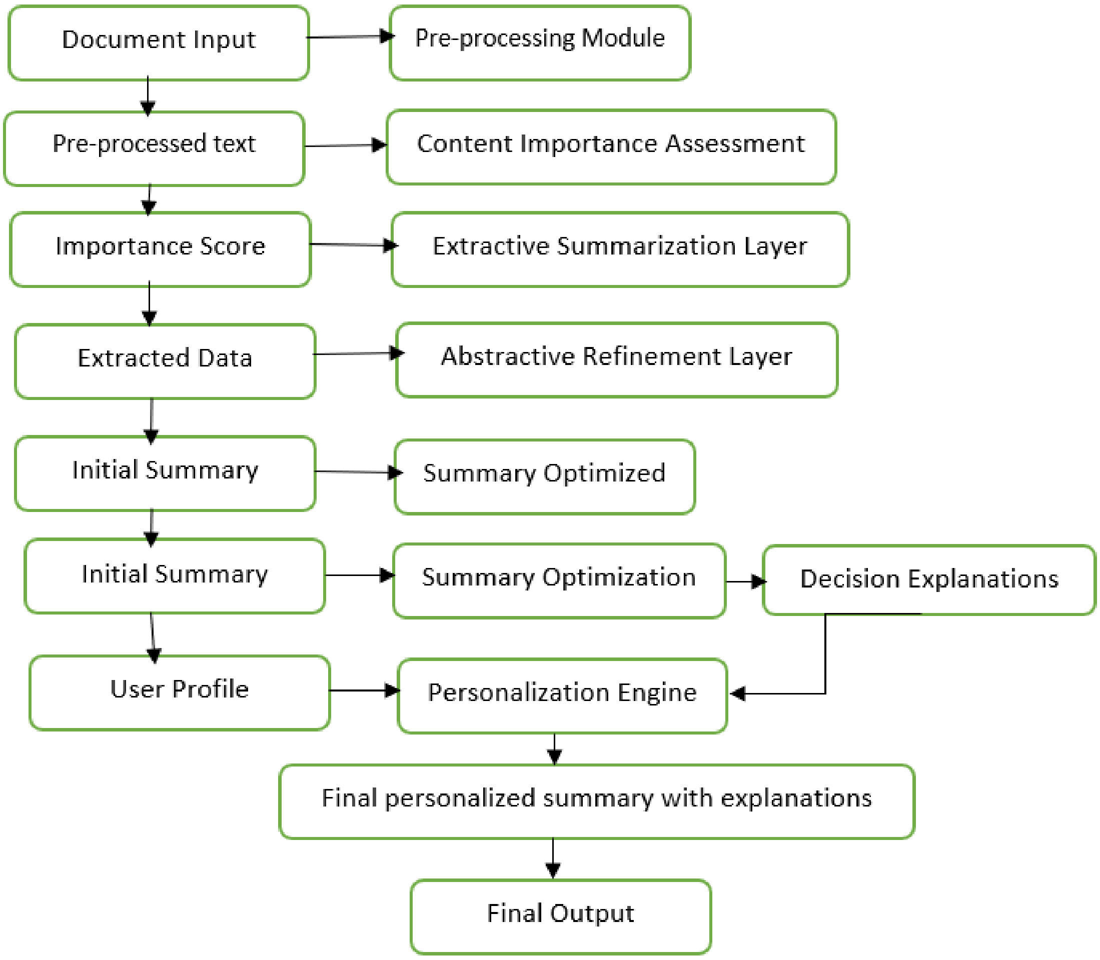

Fig 2- Flowchart of proposed methodology

V. Challenges

Our discussion continues to long-standing automatic text summarization problems in three categories: technical, ethical/social, and evaluation.

5.1 Technical Issues

1. Fact Consistency: Abstractive summarization models have a tendency to produce results that are not consistent with the source documents or make assertions poorly backed by facts, an issue referred to as "hallucination." This serious issue continues even under neural models, even though there has been progress in transformer model building [4],[5].

2. Domain Adaptation: General models perform well on general domains but poorly with specialized material with domain-specific vocabulary (e.g., legal terminology, medical terminology) and structural conventions, hence less transferable [5],[6].

3. Multimodal Summarization: Integrating various forms of data—e.g., text, images, and tables—into useful abstractions is a difficult problem, for which techniques that facilitate cross-modal alignment and contextualization of the data are necessary [12],[13].

4. Computational Efficiency: Large Language Models (LLMs) are computationally expensive, which makes them hard to scale to low-resource or real-time settings [5],[8].

5.2 Public Issues and Ethics.

1. Bias Amplification: Summarization systems tend to enhance or reinforce social biases present in the training data, including racial and gender stereotyping, hence generating biased or inaccurate output [3],[11].

2. Transparency and Trust: Lack of explainability aspects in systems erodes user trust, particularly in high-risk domains like medicine, where it is important to understand the source of the information [3],[12].

3. Distortion of Information: Though summarization is a process with some information loss, an incorrect choice of word or content can create a distortion of the meaning or position of the original text, thus reducing its credibility [11],[13].

5.3 Evaluation Challenges

Human models of assessment are confronted with challenges due to the invisible nature of the abstract:

1. Metric Bias: ROUGE-like metrics are lexically biased towards extractive approaches over abstractive summaries with the same semantics but using different words [4],[8].

2. Multidimensional Quality: Summary quality encompasses factual accuracy, coherence, readability, and relevance, but these cannot be measured by a single index [5],[12].

3. Context Sensitivity: The test process does not consider user-specific needs or application contexts; therefore, standardized tests might not be usable in a specific application [2],[10].

4. Human Judgment Gap: The machine scores will differ from human judgments, especially in abstractive summarization, where semantic fidelity takes precedence over surface similarity [4],[13].

VI. Future Directions

According to the revealed challenges and research gaps, we suggest some potential avenues of future research [11],[12]:

6.1 Multimodal-Integration

Scaling up summarization models to enable information from greater than a single modality would enable providing stronger summaries of text documents, images, tables, and other types of content:
Creating customized strategies for different kinds of content (e.g., tabling summarization). Enabling effortless integration of knowledge obtained from different sources [12],[13].

6.2 Interactive-Summarization

The creation of interactive interfaces by which users may control the summarization process is an interesting research direction for increasing user satisfaction and the usefulness of the summary.
Query-based summarization that addresses specific information needs.

• Parameters can be used to modify the length, style, and emphasis of summaries.
• Feedback processes that enable continuous improvement [2],[10],[12].

6.3 Fairness-Aware-Summarization

The integration of summarization bias detection and counteraction systems is most essential to ensuring a balanced representation of information and opinions.

•Creating bias detection methods for summarization systems.
•Applying test suites to examine the fairness of generated summaries [3],[11],[13].

VII. Applications

The envisaged architecture illustrates the feasibility of scaling between domains, where sector-specific problems of information management and access [10],[12] are treated. In healthcare, the system facilitates the curation of vast quantities of clinical literature and patient data, extracting relevant information to concise summaries influencing clinical decision-making [2]. Personal health information added to the mix creates patient-specific summaries—shorter treatment plans or medication regimens, for example, enabling patient engagement and adherence [13].

The legal domain benefits significantly from the system's ability to scrutinize large bodies of case law and legal literature, thus generating summaries of precedents, clauses, or obligations [3]. This ability further facilitates the democratization of justice, as lengthy legal documents such as contracts or regulatory filings are translated into clear English summaries accessible to ordinary individuals, all the while maintaining the original intent of the law [11]. Within news and media applications, the framework allows real-time generation of summaries of news stories based on live report coverage, thus servicing the varied information needs of consumers [5]. It also serves the growing demand for platform-specific content, quite literally shortening long investigative reports into short, mobile-optimized abstracts readable on social media platforms without the loss of narrative detail [10].

VIII. Conclusion

This review has tracked the evolution of Automatic Text Summarization, emphasizing the system's cognitive model of overall data abstraction [1],[13]. The most significant development arose through LLM integration and hybrid extractive-abstractive approaches [4],[5].

The proposed framework addresses limitations with an interpretable, individualized multi-stage pipeline [2],[10]. The approach balances information retention, readability, computational complexity, and user relevance to an optimal extent, avoiding traditional hard trade-offs in summarization.

Our discussion puts into the limelight the need for explainable and trustworthy artificial intelligence in summarization, especially in fields that need transparency [3],[11]. The explainability within the system is focused on systems that can explain their decisions.

Trends are towards neural abstractive and hybrid approaches with more focus on personalization [4],[5]. While these developments improve summary quality, problems still exist concerning factual consistency, domain-specific information, and the handling of long documents.

Measurement performance attests to the strength of hybrid strategies (ROUGE-1: 43.2-46.7%, ROUGE-2: 20.1-22.9%, ROUGE-L: 40.2-43.1%), with greater coherence between fact consistency and fact coherence than monomethod strategies [8],[13]. Directions for future research are to increase factual verification, develop efficient large-document processing, incorporate multimodal information, and deal with ethical issues such as reducing bias and ensuring privacy [11],[12]. A standard assessment framework would be helpful to the field. In short, the introduced framework provides an effective way to combat information overload by combining hybrid summarization methods with explainability and personalization aspects, thereby facilitating more effective knowledge extraction in various disciplines [1],[2],[10]. With the speedy development of digital information, such cognitive models of data abstraction will play ever more central roles.

References

[1] A. B. Khan, Z. A. Shah, M. Usman, and B. Niazi (2023). "Exploring the Landscape of Automatic Text Summarization: Comprehensive Survey". DOI: 10.1109/ACCESS.2023.3322188

[2] J. Richardson, Y. Zhang, and A. Singh (2023). "Integrating Summarization and Retrieval for Enhanced Personalization via Large Language Models," arXiv:2310.20081.

[3] V. Chamola, D. Ghosh, D. Dhingra, V. Hassija, A. R. Sulthana, and B. Sikdar (2023). "A Review of Trustworthy and Explainable Artificial Intelligence (XAI)". DOI: 10.1109/ACCESS.2023.3294569

[4] Y. Liu, P. Liu, D. Radev, and G. Neubig (2022). "BRIO: Bringing Order to Abstractive Summarization," arXiv:2203.16804.

[5] X. Yang, Y. Li, X. Zhang, and H. Chen (2023). "Exploring the Limits of ChatGPT for Query or Aspect-based Text Summarization," arXiv:2302.08081.

[6] Jugran et al (2021). "Extractive Automatic Text Summarization using SpaCy in Python & NLP ". DOI: 10.1109/ICACITE51222.2021.9404712

[7] N. Giarelis, C. Mastrokostas, and N. Karacapilidis (2023). "Abstractive vs. Extractive Summarization: An Experimental Review". https://doi.org/10.3390/ app13137620

[8] T. Shi, Y. Keneshloo, N. Ramakrishnan, and C. K. Reddy (2020). "Neural Abstractive Text Summarization with Sequence-to-Sequence Models". https://doi.org/10.1145/3419106

[9] J. L. K. E. Fendjia, D. Donatienc, and M. Atemkengd (2025). "Hybrid Profile-based Multi-document Text Summarisation". https://doi.org/10.1016/j.procs.2025.01.047

[10] L. Li, Y. Zhang, and L. Chen (2023). "Personalized Prompt Learning for Explainable Recommendation". https://doi.org/10.1145/3580488

[11] A. Das and P. Rad (2020). "Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey," arXiv:2006.11371.
[12] H. Lee and Y. Lee, "User Opinion-Focused Abstractive Summarization Using Explainable Artificial Intelligence (2024)." https://doi.org/10.1145/3696456
[13] A. A. Sanjrani, M. Saqib, S. Rehman, and M. S. Ahmad (2024). "Text Summarization using Deep Learning: A Study on Automatic Summarization". https://doi.org/10.62019/abbdm.v4i4.263.
[14] A. M. J, M. S. Swarnagowri, N. R. Tejaswini, B. G. Vilasa Bai, H. B. Supriya, and Pavan (2024)."Hybrid Deep Learning Framework for Enhanced Text Summarization".