Analysis of Research Trends and a Unified Framework for Comparative Text Summarization

Ann Baby[1,3], Basant Kumar[1,2], Aiswarya Nair[3]

[1] Lincoln University College, Malaysia;

[2] Modern College of Business and Science, Muscat, Sultanate of Oman

[3] Rajagiri College Of Social Sciences, India

Email ID – annbaby2k@yahoo.co.in

Abstract— The rapid expansion of unstructured text has intensified the need for effective and deployable text summarization systems. While extractive, abstractive, and large language model (LLM)–based approaches have advanced significantly, existing studies often evaluate these paradigms in isolation and lack standardized, reproducible frameworks. This study addresses this gap by combining a bibliometric analysis of LLM-related natural language processing (NLP) research with the development of a unified comparative summarization framework. A bibliometric analysis of 1,014 Scopus-indexed publications reveals key trends in LLM-based NLP research, including the dominance of conference-driven dissemination, concentration of research output among a small number of authors and institutions, and a strong geographic emphasis on Asia and North America. These findings highlight the rapid evolution of the field and the scarcity of deployable, standardized evaluation systems.

Keywords—Text Summarization, large language model (LLM), natural language processing (NLP), Extractive Summarization, Abstractive Summarization, Large Language Models, Transformer Models

---

1. INTRODUCTION

1.1 Background and Motivation

The rising digital text volume created through online sources, professional documents, and enterprise systems has increased the demand for effective ways to collect and use important information from large quantities of unstructured data. Text summarisation's objective is to summarise text documents of considerable length into a single summary that maintains the most vital information (facts) as well as semantic (meaning) linkages between phrases within a text. The variety of content types available within the text summarisation space includes many industry verticals, including news aggregation, legal document analysis, healthcare reporting, and applicant tracking systems in which a process will gather and produce numerous applications/resumes quickly to allow for timely decisions concerning applicants.

1.2 Evolution of Text Summarization Techniques

Historically, extractive summarisation was an early approach used to summarise documents. With extractive summarisation, the most relevant sentences were chosen from the original document according to the statistics or structure of the document (how often a keyword was used, was it at the end of the sentence, etc.). Although extractive methods had utilitarian aspects (very low computational cost, audibility), with the exception of certain cases where scientific theory allowed for statistical comparisons between relevant sentences, these methods also lacked coherence and an abstract view of the content. The evolution of deep learning-based methods, primarily transformer networks, created a major shift to utilise a more abstract or implicit view on summarising documents by allowing systems such as BART, T5, and PEGASUS to develop summarising systems comparable to traditional extractive methods that could create fluent summaries by paraphrasing and reorganising the content. The newest generation of summarising systems is based on large language models (LLM's), using the instruction-following aspect of LLM's and contextual inference to create summarisation systems that generate high-quality summaries that utilise inferences derived from the input data and have high coherence across all types of input.

1.3 Research Gap and Problem Statement

While the development of summarization models has made great strides, much of the work in the literature is limited by several key issues. A majority of the work has focused on a single type of summarization (i.e., "extractive" vs. "abstractive" or "language-model-based"), making it difficult to gain broad-based comparative insights between these three categories of summarization methods. Additionally, evaluation methodologies are often inconsistent, as metrics are used only within the context of specific experimental frameworks rather than under a common framework. For example, the evaluation methodologies used to evaluate a particular method may not apply to the evaluation methodologies used for an entirely different method. Furthermore, many of the methods that have been proposed in the literature have not yet been incorporated into real-world summarization systems, thus limiting the ability of researchers to reproduce and use the results of these studies in practice. All of these issues point to the need for a consolidated framework in which to systematically compare different summarization paradigms with respect to their use in a deployable setting.

1.4 Objectives and Contributions

This research aims to address the identified gaps by proposing a unified and deployable framework for comparative text summarization. The key contributions of this work are:

- To analyse research trends in text summarisation, particularly the evolution from extractive and transformer-based abstractive methods to LLM-based approaches, as identified through bibliometric analysis of recent NLP literature.
- To identify gaps in existing summarisation research, including the dominance of isolated paradigms, inconsistent evaluation practices, and the lack of deployable and reproducible systems highlighted by publication and collaboration patterns.
- To examine how research concentration and publication practices (e.g., conference-driven dissemination and regional clustering of authors and institutions) influence the development, evaluation, and real-world adoption of summarisation techniques.

1.5 Paper Organization

The remaining sections of this paper are organised as follows: Section 2 contains a detailed review of research on text summarisation; Section 3 provides a bibliometric analysis of research on text summarisation to provide context for examining trends and gaps in the literature; Section 4 explains the bibliometric analysis; and Section 5 concludes the paper and offers conclusions and future work.

2. LITERATURE REVIEW

Text summarization research has evolved significantly over the past decades, transitioning from rule-based extractive techniques to transformer-based abstractive models and, more recently, large language model (LLM)–driven approaches. This section reviews existing literature by organizing prior work into extractive summarization, abstractive transformer-based summarization, LLM-based summarization, evaluation methodologies, and applied systems in recruitment and document analysis.

2.1 Extractive Text Summarization Approaches

Historically, researchers have developed numerous methods for summarizing text using extractive approaches, allowing for the identification and selection of significant sentences from the original text without altering the underlying meaning. The traditional approach to building extractive models utilizes statistical features (e.g., sentence scores), frequency-based heuristics, and sentence-length ratios, thereby allowing for the retention of accurate and meaningful information. As a result of a thorough review of the various types of extractive summarization methodologies, the researchers found that extractive summarization offers advantages in transparency and computational efficiency; however, these processes are limited with respect to the ability to create semantic abstractions and consistency in meaning [4].

In addition, numerous studies conducted in the field have illustrated the utility of extractive summarization in domain specificity (e.g., legal and financial documentation) where the importance of factual accuracy is paramount. Frequency-based scoring techniques for evaluating sentence relevance and comparing content similarities have produced highly precise and accurate summaries, although it is acknowledged that these techniques do not provide the depth of content understanding that can be achieved by semantically based techniques [8]. Despite their limitations, extractive techniques continue to serve as baseline methodologies and are employed in hybrid approaches to text summarization..

2.2 Abstractive Summarization Using Transformer Models

As an important milestone in progressing to abstractions, transformer based architectures have allowed for the creation of models that can generate entirely new sentences that encapsulate the semantic information contained in the input. In particular, pretrained transformer models like BART, T5 and PEGASUS have played a key role in shaping the trajectory of the field for developing modern systems of abstraction of textual data. In comparative evaluations conducted using multiple datasets, PEGASUS displays the best results when summarising lengthy forms of written material; conversely, the models BART and T5 have shown superior performance in summarising conversational and concise forms of texts [9].

The integration of transformer networks and other technologies into applied computational systems has provided empirical support showing that producing abstractions of a document's content can significantly enhance specific downstream-systems. When using a series of different pretrained transformer models to summarize job listings, researchers found that recommendation accuracy and ranking quality improved considerably, with those models based on BERT ranking as the strongest in terms of ROUGE and other generic recommendation performance indicators [7]. However, there are challenges that are posed to using these systems due to hallucinations, where they may generate abstractions that include information not contained in the source text. Recent literature suggests that not every hallucination is damaging, and it proposes entity-aware approaches to identify between what constitutes "good" abstraction and factual inaccuracies, particularly in relation to legal summary tasks

[5]. With these studies indicating the requirement for evaluation and controlled implementation of transformational networks for producing abstracts.

2.3 Large Language Model–Based Summarization

Abstractive summarization has been advanced with the introduction of LLMs that use their instruction following capabilities and reasoning abilities. By generating fluent, contextualised summaries that often equal or exceed the coherence of those produced by traditional transformer architecture, LLM-based summation systems offer greater functionality and improved reliability in comparison to other solutions currently available for this purpose. A more systematic investigation into LLM architecture and training methodologies has provided additional information about the relative trade-offs between size, performance, and ease of use/reasoning for real world application [6].

Empirical investigations indicate that while more compact LLMs like TinyLLaMa are still feasible in terms of computational resources and thus suitable for deployment within the constraints of limited infrastructure, they have also demonstrated competitive quality with respect to source content synthesis at less than 20% of the cost of LLMs based on 10 million parameters. Moreover, the development of LLMs with greater levels of abstraction (via their increased capacity for generating higher-level representations) has raised additional concerns over how well LLMs might alleviate issues related to hallucination and factual consistency, further supporting the necessity of continued comparative evaluations of these systems to determine their relative strengths and weaknesses when used as extractors or transformers of original text.

2.4 Evaluation Metrics for Text Summarization

Assessing Summarization Quality remains a difficult task, especially when evaluating Extractive vs. Abstractive paradigms using different ROUGE metrics. ROUGE-1, ROUGE-2 and ROUGE-L have all become very popular for their simplicity and scalability. ROUGE was able to demonstrate through an empirical study that it provides a consistent and non-biased way to evaluate Extractive and Abstractive Summaries. Although there may be some limitations when it comes to capturing Semantic Coherence and Factual Accuracy completely, this study has proved that ROUGE is still a viable option for evaluating Summarization techniques [10].

Many other studies have also looked at different metrics to evaluate Summarization, including BLEU, METEOR, and Semantic Similarity, and have suggested various Hybrid Evaluation approaches that work around the limitations associated with n-Grams based metrics [3]. Despite all of the new possibilities created by these recent studies, ROUGE is the prevailing evaluation metric being used to assess Summarization in most cases, and is especially prevalent when conducting a Survey Study.

2.5 Summarization Systems for Recruitment and Resume Analysis

The application of text summarisation to support recruitment systems has been a key development in applied research. Text summarisation has been applied to help recruit large volumes of resumes and job descriptions. Recruitment resume optimisation tools that use Natural Language Processing (NLP) and Machine Learning (ML) to improve the visibility of candidates using Applicant Tracking System (ATS) compatibility and keyword matching and ranking, fail to provide the candidate with the opportunity to see how their resume is semantically summarised.[1] Enhanced recruitment systems integrate autores publications, screening, scoring, and recommendation into end-to-end recruitment pipelines. Less complex multi-agent recruitment frameworks have demonstrated substantial efficiencies but lack transparency and have been unable to make side-by-side comparisons across summarisation methodologies.[2] Most recruitment systems use a single type of summarisation strategy that fails to provide insight into the trade-offs between various summarisation methods.

2.6 Systematic Reviews and Identified Research Gaps

A series of systematic reviews of relevant literature have recently created a clear overview of the body of work concerning Abstractive Summarization, clarifying that Transformer Models are currently the best available models to perform this task and that ROUGE evaluation metrics are commonly used in the vast majority of the research conducted in this area. The limitations of the existing body of work identified by the recent systematic literature reviews have consistently been the same issues across all recent works: "hallucinations"; high Computational Resources Required; "non-interpretability" of operations; and no standardized evaluation framework.

In addition, although there is an enormous amount of work on the topic of Abstractive summarization, almost all of it is focused on very niche (i.e., isolated paradigms) or strongly domain-specific applications. There is currently an extremely limited number of deployable frameworks available to allow for the side-by-side comparison (using standardized conditions) of Extractive / Abstractive / Large Language Model (LLMs) summarization algorithms, with a clear need for the development of such frameworks. In response to the identified need, this paper presents a unified framework for systematically comparing the three summarization paradigms, using as an example a real-world application of Resume Summarization.

3. Bibliometric Analysis

3.1 Data Collection and Methodology

This bibliometric analysis examined the relationships between large language models and their use in NLP and related text-processing disciplines (text processing). This data was collected from Scopus using an established search strategy that identified bibliometric studies involving "large-language-model", "foundation model", or "generative AI" in the context of natural language processing, text and/or language type processing. Studies related to robotics and autonomous systems were excluded so as to maintain a continually thematically coherent body of literature. The analyses totalled 1,014 documents written in English and were selected for this bibliometric analysis.

The Bibliometric analyses were conducted with Biblioshiny (Bibliometrix in R) and VOSviewer, which are commonly employed software applications for conducting quantitative and network-based analyses of the bibliometric literature.

3.2 Most Relevant Authors

Figure 1 shows the main researchers based on their number of publications as shown in the search. It is evident from these results that a small group of researchers publish a disproportionate number of actual publications, which shows the existence of a small but important group of researchers that are pushing forward research in LLM-focused NLP. The groups of authors listed as the most prolific authors, such as Li Y and Wang Y, have published a high number of publications compared to other groups within the dataset examined. As such, these authors are likely representative of the most active research groups in LLM-based NLP. Therefore, the high concentration of publications associated with a single author shows that research into LLMs has a small number of very active and productive researchers contributing to developing new methods. In a rapidly changing field of study, it is common for the leading researchers to have a major influence on the direction of methodology.

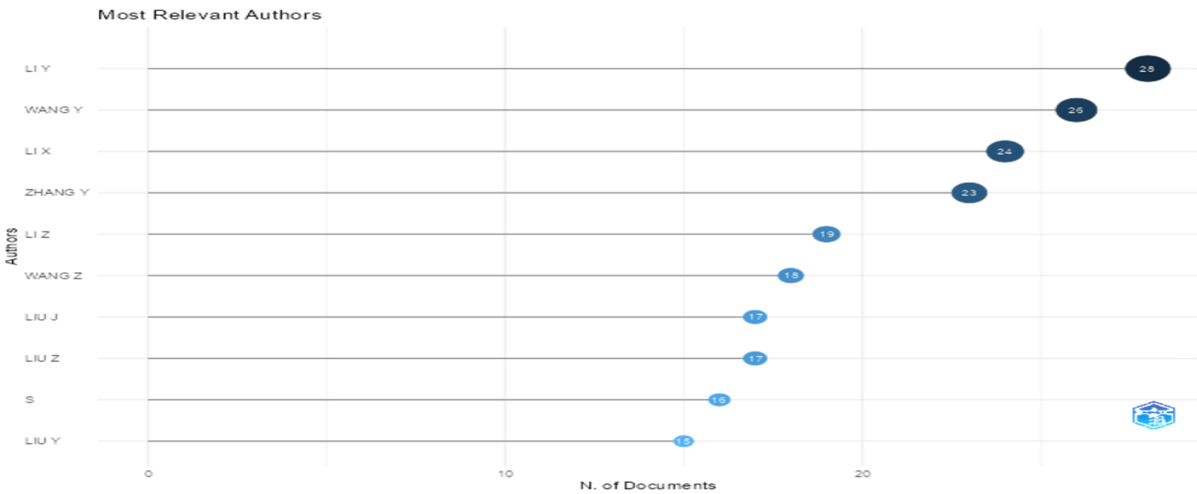

Figure 1. Most relevant authors based on the number of publications in LLM-related NLP research.

3.3 Most Relevant Sources

Figure 2 provides an overview of how published work has been dispersed among different journals/publications.

From the findings, it appears that Lecture Notes in Computer Science is the predominant publication source followed closely by Communications in Computer and Information Science then Lecture Notes in Networks and Systems. The prevalence of conference-based and hybrid-type Publishing is evidence of how quickly LLM-related research develops and LLM researchers' need for rapid communication of their results. This indicative trend is supported further by the journal-based publications like Expert Systems with Applications, Information Fusion, Pattern Recognition, etc., indicating newer LLM-based NLP research has reached a higher level of maturity within the field and has become applied. One important point to make from this analysis is how it demonstrates there is now a good mix between basic research being conducted in the area of LLM-based NLP versus that of conducting both applied and applied-based research.

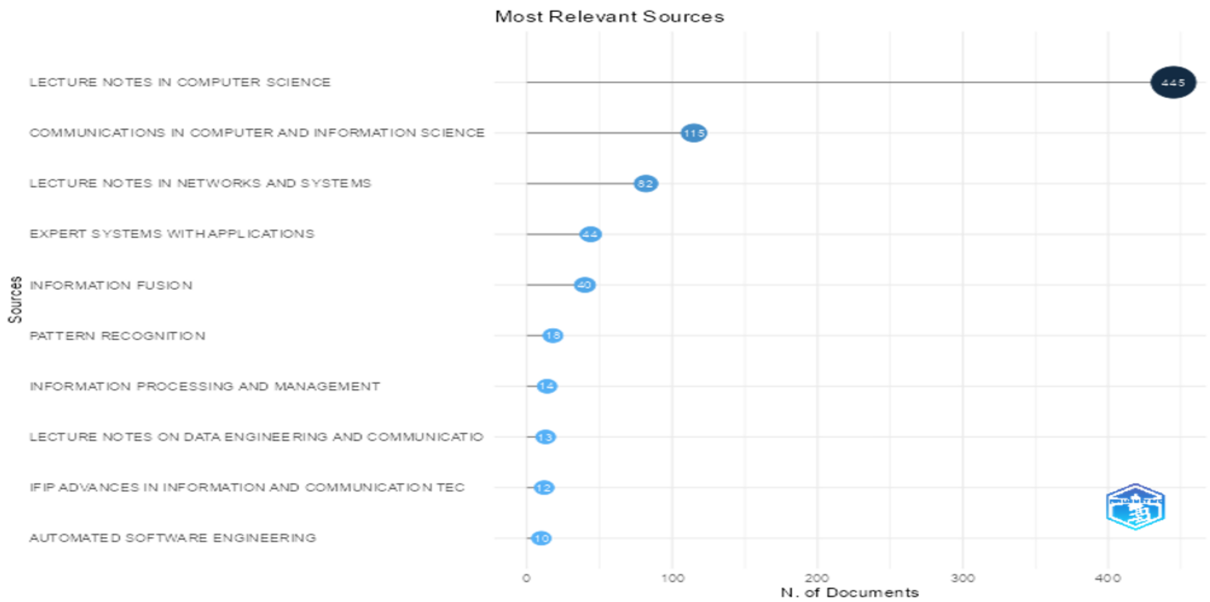

Figure 2. Most relevant publication sources in LLM and NLP research.

3.4 Institutional Collaboration Network

Using VOSviewer to create an institutional collaboration network, we can see in Figure 3 that there are significant clusters associated with some of the highest profile universities of China such as Tsinghua University, Peking University, Shanghai Jiao Tong University, Chinese University of Hong Kong, The Hong Kong University of Science and Technology. The high level of interconnectivity between these major universities located within China and Hong Kong also suggests that there is a very high rate of regional collaboration taking place in these countries, while the links to other parts of the world (particularly Europe and North America) are indicative of a growing global interconnectivity among researchers. In totality, the clustering of institutions illustrates that while LLM research in NLP is global, it is also heavily concentrated within Asia, which has become the principal contributor to this type of research.

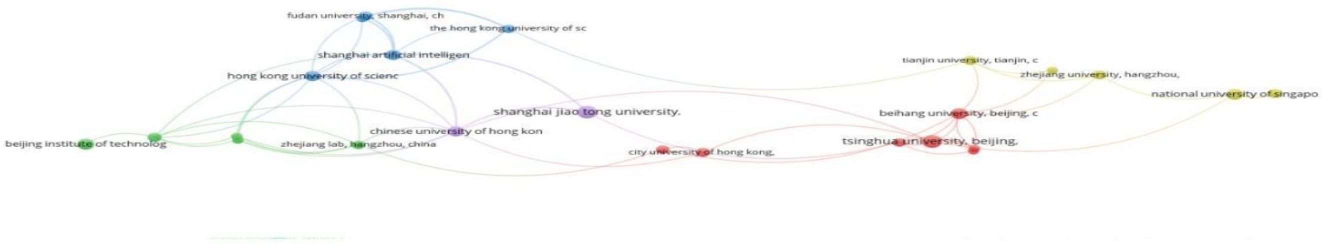

Figure 3. Institutional collaboration network in LLM-based NLP research.

3.5 Geographic Distribution of Research Output

Figure 4 illustrates the geographical distribution of research outputs at a country level using a density map. It was determined through this analysis that the USA and China produced much of the research with India, Germany, the UK, and France as the other major contributors. The high intensity of research in these areas suggest there is a high level of investment from both academia and industry in LLM technology within these regions. In addition, the wide dispersion of LLM-based NLP research across the globe suggests that many countries are interested in this field; however, the limited number of countries with a strong level of LLM-based NLP research represents an imbalance of both research capacity and availability of resources for conducting LLM-based NLP research.

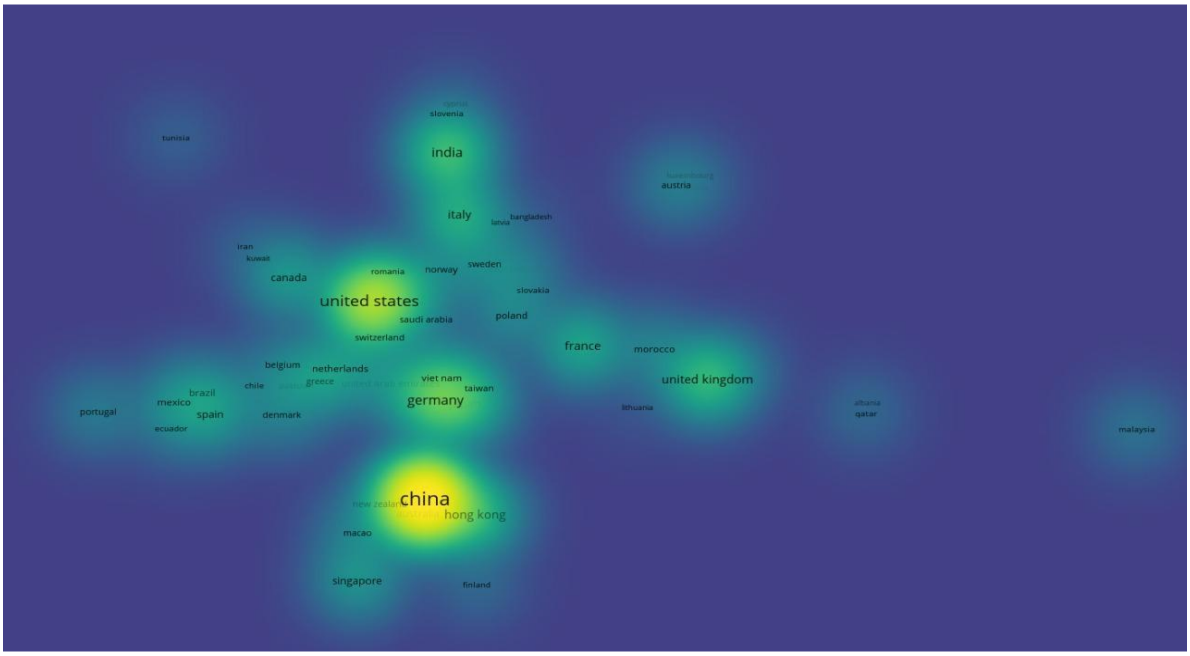

Figure 4. Geographic distribution of publications in LLM-related NLP research.

3.6 Discussion and Implications for the Present Study

The bibliometric results yield multiple important trends that are significant. The first trend suggests that LLM-based NLP methodological innovation predominantly occurs through a small number of key research teams, which accounted for most of all authors and institutions publishing within this fast-evolving area. The second trend indicates that most publication venues for the LLM-based NLP field are oriented toward conferences due to the rapid development of new techniques, resulting in a lack of comparable systems and evaluations of systems used to summarize text. The third trend indicates that the majority of LLM-based NLP research outputs are from a limited geographical area, highlighting the lack of access to computational resources directly correlates with the capability to use a large LLM system for a large-scale deployment, thus influencing overall deployment capability.

The above three trends are what support the current study. While the majority of current literature provides substantial information about the development of the model and theoretical progression for LLM-based NLP research, it remains that there are limited systematic frameworks available that would allow for a fair/equal, side-by-side comparison between extractive, abstractive, and LLM-based summarization techniques that would simulate the resource-constrained environment typical of most production systems. Therefore, our goal is to develop a cohesive, unified framework that allows for a systematic side-by-side comparison under specifications reflective of the end user.

4. Future scope

Although the proposed framework demonstrates the feasibility and value of comparative text summarization work, several avenues for future work and system enhancement remain available. First, the evaluation methodology could be expanded beyond the use of ROUGE as a means of metric evaluation and include the use of semantic similarity metrics such as BERTScore and fact consistency metrics in order to provide a more complete picture of summary quality, particularly for highly abstracted and LLM-based outputs. Second, the addition of hallucination mitigation strategies, including entity-aware summarization and retrieval-augmented generation, will allow us to effectively enhance the level of factual reliability in generative models. From a systems perspective, deploying the advanced resume intelligence framework in GPU-compliant environments would enable the use of larger transformer and LLM models, support the use of batch processing for multiple documents, and provide richer evaluation and analytical capabilities. Additionally, incorporating human-in-the-loop evaluation could enhance reliability and trust in summarization outputs, particularly in high-stakes applications such as recruitment and decision support. Finally, the framework can be adapted to additional domains such as healthcare, legal analysis, and scientific literature summarization, further extending its applicability and impact.

5. CONCLUSION

A unified framework that allows for a side-by-side comparison of extractive-based, abstractive-based, and LLM-based text summarization techniques has been developed within this paper. In light of the disconnection between the current disparate approaches to researching summarization methods, as well as the fact that little effort has been put forth to this point in terms of creating deployable systems to perform a comparative analysis of summarisation techniques, a comprehensive environment incorporating multiple summarisation methodologies has been created within a single space (Hugging Face) that utilises standard ROUGE evaluations applied equally across the same conditions of input data. Resume collection has served as a representative example of a practical implementation, and the iterative routing process (growth process) employed allowed the authors to clarify the tradeoffs between system complexity, computational feasibility, and summarisation quality during the course of this study. Results yielded and qualitative analysis indicated that the extractive model(s) deliver maximum levels of consistency and interpretability, that the abstractive approach produces a reasonable balance between both conciseness and semantic abstraction and that the LLM-based paradigm delivers superior fluency using an increased level of abstraction and by adding a potential for hallucinations. The theoretical works in summarization will now be connected with the practical implementations of summarization by allowing all three approaches to be evaluated simultaneously in a deployable environment.

References

[1] R. Sharma, A. Verma, and S. Gupta, "Resume optimization and applicant tracking system scoring using natural language processing," *International Journal of Computer Applications*, vol. 185, no. 12, pp. 1–7, 2023.

[2] A. K. Mishra and P. Singh, "An intelligent recruitment system using multi-agent artificial intelligence," *Procedia Computer Science*, vol. 167, pp. 1531–1540, 2020.

[3] S. Bhandari, A. Gour, and P. Kumar, "Improving evaluation metrics for text summarization using semantic similarity measures," *Information Processing & Management*, vol. 58, no. 6, Art. no. 102670, 2021.

[4] M. Allahyari, S. Pouriyeh, M. Assefi, S. Safaei, E. D. Trippe, J. B. Gutierrez, and K. Kochut, "Text summarization techniques: A brief survey," *Journal of Big Data*, vol. 4, no. 1, pp. 1–33, 2017.

[5] S. Maynez, S. Narayan, B. Bohnet, and R. McDonald, "On faithfulness and factuality in abstractive summarization," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics (ACL)*, Seattle, WA, USA, 2020, pp. 1906–1919.

[6] W. X. Zhao, K. Zhou, J. Li, et al., "A survey of large language models," *ACM Computing Surveys*, vol. 55, no. 12, pp. 1–35, 2023.

[7] S. Jain and G. Gupta, "Enhancing e-recruitment systems using transformer-based text summarization," *Expert Systems with Applications*, vol. 185, Art. no. 115608, 2021.

[8] R. Kumar and S. K. Jain, "NLP-driven extractive summarization for legal and financial documents," *Procedia Computer Science*, vol. 165, pp. 543–550, 2019.

[9] J. Zhang, Y. Zhao, M. Saleh, and P. Liu, "PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization," in *Proc. 37th Int. Conf. Mach. Learn. (ICML)*, 2020, pp. 11328–11339.

[10] C.-Y. Lin, "ROUGE: A package for automatic evaluation of summaries," in *Proc. ACL Workshop on Text Summarization Branches Out*, Barcelona, Spain, 2004, pp. 74–81.

[11] A. Patel, R. Shah, and M. Joshi, "A document summarization system using BART and T5," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 9, pp. 455–462, 2021.

[12] S. Gupta and A. K. Singh, "Neural network-based text summarization models: A comparative study," *Journal of Artificial Intelligence Research*, vol. 69, pp. 933–960, 2020.

[13] D. Verma, P. Mehta, and R. Patel, "ROUGE score-based evaluation of text summarization models," *International Journal of Engineering Research & Technology*, vol. 10, no. 6, pp. 227–231, 2021.

[14] K. Nair, S. Menon, and A. Thomas, "ATS score prediction and resume–job matching using machine learning," *International Journal of Computer Science and Information Security*, vol. 20, no. 2, pp. 45–52, 2022.

[15] Y. Zhang, D. Li, X. Chen, and Z. Liu, "Large language models for text summarization: A survey," *arXiv preprint arXiv:2401.01234*, 2024.

[16] A. Altmami, M. Al-Zubaidi, and S. R. Alsharif, "Single-document abstractive text summarization: A systematic literature review," *Information*, vol. 16, no. 3, Art. no. 333, 2025.