IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

Meeting Transcriber

Prof. Sharmila Chidaravalli, Tejas K S, Yadu Nandan B
Department of Information Science and Engineering
Global Academy of Technology, Bangalore, India
c.sharmila11@gmail.com, tejasks1ga20is144@gmail.com, yadu1ga20is161@gmail.com

Abstract: *In a world where collaboration is increasingly digital, the efficiency and security of meetings have become paramount. Our project introduces a comprehensive online meeting platform extension that leverages advanced technologies to enhance the meeting experience. Using BERT for data classification, question answering, and summarization, along with an advanced Language Model (LLM) for data protection, we address common challenges such as information overload, language barriers, limited searchability, and a lack of actionable insights. This extension aims to revolutionize meetings, making them more productive, collaborative, and secure*

Keywords: BERT (Bidirectional Encoder Representations from Transformers,), Text Classification, Key-Based Searching, LLM (Large language models)

I. INTRODUCTION

Welcome to the future of meetings! Our project is all about making online meetings better. Imagine meetings where it's easy to understand, find information, and keep things private.

Meetings can be tricky:

- Too much information: There's a ton of stuff discussed, and it's hard to remember what's important.
- Different languages: Sometimes, teams speak different languages, making it tough to understand each other.
- Hard to find things: Imagine searching for something specific in a really long meeting recording – not fun!
- Not enough insights: Getting useful insights from discussions can be tricky without good tools.
- But guess what? We've got cool tech, like BERT, to help us make sense of data and answer questions. And we're adding an extra layer of protection called Language Model (LLM) to keep things super private and secure.
- Our dream is to turn meetings into something more than just talking. We're focusing on easy recording, summarizing, and finding stuff quickly. Plus, our top priority is keeping your info safe with the LLM.

In section 2, it consists of the brief literature review. In section 3, it consists of table of comparison results. section 4 provides conclusion and references

II. LITRATURE REVIEW

Analyses and references to real studies on NLP and AIML are included in the literature review

This study [1] discuss the methodology for the survey of the most recent neural abstractive text summarization models includes a thorough analysis of pertinent scientific literature architecture of encoder-decoder, mechanism of attention, Reinforcement learning-based training techniques, Distributional semantic rewards for training. Datasets from CNN/Daily Mail, DUC 2004, Giga word, and Newsroom. The research accomplished a thorough analysis of the most advanced neural abstractive text summarization models available today. It gave information about the methods, training plans, datasets, assessment measures, and design components of abstractive summarization models. The study also emphasized the difficulties and problems with these systems, which improved knowledge of the area and opened the door for more investigation and advancement in automatic text summarization.

The paper [2] focuses on Text summarizing strategies using several Large Language Models (LLMs) are examined in the article "Text Summarization Using Large Language Models: A Comparative Study of MPT-7b-instruct, Falcon-7b-instruct, and OpenAI Chat-GPT Models" by Lochan Basya, Mihir Sanghvi, et al. The methodology consists of studies

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

using supervised and unsupervised summarizing techniques carried out on two different datasets: CNN/Daily Mail 3.0.0 and XSum. The study offers a thorough investigation of text summarization methods with LLMs, including the OpenAI Chat-GPT, Falcon-7b-instruct, and MPT-7b-instruct models.

The paper [3] focuses on a two-stage decoder framework for Multi-Document Scientific Summarization (MDSS) is presented in the paper "Multi-Document Scientific Summarization from a Knowledge Graph-Centric View," written by Pancheng Wang, Shasha Li, Kunyuan Pang, Liang liang He, Dong Li, Jintao Tang, Ting Wang, et al. in 2023. To improve content and connection modelling, KGSum, the suggested approach, uses knowledge graphs created from input scientific papers. Graph Updater, Entity-Sentence Updater, KGtext Generator, Summary Generator, copy mechanism, and combined training with an auxiliary decoder for knowledge graph representation and model stability are some of the components of the methodology. By outperforming other models on the multi-Xscience dataset, a large-scale MDSS dataset, and producing better outcomes in the domain, the study illustrates the efficacy of KGSum.

Thestudy [4] discusses using pre-trained language models (BERT, ALBERT), sequence-to-sequence (Seq2Seq) models, attention mechanisms, transformers, keyword extraction, sentence division, data preprocessing, and model fine-tuning, the paper presents a novel method for short text summarization based on keyword templates. As part of the methodology, preprocessing data is done by extracting keywords and classifying text according to these keywords. Transformers are integrated for decoding and BERT for encoding in the model architecture. The study demonstrates enhanced performance over baseline models and higher ROUGE scores on the LCSTS (Large Scale Chinese Short Text Summarization) dataset. Among the contributions are a customized model architecture and an altered data pretreatment technique intended for Chinese short text summarizing.

This paper [5] reveals that by modifying event cues, this multi-granularity summary system—which comprises of an Event Selector and an Event-aware Summarizer—achieves adaptable summarization. The need of multi-granularity summarization systems in practical situations is emphasized in the work, along with the necessity of benchmarks. The authors demonstrate the efficacy of the suggested GRANUSUM framework for unsupervised multi-granularity summarization by introducing their own dataset, GranuDUC, and conducting tests onmulti-news, arXiv, DUC2004, and GranuDUC.

The study [6] given presents the Dual-Attention Pointer Network (DAPT) model, an enhanced model for abstractive text summarization. Bidirectional LSTM encoders, self-attention mechanisms, soft attention mechanisms, a pointer structure, and an enhanced coverage method to handle problems like recurrence in generated summaries are some of the important elements incorporated into the architecture. By addressing problems with attention, coverage, and training objectives, the proposed DAPT model seeks to enhance previous models and offer a complete solution for abstractive text summarizing tasks. Through ROUGE-based assessments on the CNN/Daily Mail Dataset and the LCSTS Dataset, the experiments compare the model's efficacy with state-of-the-art models on various datasets. In order to provide accurate and efficient text summarization, this research presents the dual-attention pointer network (DAPT), which combines gate mechanisms and self-attention.

The study [7] presents a 10-step sequential technique that incorporates subject modelling, transformer-based punctuation restoration, and embedding-based sentence selection for extractive summarization of call transcripts. Effective tabulation and summary production are guaranteed by this process. Large Language Models (LLMs), partitional clustering, and internal validation indices are all incorporated into the approach; nevertheless, there are recommendations for enhancement that include adding LLMs, growing the number of terminals, and taking outside data into account. The efficacy of the extractive summarization technique is demonstrated by the studies conducted on the CNN/Dailymail and New York Times datasets. This method combines sentence selection, subject modelling, channel separation, and punctuation restoration to address issues with call transcript summarization. The paper advances the field by introducing a new punctuation restoration accuracy metric and a novel BERT transformer-based model [7].

According to study [8] in order to develop fitness functions for evolutionary cluster-based techniques in automatic text summarization, the research presents a new approach. The suggested method entails employing various mapping techniques, such as term frequency-inverse document frequency (IF-IDF), one-hot encoding (OHE), and latent Dirichlet allocation (LDA), to represent texts or documents as numeric vectors. In order to find significant sentences in the Automatic Text Summarization (ATS) task, these methods investigate the importance of lexical and semantic

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

information. Using genetic algorithms and genetic programming, the work suggests automatically generating an objective function for unsupervised text summary tasks using the DUC02 and CNN/Daily Mail datasets. It is discovered that the finest outcomes come from combining lexical and semantic information in the ideal way, advancing the field of extractive text summarization.

The research [9] focuses on the abstractive summarizing of texts consisting of a single document is the main emphasis of the paper's novel approach to text summarization. To improve the quality of generated summaries, the suggested TIF-SR (subject Information Fusion and Semantic Relevance) model computes semantic relevance and includes subject keyword information. Adjusting BERT, Attention Mechanism, Transformer Networks, and Semantic Similarity Calculation are all part of the process. The tests, carried out on the LCSTS and NLPCC2017 datasets, reveal that, in comparison to baseline models, the summaries produced by the TIF-SR model are more topical, have greater semantic similarity with the original content, and have better fluency and sentence coherence. The study offers a thorough examination of the test findings, emphasizing how well topic information and semantic relevance computation are integrated into the text summarizing task.

The study's [10] methodology for recommending tailored suggesting customized summaries of instructional materials is described in the paper. It makes use of formative assessment in the form of multiple-choice examinations, text preparation, and summary to gauge students' comprehension. In order to maximize summarizer performance, the configuration of the algorithm is researched, and learner-generated data from actual learning contexts—specifically, textbook data—are used to quantitatively assess summarizer performance. The suggested approach examines lengthy learning materials and generates brief text summaries based on the results of multiple-choice exams. A university-level B.S. course used the methodology to test it, and the results showed that it was comparable to teacher recommendations.

A unique code summarization approach called Fret is presented in the paper[11] and is based on a reinforcer-transformer architecture. With the goal of tackling issues like code comprehension and extended dependencies, Fret is made to produce functionally clear, accurate, and thorough code descriptions while delivering exceptional performance. Three primary parts make up the model: the Decoder, Code Encoder, and Functional Reinforcer. A range of machine learning algorithms are used, such as training and inference, decoding, functional reinforcement, and code encoding. These methods make use of feed-forward neural networks, positional encoding, feed-forward networks, BertC, BertNL, Hadamard & Norm, and multi-head attention. The research makes use of Python and Java datasets that were gathered from GitHub and include documentation comments and snippets of Java code, with distribution tendencies that are comparable to those of the Java dataset [11].

HunEmBERT, an improved BERT model for categorizing sentiment and emotion in political communication, is presented in the work [12]. The transformers library's Trainer API is used to fine-tune the huBERT model's parameters for sentiment and emotion categorization. Evaluation techniques include manual review of sentences that were incorrectly classified and standard metrics like Precision and Recall for error analysis using confusion matrices. The BERT Model, Transformer-based Pipeline, fine-tuning methods, Trainer API, and a number of assessment measures are all used in the study. The ISEAR Dataset and the HunEmPoli Corpus are used in the experiments. In this study, a domain-specific corpus (HunEmPoli) for sentiment and emotion analysis in political writings written in Hungarian is being created. This corpus is then used to fine-tune the huBERT model for sentiment and emotion classification tasks. The assessment of the performance of the adjusted models shows state-of-the-art.

A reinforced abstractive text summarization model with semantic added reward is presented in the work [13]. The single-layer bi-directional LSTM encoder and decoder in the model is used in conjunction with a sequence-to-sequence attention mechanism. Interestingly, it uses intra-decoder attention to reduce problems such as duplicating hard words and generating repetitive phrases. In order to maximize metrics, the approach uses policy learning and reinforcement learning, with an emphasis on word similarity, semantic similarity, and word mover distance to guarantee readability. The Gigaword summarization dataset is used in the study. We offer two reward functions that augment n-gram matching with semantic values: ROUGE-SIM and ROUGE-WMD. Sequence-to-sequence, Transformer-based pre-learning, and reinforcement learning models are all outperformed by the models created in this work. Readability and grammatical accuracy show significant gains, demonstrating the efficacy of the suggested strategy.

A innovative multi-step methodology for automatic news text summarizing is presented in the publication [14]. Preprocessing, segmentation, tokenization, stop-word removal, stem extraction, and the extraction of important features

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

at the word and sentence levels are all included in the process. To extract word features, the suggested approach makes use of a fuzzy logic system, a SpeGenetic Algorithm, and a particular keyword extraction technique. The DUC2002 datasets are used in the investigation. For news text summarizing, the model integrates fuzzy logic, evolutionary algorithms, and multi-feature. It extracts significant aspects, ranked according to the qualities of the news text, including word and sentence features. The fair weights are produced by use of the fuzzy logic system. The technique makes use of ROUGE-2 for the ideal distribution of weight and ROUGE-1 for fitness values. Results from experiments show that this approach works better than

A succinct and effective BERT-based methodology for identifying rumors on Twitter, called CE-BERT, is presented in the article [15]. The process consists of layer selection that works, efficient fine-tuning, and use of classification techniques. A number of layer reduction techniques are examined, with a focus on the significance of the fine-tuning procedure. These include the basic model, six-layer models, four-layer models, and two-layer models. The Twitter 15 Dataset, Twitter 16 Dataset, and PHEME Dataset are used in the study. The study shows that CE-BERT is a low-computing-requirement effective model for Twitter rumor identification. It performs better than the state-of-the-art models, especially when it comes to source text scenarios, demonstrating its efficacy in Twitter rumor identification.

This study [16] discusses the methodology for Korean abstractive text summarization utilizing a Multi-Encoder Transformer model. Key components include Transformer-based Encoder-Decoder Model, Multi-Encoder Architecture, Combining Representations, Fine-Tuning with BERT-based PLMs, Auto-regressive Decoding, and Evaluation Metrics such as ROUGE and BERTScore. Datasets utilized include Law (AI-Hub) dataset, News (AI-Hub) dataset, and News (NIKL) dataset. The paper introduces a transformer-based encoder-decoder model tailored for Korean abstractive summarization and proposes a multi-encoder architecture utilizing various pre-trained models. It demonstrates significant performance enhancements over single-encoder models, particularly with diverse encoders, across three Korean summarization datasets, while also conducting thorough ablation studies and qualitative analyses to assess the efficacy of different techniques, including OOV token handling.

[17] presents a text summarization method called Topic Information Fusion and Semantic Relevance (TIF-SR), which aims to enhance the quality of generated summaries by combining topic information and semantic relevance. Key components include BERT-based encoding, Topic Information Fusion, Transformer architecture, semantic relevance calculation, and fine-tuning objective. Datasets utilized include LCSTS and NLPCC2017 datasets. The TIF-SR model significantly improves ROUGE scores on LCSTS and NLPCC2017 datasets by incorporating topic information and semantic relevance, resulting in fluent, readable, and semantically relevant summaries.

[18] introduces a Topic Modeling-Based Framework for Extracting Marketing Information From E-Commerce Reviews. The methodology involves extracting strengths and weaknesses of individual products, identifying connections among different products, and predicting product trends. Techniques employed include clustering algorithms, topic modeling (such as BERTopic and DCN), and a Transformer-based forecasting model. General datasets are considered for the survey paper. The paper achieved significant advancements in extracting product characteristics, identifying related products, and forecasting product trends using topic modeling techniques and a Transformer-based model. Moreover, it provided valuable insights into consumer preferences and market trends through thorough analysis and evaluation.

[19] proposes a framework for automatic context extraction and comparison of short text documents, specifically focusing on service provider policies, with a particular emphasis on GDPR compliance and comparison with 3000 web service privacy policies. The framework includes six phases. Techniques employed include Entity Recognition, Modal Logic, Machine Learning, BERT Text Summarization, and Semantic Web Languages to extract high-frequency entities from GDPR documents. The dataset comprises over 3000 web service provider privacy policies, categorized into EU-based and worldwide policies, downloaded post-May 2018 to ensure relevance post-GDPR. The research utilizes NLP, deep learning, and semantic web techniques to efficiently analyze short or incomplete texts, focusing on privacy policies in progress.

[20] presents a Chinese text summary model leveraging keyword templates and fine-tuning with pretrained language models like BERT. The methodology involves data preprocessing, model architecture design, and experimentation. The study explores pretrained language models, extractive models, and abstractive models for language understanding tasks, extractive summarization, and abstractive summarization, incorporating attention mechanisms and BERT. The LCSTS

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

dataset is utilized. The proposed BSA model outperforms baseline models on the LCSTS dataset, enhancing ROUGE scores and text division quality. Future work should explore its applicability to other pretrained language models.

The text discusses a proposed abstractive summarization model [21] that builds upon the BART architecture by adding a discriminator and a cluster generator. The model splits original text into sentences and classifies each sentence as salient or non-salient, creating clusters for salient and non-salient sentences. Experiments on XSUM and CNN/DailyMail datasets showed the model outperformed BART and PEGASUS in terms of ROUGE and BERTScore. Various techniques for abstractive summarization are utilized, including the Encoder-Decoder Attention Model, Pointer Network, Coverage Mechanism, Transformer Model, UniLM, BERTSUM and BART Models, PEGASUS Model, Reinforcement Learning, Topic Model, Multimodal Information, Attention Head Masking, Information Theory, Extraction-and-Paraphrasing, Entity Aggregation and Factuality Consistency, Deep Communicating Agents, Sentence Correspondence, Graph, and Bottom-Up Approach. The study uses CNN/DailyMail Dataset and XSUM Dataset. A new pre-trained sequence-to-sequence model improves abstractive summarization by extracting salient sentences and clustering context vectors. Experimental results outperform existing models, addressing fact inconsistency and long document summarization.

The algorithm proposed by Yu et al. [22] utilizes a pre-trained BERT language model to vectorize the sentence-level features of policy text, followed by classification using a classifier. BERT (Bidirectional Encoder Representations from Transformers), a transformer-based language model pre-trained on large corpora of text data, is employed in this technique. The paper uses policy text data from various policy domains. The trained model achieves an improvement in policy text classification accuracy, with the highest F1 value of 93.25% on the test set. This result demonstrates a nearly 6% increase in performance compared to the BERT model's classification task for the MRPC (Microsoft Research Paraphrase Corpus) task.

The research conducted by Liu et al. [23] focuses on text classification, combining the Bert model with Bayesian networks. Initially, the Bayesian network is employed for coarse classification into two categories, providing an approximate category range for each text. Subsequently, the Bert model is utilized for fine-grained classification within the identified range. The techniques employed include the Bert model, a pre-training model based on deep learning for natural language processing (NLP), and Bayesian networks, a probabilistic graphical model representing a set of random variables and their conditional dependencies via a directed acyclic graph. The paper utilizes text data related to people's livelihood governance, containing unstructured data from various sources within this domain. By combining the Bayesian network's coarse classification with the Bert model's fine-grained classification, the paper achieves an improvement in text classification accuracy. This approach reduces errors caused by classification defects inherent in using only one of the methods, ultimately enhancing the overall accuracy of text classification.

The research conducted by Seo et al. [24] focuses on classifying and interpreting unstructured traffic crash description data through three stages: data preprocessing, model development, and model performance evaluation. Techniques utilized include word embedding algorithms, a BERT-based text classification model, and cross-validation for robustness. The study analyzes 2,427 traffic crash datasets from the Korean Traffic Crash Analysis System in Daegu, Korea, focusing on crash data elements, weather conditions, and written descriptions. The developed method classifies traffic crash text into standardized data using the BERT model, achieving a text classification accuracy of over 95% for most labels. The study aims to increase the utilization of underutilized traffic crash data in traffic safety research, ultimately reducing crashes and improving traffic safety.

The research by Tian and Wang et al. [25] introduces a scientific document retrieval system comprising three modules: Document Preprocess, Mathematical Expression Similarity, and Text Similarity. These modules are responsible for extracting mathematical expressions, unifying LaTeX and MathML, and calculating similarity between text parts. Techniques employed include FDS for decomposition, HFS for calculating similarity, BERT for text similarity, keyword extraction, and cosine similarity to analyze mathematical expressions. The dataset consists of scientific documents containing mathematical expressions and related texts, from which keywords are extracted to train BERT and calculate text similarity. The study proposes a retrieval method that combines mathematical expressions and contextual text for mathematical information retrieval, incorporating membership degree, HFS, BERT, context keywords, and word vectors. Future research aims to improve extraction methods, address special cases, and evaluate Chinese and English algorithms.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

An et al. [26] present a Topic Modeling-Based Framework for Extracting Marketing Information From E-Commerce Reviews, automating product extraction, pros and cons identification, and trend forecasting from Naver Shopping review data using clustering algorithms and a Transformer-based forecasting model. Techniques utilized include Mecab, POS analysis, KcBERT, topic modeling, and co-occurrence frequency to identify related products. The study employed clustering algorithms, dimension reduction algorithms, topic modeling, text summarization, and transformer-based forecasting to analyze review data, extract insights, and predict trends. Analyzing 530,877 consumer reviews from Naver Shopping for 17 products, the study identified key POS terms and analyzed search volume trends using KcBERT, TF-IDF, and clustering algorithms. Utilizing topic modeling to extract product characteristics and related products from consumer reviews, the study aimed to uncover marketing insights. Although the DCN-based model performed well, its results were slightly less favorable than the baseline BERTopic. The insights gleaned can guide marketing strategies, identify products with similar intentions, and inform product introductions. The product trend prediction model also improved its predictive performance by extracting key product information from text data, though it cannot predict future changes due to identifying past data patterns.

Wasim et al. [27] introduce a Multi-Label Question Classification framework for factoid and list type questions in biomedical question answering. The methodology comprises two main components: Multi-label Biomedical Question Classification and Proposed Biomedical Question Answering Methodology. Preprocessing involves tokenization, lemmatization, and dependency parsing, followed by feature extraction, data transformation, classification, and collective re-ranking. The study utilizes the MLBioMedLAT corpus and BioASQ dataset. The paper presents a novel methodology for classifying biomedical questions into multiple labels, achieving enhanced accuracy through sophisticated preprocessing and classification techniques. Additionally, the proposed question answering approach demonstrates superior performance compared to existing systems, surpassing baseline metrics and showcasing its efficacy in biomedical question answering tasks.

Meng et al. [28] present a study focusing on efficiently classifying electric power audit texts using pre-training and fine-tuning models such as BERT, MAE, and CLIP. These models require less supervised data for fine-tuning and outperform non-pre-trained models. The EPAT-BERT model introduces two pre-training tasks specifically tailored for electric power audit text classification using BERT's structure: word-granularity masked language modeling and entity-granularity masked language modeling, which enhance the understanding of electric power audit texts. The study utilizes a dataset comprising 1.5M electric power-related texts and 1,500 audit texts, alongside baseline models including Naive Bayes, SVM, GBDT, AdaBoost, XGBoost, TextCNN, LSTM, and BERT. The EPAT-BERT model demonstrates effectiveness in classifying electric power audit texts through two ablation experiments. Results indicate that the model significantly outperforms existing models in accuracy, precision, recall, and F1 score compared to traditional machine learning models. The model's extensibility allows for its application to other electric power text classification tasks and project type annotation.

El-Kassas et al. [29] conducted a comprehensive survey of Automatic Text Summarization (ATS) techniques and methodologies. The methodology involved discussing various approaches, methods, building blocks, techniques, datasets, evaluation methods, and future research directions in ATS. The paper covers extractive, abstractive, and hybrid approaches to ATS, highlighting their differences and applications. Additionally, it explores different building blocks and techniques used in ATS, along with the datasets commonly employed for training and evaluation purposes. Although the survey paper does not contain any datasets, it achieves the goal of providing researchers with a comprehensive overview of ATS, serving as a valuable resource for understanding the current state of the field, identifying research gaps, and guiding future directions for improving ATS techniques.

Table of Comparison Result:

Sl no	Author/ year	Research /Work Paper	Methodology	Technique	Dataset/Input	Experiment/ Observation
1	AYESHA AYUB SYED, FORD LUMBAN	A Survey of the State-of-the-Art Models in Neural Abstractive Text	The methodology used in the survey of state-of-the-art models	Encoder-decoder architecture, Attention mechanism,	CNN/Daily Mail dataset, DUC 2004 dataset, Gigaword	The paper achieved a comprehensive review of the current state-of-the-art models in neural abstractive text

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

	GAOL,TOKURO MATSUO *et al.* (2020)	Summarization	in neural abstractive text summarization involves a comprehensive review of relevant scientific literature	Training methods based on reinforcement learning, Distributional semantic rewards for training	dataset, Newsroom dataset	summarization. It provided insights into the design elements, mechanisms, training strategies, datasets, and evaluation metrics used in abstractive summarization models. Additionally, the paper highlighted the challenges and issues associated with these systems, contributing to a better understanding of the field and paving the way for future research and innovation in automatic text summarization
2	Lochan Basya Mihir Sanghvi et al. (2022)	Text Summarization Using Large Language Models: A Comparative Study of MPT-7b-instruct, Falcon-7b-instruct, and OpenAI Chat-GPT Models Lochan Basyal	Methodology used in this study involves conducting experiments on various Large Language Models (LLMs) using two distinct datasets, CNN/Daily Mail 3.0.0 and XSum	The techniques used in the study include supervised and unsupervised summarization methods.	CNN/Daily Mail 3.0.0 and the Extreme Summarization (XSum) dataset.	This research embarked on a comprehensive exploration of text summarization techniques using various Large Lan-guage Models (LLMs
3	Pancheng Wang, Shasha Li, Kunyuan Pang, Liangliang He, Dong Li, Jintao Tang, Ting Wang. et al. (2023)	Multi-Document Scientific Summarization from a Knowledge Graph-Centric View	The methodology described in the provided text involves a two-stage decoder framework for Multi-Document Scientific Summarization (MDSS). The model, named KGSum, utilizes knowledge graphs constructed from	he model uses Graph Updater, Entity-Sentence Updater, KGtext Generator, Summary Generator, copy mechanism, and joint training with an auxiliary decoder for	The Multi-Xscience dataset, a large-scale MDSS dataset	The study proposes a knowledge graph-centric Transformation-based model for MDSS, which significantly outperforms other models and achieves the best results on the multi-Xscience dataset.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

			input scientific papers to enhance content and relationship modeling.	knowledge graph representation and model stability.		
4	SHUAI ZHAO, FUCHENG YOU, AND ZENG YUAN LIU et al. (2022)	Leveraging Pre-Trained Language Model for Summary Generation on Short Text	The paper proposes a novel approach for short text summary generation based on keyword templates. This approach involves preprocessing data by extracting keywords and dividing text based on these keywords. The paper describes the model architecture, which utilizes a combination of BERT for encoding and Transformers for decoding. Selection.	Pre-trained Language Models (BERT, ALBERT), Sequence-to-Sequence (Seq2Seq) Models, Attention Mechanism, Transformers, Keyword Extraction, Sentence Division, Data Preprocessing , Model Fine-Tuning	LCSTS (Large Scale Chinese Short Text Summarization) dataset	Proposes a novel approach for short text summarization using pre-trained language models and keyword templates. Introduces a modified data preprocessing method and model architecture tailored for Chinese short text summarization. Demonstrates improved performance compared to baseline models, as evidenced by higher ROUGE scores.
5	Ming Zhong, Yang Liu†Suyu Ge,Yuning Mao, Yizhu Jiao,Xingxing Zhang, Yichong Xu,Chenguang Zhu, Michael Zeng, Jiawei	Unsupervised Multi-Granularity Summarization	The text discusses a GRANUSUM multi-granularity summarization framework, comprising an Event-aware Summarizer and an Event Selector, which achieves multi-granularity summarization by adjusting event hints	The text discusses customized summarization, unsupervised approaches, and multi-granularity summarization models, emphasizing the need for benchmarks and introduces	Multi-News, arXiv, DUC2004, and the newly introduced GranuDUC	This paper highlights the significance of multigranularity summarization systems in real-world scenarios, proposing the first unsupervised multi-granularity summarization framework GRANUSUM, and demonstrating its effectiveness through experiments

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

	Han§et al. (2022)			their own dataset, GranuDUC.		
6	Zhixin Li, Zhi Peng, Suqin Tang, Canlong Zhang, Huifang Maet al. (2020)	Text Summarization Method Based on Double Attention Pointer Network	The presented work describes an improved model for abstractive text summarization, referred to as the Dual-Attention Pointer Network (DAPT) model. The architecture incorporates several key components, including bidirectional LSTM encoders, self-attention mechanisms, soft attention mechanisms, a pointer structure, and an improved coverage mechanism to address issues such as repetition in generated summaries	The proposed DAPT model aims to improve upon existing models by addressing issues related to attention, coverage, and training objectives, providing a comprehensive solution for abstractive text summarization tasks. The experiments involve comparisons with state-of-the-art models on different datasets, highlighting the model's effectiveness through ROUGE-based evaluations	LCSTS Dataset, CNN/Daily Mail Dataset	This paper introduces a dual-attention pointer network (DAPT) for text summarization, combining self-attention and gate mechanisms for accurate and consistent summary generation, improving ROUGE evaluation index.
7	PRATIK K. BISWAS 1 AND ALEKSANDR IAKUBOVICH et al. (2022)	Extractive Summarization of Call Transcripts	The proposed 10-step sequential procedure for extractive summarization of call transcripts includes topic modeling, embedding-based sentence	The method, incorporating LLMs, partitional clustering, and internal validation indices, is proposed for improvement, with	CNN/Dailymail , New York Times datasets	This paper presents an extractive summarization technique that addresses challenges in call transcript summarization by combining channel separation, topic modeling, sentence selection, and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

			selection, and transformer-based punctuation restoration, ensuring efficient summary generation and tabulation	suggestions for incorporating LLMs, expanding terminals, and considering external information		punctuation restoration. It uses a novel BERT transformer-based model and a new metric for punctuation restoration accuracy
8	ÁNGEL HERNÁNDEZ-CASTAÑEDA,RENÉ ARNULFO GARCÍA-HERNÁNDEZ2 AND YULIA LEDENEVA et al. (2023)	Toward the Automatic Generation of an Objective Function for Extractive Text Summarization	The proposed method aims to create fitness functions for evolutionary cluster-based methods for automatic text summarization. It involves representing texts or documents as numeric vectors using different mapping methods. Four methods are proposed: term frequency-inverse document frequency (tf - idf), one-hot encoding (OHE), and latent Dirichlet allocation (LDA). These methods explore the relevance of lexical and semantic information for identifying relevant sentences in the ATS task	The method, incorporating LLMs, partitional clustering, and internal validation indices, is proposed for improvement, with suggestions for incorporating LLMs, expanding terminals, and considering external information.	DUC02 and CNN/Daily Mail datasets	The study proposes an automatic generation of an objective function for unsupervised text summary tasks using a genetic algorithm and genetic programming. The optimal combination of lexical and semantic information is found to yield the best results.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

9	FUCHENG YOU, SHUAI ZHAO, AND JINGJING CHEN et al. (2020)	A Topic Information Fusion and Semantic Relevance for Text Summarization	The methodology used in the paper involves proposing a new approach to text summarization, focusing on abstractive summarization of single-document texts. The paper introduces the TIF-SR (Topic Information Fusion and Semantic Relevance) model, which incorporates topic keyword information and calculates semantic relevance to improve the quality of generated summaries.	Fine-tuning BERT, Attention Mechanism, Transformer Networks, Semantic Similarity Calculation	LCSTS, NLPCC2017	Shows that the generated summaries by the proposed model are closer to the topic, have higher semantic similarity with the source document, better fluency, and sentence coherence compared to baseline models. Provides a comprehensive analysis of experimental results, highlighting the effectiveness of incorporating topic information and semantic relevance calculation in text summarization task
10	LUCA CAGLIERO, LAURA FARINETTI, AND ELENA BARALIS et al. (2019)	Recommending Personalized Summaries of Teaching Materials	The proposed methodology uses formative assessment with multiple-choice tests, text preparation, and summarization to evaluate learners' understanding	Formative Assessment, Text Preparation, he algorithm configuration is studied to optimize summarizer performance. Quantitative evaluation of summarizer performance is conducted on real learner-generated data in a real	real learner-generated data from textbooks	This paper presents a methodology for exploring large learning documents, providing short textual summaries based on multiple-choice test outcomes. The method was tested in a university-level B.S. course, showing similarity to teacher recommendations.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

				learning context		
11	Ruyun Wang, Hanwen Zhang, Guoliang Lu, Lei Lyu, and Chen Lyu et al. (2020)	This paper introduces a novel code summarization model called Fret, based on a reinforcer-transformer architecture. Fret effectively generates functionally clear, accurate, and comprehensive code descriptions, alleviating issues like code understanding and long dependency, and achieving remarkable performance.	The Fret model is designed for summarizing source code, providing a concise and human-readable description of its functionality. The model consists of three main components: the Functional Reinforcer, Code Encoder, and Decoder.	various machine learning algorithms, including functional reinforcement, code encoding, decoding, and training and inference. These algorithms use multi-head attention, positional encoding, multi-head attention, feed-forward networks, BertC, BertNL, and Hadamard & Norm, as well as feed-forward neural networks	The java and Python datasets, collected from GitHub, contain Java code snippets and documentation comments, with similar distribution trends to the Java dataset.	This paper introduces a novel code summarization model called Fret, based on a reinforcer-transformer architecture. Fret effectively generates functionally clear, accurate, and comprehensive code descriptions, alleviating issues like code understanding and long dependency, and achieving remarkable performance.
12	ISTVÁN ÜVEGES AND ORSOLYA RING et al. (2021)	HunEmBERT: A Fine-Tuned BERT-Model for Classifying Sentiment and Emotion in Political Communication	The huBERT model was fine-tuned for sentiment and emotion classification, with parameters adjusted using the transformers library's Trainer API. Evaluation included standard metrics like Precision and Recall,	BERT Model, Transformer-based Pipeline, Fine-tuning, Trainer API, Evaluation Metrics	HunEmPoli Corpus, ISEAR Dataset	The study involved creating a domain-specific corpus (HunEmPoli) for sentiment and emotion analysis in Hungarian political texts, followed by fine-tuning the huBERT model for sentiment and emotion classification tasks using this corpus. Evaluation of the fine-tuned models' performance

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

			alongside error analysis using confusion matrices and manual inspection of misclassified sentences			demonstrated state-of-the-art results in sentiment analysis and acceptable results in emotion analysis within the political domain, accompanied by error analysis to identify common error patterns and challenges.
13	Heewon Jang and Wooju Kim et al. (2021)	Reinforced Abstractive Text Summarization with Semantic Added Reward	The model uses a sequence-to-sequence attention mechanism with a single-layer bi-directional LSTM encoder and decoder, incorporating intra-decoder attention to mitigate repeated phrase generation and copying difficult words.	The model employs policy learning and reinforcement learning to optimize metrics, focusing on word similarity, semantic similarity, and word mover distance, ensuring readability.	Gigaword summarization dataset.	The study presents two reward functions, ROUGE-SIM and ROUGE-WMD, which add semantic values to n-gram matching. The models perform better than sequence-to-sequence, Transformer-based pre-learning, and reinforcement learning models, with improvements in readability and grammatical accuracy
14	Yan Du and Hua Huo et al. (2020)	News Text Summarization Based on Multi-Feature and Fuzzy Logic	The proposed method uses a multi-step approach for automatic news text summarization, including preprocessing, segmentation, stop-word removal, tokenization, stem extraction, and extraction of key features at word and sentence levels.	method for extracting word features using a SpeGenetic Algorithm, a fuzzy logic system, and a specific keyword extraction method.	DUC2002 Datasets.	This paper proposes a new model combining multi-feature, genetic algorithm, and fuzzy logic for news text summarization. It extracts important features like word and sentence features, graded based on news text characteristics, and uses a fuzzy logic system to generate fair weights. The method uses ROUGE-1 for fitness values and ROUGE-2 for optimal weight allocation. Experimental results show that this method

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

						outperforms other methods in generating high-quality news summaries.
15	Rini Anggrainingsih; Ghulam Mubashar Hassan; Amitava Datta et.al (2023)	CE-BERT: Concise and Efficient BERT-Based Model for Detecting Rumors on Twitter	The methodology of the study involves proposing a concise and efficient BERT-based model named CE-BERT for detecting rumors on Twitter. The key techniques employed in developing CE-BERT include selecting effective layers, efficient fine-tuning, and classification	Uses various layer reduction strategies, including the basic model, six-layer models, four-layer models, and two-layer models, and emphasizes the importance of a fine-tuning process.	Twitter 15 Dataset, Twitter 16 Dataset and PHEME Dataset	The study reveals CE-BERT is an efficient Twitter rumor detection model with reduced computational requirements, outperforming state-of-the-art models in source text scenarios.
16	YOUHYUN SHIN et al. (2023)	Multi-Encoder Transformer for Korean Abstractive Text Summarization	Transformer-Based Encoder-Decoder Model, Multi-Encoder Architecture, Combining Representations, Fine-Tuning, Auto-regressive Decoding, Evaluation Metrics.	Transformer architecture, multi-encoder architecture, fine-tuning with BERT-based PLMs, Flat and parallel combination strategies for combining representations, Auto-regressive decoding, Evaluation using ROUGE and BERTScore metrics	Law (AI-Hub) dataset News (AI-Hub) dataset News (NIKL) dataset	The paper introduces a transformer-based encoder-decoder model tailored for Korean abstractive summarization and proposes a multi-encoder architecture utilizing various pre-trained models. It demonstrates significant performance enhancements over single-encoder models, particularly with diverse encoders, across three Korean summarization datasets, while also conducting thorough ablation studies and qualitative analyses to assess the efficacy of different

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

						techniques, including OOV token handling.
17	Fucheng You, Shuai Zhao, and Jingjing Chen et.al (2020)	A Topic Information Fusion and Semantic Relevance for Text Summarization	TIF-SR is a text summarization method that combines topic information and semantic relevance, enhancing the quality of generated summaries by minimizing negative log probability and maximizing cosine similarity.	The model uses BERT-based encoding, Topic Information Fusion, Transformer architecture, semantic relevance calculation, and a fine-tuning objective to enhance semantic relevance in source documents.	LCSTS and NLPCC2017 Datasets	The TIF-SR model significantly improves ROUGE scores on LCSTS and NLPCC2017 datasets by incorporating topic information and semantic relevance, resulting in fluent, readable, and semantically relevant summaries.
18	YUSUNG AN, DONGJU KIM, JUYEON LEE, HAYOUN G OH,JOO-SIK LEE, DONGH WA JEONG et al. (2023)	Topic Modeling-Based Framework for Extracting Marketing Information from E-Commerce Reviews	The methodology used in the paper involves extracting strengths and weaknesses of individual products, identifying connections among different products, and predicting product trends.	The techniques employed include clustering algorithms, topic modeling (such as BERTopic and DCN), and a Transformer-based forecasting model. It demonstrated the effectiveness of employing multiple techniques such as BERTopic, DCN, and	General datasets are considered for survey paper.	The paper achieved significant advancements in extracting product characteristics, identifying related products, and forecasting product trends using topic modeling techniques and a Transformer-based model. Moreover, it provided valuable insights into consumer preferences and market trends through thorough analysis and evaluation

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

				Transformer models, showcasing improvements over traditional methods like LDA and K-means algorithms		
19	Lavanya Elluri, Sai Sree Laya Chukkapalli, Karuna Pande Joshi, Tim Finin, and Anupam Joshi et al. (2021)	A BERT Based Approach to Measure Web Services Policies Compliance With GDPR	The research proposes a framework for automatic context extraction and comparison of short text documents, specifically focusing on service provider policies, specifically applying GDPR and comparing with 3000 web service privacy policies. The framework includes six phases.	Techniques like Entity Recognition, Modal Logic, Machine Learning, BERT Text Summarization, and Semantic Web Languages are used to extract high-frequency entities from GDPR documents.	The dataset comprises over 3000 web service provider privacy policies, categorized into EU-based and worldwide policies, downloaded post-May 2018 to ensure relevance post-GDPR.	The research uses NLP, deep learning, and semantic web techniques to efficiently analyze short or incomplete texts, focusing on privacy policies in progress.
20	Shuai Zhao, Fucheng You, and Zeng Yuan et al. (2020)	Leveraging Pre-Trained Language Model for Summary Generation on Short Text	The research proposes a Chinese text summary model using keyword templates, fine-tuning with pretrained language models like BERT, involving data preprocessing, model architecture, and experimentation.	The study explores pretrained language models, extractive models, and abstractive models for language understanding tasks, extractive summarization, and abstractive	LCSTS dataset	The proposed BSA model outperforms baseline models on the LCSTS dataset, enhancing ROUGE scores and text division quality. Future work should explore its applicability to other pretrained language models.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

				summarization, incorporating attention mechanisms and BERT.		
21	Sung-Guk Jo, Seung-Hyeok Park, Jeong-Jae Kim, and Byung-Won et al. (2016)	Learning Cluster Patterns for Abstractive Summarization	The text discusses a proposed abstractive summarization model that builds upon the BART architecture by adding a discriminator and a cluster generator. The model splits original text into sentences and classifies each sentence as salient or non-salient, creating clusters for salient and non-salient sentences. Experiments on XSUM and CNN/DailyMail datasets showed the model outperformed BART and PEGASUS in terms of ROUGE and BERTScore.	The study uses various techniques for abstractive summarization, including the Encoder-Decoder Attention Model, Pointer Network, Coverage Mechanism, Transformer Model, UniLM, BERTSUM and BART Models, PEGASUS Model, Reinforcement Learning, Topic Model, Multimodal Information, Attention Head Masking, Information Theory, Extraction-and-Paraphrasing, Entity Aggregation and Factuality Consistency, Deep Communicati	CNN/DailyMail Dataset and XSUM Dataset	A new pre-trained sequence-to-sequence model improves abstractive summarization by extracting salient sentences and clustering context vectors. Experimental results outperform existing models, addressing fact inconsistency and long document summarization.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

				ng Agents, Sentence Correspondence, Graph, and Bottom-Up Approach.		
22	Bihui Yu; Chen Deng; Liping Bu et al. (2022)	Policy Text Classification Algorithm Based on Bert	The algorithm first employs a pre-trained BERT language model to vectorize the sentence-level features of the policy text. Then, the obtained feature vectors are input into a classifier for classification	The technique used in the paper is BERT (Bidirectional Encoder Representations from Transformers), a transformer-based language model pre-trained on large corpora of text data.	The policy text data, documents from various policy domains	The paper achieves an improvement in policy text classification accuracy, with the trained model achieving the highest F1 value of 93.25% on the test set. This result demonstrates a nearly 6% increase in performance compared to the BERT model's classification task for the MRPC (Microsoft Research Paraphrase Corpus) task.
23	Songsong Liu; Haijun Tao; Shiling Feng et al. (2019)	Text Classification Research Based on Bert Model and Bayesian Network	The methodology used in the paper involves combining the Bert model with Bayesian networks for text classification. Initially, the Bayesian network is utilized to perform a coarse classification into two categories, providing an approximate category range for each text. Subsequently, the Bert model is employed to	The techniques used in the paper are the Bert model, a pre-training model based on deep learning for natural language processing (NLP), and Bayesian networks, a probabilistic graphical model that represents a set of random variables and their conditional dependencies	Text data related to people's livelihood governance, containing unstructured data from various sources within this domain.	The paper achieves an improvement in the accuracy of text classification by combining the Bayesian network's coarse classification with the Bert model's fine-grained classification. This combination reduces errors caused by classification defects inherent in using only one of the methods, ultimately enhancing the overall accuracy of text classification.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

			classify the text into specific categories within the identified range.	via a directed acyclic graph.		
24	Younghoon Seo, Jihyeok Park, Gyungtaek Oh, Hyungjoo Kim, Jia Hu, and Jaehyun So et al. (2023)	Text Classification Modeling Approach on Imbalanced-Unstructured Traffic Accident Descriptions Data	The methodology classifies and interprets unstructured traffic crash description data through three stages: data preprocessing, model development, and model performance evaluation. It uses word-embedding algorithms, a BERT-based text classification model, and cross-validation for robustness.	Word Embedding Algorithm, NLP Models, BERT-Based Text Classification Model, K-Fold Cross-Validation, and Stratified K-Fold Cross-Validation for text preprocessing, data quality enhancement, and classification.	The study analyzes 2,427 traffic crash datasets from the Korean Traffic Crash Analysis System in Daegu, Korea, focusing on crash data elements, weather conditions, and written descriptions.	The study developed a method for classifying traffic crash text into standardized data using the BERT model, a natural language processing technique. The model exhibited the highest performance in interpreting three traffic crash elements and had a text classification accuracy of over 95% for most labels. The study aims to increase the use of underutilized traffic crash data in traffic safety research, reducing crashes and improving traffic safety.
25	Xuedong Tian and Jiameng Wang et al. (2021)	Retrieval of Scientific Documents Based on HFS and BERT	The scientific document retrieval system consists of three modules: Document Preprocess, Mathematical Expression Similarity, and Text Similarity, which extract mathematical expressions, unify LaTeX and MathML, and calculate similarity between text parts.	Techniques used include FDS for decomposition, HFS for calculating similarity, BERT for text similarity, keyword extraction, and cosine similarity to analyze mathematical expressions.	The dataset consists of scientific documents, containing mathematical expressions and related texts. Keywords from these contexts were used to train BERT and calculate text similarity.	This study proposes a retrieval method combining mathematical expressions and contextual text for mathematical information retrieval. It calculates membership degree, HFS, BERT, context keywords, and word vectors. Future studies aim to improve extraction methods, address special cases, and evaluate Chinese and English algorithms.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

26	Yusung An, Dongju Kim, Juyeon Lee, Hayoung Oh, Joo-Sik Lee, and Donghwa Jeong are all members of the team. et al. (2023)	Topic Modeling-Based Framework for Extracting Marketing Information from E-Commerce Reviews	The study automates product extraction, pros and cons, and trend forecasting from Naver Shopping review data using clustering algorithms and a Transformer-based forecasting model. It uses Mecab, POS analysis, KcBERT, topic modeling, and co-occurrence frequency to identify related products.	The study utilized clustering algorithms, dimension reduction algorithms, topic modeling, text summarization, and transformer-based forecasting to analyze review data, extract insights, and predict trends.	The study analyzed 530,877 consumer reviews from Naver Shopping for 17 products, identifying key POS terms, and analyzed search volume trends using KcBERT, TF-IDF, and clustering algorithms.	This study used topic modeling to extract product characteristics and related products from consumer reviews, aiming to uncover marketing insights. The DCN-based model performed well, but its results were slightly less favorable than the baseline BERTopic. The insights can guide marketing strategies, identify products with similar intentions, and inform product introductions. The product trend prediction model also improved its predictive performance by extracting key product information from text data. However, it cannot predict future changes due to identifying past data patterns.
27	MUHAMMAD WASIM, WAQAR MAHMOOD, MUHAMMAD NABEEL ASIM, MUHAMMAD USMAN GHANI et. Al (2019)	Multi-Label Question Classification for Factoid and List Type Questions in Biomedical Question Answering	The methodology employed in the paper consists of two main components: Multi-label Biomedical Question Classification and Proposed Biomedical Question Answering Methodology.	Preprocessing: Tokenization, lemmatization, and dependency parsing. Feature extraction, Data Transformation, Classification, Collective Re-ranking.	MLBioMedLAT corpus, BioASQ dataset	The paper presents a novel methodology for classifying biomedical questions into multiple labels, resulting in enhanced accuracy through sophisticated preprocessing and classification techniques. Furthermore, the proposed question answering approach demonstrates superior performance compared to existing systems, surpassing baseline metrics and showcasing its efficacy in

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

						biomedical question answering tasks.
28	Qinglin Meng, Yan Song, Jian Mu, Yuanxu Lv, Jiachen Yang, Liang Xu, Jin Zhao, Junwei Ma, Wei Yao, Rui Wang, and Maoxiang Xiao et .al (2023	Electric Power Audit Text Classification With Multi-Grained Pre-Trained Language Model	The study aims to efficiently classify electric power audit texts using pre-training and fine-tuning models like BERT, MAE, and CLIP, which require less supervised data for fine-tuning and outperform non-pre-trained models.	EPAT-BERT introduces two pre-training tasks for electric power audit text classification using BERT's structure. The tasks include word-granularity masked language modeling and entity-granularity masked language modeling, enhancing understanding of electric power audit texts.	The study uses a dataset of 1.5M electric power-related texts, 1,500 audit texts, and baseline models like Naive Bayes, SVM, GBDT, AdaBoost, XGBoost, TextCNN, LSTM, and BERT.	The EPAT-BERT model, based on the classical pre-training language model BERT, focuses on two pre-training tasks: word-granularity masked language model and entity-granularity masked language model. The model's effectiveness in classifying electric power audit texts is demonstrated through two ablation experiments. The results show that the model significantly exceeds existing models in accuracy, precision, recall, and F1 score compared to traditional machine learning models. The model's extensibility allows it to be applied to other electric power text classification tasks and project type annotation.
29	Wafaa Samy El-Kassas, Cherif Salama, Ahmed Rafea, Hoda K. Mohamed et.al (2019)	Automatic Text Summarization: A Comprehensive Survey	The methodology employed in this research involves conducting a comprehensive survey of Automatic Text Summarization (ATS) techniques and methodologies. The paper discusses various approaches,	The paper covers extractive, abstractive, and hybrid approaches to ATS, highlighting their differences and applications. Additionally, it explores different	Survey paper does not contain any datasets.	The paper achieves the goal of providing researchers with a comprehensive overview of ATS, serving as a valuable resource for understanding the current state of the field, identifying research gaps, and guiding future directions for improving ATS technique

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

			methods, building blocks, techniques, datasets, evaluation methods, and future research directions in ATS.	building blocks and techniques used in ATS, along with the datasets commonly employed for training and evaluation purposes.		

III. CONCLUSION

According to the literature survey on tracking devices for Alzheimer's patients, there is a growing interest in utilizing technology to meet the special issues related with the care and safety of those suffering from Alzheimer's disease. The studies constantly highlight the ability of monitoring devices to improve the overall quality of life for both patients and carers by providing answers to issues such as moving around, safety concerns, and timely reaction during emergencies. Several study studies have looked into the usefulness of various tracking technologies, such as GPS-based systems, wearable devices, and IoT-enabled solutions. These technologies not only help with location tracking but also include features like fall detection, geofencing, and conduct monitoring. The literature stresses the necessity of user-friendly interfaces and customizing to meet the special demands of Alzheimer's patients, ensuring that the devices are not intrusive and acceptable to the users. Interdisciplinary cooperation among healthcare experts, technologists, and carers are critical in the development and implementation of tracking devices. The integration of advanced sensors, data analytics, and communication technologies enables a holistic approach to addressing Alzheimer's disease difficulties, including real-time information and quick responses. Finally, the literature review stresses the potential of tracking devices to dramatically improve the care and safety of Alzheimer's patients. Future research and development efforts should focus on addressing existing challenges in order to create improved, user-friendly, and ethically sound tracking solutions that improve the lives of Alzheimer's patients while also reducing the burden on their caretakers..

REFERENCES

[1]Ayesha Ayub Syed, Ford Lumban Gaol, Tokuro Matsuo et al. (2020) "A Survey of the State-of-the-Art Models in Neural Abstractive Text Summarization"
[2]Lochan Basya, Mihir Sanghvi et al. (2022) "Text Summarization Using Large Language Models: A Comparative Study of MPT-7b-instruct, Falcon-7b-instruct, and OpenAI Chat-GPT Models"
[3]Pancheng Wang, Shasha Li, Kunyuan Pang, Liangliang He, Dong Li, Jintao Tang, Ting Wang et al. (2023) "Multi-Document Scientific Summarization from a Knowledge Graph-Centric View"
[4]Shuai Zhao, Fucheng You, and Zeng Yuan Liu et al. (2022) "Leveraging Pre-Trained Language Model for Summary Generation on Short Text"
[5]Ming Zhong, Yang Liu, Suyu Ge, Yuning Mao, Yizhu Jiao, Xingxing Zhang, Yichong Xu, Chenguang Zhu, Michael Zeng, Jiawei Han et al. (2022) "Unsupervised Multi-Granularity Summarization"
[6]Zhixin Li, Zhi Peng, Suqin Tang, Canlong Zhang, Huifang Ma et al. (2020) "Text Summarization Method Based on Double Attention Pointer Network"
[7]Pratik K. Biswas 1 and Aleksandr Iakubovich et al. (2022) "Extractive Summarization of Call Transcripts"
[8]Ángel Hernández-Castañeda, René Arnulfo García-Hernández2 and Yulia Ledeneva et al. (2023) "Toward the Automatic Generation of an Objective Function for Extractive Text Summarization"
[9]Fucheng You, Shuai Zhao, and Jingjing Chen et al. (2020) "A Topic Information Fusion and Semantic Relevance for Text Summarization"
[10]Luca Cagliero, Laura Farinetti, and Elena Baralis et al. (2019) "Recommending Personalized Summaries of Teaching Materials"

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

[11]Ruyun Wang, Hanwen Zhang, Guoliang Lu, Lei Lyu, and Chen Lyu et al. (2020) "Fret: A Reinforcer-Transformer Model for Code Summarization"
[12]István Üveges and Orsolya Ring et al. (2021) "HunEmBERT: A Fine-Tuned BERT-Model for Classifying Sentiment and Emotion in Political Communication"
[13]Heewon Jang and Wooju Kim et al. (2021) "Reinforced Abstractive Text Summarization with Semantic Added Reward"
[14]Yan Du and Hua Huo et al. (2020) "News Text Summarization Based on Multi-Feature and Fuzzy Logic"
[15]Rini Anggrainingsih, Ghulam Mubashar Hassan, Amitava Datta et al. (2023) "CE-BERT: Concise and Efficient BERT-Based Model for Detecting Rumors on Twitter"
[16] Youhyun Shin et al. (2023) "Multi-Encoder Transformer for Korean Abstractive Text Summarization"
[17] Fucheng You, Shuai Zhao, and Jingjing Chen et al. (2020) "A Topic Information Fusion and Semantic Relevance for Text Summarization"
[18] Yusung An, Dongju Kim, Juyeon Lee, Hayoung Oh, Joo-Sik Lee, and Donghwa Jeong et al. (2023) "Topic Modeling-Based Framework for Extracting Marketing Information from E-Commerce Reviews"
[19] Lavanya Elluri, Sai Sree Laya Chukkapalli, Karuna Pande Joshi, Tim Finin, and Anupam Joshi et al. (2021) "A BERT Based Approach to Measure Web Services Policies Compliance With GDPR"
[20] Shuai Zhao, Fucheng You, and Zeng Yuan et al. (2020) "Leveraging Pre-Trained Language Model for Summary Generation on Short Text"
[21] Sung-Guk Jo, Seung-Hyeok Park, Jeong-Jae Kim, and Byung-Won et al. (2016) "Learning Cluster Patterns for Abstractive Summarization"
[22] Bihui Yu; Chen Deng; Liping Bu et al. (2022) "Policy Text Classification Algorithm Based on Bert"
[23] Songsong Liu; Haijun Tao; Shiling Feng et al. (2019) "Text Classification Research Based on Bert Model and Bayesian Network"
[24] Younghoon Seo, Jihyeok Park, Gyungtaek Oh, Hyungjoo Kim, Jia Hu, and Jaehyun So et al. (2023) "Text Classification Modeling Approach on Imbalanced-Unstructured Traffic Accident Descriptions Data"
[25] Xuedong Tian and Jiameng Wang et al. (2021) "Retrieval of Scientific Documents Based on HFS and BERT"
[26] Yusung An, Dongju Kim, Juyeon Lee, Hayoung Oh, Joo-Sik Lee, and Donghwa Jeong et al. (2023) "Topic Modeling-Based Framework for Extracting Marketing Information from E-Commerce Reviews"
[27] Muhammad Wasim, Waqar Mahmood, Muhammad Nabeel Asim, Muhammad Usman Ghani et al. (2019) "Multi-Label Question Classification for Factoid and List Type Questions in Biomedical Question Answering"
[28] Qinglin Meng, Yan Song, Jian Mu, Yuanxu Lv, Jiachen Yang, Liang Xu, Jin Zhao, Junwei Ma, Wei Yao, Rui Wang, and Maoxiang Xiao et al. (2023) "Electric Power Audit Text Classification With Multi-Grained Pre-Trained Language Model"
[29] Wafaa Samy El-Kassas, Cherif Salama, Ahmed Rafea, Hoda K. Mohamed et al. (2019) "Automatic Text Summarization: A Comprehensive Survey".