A Comprehensive Study of Resume Summarization using Large Language Models

Akshata Upadhye
Cincinnati, OH, USA

ABSTRACT
Due to a large number of applications received for a job posting, the recruiters and hiring teams can afford to spend very less time reviewing each resume. Due to the time constraint, it could be very helpful to the recruiters and the hiring teams if the key information from a resume could be summarized to provide a quick overview of the candidate's skills and experiences for initial screening. Therefore, this research focuses on exploring resume summarization through the utilization of various Language Models. This study explores the efficiency of various models like the BERT, T5 and BART for extractive and abstractive summarization in comprehensively summarizing diverse resumes. The research investigates the potential of LLMs in capturing important information, skills, and experiences, aiming to enhance the efficiency of the hiring process. By leveraging the power of these language models, the goal of this research is to contribute to the evolution of resume summarization techniques, offering a more context-aware approach for recruiters and the hiring teams.

Keywords
Natural Language Processing, Large Language Models, Extractive Summarization, Abstractive Summarization, BERT, T5.

1. INTRODUCTION
Employers often receive a large number of applications when they post job openings on sites like LinkedIn. Therefore, with a large number of applications being received, there are hundreds of resumes which at times can cause a delay in reviewing the candidates. In such a scenario, summarizing resumes will allow recruiters and the hiring managers to quickly review and screen the candidates, saving them time to focus on the most relevant applicants. Additionally, a well summarized resume allows the recruiters to identify candidates who closely match the requirements of the job since it provides a snapshot of the candidate's key qualifications, skills, and experiences. A lengthy resume can be overwhelming for the recruiters and therefore a summary can be helpful in capturing the essence of the candidate's professional background for the recruiters without getting bogged down in excessive details.

Given the volume of job applications received by the employers, manually summarizing each resume can be time-consuming. With the advancements in the field of Natural Language Processing, it is possible to automate the resume summarization process. Therefore, automated tools can quickly process and summarize resumes in large batches, allowing recruiters to focus their time on the most promising candidates. In addition to the time efficiency, automated tools also ensure a standardized approach to resume summarization which helps reduce any possibility for human bias thus promoting fair and objective candidate evaluations. Additionally, as the automated tools can summarize resumes in large batches, they enable the organizations to manage high volumes of applications without compromising the quality of the summarization process. Finally, by automating the resume summarization the HR professionals can allocate their time more strategically to focus on the promising candidates for the job. Hence automated resume summarization can be used as aid to enhance efficiency and objectivity in the early stages of the recruitment process.

2. RELATED WORK
This section is used to describe the existing techniques and applications for text summarization. Due to the recent developments in Natural Language Processing and Deep Learning research various methods to summarize text have emerged and continue to emerge. This section will present the different techniques for text summarization and the applications in which they are currently used.

2.1 Extractive Text Summarization
Extractive summarization technique is a method which creates a concise summary by extracting and combining the most important sentences from the source text. Therefore, in this method new sentences are not generated as a part of the summary.

The paper [1] explores diverse mechanisms used in the extractive text summarization process in order to provide a comprehensive comparison of various approaches in extractive summarization. Although text summarization has been an ongoing effort, in the recent times there has been a shift in the applications of text summarization in diverse content such as advertisements, blogs, emails, and news articles and simple sentence elimination has proven effective across such applications. The review paper also addresses the gap and numerous challenges of extractive text summarization.

The paper [2] explores the vast amount of research in text summarization and categorizes the text summarization into different groups based on the techniques used. The paper examines the relationship between text mining and summarization and highlights the crucial design criteria for text summarization systems and introduces diverse approaches while emphasizing the essential parameters used for ranking the important sentences. Finally, the study concludes by discussing the important evaluation methods used in text summarization.

The paper [3] discusses the significance of effective lecture summarization that could be very helpful for university students' study and memory enhancement. The paper discusses some of the existing research in this area such as TextRank and also states the drawbacks of this approach. Additionally, the paper introduces the BERT model which is able to improve the summarization by utilizing contextual information. The paper presents a lecture summarization service utilizing BERT, tailoring summaries based on user configurations which is a notable advancement in the field of text summarization.

2.2 Abstractive Text Summarization

Abstractive text summarization can be used to generate a concise summary of the text by using Natural Language Processing techniques to understand and interpret the important aspects of a text. These techniques usually generate large amounts of text by creating acceptable representations based on the input text which is then summarized into suitable output. Therefore, the summary created using this technique may contain new phrases or sentences.

The paper [4] highlights the significance of text summarization in NLP and data mining applications. The paper explores various extractive summarization techniques being available for various languages, including Bengali. However, the researchers found a gap exists in the domain of Bengali abstractive text summarization. Therefore, the paper explores diverse abstractive data mining approaches for summarizing text documents in various languages and provides a comparative analysis of these works. Finally, the researchers propose of a novel abstractive summarization technique specifically tailored for the Bengali language.

The survey presented in the paper [5] examines diverse methods of abstractive summarization to provide a comprehensive comparison of various techniques. It also highlights the advantages of abstractive summarization such as their ability to generate cohesive, coherent, and information-rich summaries with reduced redundancy. The review also addresses the current challenges and suggests future research directions particularly in terms of space and time complexity, which are yet to be fully explored.

The study in paper [6] comprehensively explores recent advancements in Automatic Text Summarization (ATS) research specifically using abstractive techniques. The focus of this study is on deep neural sequence-to-sequence models, Reinforcement Learning (RL) approaches, and Transfer Learning (TL) approaches using Encoder-decoder architectures, particularly those employing deep sequence-to-sequence models. This survey analyzes abstractive text summarization types, datasets, techniques, architectures, challenges, solutions, contributions, evaluation metrics, research trends, and State-of-the-Art (SotA) model comparisons.

2.3 Extractive Abstractive Text Summarization

This method of generating an extractive summary to get all the key information and then applying an abstractive summarizer to generate a summary based on all the key points in a piece of text proposed in the study [7] is especially useful for longer text documents.

3. BACKGROUND

This section presents some of the pretrained language models that can be used for extractive and abstractive summarization of text and also presents some of the metrics used for evaluating the summary generated using these language models.

3.1 Pretrained Language models for text summarization

3.1.1 bert-extractive-summarizer

BERT Extractive Summarizer [3] is used to extract key information using the BERT natural language model. BERT uses RNN, Attention mechanisms, and Transformers, to understand human languages. This method aims to reduce memory usage while preserving content value and allows control over summarization aspects such as sentence count and character count which is useful in selecting prominent sentences for a comprehensive summary.

3.1.2 T5-large

T5-large language model refers to a variant of the T5 (Text-to-Text Transfer Transformer) model, which is a Transformer-based model developed by Google Research for natural language processing tasks [8]. The “large” designation indicates a larger number of parameters compared to smaller versions. T5-large is specifically designed for text-to-text tasks, meaning it can be used for various NLP tasks to convert input text into output text such as language translation, text summarization, etc. This model has achieved state-of-the-art performance on various natural language understanding and generation tasks when fine-tuned on specific datasets.

3.1.3 stevhliu/my-awesome-billsum-model

This model is obtained by finetuning the T5-small language model for abstractive summarization using the California state bill subset of the BillSum dataset uploaded on the hugging face models repository [9].

3.1.4 facebook/bart-large-cnn

This model is obtained by using BART pretrained language model which has a bidirectional encoder and an autoregressive decoder which makes it suitable for text generation tasks such as summarization. bart-large-cnn model is obtained by fine tuning the BART language model on the CNN Daily Mail dataset [10].

3.2 Evaluating the summarization

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [11] is a set of metrics used for the evaluation of machine generated text, specifically in the context of text summarization. ROUGE measures the quality of text summaries by comparing them to gold standard reference summaries. The main ROUGE metrics include:

3.2.1 Rouge 1

The Rouge 1 metric calculates the overlap of unigrams (individual words) between the machine generated summary and the reference summary. It calculates how many unigrams in the generated summary match those in the reference summary.

3.2.2 Rouge 2

Similar to Rouge 1, Rouge 2 metric evaluates the overlap at the bigram level. It considers pairs of consecutive words in the machine generated and reference summaries.

3.2.3 Rouge L

ROUGE-L measures the longest common subsequence between the machine generated summary and the reference summary. Therefore, it takes into account the longest sequence of words that appears in both summaries while allowing for some flexibility in word order.

3.2.4 Human Evaluation

Apart from the metrics discussed, human annotators can validate the summarization by reading the original text versus the summary and then rate it on various criteria such as relevance, coherence, conciseness, and readability. This is the most reliable evaluation method, but it comes at a cost and is time-consuming.

4. METHODOLOGY

This section describes the proposed methodology for building a resume classifier as shown in Fig 1.

4.1 Preprocessing the resumes
For this step every model's respective text tokenizer will be loaded from the repository and used to tokenize the resumes before summarization.

4.2 Generating abstractive summary
In this step the T5-large, stevhliu/my-awesome-billsum-model and facebook/bart-large-cnn pretrained, finetuned language models are loaded from the hugging face repository and used to generate abstractive summary from the original resume.

4.3 Generating extractive summary
In this stage the bert-extractive-summarizer pretrained model will be loaded and used to generate an extractive summary for every resume. The ratio parameter will be set to 0.5 to generate a summary with 50% length of the original text. This helps capture the relevant sentences from the original resume.

4.4 Generating extractive-abstractive summary
In this phase the T5-large, stevhliu/my-awesome-billsum-model and facebook/bart-large-cnn pretrained, finetuned language models are loaded from the hugging face repository and used to generate abstractive summary from the extractive summary of every resume generated using the bert-extractive-summarizer pretrained model. This helps generate a summary covering the important points within the whole resume.

4.5 Evaluating machine generated summaries
Once the summaries are generated using both the techniques using various language models the summaries are evaluated by using the ROUGE metrics and with human judgement.

5. DATASET
For this research a resume dataset from Kaggle containing 962 resumes has been used. A sample of 100 random resumes has been taken from this dataset to test the performance of various summarization techniques, pretrained language models and finetuned pretrained language models.

6. RESULTS AND INTERPRETATION
In this section the results of summarizing resumes will be presented with an in-depth analysis of ROUGE metrics and a gold standard resume summary, and the corresponding summary generated using various techniques.

Gold Standard Summary
Experienced Data Science Assurance Associate at Ernst and Young LLP with 24 months of expertise in JavaScript, jQuery, and Python. A core team member in developing an automated review platform tool, specializing in Technology Assisted Review (TAR) for Fraud Investigations and Dispute Services. Proficient in various data science and analytic projects, including text analytics, chatbot development, and information governance, using tools such as Python, scikit-learn, NLTK, spacy, JavaScript, SQL, Elastic Search, Kibana, and more.

Fig 1: Gold Standard Summary

The summary presented in figure 1 is the gold standard summary for the one of the resumes from the dataset. As can be observed, the resume seems to belong to a Data Science candidate with relevant skills and experience. Now let's dive into the analyzing the machine generated summaries using extractive, abstractive and extractive-abstractive summarization techniques for the same resume represented in the gold standard summary.

BERT extractive summary
Skills Programming Languages: Python (pandas, numpy, scipy, scikit-learn, matplotlib), Sql, Java, JavaScript/JQuery. Database Visualizations: Mysql, SqlServer, Cassandra, Hbase, ElasticSearch D3.js, DC.js, Plotly, kibana, matplotlib, ggplot, Tableau. Others: Regular Expression, HTML, CSS, Angular 6, Logstash, Kafka, Python Flask, Git, Docker, computer vision - Open CV and understanding of Deep learning. Core member of a team helped in developing automated review platform tool from scratch for assisting E discovery domain, this tool implements predictive coding and topic modelling by automating reviews, resulting in reduced labor costs and time spent during the lawyers review. Worked on analyzing the outputs and precision monitoring for the entire tool. TAR assists in predictive coding, topic modelling from the evidence by following EY standards. Tools & Technologies: Python, scikit-learn, tfidf, word2vec, doc2vec, cosine similarity, NaÃƒÂ¯ve Bayes, LDA, NMF for topic modelling, Vader and text blob for sentiment analysis. Matplot lib, Tableau dashboard for reporting. MULTIPLE DATA SCIENCE AND ANALYTIC PROJECTS (USA CLIENTS) TEXT ANALYTICS - MOTOR VEHICLE CUSTOMER REVIEW DATA Received customer feedback survey data for past one year. Created customized tableau dashboards for effective reporting and visualizations. This chat bot serves entire product related questions. Giving overview of tool via QA platform and also give recommendation responses so that user question to build chain of relevant answer. The integrated Information Governance portfolio synthesizes intelligence across unstructured data sources and facilitates action to ensure organizations are best positioned to counter information risk. which frequently targeted during cyber-attacks. Ã¢Â€Â¢ FAP is a Fraud Analytics and investigative platform with inbuilt case manager and suite of Analytics for various ERP systems.

Fig 2: BERT extractive summary

6.1 Extractive summary results
The figure 2, presents the resume summary generated using the BERT extractive summarizer pretrained model. As can be seen the model does a good job of combining the most significant sentences to form a summary. Therefore, the ROUGE metrics as shown in the table I have values above 0.6 which indicates a good match with the reference text

Table 1. ROUGE metrics for extractive summarization

Model	ROUGE 1	ROUGE 2	ROUGE L
bert extractive summarizer	0.664	0.605	0.664

6.2 Abstractive summary results
Figures 3 and 4, represent the resultant summaries generated using the T5-large and stevhliu/my-awesome-billsum-model pretrained language models respectively by using the original resume as the input. As it can be seen these summaries only cover the context of the resume partially, because the resume input gets truncated due to the model's maximum acceptable token input length. Whereas in the figure 5, representing the summary generated by the facebook/bart-large-cnn pretrained model, it can be observed that the summary is relatively more comprehensive when compared to the other two models. This is due to facebook/bart-large-cnn's longer acceptable token input length. Therefore, the facebook/bart-large-cnn is able to provide a better overview of the candidate's resume. The ROUGE metrics for these models vary in the range 0.2 - 0.4 which is good considering the short summary.

Table 2. ROUGE metrics for abstractive summarization

Model	ROUGE 1	ROUGE 2	ROUGE L
T5-large	0.274	0.187	0.274

my-awesome-billsum-model	0.411	0.344	0.411
bart-large-cnn	0.368	0.268	0.366

6.3 Abstractive Extractive summary results

Abstractive-Extractive summarization will help overcome the problem of abstractive summary generated using the first few lines of the resume due to models limited input token size. By using this approach, the extractive summarizer will be able to reduce the length of the input resume by keeping the most relevant sentences in the extractive summary. This extractive summary is then fed to the abstractive summarization models to generate a comprehensive yet short summary of the resume. As it can be seen in figure 6, the summary is generated using T5-large pretrained model using the extractive summary generated using bert-extractive-summarizer model as the input. It can be seen that this summary covers relatively more context for various projects mentioned by the candidate when compared to the summary in figure 3 generated using the T5 abstractive summarization on the entire resume.

Table 3. ROUGE metrics for extractive abstractive summarization

Model	ROUGE 1	ROUGE 2	ROUGE L
T5-large	0.410	0.302	0.410
my-awesome-billsum-model	0.577	0.497	0.576
bart-large-cnn	0.495	0.390	0.494

Similarly in figure 7 which represents abstractive summary generated with stevhliu/my-awesome-billsum-model using the extractive summary as input, relatively more context is covered but it mostly contains the skills which might not be entirely useful as a summary. In figure 8 which represents abstrative summary generated with facebook/bart-large-cnn model using the extractive summary as input, it can be seen that the summary is quite comprehensive and covers the important points from various projects within the resume. Therefore, the summary generated using facebook/bart-large-cnn after extractive summarization, is giving a good idea about the overall experience of the candidate in a few sentences which can be very helpful for a recruiter and the HR teams, while reviewing a bunch of resumes.

The ROUGE metrics for this hybrid approach vary in the range 0.3 - 0.6 which is really good considering the short length of the summary.

T5 large abstractive summary
JAVASCRIPT- Exprience - 24 months jQuery-Exprience - 24 months. Developed the classifier models in order to identify "red flags" and fraud-related issues. TAR assists in predictive coding, topic modelling from the evidence by following EY standards. Developed the classifier models in order to identify "red flags" and fraud-related issues.

Fig 3: T5 abstractive Summary

stevhliu/my_awesome_billsum_model abstractive summary
TAR (Technology Assisted Review) assists in accelerating the review process and run analytics and generate reports. TAR assists in predictive coding, topic modelling from the evidence by following EY standards. TAR assists in predictive coding, topic modelling from the evidence by automating reviews, resulting in reduced labor costs and time spent during the lawyers review. TAR assists in predictive coding, topic modelling from the evidence by following EY standards.

Fig 4: stevhliu/my-awesome-billsum-model abstractive Summary

facebook/bart-large-cnn abstractive summary
FAP is a Fraud Analytics and investigative platform to review all red flag cases. The integrated Information Governance portfolio synthesizes intelligence across unstructured data sources and facilitates action to ensure organizations are best positioned to counter information risk. TAR (Technology Assisted Review) assists in accelerating the review process and run analytics and generate reports. CHATBOT is a user friendly chatbot for one of our Products which handle simple questions.

Fig 5: facebook/bart-large-cnn abstractive Summary

T5 large extractive abstractive summary
Core member of a team helped in developing automated review platform tool. TAR assists in predictive coding, topic modelling from the evidence by following EY standards. created customized tableau dashboards for effective reporting and visualizations. TEXT ANALYTICS - MOTOR VEHICLE CUSTOMER REVIEW DATA Received customer feedback survey data for past one year.

Fig 6: T5 extractive abstractive Summary

stevhliu/my_awesome_billsum_model extractive abstractive summary
Database Visualizations: Mysql, SqlServer, Cassandra, Hbase, ElasticSearch D3.js, DC.js, Plotly, kibana, matplotlib, ggplot, Tableau. Other: Regular Expression, HTML, CSS, Angular 6, Logstash, Kafka, Python Flask, Git, Docker, computer vision - Open CV and understanding of Deep learning.

Fig 7: stevhliu/my-awesome-billsum-model extractive abstractive Summary

facebook/bart-large-cnn extractive - abstractive summary
Core member of a team helped in developing automated review platform tool from scratch for assisting E discovery domain. TAR assists in predictive coding, topic modelling from the evidence by following EY standards. FAP is a Fraud Analytics and investigative platform with inbuilt case manager and suite of Analytics for various ERP systems. The integrated Information Governance portfolio synthesizes intelligence across unstructured data sources and facilitates action.

Fig 8: facebook/bart-large-cnn extractive abstractive Summary

7. CONCLUSION

In conclusion, this paper has explored various natural language processing techniques for automating resume summarization, specifically by leveraging Large Language Models (LLMs) such as BERT, BART and T5 and their fine-tuned variants. Through a thorough exploration of these models, their potential to revolutionize the summarization process has been highlighted and their limitations have been discussed in this paper by using examples and various metrics. By demonstrating the effectiveness of LLMs in capturing the intricacies of resumes, including skills, experiences, and qualifications this study opens new avenues which can help provide the recruiters and the hiring teams with a context-aware tool for initial candidate pool evaluation for enhancing the efficiency and accuracy of the hiring process. Additionally, a demonstration of using a hybrid approach utilizing both extractive and abstractive summarization was presented with good results to generate a comprehensive short summary of the resume. Therefore, with the development and research in the field of natural language processing and the integration of LLMs into resume summarization showcases a promising direction for future research and practical applications in talent acquisition.

8. REFERENCES

[1] Moratanch, N., and S. Chitrakala. ”A survey on extractive text summarization.” In 2017 international conference on computer, communication and signal processing (ICCCSP), pp. 1-6. IEEE, 2017.

[2] Rahimi, Shohreh Rad, Ali Toofanzadeh Mozhdehi, and Mohamad Abdolahi. ”An overview on extractive text summarization.” In 2017 IEEE 4th international conference on knowledge-based engineering and innovation (KBEI), pp. 0054-0062. IEEE, 2017.

[3] Miller, Derek. ”Leveraging BERT for extractive text summarization on lectures.” arXiv preprint arXiv:1906.04165 (2019).

[4] Yeasmin, Sabina, Priyanka Basak Tumpa, Adiba Mahjabin Nitu, Md Palash Uddin, Emran Ali, and Masud Ibn Afjal. ”Study of abstractive text summarization techniques.” American Journal of Engineering Re- search 6, no. 8 (2017): 253-260.

[5] Moratanch, N., and S. Chitrakala. ”A survey on abstractive text summarization.” In 2016 International Conference on Circuit, power and computing technologies (ICCPCT), pp. 1-7. IEEE, 2016.

[6] Alomari, Ayham, Norisma Idris, Aznul Qalid Md Sabri, and Izzat Alsmadi. ”Deep reinforcement and transfer learning for abstractive text summarization: A review.” Computer Speech and Language 71 (2022): 101276.

[7] Teo, Louis. “The Secret Guide to Human-like Text Summariza- tion.” Medium, July 28, 2022. https://towardsdatascience.com/the-secret- guide-to-human-like-text-summarization-fcea0bfbe801.

[8] Kale, Mihir, and Abhinav Rastogi. ”Text-to-text pre-training for data- to-text tasks.” arXiv preprint arXiv:2005.10433 (2020).

[9] Stevhliu. “Stevhliu/Myawesomebillsummodel · Hugging Face.” stevhliu/my-awesome-billsum-model · Hugging Face. Accessed December 20, 2023. https://huggingface.co/stevhliu/my-awesome- billsum-model.

[10] Lewis, Mike, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Ab- delrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettle- moyer. ”Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension.” arXiv preprint arXiv:1910.13461 (2019).

[11] Lin, Chin-Yew, and Eduard Hovy. ”Automatic evaluation of summaries using n-gram co-occurrence statistics.” In Proceedings of the 2003 human language technology conference of the North American chapter of the association for computational linguistics, pp. 150-157. 2003.