Customizing Legal Document Summarization: A Case Study In Patent Documents

Course code/Name: Info-5731/ Computational Methods for Information Systems.

(Section 020,022-Spring 2024)

Madhu surusetti
Madhusurusetti@my.unt.edu
11657301 University of North Texas
Denton, Texas, USA

Anusree Putta
AnusreePutta@my.unt.edu
11564538 University of North Texas
Denton, Texas, USA

Nagubudi Venugopal Bhavyanth
NagubudiVenugopalBhavyanth@my.unt.edu
11596945 University of North Texas
Denton, Texas, USA

Grishma Tallapareddy
Grishmatallapareddy@my.unt.edu
11653727 University of North Texas
Denton, Texas, USA

Abhiram Nallamothu
Abhiramnallamothu@my.unt.edu
11645841 University of North Texas
Denton, Texas, USA

Anudeep Oggu
Anudeepoggu@my.unt.edu
11596825 University of North Texas
Denton, Texas, USA

ACM Reference Format:
Madhu surusetti, Anusree Putta, Nagubudi Venugopal Bhavyanth, Grishma Tallapareddy, Abhiram Nallamothu, and Anudeep Oggu. 2024. Customizing Legal Document Summarization: A Case Study In Patent Documents: Course code/Name: Info-5731/ Computational Methods for Information Systems. (Section 020,022-Spring 2024). In . ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Abstract

In the legal profession, technology could speed up decision-making and enhance access to legal knowledge. Patent documents, a kind of legal document that explains an innovation, are intended to be used with the system. To assist the LLMs in writing summaries, the system employs a range of questions and prompts. With the technique, summaries concentrate on distinct areas of the patent document, like the claims or the description. Although the system is still in the early stages of development, anyone who wants to access legal information may find it to be a useful resource. Large language models (LLMs) are used by the system to generate summaries of patent documents. Numerous metrics are used to assess the summaries, including cosine similarity, METEOR score, BLEU score, and ROUGE score. Summaries can be created by the system to match the user's unique requirements.

The study investigates producing summaries based on user choices, such as the preferred amount of detail and section emphasis. Various types of summarization are discussed to accomplish this customization.

2 Github link of the project

https://github.com/AbhiramNallamothu/Project

3 Keywords

Dataset, Bleu Score, Bert Score, Annotation with Large Language Models (LLMs), Evaluation Metrics, Customizable Summaries, Human Evaluation, Summac Score, and patent document.

4 Introduction

Legal documents, particularly patents, are notorious for their complexity, characterized by complicated language and specialized content. As members of the research team at the University of North Texas, we have encountered the challenges of gathering details and crafting personalized summaries for similar documents. This experience has propelled us into the realm of legal document summarization, where we are exploring innovative approaches to simplify and tailor summaries to individual needs. In fields such as law, technology, and innovation, a deep understanding of patents is essential. However, conventional summarization techniques often struggle to extract the most pertinent information from these documents, falling short of meeting users' diverse expectations. This gap in current research underscores the necessity for a personalized and effective document summary system. In addressing this gap, our research endeavors to develop

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

novel methods for customizing legal document summarization, with a particular emphasis on patents. Our vision is to provide easily understandable summaries that are not only more readable but also tailored to the preferences of different users. Through this, we aim to enhance the accessibility of legal information and improve decision-making rationality across various fields.

4.1 Research questions

In our quest to improve legal document summarization, we aim to answer several key questions:

- How can we develop summarization techniques that effectively capture the complexities of legal language, particularly in patents?
- What strategies can be employed to tailor summaries to individual user preferences and needs?
- How do our proposed methods compare to existing approaches in terms of accuracy, customizability, and user satisfaction?

4.2 Research goal

The main purpose of our study is to move away from the traditional summarization of legal documents, especially patents. Our purpose is to practice the techniques that will grant legal information accessibility and the user will also get summaries personalized to meet their requirements. Ultimately, we are aiming to reduce the gap between legal terminology and simplified data, which will ease both comprehension and choice-making across different spheres. This journey set forth will eventually lead to a new kind where legal document summarization does not only yield boring results but a transformation process that enables the users with actionable insights and lessens the hassle of decision-making in an ever-changing legal environment.

5 Related work

Customizing lawful record summarization, especially within the setting of obvious reports, presents a critical challenge due to the complicated nature of legitimate dialect and the specificity of obvious substance. Conventional summarization procedures frequently come up short of catering to personal client inclinations and interface, resulting in non-specific run-downs that will not completely meet the desires of differing clients. This writing survey dives into different approaches and systems pointed to this challenge.

1. [10] presented an Optimization System for Intuitively Personalized Summarization (IPS), which emphasizes producing outlines in an intuitive and personalized way. Their ponder highlights the significance of capturing client interface through intelligent clicks and joining personalization by modeling peruser inclinations. The assessment comes about to illustrate IPS's capacity to create rundowns with higher client fulfillment compared to conventional methods.
2. [8] proposed a Controllable Abstractive. Summarization show, which coordinates client inclinations such as wanted length and fashion into the summarizing handle. Their neural summarising permits clients to indicate high level traits, coming about in rundowns that superior adjust with personal preferences.
3. [1]centered on obvious summarization procedures, recognizing the challenges posed by the characteristics of the obvious class. Their approach utilizes lexical chains and full-fledged normal dialect-era procedures to create comprehensive outlines of obvious records, considering the unequal conveyance of substance and unique lexicon characteristic of patents.
4. [4] investigated extractive summarization of lawful writings, especially judgments from the UK House of Rulers. Their inquiry illustrates the potential of summarization innovation for lawful data admin frameworks, emphasizing the utility of programmed summarization in organizing and fitting rundowns to distinctive client types.
5. [6] addressed the limitations of generic summaries by proposing a model for generating personalized summaries of online articles tailored to individual user preferences. Their approach leverages user data available on the internet to model user backgrounds and preferences, leading to higher user satisfaction compared to generic summaries.
6. [2] advocated for personalized summarization to maximize the relevance of recommendations delivered by personalized information systems. Their study in the digital newspaper domain demonstrated that personalized summaries outperformed generic-personalized summaries in identifying documents that meet user preferences.
7. [3] presented a summarization framework within the PERCIVAL medical digital library, focusing on methods for defining and generating customized summaries for medical professionals and laypersons. Their approach combines machine-generated and extracted content to meet identified user needs, showcasing the potential of customization in a medical digital library.
8. [8]explored text mining methods for patent analysis, emphasizing the importance of automating the analysis process due to the lengthy and technical nature of patent documents. Their approach includes text segmentation, summary extraction, feature selection, and cluster generation, demonstrating effectiveness in patent analysis and mapping.

In summary, the reviewed literature highlights various approaches to customizing legal document summarization, particularly within the context of patent documents. These approaches include intuitively personalized summarization, controllable abstractive summarization, and the use of text-mining techniques to address the challenges posed by legal language and document specificity

6 Data collection

We use patent databases and repositories as part of our data collection strategy to guarantee a wide range of patent papers from various sectors. Data like the names of the authors, patent id, claims, and descriptions are gathered. For a thorough analysis, we will also gather statistics on the dataset's size, average document length, and distribution across various patent categories. Eliminating redundancies, standardizing formats, and guaranteeing metadata consistency are all part of data cleaning. To guarantee correctness, quality checks will be carried out by removing any unnecessary or inaccurate data.

7 Methodology

7.1 Overview of methodology

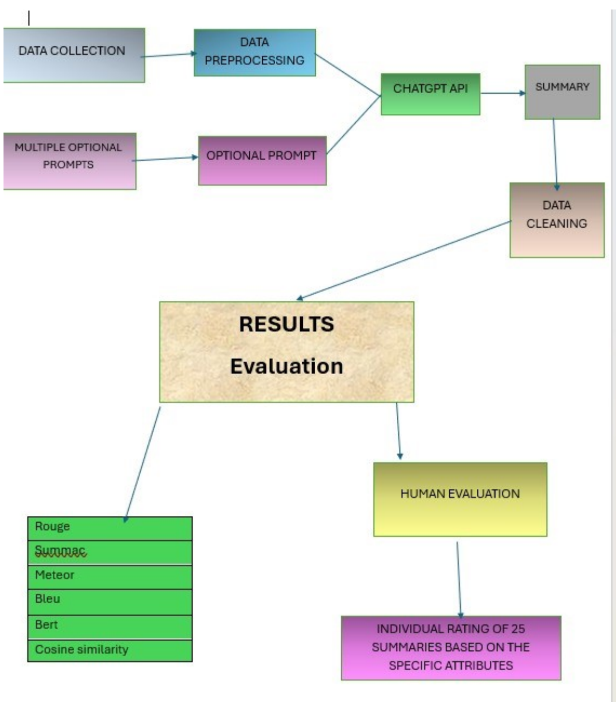

Figure 1: Overview of methodology

The process starts by gathering data from various sources. This data is then cleaned up and prepared through preprocessing steps like formatting, removing errors, etc. Next, this preprocessed data is fed into a language model called ChatGPT API. This AI model is designed to understand and process natural language. At this stage, there are options to provide additional guidance to the ChatGPT model in the form of prompts. These prompts can help steer the model to generate summaries in a particular way or focus on certain aspects of the data. It uses the input data to automatically generate summaries.

After the summaries are generated by the ChatGPT model, there may be additional cleaning or processing steps applied to the summaries themselves to improve their quality or formatting. Once the final summaries are ready, they go through an evaluation process. This evaluation involves multiple automated scoring metrics like ROUGE, Summac, Meteor, Bleu, Bert, and Cosine similarity. These metrics analyze different aspects of the summaries and provide scores based on their quality, relevance, and accuracy.

In addition to the automated scoring, there is also a human evaluation component. In this step, people review and rate a set of 25 patent documents based on specific attributes or criteria that are important for the particular use case or application. The human evaluation is crucial because while automated metrics can provide objective scores, they may not capture all the nuances and contextual factors that humans can assess. The human ratings, combined with the automated scores, provide the data. Once the final summaries are ready, they go through an evaluation process. This evaluation involves multiple comprehensive evaluations of the summary quality. Overall, this process aims to leverage advanced language models like ChatGPT to generate summaries from data of patent documents based on customization, while also incorporating robust evaluation techniques, including both automated metrics and human assessments, to ensure the summaries meet the desired quality standards.

7.2 Data description and preprocessing

The dataset consists of an extensive amount of carefully curated patent documents that are all saved in a structured text format (.txt). Similar to intricate blueprints for inventions, these documents provide an immense amount of details, including the patent's title, patent ID, claims, and a thorough description. Consider every section of a patent filing as a mini-book with a specific purpose to better understand this. The patent's topic matter is named in the title, which serves as its cover. Similar to a book's ISBN, the patent number serves as a unique code that allows for precise tracking and referencing. Similar to the book's back matter, the abstract provides a concise summary of the patent's main points. It simplifies the invention's main ideas into a few paragraphs, which helps people understand its importance. The core of the patent is the claims section, which provides a detailed description of the invention's rights and the scope of the patent. It is the document's legal core, describing the precise elements or characteristics of the invention that protect it from violation by patent law.

Let's explore each section in more detail now. Similar to the book's back matter, the abstract provides a concise summary of the patent's main points. It simplifies the invention's main ideas into a few paragraphs, which helps people understand its importance.

Yet, classifications group patents into certain technological domains or topics, much like the index or category sections in a library catalog. This simplifies researchers' search for patents that involve their areas of study. Finally, the description section offers an in-depth explanation of the invention, just like the chapters in a book. It covers every aspect—from the history and technical specifics to the plans, samples, and possible uses. The invention's reality is covered in this part, which also explains what it means and how it works.

Let us take an example from a patent called "System and method for prediction-based lossless encoding." In applications where data integrity is critical, this invention presents a revolutionary technique for reducing data without losing any of it. To successfully estimate and encode signal values, it makes use of complex methods such as modified Least Mean Squares and residual coding. While the classifications part would group this method under pertinent categories like signal processing or data compression, the claims section would detail exactly which aspects of this process are protected.

However, the dataset has plenty of additional data in addition to having a collection of documents. Moreover, complete dataset statistics, including the overall count of documents, mean document length, and patent distribution among various classes, provide important insight into the dataset's design and range

Each patent document is stored as a text file with a corresponding patent number, facilitating easy access and reference for further research and analysis. This comprehensive data description provides valuable insights into the structure and significance of the patent dataset for various research and analytical purposes

7.3 Data Annotation with LLMs

Annotation process using Large Language Models (LLMs). This is similar to adding notes or annotations to a book to aid readers in understanding the text. We designed prompts to help the LLMs write individual summaries, starting with a few sample patents. These prompts specified things like specificity, abstractiveness, and main focus. We provided an interface with these questions and patents, and it generated summaries according to our instructions.

By carefully evaluating a few summaries to make sure they fulfilled requirements which include coverage, accuracy, clarity, and customizability. We changed the questions and tried again if a summary fell below the mark until we were able to produce summaries that effectively expressed the main ideas of the patents.

The iterative method assures that the findings are accurate and clear, while also being adaptable to suit the different needs of analysts, researchers, and intellectual property experts. To put it simply, it's similar to sharpening a map so that it leads you exactly to the prize you're looking for, which is, in this case, the useful data included in every patent document.

	A	B	C	D	E
1	Customization	Value	Prompt	Source Patent ID	Rating
2	Specificity	Low	Summarize the patent document	US9667499	4
3	Specificity	Medium	Provide a detailed summary of the patent document	US9667499	3
4	Specificity	High	Include comprehensive details from the patent document	US9667499	5
5	Abstractiveness	Low	Condense the key points of the patent document	US7187715	4
6	Abstractiveness	Medium	Capture important information from the patent document	US7187715	3
7	Abstractiveness	High	Provide a thorough summary covering all aspects of the patent document	US7187715	5
8	Main Focus	Invention Description	Focus on describing the invention presented in the patent	US9667499	4
9	Main Focus	Claims	Highlight the claims made in the patent document	US9667499	3
10	Main Focus	Novelty	Emphasize the novelty and unique features of the invention	US9667499	5
11	Specificity	Low	Provide a brief overview of the patent document	US7187715	3
12	Specificity	Medium	Summarize key aspects of the patent document	US7187715	4
13	Specificity	High	Detail all relevant information from the patent document	US7187715	5
14	Abstractiveness	Low	Extract main ideas from the patent document	US9667499	3
15	Abstractiveness	Medium	Cover significant points from the patent document	US9667499	4
16	Abstractiveness	High	Provide an extensive summary of the patent document	US9667499	5
17	Main Focus	Invention Description	Describe the invention outlined in the patent	US7187715	4
18	Main Focus	Claims	Summarize the claims made in the patent document	US7187715	3
19	Main Focus	Novelty	Highlight the novelty and inventive aspects of the patent	US7187715	5
20	Specificity	Low	Summarize key points of the patent document	US9667499	3
21	Specificity	Medium	Provide a comprehensive overview of the patent document	US9667499	4
22	Specificity	High	Include detailed information from the patent document	US9667499	5
23	Abstractiveness	Low	Condense the main ideas from the patent document	US7187715	3
24	Abstractiveness	Medium	Cover important aspects of the patent document	US7187715	4
25	Abstractiveness	High	Provide a detailed summary of the patent document	US7187715	5

Table 1: A few samples of optional prompts

7.4 Generating a customizable summarization dataset for patent documents

Using the best possible prompts from the annotation process, a customizable summarizing dataset for patent documents was created. Customized summaries were created for each of the 1100 patent documents chosen for the collection using the GPT-3.5 API.

The resulting dataset was saved in a structured format like CSV or JSON and contained the created summary, the related patent ID, and the patent document. Generated data set details

- Dataset Name: INFO5731Group8.xls
- Dataset Description: This dataset contains a collection of patent id, patent document, and their corresponding summaries.
- Number of Records: 1100
- Number of Columns/Fields: 3 (patent id ,patent document, Summary)
- Average Length of Summaries: 15000 words
- Average Length of Generated Summaries: 5000 words

7.5 Evaluate the quality of the generated summarization dataset

The resultant summary dataset was subjected to both automated and manual assessment techniques for quality evaluation. The summaries were evaluated using automated criteria, including Bert score, SummaC, Meteor score, Cosine similarity, and ROUGE score, in comparison to the original patent materials. A manual evaluation of a sample consisting of approximately 25 documents was conducted on many aspects, such as clarity, accuracy, coverage, and customizability. The goal of this thorough assessment was to guarantee the efficacy and quality of the summaries that were produced for real-world applications.

The overlap between n-grams in the patent document and the summary is measured by ROUGE scores, which provide information on how effective the summarization procedure was. By comparing the summary to the reference patent document and taking into account sentence length and n-gram matching, BLEU scores evaluate the quality of the summary. METEOR scores evaluate word-level matches and alignment to determine how semantically comparable the patent text and the summary. Moreover, cosine similarity computes the similarity between the patent document's and the summary's TF-IDF representations, offering an additional viewpoint on their textual similarity.

A thorough assessment of the effectiveness of the summary algorithm and the degree of correspondence between the summaries of patent documents and their original texts is obtained by looking at the distribution of these scores.

This study makes it easier to evaluate the efficacy of the summary process and pinpoint possible areas for improvement. The given data frame contains the calculated similarity scores for patent document summaries compared to the corresponding patent papers, including BLEU, METEOR, and cosine similarity metrics. These ratings function as numerical gauges of the degree of similarity and efficacy of the summarizing procedure.

8 Experiment and data analysis

As part of our plan, we will construct and assess models using advanced summarization models such as GPT-3.5. To create summaries that satisfy user choices regarding abstraction level, section focus, and level of detail, we propose a continuous improvement approach. These summaries will be manually evaluated according to standards like flexibility, clarity, accuracy, and coverage. SummaC, BERT, and ROUGE scores are examples of automated metrics that will be used in the quantitative evaluation process. Through the use of a well-selected dataset, we want to set new standards for flexible summarizing methods.

9 Results and Discussion

9.1 Rouge score

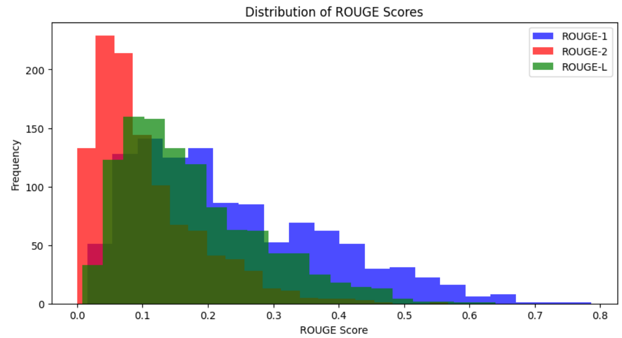

Figure 2: Distribution of rouge score

The 'rouge scores' column consists of dictionaries containing ROUGE scores, including ROUGE-1 scores, ROUGE-L scores, and their corresponding precision, recall, and F1 scores. These scores evaluate the similarity between the patent document summaries and the original documents. From the visualization, we can deduce that with an increase in rouge value leads to a decrease in frequency in the similarity of the patent document and its summary.

- ROUGE-1 Analysis: This metric evaluates the overlap of unigrams between the summary and the original text. The distribution shows a concentration of scores around lower to mid-range values, indicating a general proficiency in capturing key individual terms within summaries, but also room for improvement in optimizing word selection to enhance representativeness.
- ROUGE-2 Analysis: Assessing bi-gram overlap, the ROUGE-2 scores predominantly cluster at lower values. This pattern suggests difficulties in effectively capturing consecutive word pairs, reflecting challenges in preserving the nuanced relationships and specific phrasings present in the original texts.
- ROUGE-L Analysis: Focused on the longest common subsequence, this metric displays a distribution that suggests a moderate capability of the summaries to maintain the sequence structure of the source texts. The scores indicate a balanced performance, capturing the flow and overall structure, yet with potential for further refinement.

9.2 Summac score

The SummaC score evaluates the coherence of a text summary, specifically focusing on how logically connected and understandable the summary is as a standalone text. It is particularly useful for assessing the continuity and flow of generated summaries. The scores mostly cluster around the 0.0 to 0.2 range with the peak just slightly above 0.0. This indicates that while a significant number of summaries possess basic coherence, there is a substantial variation in the degree of logical continuity achieved. The presence of scores below zero is a critical indicator of summaries that potentially distort the logical sequence or omit crucial linking information, making them less coherent.

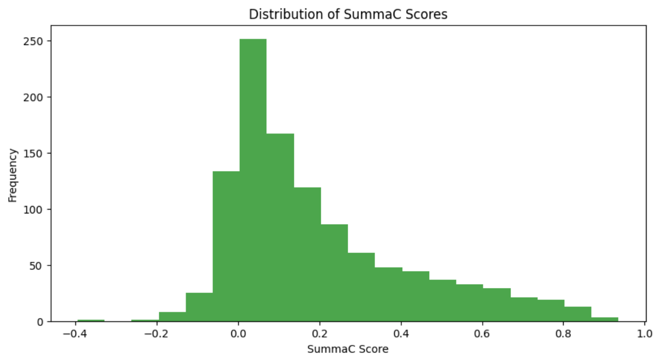

Figure 3: Distribution of Summac score

The distribution of SummaC scores shows the current performance of summarization techniques in maintaining the document summaries. By focusing on improvements in how we can process and reconstruct the logical flow of texts, there is potential to significantly enhance the quality and usability of generated summaries, as evidenced by the potential to shift the bulk of scores towards higher values.

9.3 Bert Score

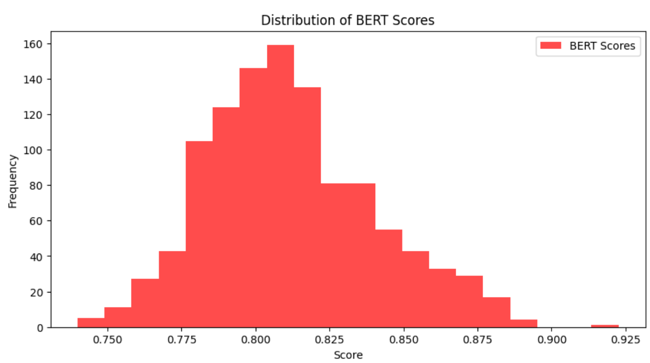

Figure 4: Distribution of Bert score

With the maximum frequency around the 0.850 level, the scores are mostly centered between 0.775 and 0.900. This central peak shows that most summaries are semantically aligned with their reference texts to a considerable extent, which is encouraging for the efficiency of the summary algorithms used. When the summarization achieves extremely high semantic accuracy, it is shown as a tail that extends towards higher scores (around 0.925). These are fewer in number, but they demonstrate how, under some scenarios, the summarizing algorithm may yield extraordinarily aligned summaries. Scores as low as 0.750 are present, indicating that the summarizing algorithms' performance is not constant. These lower scores may pinpoint specific cases where the summarization fails to capture key semantic elements of the original text, necessitating further investigation and refinement of the models.

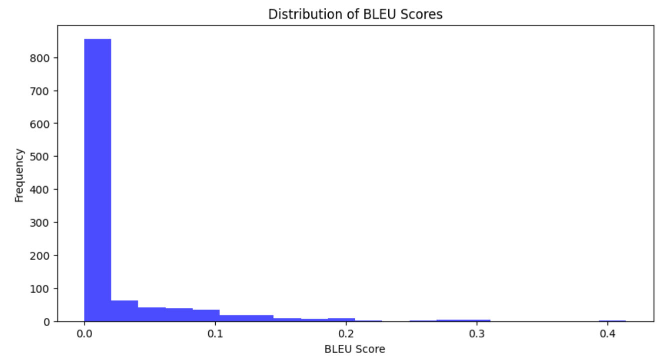

Figure 5: Distribution in Bleu score

The BLEU score, ranging from 0 to 1, quantifies the overlap of n-grams between the summary and the patent document. Higher BLEU scores signify a higher level of similarity, reflecting a more faithful summarization of the original document. Conversely, lower scores suggest a lesser degree of concordance, potentially indicating areas for improvement in the summarization algorithm.

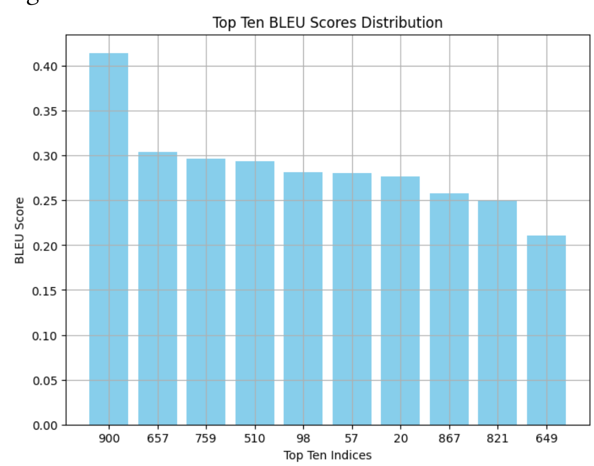

Figure 6: Top ten Bleu distribution

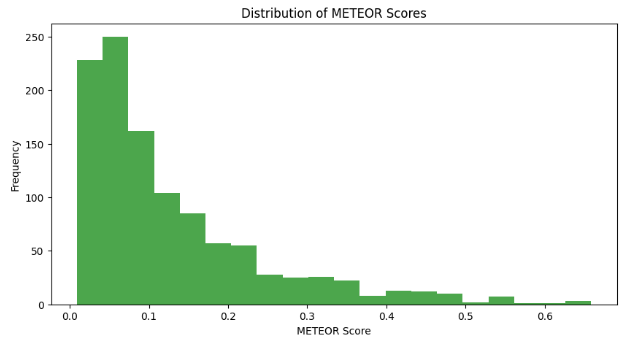

Figure 7: Distribution in Meteor score

Similarly, the METEOR score provides insight into the semantic congruity between the summary and the patent document, considering word-level matches and alignment. A higher METEOR score denotes a closer semantic resemblance, indicating a more accurate representation of the original content. On the contrary, lower scores may imply discrepancies in meaning or inadequate summarization.

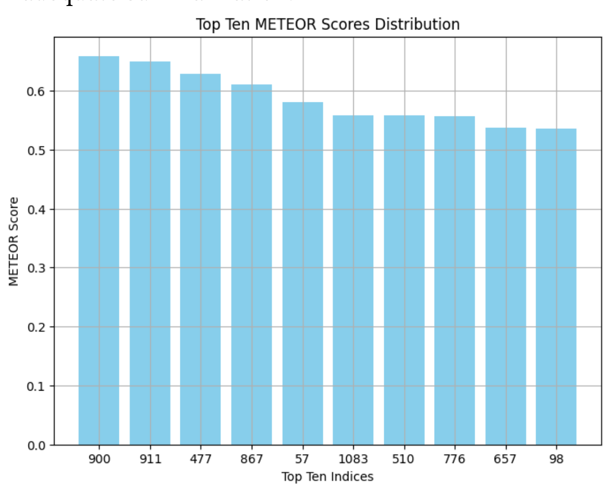

Figure 8: Top ten Meteor distribution

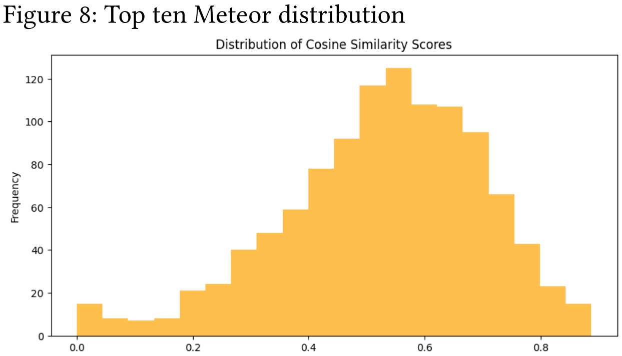

Figure 9: Distribution in Cosine similarity score

Furthermore, the cosine similarity score measures the textual resemblance between the summary and the patent document based on their TF-IDF representations. A cosine similarity score close to 1 signifies a high degree of textual similarity, suggesting a well-summarized document. Conversely, scores closer to 0 indicate greater dissimilarity, highlighting potential areas where the summarization process may fall short of capturing the essence of the original text. From the visualization, we can conclude that the cosine similarity between the patent doc and summary increases as the number of documents increases.

Overall, these computed similarity scores serve as valuable evaluative tools, guiding the assessment of the summarization algorithm's performance and providing insights into its efficacy in producing concise and faithful summaries of patent documents. They enable researchers and practitioners to identify strengths and weaknesses in the summarization process, facilitating iterative improvements and enhancements to achieve more accurate and informative summaries.

9.4 Human EValuation

In our data set, we took the Patent id as a variable and We used the 25 Patent IDs to construct the 75 summaries for this human evaluation. We have produced three summaries for each patent number. We have rated each patent number according to its customizability, clarity, accuracy, and coverage. We have also produced visualizations using the TABLUE software.

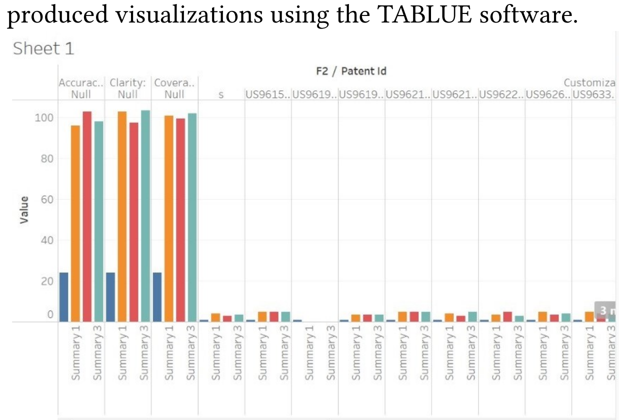

Figure 10: Human evaluation representation using tableau

From the graph, we can see the Exploratory data analysis. We can see the F2, Patient Id, and Measure Names in the x-axis and Measure Values in the Y-axis. Where the chart visualizes the frequency of the sum of the values in Accuracy, Clarity, and Coverage in the first three tiles. It also shows the frequency of summary 1, summary 2, and summary 3 for each patient. From the first graph we can observe that summary has a better rating for accuracy, summary 1 and summary 2 provide a better rating for clarity, and summary 3 produces a better rating for coverage.

9.5 Discussion

The study's discussion section offers a thorough examination of the results and their wider implementation. It focuses specifically on patent documents and customizes legal document summaries. Analyzing similarity measures such as cosine similarity, ROUGE, BLEU, and METEOR scores provides important information about how effective the summarizing process is. These metrics' distribution patterns provide insight on the quality of the summarization, highlighting both its advantages and disadvantages. Interestingly, a decline in frequency along with higher ROUGE scores indicates that although some summaries are in close agreement with the patent materials, others might need further work. In a similar vein, the distribution of scores like BLEU and METEOR highlights the significance of semantic congruency and textual similarity, directing improvements in the performance of the summarization algorithm. This conversation clarifies the technical components of evaluation and guides future research paths toward improving legal document summarizing methods. Additionally, the paper explores the opportunities and difficulties associated with creating personalized legal document summaries. The approach takes user preferences into account, emphasizing the need of customizing summaries to meet specific needs. These preferences include abstraction level, section focus, and specificity. Through the use of sophisticated summarizing models such as T5, BART, Pegasus, and GPT-3.5, the research seeks to create guidelines for adaptable summarization that connects legal jargon with streamlined data. Nonetheless, differences in similarity measures highlight the intricacies of legal language and the need for incremental advancements in summarizing techniques. This conversation lays the groundwork for further investigation into more sophisticated summarizing methods that strike a compromise between precision and comprehensibility while catering to a range of user requirements and preferences. The study's conclusions also have consequences for other stakeholders who work in fields that depend on legal information, such as scholars, practitioners, and policymakers. The study enhances decision-making rationality in several domains by offering practical insights into the area of improvement and efficacy of the summarizing process. The work yielded a customizable summary dataset that has the potential to improve legal information accessibility and expedite decision-making in a dynamic legal landscape. This conversation highlights how individualized summarizing techniques can completely change how people access, comprehend, and use legal material, opening the door to a more effective and inclusive legal system. Ultimately, the discussion section offers a comprehensive examination of the consequences that flow from the study's conclusions about the customization of legal document summaries, with a specific emphasis on patent documents. It underlines the significance of assessing the efficacy of summaries using a variety of similarity measures and the necessity of customized summarization strategies to satisfy a range of user needs. The study's conclusions provide insightful information for boosting decision-making procedures in the legal field and refining summarization algorithms. Furthermore, the conversation emphasizes how tailored summarization techniques can revolutionize the legal field by creating a more approachable and effective legal environment.

10 Conclusion and Limitations

In conclusion, the study's thorough analysis of similarity metrics provides deep insights into the efficacy of legal document summarization, particularly focusing on patents. By scrutinizing metrics such as ROUGE, BLEU, METEOR scores, and cosine similarity, the research offers a nuanced understanding of summarization quality. This nuanced understanding reveals both strengths and areas for improvement in the summarization process. For instance, the observed patterns in distribution underscore the necessity for continued refinement to achieve optimal alignment between summaries and patent documents. Moreover, the study's methodology, incorporating user preferences and advanced summarization models, lays the groundwork for future advancements in customizable summarization techniques.

The discussion further highlights the challenges and opportunities inherent in tailoring legal document summarization to meet diverse user needs. While the study's approach demonstrates promise in bridging the gap between complex legal terminology and simplified data, the observed discrepancies in similarity metrics underscore the intricacies of legal language. This acknowledgment prompts a call for further research aimed at developing more nuanced summarization strategies that balance comprehensibility with accuracy.

Ultimately, the study's findings hold significant implications for various stakeholders, including researchers, practitioners, and policymakers. By providing actionable insights into summarization effectiveness, the research contributes to enhancing decision-making processes across different domains reliant on legal

information. Moreover, the customizable summarization dataset generated through the study offers a valuable resource for improving accessibility to legal information and streamlining decision-making in an ever-evolving legal landscape.
In essence, the study marks a significant step towards revolutionizing the accessibility and usability of legal information through personalized summarization techniques. As such, it sets the stage for a more efficient, inclusive, and informed legal environment, wherein individuals and organizations can navigate complex legal documents with greater ease and confidence.
All things considered, the study offers a useful starting point for creating individualized patent document summaries. However, for practical uses and broader acceptance in the legal field, resolving the aforementioned restrictions is essential.

10.1 Limitations

Pay attention to particular metrics: ROUGE, BLEU, METEOR, cosine similarity, and SummaC scores were among the automated metrics that were used in the evaluation. Even while these measures are insightful, they might not provide a thorough evaluation of the summaries' quality, especially when it comes to legal papers that call for subtleties and particular information.
Accuracy of Large Language Models (LLMs): The success of the prompting strategies and the caliber of the training data determine how accurate the summaries produced by LLMs are. The summary may contain biases or constraints related to the training data.
Subjectivity in Human Evaluation: Although an essential procedure, subjectivity may arise in the human evaluation process. For manual review, a small sample size of 25 summaries might not offer a comprehensive view of all the summaries.
Difficulties with Legal Language: Legal language is renowned for its intricate and distinctive wording. The legal vocabulary may be difficult for the summarization models to accurately convey, which could result in misunderstandings.
Limited customization: The research investigates customization according to the degree of detail, section focus, and abstraction level. It might not, however, take into account more extensive user choices, including certain legal notions or claim components.
Generalizability to other legal documents: Although the results are centered on patent documents, more research is necessary to see whether the conclusions apply to other legal documents such as contracts or court decisions.
[7] [8] [10] [5] [9].

11 Author contribution

Timeline	Team members
Madhu surusetti	Data preprocessing , Data annotation with LLM
Anusree Putta	Generating a customizable summarization dataset for patent documents, code for data set generation
Anudeep Oggu	Evaluating the quality of the generated summarization dataset
Abhiram Nallamothu	Experiment and data analysis plan, Code of results
Nagubudi Venugopal Bhavyanth	Code for results
Grishma Tallapareddy	Human Evaluations

Table 2: Author contribution The task and team members contributions of the project are represented in the table format
The task and timeline of the project are below represented in the table format

12 Acknowledgment

As a team, we would like to thank our mentor Huyen Thi Ngoc Nguyen, and Professor. Huyen Thi Ngoc Nguyen for their guidance and support throughout this project.

References

[1] Joan Codina-Filbà, Nadjet Bouayad-Agha, Alicia Burga, Gerard Casamayor, Simon Mille, Andreas Müller, Horacio Saggion, and Leo Wanner. 2017. Using genre-specific features for patent summaries. *Information Processing & Management* 53, 1 (2017), 151–174.
[2] Alberto Díaz and Pablo Gervás. 2007. User-model based personalized summarization. *Information Processing & Management* 43, 6 (2007), 1715–1734.
[3] Noemie Elhadad, M-Y Kan, Judith L Klavans, and Kathleen R McKeown. 2005. Customization in a unified framework for summarizing medical literature. *Artificial intelligence in medicine* 33, 2 (2005), 179–198.
[4] Ben Hachey and Claire Grover. 2006. Extractive summarisation of legal texts. *Artificial Intelligence and Law* 14 (2006), 305–345.
[5] Huan Yee Koh, Jiaxin Ju, Ming Liu, and Shirui Pan. 2022. An empirical survey on long document summarization: Datasets, models, and metrics. *ACM computing surveys* 55, 8 (2022), 1–35.
[6] Chandan Kumar, Prasad Pingali, and Vasudeva Varma. 2008. Generating personalized summaries using publicly available web documents. In *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, Vol. 3. IEEE, 103–106.
[7] Jonathan Pilault, Can Liu, Mohit Bansal, and Markus Dreyer. 2023. On conditional and compositional language model differentiable prompting. *arXiv preprint arXiv:2307.01446* (2023).
[8] Yuen-Hsien Tseng, Chi-Jen Lin, and Yu-I Lin. 2007. Text mining techniques for patent analysis. *Information processing & management* 43, 5 (2007), 1216–1247.
[9] Ashok Urlana, Pruthwik Mishra, Tathagato Roy, and Rahul Mishra. 2023. Controllable Text Summarization: Unraveling Challenges, Approaches, and Prospects–A Survey. *arXiv preprint arXiv:2311.09212* (2023).
[10] Rui Yan, Jian-Yun Nie, and Xiaoming Li. 2011. Summarize what you are interested in: An optimization framework for interactive personalized summarization. In *Proceedings of the 2011*

conference on empirical methods in natural language processing. 1342–1351.