

Generative Artificial Intelligence for Automated Document Summarization and Knowledge Extraction

Dr. Émilie Dubois

Assistant Professor, Toulouse Institute of Intelligent Systems, France

Abstract

The rapid expansion of digital information has created increasing demand for computational systems capable of condensing lengthy documents and identifying useful knowledge from large text collections. Conventional extractive summarization methods primarily select important passages from source documents, whereas generative approaches attempt to construct new textual representations of salient information. At the same time, knowledge-extraction methods identify entities, attributes, and relationships that can support structured information retrieval and downstream analytical tasks. Integrating these capabilities offers a methodological basis for more comprehensive automated document understanding. This hypothetical methodological study examines a generative artificial intelligence framework for automated document summarization and knowledge extraction. The proposed architecture combines document encoding, attention-guided sequence generation, semantic representation, relation extraction, and post-generation consistency assessment. A conceptual dataset of 600 documents is divided into hypothetical training, validation, and testing partitions. Three model configurations are considered: an extractive baseline, an attention-based generative summarizer, and an integrated generative summarization and knowledge-extraction architecture.

The illustrative analysis compares the configurations using ROUGE-1, ROUGE-2, ROUGE-L, and a hypothetical factual-consistency measure. A separate knowledge-extraction analysis evaluates entity-property, binary-relation, and event-context extraction using precision, recall, and F1 score. Under the assigned values, the integrated architecture records higher summarization overlap while maintaining stronger hypothetical factual consistency than the standalone generative configuration. Knowledge-extraction performance varies across relation categories, with structured entity-property relations producing the highest illustrative F1 score. The study concludes that generative document processing may provide a useful methodological framework for combining natural-language summarization with structured knowledge extraction. However, generated summaries can omit, distort, or introduce information unless their outputs are evaluated against the source documents. The study therefore emphasizes reproducible assessment, factual verification, transparent model comparison, and separation between linguistic fluency and knowledge accuracy.

Keywords: Generative Artificial Intelligence, Automated Summarization, Knowledge Extraction, Natural Language Processing, Abstractive Summarization, Information Extraction, Sequence-to-Sequence Learning, Attention Mechanism, Relation Extraction, Document Intelligence.

1. Introduction

The continuing growth of digital documents has created substantial challenges for individuals and organizations attempting to locate, interpret, and reuse relevant information. Reports, research articles, organizational records, technical documents, and other textual resources frequently contain information distributed across lengthy passages. Manual reading can become inefficient when document collections are large. Automated document-processing techniques therefore aim to identify salient content and represent it in more concise forms while preserving information that remains important for subsequent interpretation and analytical use.

Automatic summarization has traditionally included extractive and abstractive approaches. Extractive methods identify sentences or textual segments considered important and reproduce them within a shorter representation. Graph-based methods such as LexRank and TextRank contributed important approaches for ranking textual units according to structural and semantic relationships (Erkan & Radev, 2004; Mihalcea & Tarau, 2004). Although extraction can retain wording directly from the source, it may produce summaries that contain redundant information or lack coherent transitions between selected sentences.

Generative or abstractive summarization addresses a different objective by constructing new sentences rather than exclusively reproducing original passages. Advances in neural language modeling, sequence-to-sequence learning, and attention mechanisms provided important foundations for such systems. Sutskever et al. (2014) demonstrated sequence-to-sequence learning with neural networks, while Rush et al. (2015) and Chopra et al. (2016) applied neural attention and recurrent architectures to abstractive summarization. These approaches established that summary generation could be treated as a learned conditional language-generation task.

Document understanding also requires more than producing concise text. Structured knowledge extraction identifies entities, properties, events, and semantic relationships that may be useful for search, reasoning, or knowledge-base construction. Banko et al. (2007) introduced open information extraction for discovering relational tuples without requiring a fixed set of manually specified relations. Subsequent work on distant supervision and relation extraction expanded the ability to identify knowledge from large text collections (Mintz et al., 2009; Surdeanu et al., 2012).

The present study combines these two research directions within a hypothetical generative document-analysis framework. The term generative artificial intelligence is used here in a broad methodological sense to describe computational systems capable of producing new textual sequences from source information. The study does not claim the implementation of a modern large-scale generative model. Instead, it develops an illustrative architecture grounded in sequence generation, attention-based summarization, semantic representation, and relation extraction, followed by transparent hypothetical evaluation.

2. Objectives of the Study

The principal objective of this study is to examine how generative artificial intelligence may support automated document summarization and structured knowledge extraction within one analytical framework. The investigation focuses on summary generation, factual preservation, relation extraction, and methodological requirements for evaluating generated document representations.

2.1 Examination of Generative Document Summarization

The first objective is to examine how generative sequence models may condense lengthy documents into shorter textual representations. The proposed framework considers document encoding, attention-based content selection, and recurrent sequence generation. These mechanisms conceptually allow the system to identify salient information while producing sentences that are not restricted to direct extraction. The study evaluates the methodological advantages and limitations of this approach relative to conventional extractive summarization, particularly with respect to conciseness, semantic coverage, and factual preservation.

The objective also considers the distinction between linguistic fluency and informational reliability. A grammatically coherent generated summary may still omit essential information or introduce statements not supported by the source document. Consequently, summary quality should not be assessed solely through readability or lexical overlap. The proposed methodology incorporates a conceptual factual-consistency assessment alongside established summary-overlap measures. This approach emphasizes that useful generative summarization requires both appropriate content compression and defensible correspondence with source information.

2.2 Evaluation of Knowledge Extraction Capability

The second objective is to examine how document-processing models may identify structured knowledge from unstructured text. The hypothetical architecture represents extracted information through entity-property associations, binary relations, and event-context relationships. These categories illustrate different levels of extraction complexity. Entity properties may be expressed relatively directly, whereas event-context relations can depend on sentence structure, contextual information, and interactions among multiple textual elements. The objective therefore evaluates extraction performance separately across relation categories.

Another purpose is to examine whether summary generation and knowledge extraction can function as complementary analytical components. Extracted relations may provide structured information useful for checking whether important facts remain represented within a generated summary. Conversely, summarization may identify salient textual regions from which structured knowledge can be extracted more efficiently. The study does not assume that either mechanism automatically improves the other. Instead, it establishes a framework for comparing their outputs and investigating potential integration empirically.

2.3 Identification of Reliability and Implementation Challenges

The third objective is to identify methodological difficulties associated with generative document understanding. Relevant problems include source-document length, ambiguous language, rare terminology, incomplete relation labels, summary omissions, unsupported generated statements, and evaluation limitations. These conditions may affect both summarization and knowledge extraction. Identifying them is important because fluent generation can conceal factual errors that would be easier to recognize in direct extractive approaches. Reliable system design therefore requires explicit validation against source information.

The study additionally aims to establish requirements for reproducible evaluation. Model comparisons should use identical document partitions, clearly specified preprocessing procedures, common reference summaries, and consistently labeled extraction targets. Training information must

remain separate from evaluation observations. Summary-generation metrics should also be interpreted alongside factual assessment and structured extraction performance. These requirements reduce the possibility that apparently improved language generation is incorrectly interpreted as stronger document understanding.

3. Review of Literature

3.1 Development of Automated Text Summarization

Erkan and Radev (2004) introduced LexRank, which uses graph-based lexical centrality to identify salient sentences within document collections. Mihalcea and Tarau (2004) developed TextRank as a broader graph-based ranking approach applicable to text-processing tasks. These methods demonstrated the effectiveness of modeling relationships among textual units rather than evaluating sentences independently. Lin (2004) subsequently formalized ROUGE as a family of overlap-based measures widely used for evaluating automatic summaries against reference texts.

Barzilay and McKeown (2005) examined sentence fusion for multidocument summarization, demonstrating methods for combining information from related sentences into more compact representations. Haghighi and Vanderwende (2009) investigated generative content models for multidocument summarization, while Nenkova and McKeown (2011) provided a broad synthesis of automatic summarization methods and evaluation concerns. These developments established important foundations for distinguishing sentence extraction, compression, content selection, and generation as related but methodologically different components of automatic summarization.

3.2 Neural Generation and Abstractive Summarization

Bengio et al. (2003) demonstrated neural probabilistic language modeling, establishing an influential foundation for distributed linguistic representations. Sutskever et al. (2014) introduced sequence-to-sequence learning using neural networks, while Bahdanau et al. (2015) developed attention-based neural translation that enabled dynamic focus on different parts of an input sequence. These contributions provided essential methodological foundations for neural text generation because they established mechanisms for encoding variable-length sequences and producing context-dependent output tokens.

Rush et al. (2015) developed a neural attention model for abstractive sentence summarization, demonstrating an end-to-end data-driven generation approach. Chopra et al. (2016) extended abstractive summarization using attentive recurrent neural networks. Nallapati et al. (2016) subsequently examined sequence-to-sequence recurrent architectures for multidocument and multisentence summarization, including mechanisms addressing important words and rare vocabulary. Together, these studies established neural abstractive summarization as an important research direction before the emergence of later large-scale generative language systems.

3.3 Information and Knowledge Extraction

Banko et al. (2007) introduced open information extraction as an approach for identifying relational tuples from large text collections without requiring every relation to be manually specified in advance. Mintz et al. (2009) developed distant supervision for relation extraction by aligning external relational knowledge with textual occurrences. Riedel et al. (2010) subsequently modeled relations and their textual mentions without requiring manually labeled sentences, strengthening the connection between weak supervision and large-scale information extraction.

Fader et al. (2011) further developed open information extraction by identifying linguistic patterns corresponding to relational expressions. Surdeanu et al. (2012) proposed multi-instance multi-label learning for relation extraction, addressing uncertainty regarding which relation is expressed by a particular entity pair. These contributions demonstrated that knowledge extraction requires appropriate representations of entities, relations, and context. They also established that relation-extraction performance depends substantially on supervision quality, relation complexity, and evaluation methodology.

4. Research Methodology

4.1 Research Design and Generative Framework

The present study adopts a hypothetical methodological research design centered on automated document summarization and structured knowledge extraction. The proposed architecture includes a document encoder, attention mechanism, recurrent generative decoder, semantic representation component, and relation-extraction layer. The summarization component produces a concise textual representation of each source document, while the extraction component identifies selected structured facts. A consistency stage conceptually compares generated summary content with extracted source-document knowledge to identify unsupported or missing factual elements.

Three conceptual model configurations are evaluated. The first is an extractive baseline based on sentence importance. The second is an attention-based generative summarizer producing newly constructed summary sentences. The third integrates generative summarization with structured knowledge extraction and a conceptual factual-consistency check. None of these configurations has been implemented for the present manuscript. The architecture and results are hypothetical and are designed to demonstrate how an empirical comparison could be structured.

4.2 Hypothetical Dataset and Data Preparation

The hypothetical corpus contains 600 text documents representing a mixture of technical reports, institutional documents, and informational articles. Four hundred twenty documents are assigned to training, ninety to validation, and ninety to independent testing. Each document is assumed to contain a human-prepared reference summary and a set of annotated knowledge relations. These documents and annotations are fictional and do not correspond to any real dataset, organization, or published document collection.

Data preprocessing conceptually includes sentence segmentation, tokenization, vocabulary normalization, document-length management, and identification of candidate entity mentions. The knowledge annotations are divided into entity-property relations, binary relations, and event-context relations. The testing set remains unavailable during model development. Reference summaries are used only for evaluation after generation. This chronological separation of modeling and evaluation information is intended to prevent the hypothetical design from relying on target summaries when producing predictions.

4.3 Evaluation Measures and Analytical Procedure

Summary evaluation employs ROUGE-1, ROUGE-2, and ROUGE-L as lexical overlap measures. ROUGE-1 evaluates unigram overlap, ROUGE-2 considers bigram overlap, and ROUGE-L reflects longest-common-subsequence correspondence between generated and reference summaries. A

hypothetical factual-consistency score is also included to demonstrate the importance of evaluating whether generated statements remain supported by source documents. This score is illustrative and does not represent an established automated factuality metric implemented in the present study.

Knowledge extraction is evaluated using precision, recall, and F1 score. Precision represents the proportion of extracted relations that match the hypothetical reference annotations, while recall measures the proportion of reference relations successfully identified. F1 combines these dimensions through their harmonic mean. The formulas are defined as $\text{Precision} = \text{TP}/(\text{TP}+\text{FP})$, $\text{Recall} = \text{TP}/(\text{TP}+\text{FN})$, and $\text{F1} = 2\text{PR}/(\text{P}+\text{R})$. All calculations reported below are based on deliberately assigned hypothetical counts.

5. Analysis

5.1 Hypothetical Document Summarization Performance

The first analysis compares the three conceptual summarization configurations across the ninety-document hypothetical test set. The extractive baseline is assigned ROUGE-1 of 38.4 and ROUGE-2 of 14.7. The attention-based generative configuration records higher illustrative overlap but slightly lower factual consistency because unrestricted text generation is assumed to introduce a small number of unsupported statements. The integrated configuration combines generation with knowledge-oriented consistency checking and is assigned the strongest overall hypothetical performance.

The assigned values illustrate why generative summarization should be evaluated through more than lexical overlap. The standalone generative configuration records higher ROUGE values than the extractive baseline, but its hypothetical factual-consistency score declines from 95% to 90%. The integrated configuration achieves a factual-consistency score of 94% while retaining higher overlap values. These figures are created solely for demonstration and do not establish that knowledge extraction necessarily improves generative summarization in real applications.

Table 1: Hypothetical Summarization Performance Across Three Model Configurations

Model Configuration	ROUGE-1	ROUGE-2	ROUGE-L	Factual Consistency (%)
Extractive Baseline	38.4	14.7	34.8	95
Attention-Based Generative Model	41.9	17.6	38.2	90
Integrated Generative + Knowledge Framework	43.5	18.9	39.7	94

Note: All values are illustrative. No summarization model was trained or evaluated to generate these results.

5.2 Hypothetical Knowledge Extraction Performance

The second analysis examines the integrated framework across three categories of knowledge relations. A hypothetical reference collection contains 160 entity-property relations, 140 binary relations, and 100 event-context relations. The framework is assigned 132 true positives for entity-property extraction, 108 for binary relations, and 74 for event-context relations. The results illustrate decreasing performance as

contextual dependence becomes more complex, although this pattern is an assumption rather than an empirically observed property.

Entity-property extraction produces hypothetical precision of 88.0%, recall of 82.5%, and F1 score of approximately 85.2%. Binary relations produce F1 of 80.0%, while event-context extraction produces approximately 75.9%. These values illustrate the calculation of structured extraction metrics and the need to report relation-specific performance. A single aggregate score could conceal substantial differences among extraction categories and provide an incomplete representation of the system's actual knowledge-retrieval capability.

Table 2: Hypothetical Knowledge Extraction Results by Relation Category

Knowledge Category	TP	FP	FN	Precision (%)	Recall (%)	F1 Score (%)
Entity-Property Relations	132	18	28	88.0	82.5	85.2
Binary Relations	108	22	32	83.1	77.1	80.0
Event-Context Relations	74	21	26	77.9	74.0	75.9

Note: The reference counts and classification outcomes are fictional and do not represent results from an implemented information-extraction system.

6. Discussion

6.1 Generative Summarization and Content Representation

The hypothetical findings illustrate how generative summarization may provide more flexible content representation than direct sentence extraction. Sequence-based generation can conceptually combine information appearing in separate portions of a document and express it through more concise wording. Early neural summarization studies by Rush et al. (2015), Chopra et al. (2016), and Nallapati et al. (2016) demonstrated the methodological potential of attention-based and recurrent sequence generation for abstractive summarization. The present framework builds conceptually on these foundations.

However, generative flexibility creates additional reliability concerns. A model that constructs new text can potentially introduce phrases unsupported by the source document, even when the resulting summary appears linguistically coherent. Extractive systems are less flexible but retain direct textual correspondence with the original material. Consequently, improvements in overlap measures should not automatically be interpreted as improvements in factual reliability. Generative document-processing research should therefore evaluate semantic correctness and source support alongside conventional summarization metrics.

6.2 Integration of Summarization and Knowledge Extraction

The integrated framework is designed around the principle that summarization and knowledge extraction address complementary aspects of document understanding. Summarization provides a compact linguistic representation, while extraction converts selected information into structured relations. Knowledge relations can conceptually support checking whether salient entities and relationships remain represented in a generated summary. Conversely, document summarization may reduce the textual region requiring subsequent detailed analysis when only highly relevant information is needed.

The hypothetical results nevertheless do not establish that integrating these tasks produces an automatic improvement. Errors in relation extraction may themselves distort the information used during consistency assessment. Similarly, a generated summary may represent a fact accurately while using language that does not align directly with a structured extraction rule. Effective integration therefore requires evaluation at both the textual and relational levels. Independent validation of each component remains necessary before system-level improvements can be attributed to combined architecture design.

6.3 Interpretation of Automated Evaluation

ROUGE remains useful for measuring lexical correspondence between generated and reference summaries, but it cannot independently determine whether a generated statement is factually accurate. Lin (2004) developed ROUGE as an automatic summary-evaluation framework based primarily on overlap characteristics. The present hypothetical analysis therefore includes a separate factual-consistency dimension. This distinction is important because two summaries can express similar information using different wording, while high lexical overlap can coexist with misleading or incomplete factual representation.

Knowledge-extraction metrics introduce similar interpretive requirements. Precision and recall depend on the quality and completeness of reference annotations. Relations that are correct but absent from the reference set may be counted incorrectly, while ambiguous annotation guidelines can influence reported performance. Relation-specific evaluation is therefore preferable to reporting only an aggregate score. Reliable empirical research should document annotation procedures, evaluator agreement, relation definitions, and error categories alongside numerical extraction metrics.

7. Challenges

7.1 Factual Consistency and Unsupported Generation

One of the central challenges in generative summarization is maintaining accurate correspondence between generated text and the source document. Sequence-generation models optimize statistical relationships learned from training information and may produce plausible language that is not explicitly supported by the input. This problem becomes particularly important when summaries are used for research, organizational, legal, medical, or technical interpretation. A fluent summary can create misplaced confidence when unsupported statements are difficult to distinguish from accurate content.

Factual consistency therefore requires dedicated validation rather than reliance on linguistic quality. Potential safeguards include source-reference comparison, entity and relation matching, consistency checks, and human review in consequential applications. The present hypothetical framework includes a knowledge-oriented consistency component to demonstrate this principle. However, no automated consistency procedure should be assumed to identify every unsupported statement. Evaluation must consider both false acceptance of incorrect content and rejection of valid paraphrased information.

7.2 Document Length and Knowledge Complexity

Long documents present computational and representational difficulties because relevant information may be distributed across distant sections. A summary model may focus strongly on introductory or frequently repeated content while overlooking information appearing later in the document. Knowledge extraction faces a related problem when a relation depends on entities or events mentioned across

multiple sentences. Local processing may therefore be insufficient for documents whose meaning depends on broader contextual relationships.

Knowledge complexity also increases when the same entities participate in multiple relationships or when pronouns, abbreviations, and indirect references must be resolved. Event-context extraction can require understanding temporal, causal, or organizational connections extending beyond a single sentence. The lower hypothetical performance assigned to this category illustrates this methodological concern without claiming empirical confirmation. Future research should examine how hierarchical representations and broader contextual modeling influence extraction reliability across different document types.

7.3 Evaluation Quality and Model Generalization

Generative document-processing systems may perform differently across domains because vocabulary, document structure, writing style, and information density vary substantially. A model trained on news articles may not transfer directly to technical manuals or scholarly research documents. Evaluation based on a narrow corpus may therefore overstate general performance. Representative test sets should include documents reflecting the intended operational environment and should remain independent from model-development decisions.

Another challenge concerns the construction of reference summaries and knowledge annotations. Human-written summaries can legitimately differ in their selection and wording of important information, while relation annotations may vary according to interpretation. This creates uncertainty in automated evaluation. Reliable studies should describe reference-generation procedures, use appropriate multiple annotations where feasible, and conduct qualitative error analysis. Numerical metrics are valuable, but they should be interpreted together with examination of omissions, unsupported content, relation errors, and document-specific failure patterns.

8. Conclusion

Generative artificial intelligence provides a methodological framework for transforming unstructured documents into concise textual summaries and structured knowledge representations. The present hypothetical study integrated attention-based sequence generation, semantic document representation, and relation extraction within a conceptual document-intelligence architecture. The approach emphasizes that automated summarization and knowledge extraction serve different but potentially complementary functions. Effective document understanding requires both meaningful content compression and accurate representation of factual relationships contained in the source.

The illustrative analysis assigned higher ROUGE scores to the integrated generative architecture than to the extractive and standalone generative configurations. It also demonstrated different hypothetical precision, recall, and F1 values across knowledge-relation categories. These results were intentionally constructed for methodological illustration and are not empirical findings. They should therefore be interpreted only as examples of how summarization and extraction performance can be reported and compared within a reproducible experimental framework.

The study concludes that future research should evaluate generative document-processing systems through factual consistency, semantic coverage, structured extraction accuracy, and domain generalization rather than linguistic fluency alone. Particular attention should be given to unsupported generation, complex contextual relations, long-document processing, and reference-quality limitations.

Empirical experimentation using representative corpora and independently prepared annotations is necessary before the proposed framework can be considered effective for practical document summarization or knowledge-extraction applications.

References

1. Chang, Y., Li, Z., & Le, X. (2024). Guiding large language models via external attention prompting for scientific extreme summarization. In Proceedings of the Fourth Workshop on Scholarly Document Processing (pp. 226–242). Association for Computational Linguistics. <https://aclanthology.org/2024.sdp-1.22/>
2. Glickman, M., & Zhang, Y. (2024). AI and generative AI for research discovery and summarization. arXiv. <https://arxiv.org/abs/2401.06795>
3. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 7871–7880). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.703>
4. Liu, P. J., Saleh, M., Pot, E., Goodrich, B., Sepassi, R., Kaiser, Ł., & Shazeer, N. (2018). Generating Wikipedia by summarizing long sequences. In International Conference on Learning Representations. <https://arxiv.org/abs/1801.10198>
5. Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing order into texts. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (pp. 404–411). Association for Computational Linguistics. <https://aclanthology.org/W04-3252/>
6. Nallapati, R., Zhai, F., & Zhou, B. (2017). Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 31, No. 1). <https://doi.org/10.1609/aaai.v31i1.10958>
7. Onan, A., & Alhumyani, H. A. (2024). DeepExtract: Semantic-driven extractive text summarization framework using LLMs and hierarchical positional encoding. Journal of King Saud University Computer and Information Sciences, 36, Article 102178. <https://doi.org/10.1016/j.jksuci.2024.102178>
8. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21(140), 1–67. <https://www.jmlr.org/papers/v21/20-074.html>
9. Reichenpfader, D., Müller, H., & Denecke, K. (2024). A scoping review of large language model based approaches for information extraction from radiology reports. npj Digital Medicine, 7, Article 222. <https://doi.org/10.1038/s41746-024-01219-0>
10. See, A., Liu, P. J., & Manning, C. D. (2017). Get to the point: Summarization with pointer-generator networks. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 1073–1083). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P17-1099>
11. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (Vol. 30). Curran Associates, Inc. <https://arxiv.org/abs/1706.03762>

12. Van Veen, D., Van Uden, C., Blankemeier, L., Delbrouck, J.-B., Aali, A., Bluethgen, C., Pareek, A., Polacin, M., Reis, E. P., Seehofnerová, A., Rohatgi, N., Hosamani, P., Collins, W., Ahuja, N., Langlotz, C. P., Hom, J., Gatidis, S., & Chaudhari, A. S. (2024). Adapted large language models can outperform medical experts in clinical text summarization. *Nature Medicine*, 30, 1134–1142. <https://doi.org/10.1038/s41591-024-02855-5>
13. Wang, T., Yang, C., Zou, M., Liang, J., Xiang, D., Yang, W., Wang, H., et al. (2024). A study of extractive summarization of long documents incorporating local topic and hierarchical information. *Scientific Reports*, 14, Article 10140. <https://doi.org/10.1038/s41598-024-60779-z>
14. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., ... Rush, A. M. (2020). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 38–45). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.emnlp-demos.6>
15. Xu, D., Chen, W., Peng, W., Zhang, C., Xu, T., Zhao, X., Wu, X., Zheng, Y., Wang, Y., & Chen, E. (2024). Large language models for generative information extraction: A survey. *Frontiers of Computer Science*, 18, Article 186357. <https://doi.org/10.1007/s11704-024-40555-y>
16. Zhang, H., Yu, P. S., & Zhang, J. (2024). A systematic survey of text summarization: From statistical methods to large language models. *arXiv*. <https://arxiv.org/abs/2406.11289>
17. Zhang, T., Ladhak, F., Durmus, E., Liang, P., McKeown, K., & Hashimoto, T. B. (2024). Benchmarking large language models for news summarization. *Transactions of the Association for Computational Linguistics*, 12, 39–57. https://doi.org/10.1162/tacl_a_00632
18. A survey of text summarization: Techniques, evaluation and challenges. (2024). *Natural Language Processing Journal*, 7, Article 100070. <https://doi.org/10.1016/j.nlp.2024.100070>
19. A comprehensive survey for automatic text summarization: Techniques, approaches and perspectives. (2024). *Neurocomputing*, 603, Article 128280. <https://doi.org/10.1016/j.neucom.2024.128280>
20. Neural natural language processing for long texts: A survey on classification and summarization. (2024). *Engineering Applications of Artificial Intelligence*, 133, Part C, Article 108231. <https://doi.org/10.1016/j.engappai.2024.108231>