Exploring ChatGPT for Email Content Compression and Summarization

R Bhuvaneswari
Department of Computer Science and Engineering
Amrita School of Computing, Amrita Vishwa Vidyapeetham
Chennai, India
bhuvanacheran@gmail.com

Tharaniesh P R
Department of Computer Science and Engineering
Amrita School of Computing, Amrita Vishwa Vidyapeetham
Chennai, India
tharaniesh3@gmail.com

2023 4th International Conference on Communication, Computing and Industry 6.0 (C216) | 979-8-3503-2732-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/C2I659362.2023.10430953

***Abstract*— In our daily lives, email communication is essential, yet monitoring the voluminous amount of emails can be time-consuming. The goal of email summarizing techniques is to quickly and accurately summarize email content in order to increase productivity and efficiency. In this study, we investigate the use of ChatGPT, a cutting-edge language model, for email synthesis. We offer a fresh strategy to take advantage of ChatGPT's conversational features for summarization while optimizing it on a big dataset of email discussions. Our tests show that ChatGPT-based email summarization outperforms other approaches in terms of performance. The generated summaries provide a clear and succinct summary of the email content that successfully captures the important details. To assess the value and user satisfaction of the summaries produced by ChatGPT, we also perform user studies. According to the findings, ChatGPT-based email summarizing has potential as an efficient and convenient method for controlling email overload. This work demonstrates the potential of conversational language models in tackling the difficulties of email management and advances email summarizing approaches.**

Keywords—text summarization, ChatGPT, BLEU Score, ROUGE, Email summary.

I. Introduction

With millions of emails being sent and received daily, email communication has become an essential part of our daily lives. However, it can be time-consuming and difficult to manage and digest the massive amount of information included in emails, a subset of text summarising known as email summarization tries to automatically produce clear and useful summaries of email content so that consumers can quickly understand the main ideas without having to read the entire email thread.

Machine learning (ML) models have become effective tools for email summarization in recent years. Large volumes of email data can be effectively used by ML models to identify trends and extract pertinent information, allowing the creation of automated email summarising systems. These systems make use of ML algorithms to parse through email content, structure, and context to produce precise summaries.

ML-based email summarization models come with a number of advantages. (i) They are able to process and summarise the enormous amount of data that is present in email threads. (ii) ML models are able to recognise intricate correlations and patterns in email content, which enables them to produce summaries that accurately convey the most important details. (iii) By continuously learning from new data, ML models may adapt and enhance over time, guaranteeing that the summarization system stays current with changing email conversation trends.

By leveraging ML models for email summarization, we anticipate significant advancements in email management and productivity. The development of accurate and efficient email summarization systems will alleviate the burden of information overload, allowing users to focus on critical tasks and make informed decisions based on concise and relevant email summaries.

II. Related Works

In this study [1], they demonstrated the effectiveness of combining linguistic techniques with machine learning to extract high-quality noun phrases. They provided a concise summary or gist of email messages. They have conducted a series of comparative experiments, employing various machine learning algorithms, to evaluate the performance of noun phrase extraction for saliency.

In this research[2], EmailSum dataset is introduced , specifically designed for abstractive Email Thread Summarization. This dataset comprises human-annotated short summaries (less than 30 words) and long summaries (less than 100 words) of 2549 email threads. These threads consist of 3 to 10 emails and cover a diverse range of topics.

They carried out an extensive empirical investigation, delving into a range of summarization techniques, encompassing both extractive and abstractive methods, single-document and hierarchical models, as well as employing transfer and semi-supervised learning approaches. The assessment of these techniques performance involved conducting human evaluations for both short and long summary generation tasks.

To address the limited availability of annotated corpora for email thread summarization, [3] have undertaken the task of annotating a subset of the W3C corpus. T. Drawing from insights gained from previous corpus annotations, they have opted to annotate this subset with both extractive and abstractive summaries. Their annotation process involves linking sentences in the abstractive summary to sentences in the extractive summary that convey the same information. Additionally, they have chosen to label additional features that can be leveraged in machine learning-based summarization. These features encompass the fundamental speech acts of individual sentences, as well as whether a sentence serves as a meta sentence, referring to the ongoing conversation.

This work [4] introduces a summarization scheme which works on query over a single-document that utilizes an

979-8-3503-2732-8/23/$31.00 ©2023 IEEE

unsupervised deep neural network. Instead of manually engineering features, a deep auto-encoder (AE) is employed to learn the relevant features. "Significantly, this method operates entirely in an unsupervised manner, devoid of any reliance on queries throughout its training process. When contrasted with current unsupervised email extractive summarization techniques, it exhibits noteworthy enhancements, reaching performance levels on par with leading supervised approaches.

In their study, Kashapov and colleagues [5] have created a notification system that places a primary emphasis on the human perspective. This system achieves this by extracting potential psychological triggers, identifying potential malicious intent, and generating a representative summary from emails. This information is presented to users in a meaningful way, aiming to enhance their prediction and improve their understanding of evolving phishing patterns. To gain further insights into the effectiveness of this methodology, user studies were conducted and perform objective experimental analysis on a larger scale. Additionally, examining a concrete metric was proposed, such as trigger and intent density as a fraction of the total email length, to assess its correlation with determining whether an email is "phishy" or not.

In this study, two distinct approaches were proposed in [6] for email thread summarization: collective message summarization (CMS) and individual message summarization (IMS). CMS utilizes a multi-document summarization technique, while IMS treats each message within the thread as a separate summarization task over a single document. Both methods are incorporated into a holistic framework guided by the process of sentence compression. Rather than relying solely on an extractive approach, the framework in [6] incorporates various compressions using statistical and linguistic approaches to generate the compressions of the original sentences. To evaluate the effectiveness of our approaches, they conducted experiments using the challenging Enron email collection, which is known for its highly technical language.

Recent advancements in abstractive Automatic Text Summarization (ATS) research has been examined in [7] across various dimensions, including evaluation metrics, challenges, types, solutions, architectures, datasets, techniques, contributions, research trends, and comparisons of State-of-the-Art (SotA) models. As the encoder-decoder structures have been used in Deep sequence-to-sequence models, it has demonstrated promising outcomes in ATS. Notably, significant progress has been achieved through architectures leveraging transfer learning and reinforcement learning. In particular, the utilization of universal representations learned by Transformer-based Pre-Trained Language Models (PTLMs) has led to notable improvements. By examining these recent developments, this study helps researchers with a comprehensive understanding of the current landscape in abstractive ATS research.

In [9], the author includes a detailed similarity analysis using term-based and sentence-based similarities for extracting important terms and creating extractive summaries. This analysis makes use of GloVe- and embeddings based on UAE to capture the semantic meaning of sentences helping to identify their relevance and significance. Remarkably, the extractive summarizer distinguishes itself as the sole solution capable of reinstating complete punctuation in summaries generated from original call transcripts that are either poorly punctuated or entirely lacking punctuation. This achievement is made possible through the implementation of a novel BERT transformer-based model.

To tackle the summarization task, authors in [10] introduced multiple models, employing two different vector representations derived from the Bag-of-Words (BOW) and word2vec approaches. By combining these models using an Ensemble method and a voting technique, they observed a remarkable enhancement in the summarization task[17][18][19]. Furthermore, for unsupervised learning models, they leveraged Autoencoders (AE), Variational Autoencoders (VAE), and ELM-AE to acquire the latent semantic representation of documents. These models played a vital role in capturing the underlying meaning of the text, thus contributing to improved summarization performance. As many Deep learning models have gained immense popularity and demonstrated remarkable power across various applications including NLP[13][14], computer vision[15][16],Speech and in health care applications, it has been focussed to use this technology to enhance the power of productivity in the proposed model.

III. Proposed Methodology

The primary objective of this research is to develop an advanced and robust email summarization system that leverages the power of OpenAI's ChatGPT model and serves the generated summaries as an API using the Flask framework. The proposed system aims to revolutionise email management for users by providing them with a seamless and efficient way to access concise and coherent summaries for their Gmail inbox emails. The research will focus on harnessing cutting-edge natural language processing techniques to significantly enhance email productivity and empower users to quickly grasp the essential content within lengthy email threads.

A. Gmail API Setup and Authentication

To begin the process, we will obtain Gmail API credentials from Google Cloud Console, ensuring secure and authorised access to the Gmail API. The authentication process will establish a seamless connection with the user's Gmail account, granting the system the necessary permissions to fetch and process inbox emails for summarization.

B. Retrieve Inbox Emails

With the Gmail API credentials in hand, we will seamlessly leverage the powerful API to meticulously retrieve inbox emails from the user's Gmail account. The fetching process will be intelligently designed to optimise performance by efficiently fetching emails in batches, avoiding any undue strain on the API and ensuring smooth and rapid data retrieval.

C. Email Information Processing

To facilitate effective summarization, we will meticulously process the retrieved emails to extract and distil essential information. This indispensable information will include key elements such as the sender, subject, date, and a

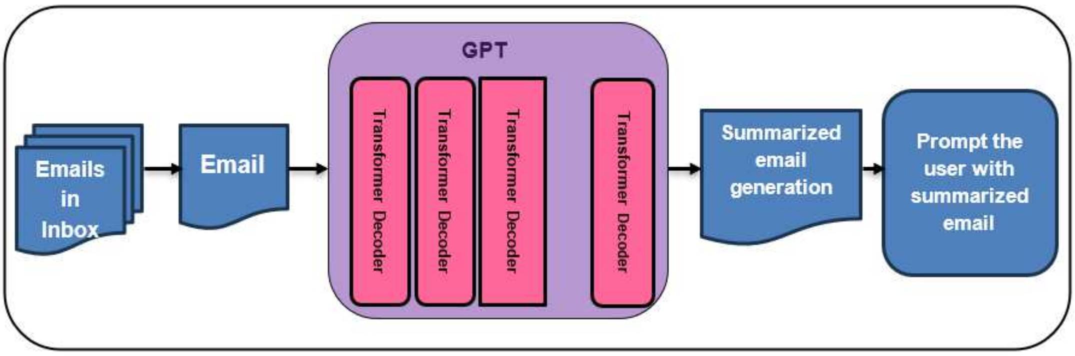

Fig. 1. The Proposed Model Development workflow

succinct snippet extracted from each email. This information will serve as the vital input for the subsequent summarization process and serve as crucial metadata to craft the final, comprehensive, and informative summarised emails.

D. *Summarization using ChatGPT API*

Before the email snippets are fed into the ChatGPT model, they undergo preprocessing to ensure clean and consistent text. This preprocessing step involves removing any unnecessary characters, special symbols, and formatting from the input long text (email snippet).

E. *Token Encoding*

To make the text suitable for the ChatGPT model, it needs to be encoded into tokens. Tokenization breaks the text into smaller units called tokens, which are the fundamental building blocks that the model can process. Each token corresponds to a word or subword piece. The tokenized text, along with the subject as a prompt, is then converted into numerical IDs to represent the text data in a format the model can understand.

F. *Summarization with ChatGPT Model*

The tokenized and encoded email snippets, along with the subject as a prompt, are fed into the ChatGPT model. The ChatGPT model is a large language model that has been pre-trained on a massive corpus of text data, enabling it to understand complex language patterns and relationships. It utilises a combination of transformer layers to process the input tokens and generate contextually relevant summaries

G. *Decoding Summary Tokens to Text*

Once the ChatGPT model processes the input tokens, it generates summary tokens as output. These summary tokens represent the concise summary of the email snippet and subject. To obtain human-readable summary text, the generated summary tokens are decoded and converted back into words or subword pieces.

H. *Post Processing Summary*

After decoding, the summary text may require some post processing to ensure coherence and grammatical correctness. Post Processing includes tasks such as removing extra spaces, capitalising the first letter of the summary, and handling punctuation.

I. *Output Summary Text*

The final output of the summarization process is the condensed and coherent summary text for each email snippet. These summaries, along with the corresponding email details (subject, sender, date), are integrated to create comprehensive and informative summarised emails. The ChatGPT model excels in generating high-quality summaries due to its ability to comprehend the context and semantic meaning of the input text. By leveraging transformer-based architectures, the model captures long-range dependencies and produces summaries that are coherent and relevant to the original content. The power of ChatGPT lies in its capacity to handle natural language in a context-aware manner, making it a potent tool for text summarization tasks. The generated summaries enable users to quickly grasp the essential information within their emails, streamlining email management, and enhancing productivity.

J. *Generate Summarised Emails*

Upon acquiring the brilliantly generated summaries from the ChatGPT model, we will seamlessly fuse the extracted email details (subject, sender, date) with their corresponding summaries. This crucial step will masterfully ensure the creation of comprehensive and highly informative summarised emails. These compact yet comprehensive emails will contain all the essential information, empowering users to quickly grasp the core content of each email without the burden of lengthy threads.

K. *Output Summarised Emails as an API*

The culmination of this remarkable research project will be the integration of a dynamic Flask web application. Functioning as an accessible API, users can effortlessly access the summarised emails through straightforward HTTP requests. The ingenious API endpoint, "/emails," will serve as a gateway, facilitating users to effortlessly request and receive the meticulously generated summarised emails as well-structured API responses.

L. Caching for Performance Optimization

To further enhance the API's response time and minimise redundant processing, the research project exhibits further ingenuity through the implementation of caching using Flask-Caching. With this clever addition, the system will significantly enhance the API's response time while minimising redundant processing. By temporarily storing the summarised emails in the cache for a specified duration, such as 60 seconds, the system will astutely fulfil repeated user requests within the cache period, drastically reducing processing overhead and ensuring a smooth user experience.

M. Custom Dataset Summarization

As an additional enhancement, the system will provide the flexibility for users to pass custom datasets containing sender and message pairs to the ChatGPT model. This feature will enable users to obtain concise summaries for any text data they wish to process through the system, adding versatility to the email summarization system.

IV. EXPERIMENTAL RESULTS

The proposed model is trained using the E-mail dataset [8]. The purposed model has been tested for the summarization of messages in the inbox using Gmail API. The developed bot access unread messages from the inbox. Then the trained model is used for the summarization of the messages passed into it. Fig 2,3,4 represents the API that access the inbox of Gmail using the credentials. Fig 5 represents one of the original messages in the inbox. Once, this message has been passed into the trained model, the summarized message can be generated and the same has been depicted in Fig 6. Fig 7 and Fig 8 represent the original and summarization emails for some other test messages in the inbox.

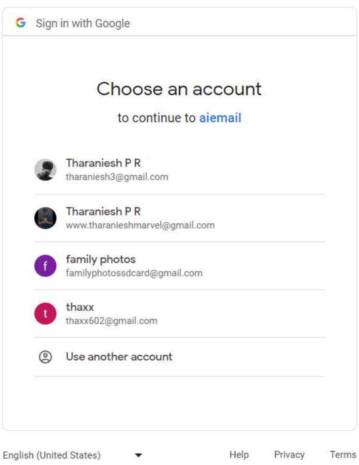

Fig. 2. Gmail Login Screen

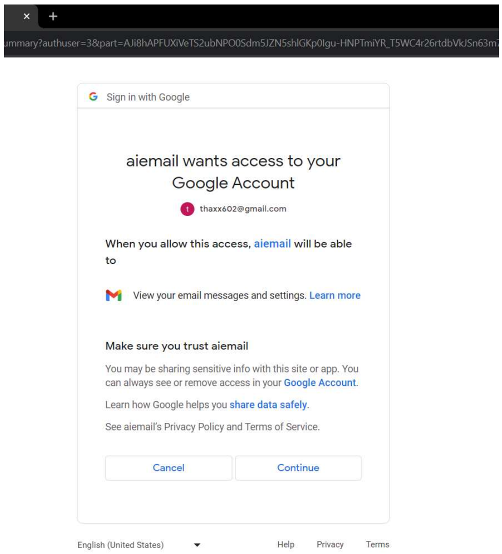

Fig. 3. Authentication of Gmail Credential using Proposed Model

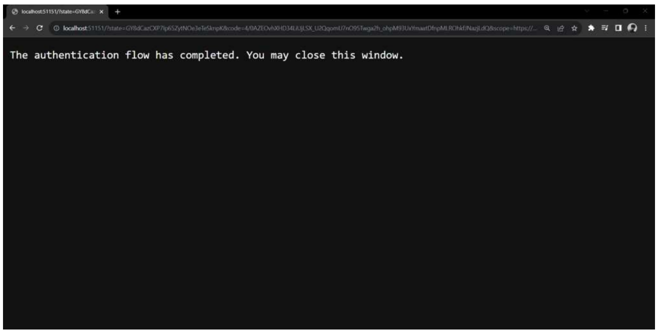

Fig. 4. Verified Authentication using Proposed Model

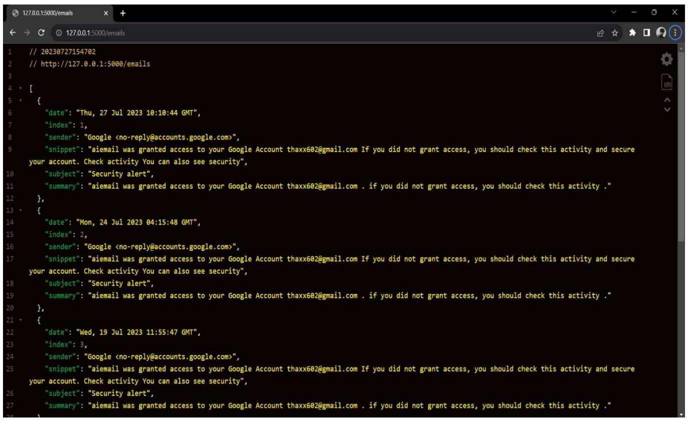

Fig. 5. Mail and its Summarized mail using Proposed Model

The proposed model has been evaluated using below performance measures.

1. Rouge Score:

- ROUGE-N: Measures the n-gram overlap between the generated summary and reference summary.

- ROUGE-L: Measures the longest common subsequence between the generated summary and reference summary.

TABLE I. ORIGINAL AND ITS SUMMARISED MESSAGES USING PROPOSED MODEL

Original Message	***Summarised Message***
Hi Champion, Looking to host your college competitions like never before? Partner with Unstop to get access to exciting features that'll make it easier for you to reach the right people and host your event seamlessly. Collaborate Now Why partner with Unstop? 🚀 Unlimited Event Listings with Complete Brand Customization for FREE 💻 Hassle-free Online Registrations: Say goodbye to IT infrastructure headaches! 💰 Secure Online Payment Collection at 0% transaction fee 📊 Real-time Management Panel to live track registrations and views of your event 👀 Proctoring for Peace of Mind: Full-screen mode for quizzes and hackathons 📧 Customized Emails at the Click of a Button In fact, we've got the promotions sorted too! 😎 💌 Reach 23,50,000+ subscribers with our Newsletters 📱 Rock Social Media with Instagram stories & Unstop App Notifications 🎖️ Commemorate Achievements with Participation Certificates and Unstop Coins Why manage this alone when you can partner with the ones who can make it easy for you? Fill out the form and make your event a BLOCKBUSTER now: https://unstop.com/our-partners/become-a-partner Regards, Team Unstop	Partner with Unstop to get access to exciting features that'll make it easier for you to reach the right people and host your event seamlessly.
1 new flutter developer internship job These job ads match your saved job alert [1] NextJS & Flutter Developer Intern Patriot Conceptions, LLC - Costa Mesa, CA $40,000 - $60,000 a year Easily apply Responsive employer The successful candidate will have a strong background in software development and a passion for creating innovative solutions	New flutter developer internship available at Patriot Conceptions, CA, offering $40,000-$60,000 per year.
Hey Tharaniesh, You left some great stuff in your cart. STILL THINKING ABOUT IT? Complete your order now, our best-sellers sell out fast! Cort AF500C Standard Series Cutaway 6 String Acoustic Guitar Quantity: 1 — Total: ₹7,590.00 Back to your cart! Confused about the choices you have made? Speak to our Product Specialist today for expert advice, special bundle deals, and the best prices guaranteed.	Order best-sellers like Cort AF500C Standard Series Cutaway 6 String Acoustic Guitar for fast delivery and expert advice.
Kaggle Hi Tharaniesh,	Kaggle Hi Tharaniesh,We're excited to announce Meta Kaggle for Code a new open source dataset made up of

Original Message	*Summarised Message*
We're excited to announce Meta Kaggle for Code a new open source dataset made up of ML code created and publicly shared by Kaggle's community. It contains hundreds of thousands of Apache 2.0 licensed Python and R notebooks used to analyze Datasets, make submissions to Competitions, and more. This represents nearly a decade of data spanning a period of tremendous evolution in the ways ML work is done. Check out the dataset Meta Kaggle for Code is also a continuation of our commitment to open data and research. This new dataset is a companion to Meta Kaggle which we originally released in 2016. Meta Kaggle enriches Meta Kaggle for Code. By joining the datasets together, you can easily understand which competitions code was run against, the progression tier of the code's author, how many votes a notebook had, what kinds of comments it received, and much, much more. We hope the new potential for uncovering deep insights into how ML code is written feels just as limitless to you as it does to us! Read our official announcement for more on why we released this dataset & how to use it. Happy Meta Kaggling! Meg Risdal & Jim Plotts on behalf of Kaggle Team	ML code created and publicly shared by Kaggle's community.

- ROUGE-S: Measures skip-bigram matches between the generated summary and reference summary.

2. F1 Score: Computes the harmonic mean of precision and recall, which evaluates the balance between the quality and completeness of the summary.

3. Precision: Measures the proportion of correctly generated summary tokens compared to the total number of generated tokens.

4. Recall: Measures the proportion of correctly generated summary tokens compared to the total number of reference summary tokens.

5. BLEU Score: Calculates the n-gram precision between the generated summary and reference summary, considering both precision and brevity penalty.

TABLE II. COMPARISON OF VARIOUS MODELS

Model	*Rouge*	*F1 Score*	*Precision*	*Recall*	*BLEUScore*
Biswas et al., [9]	Rouge 1:0.3 Rouge 2:0.26	-	-	-	0.82
Suanmali et al.,[10]	-	0.498	0.471	0.457	-
Shen et al., [11]	Rouge 2:0.245	0.202	-	-	-
Yousefi-Azar, et al., [12]	-	0.181	0.1694	0.2115	-
Proposed Model	RougeL:0.8	0.6	0.72	0.66	0.9

Table I shows the few results of the summarized email generated for the messages in the Gmail inbox. Table II shows the comparison of performance measures of various text summarization technique with the proposed model. As Rouge-L doesnot depend on the consecutive n-gram matches, it has been used as one of the performance measure in evaluating the proposed model. Moreover, it has the ability to analyse the structure of the sentence, it has been used in this work. RougeL of 0.8 clearly indicates that the model performs well and have the varaiation with only with very few words.

V. CONCLUSION

The performance metrics for text summarization techniques using ChatGPT are crucial in evaluating the quality and effectiveness of the generated summaries. Metrics such as Rouge Score, F1 Score, Precision, Recall, BLEU Score provide valuable insights into the summarization model's performance. By assessing the n-gram overlap, longest common subsequence, skip-bigram matches, precision, and recall, we can gauge the accuracy, relevance, and completeness of the generated summaries. BLEU Score enables us to evaluate the overall similarity between the generated and reference summaries, incorporating both precision and brevity. Perplexity is a valuable metric to measure the fluency and coherence of the generated summaries and assess how well the model predicts the next word. Semantic similarity allows us to assess how well the generated summaries capture the semantic meaning of the reference summaries. Moreover, considering the readability of the generated summaries is essential to ensure they are comprehensible and coherent to the readers. Overall, these performance metrics provide a comprehensive framework for evaluating the quality and performance of text summarization techniques using ChatGPT, enabling researchers to improve and optimize these models for more accurate and effective summarization.

REFERENCES

[1] Muresan, S., Tzoukermann, E. and Klavans, J.L., 2001. Combining linguistic and machine learning techniques for email summarization. In *Proceedings of the ACL 2001 Workshop on Computational Natural Language Learning (ConLL)*.

[2] Zhang, S., Celikyilmaz, A., Gao, J. and Bansal, M., 2021. Emailsum: Abstractive email thread summarization. *arXiv preprint arXiv:2107.14691*.

[3] Ulrich, J., Murray, G. and Carenini, G., 2008. A publicly available annotated corpus for supervised email summarization. In *Proc. of aaai email-2008 workshop, chicago, usa*.

[4] Yousefi-Azar, M. and Hamey, L., 2017. Text summarization using unsupervised deep learning. *Expert Systems with Applications*, *68*, pp.93-105.

[5] Kashapov, A., Wu, T., Abuadbba, S. and Rudolph, C., 2022, May. Email summarization to assist users in phishing identification. In *Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security* (pp. 1234-1236).

[6] Zajic, D.M., Dorr, B.J. and Lin, J., 2008. Single-document and multi-document summarization techniques for email threads using sentence compression. *Information Processing & Management*, *44*(4), pp.1600-1610.

[7] Alomari, A., Idris, N., Sabri, A.Q.M. and Alsmadi, I., 2022. Deep reinforcement and transfer learning for abstractive text summarization: A review. *Computer Speech & Language*, *71*, p.101276.

[8] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. URL: https://nlp.stanford.edu/pubs/glove.pdf.

[9] Biswas, P.K. and Iakubovich, A., 2022. Extractive summarization of call transcripts. *IEEE Access*, *10*, pp.119826-119840.

[10] Suanmali, L., Salim, N. and Binwahlan, M.S., 2009. Fuzzy logic based method for improving text summarization. *arXiv preprint arXiv:0906.4690*.

[11] Shen, D., Sun, J.T., Li, H., Yang, Q. and Chen, Z., 2007, January. Document summarization using conditional random fields. In *IJCAI* (Vol. 7, pp. 2862-2867).

[12] Yousefi-Azar, M. and Hamey, L., 2017. Text summarization using unsupervised deep learning. Expert Systems with Applications, 68, pp.93-105.

[13] Manojkumar, V.K., Mathi, S. and Gao, X.Z., 2023. An Experimental Investigation on Unsupervised Text Summarization for Customer Reviews. *Procedia Computer Science*, *218*, pp.1692-1701.

[14] Mohan, G.B. and Kumar, R.P., 2023. Lattice abstraction-based content summarization using baseline abstractive lexical chaining progress. *International Journal of Information Technology*, *15*(1), pp.369-378.

[15] D. M, A. R. B, B. Jagannath, R. R. Reddy and R. Bhuvaneswari, "Automating the Statutory Warning Messages in the Movie using Object Detection Techniques," 2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), Trichirappalli, India, 2023, pp. 1-7, doi: 10.1109/ICEEICT56924.2023.10157446.

[16] Bhuvaneswari, R. and Ganesh Vaidyanathan, S., 2021. Classification and grading of diabetic retinopathy images using mixture of ensemble classifiers. *Journal of Intelligent & Fuzzy Systems*, *41*(6), pp.7407-7419.

[17] Yang, X., Li, Y., Zhang, X., Chen, H. and Cheng, W., 2023. Exploring the limits of chatgpt for query or aspect-based text summarization. arXiv preprint arXiv:2302.08081.

[18] Xu, S., Zhang, X., Wu, Y. and Wei, F., 2022, June. Sequence level contrastive learning for text summarization. In Proceedings of the AAAI conference on artificial intelligence (Vol. 36, No. 10, pp. 11556-11565).

[19] Laskar, M.T.R., Hoque, E. and Huang, J.X., 2022. Domain adaptation with pre-trained transformers for query-focused abstractive text summarization. Computational Linguistics, 48(2), pp.279-320.