Automated Legal Information Retrieval and Summarization

Kannan Venkataramanan
DIME Analytics
World Bank
Washington DC, USA
kannanv93@gmail.com

Sandeep Bhupatiraju
DIME Analytics
World Bank
Washington DC, USA
sandeepbhupatiraju@gmail.com

Daniel Li Chen†
Department Name
World Bank
Washington DC, USA
daniel.li.chen@gmail.com

ABSTRACT

In this project, we conduct a comparison of various approaches to the extraction of legal entities from the judgment text. We apply various State-of-the-art (SOTA) models to extract (i) legal summary (ii) Primary Holding (iii) Facts of the case (iv) Important Question being answered (v) Conclusion of the case from the case. The analyses were carried out on the US Supreme Court judgments. We applied various natural language processing (NLP) techniques such as abstractive summarization methodologies combined with a custom pipeline for finetuning the models. We use ROUGE metrics and qualitative assessments to evaluate the performance of the model. We have successfully created a pipeline that uses SOTA models (i.e. BART) that shows a score of 0.3 for the baseline and 0.47 for the fine-tuned model. Overall, our experiments have shown that using larger models and text extraction methods did not improve the performance of the model but rather fine-tuning existing architectures with custom-pipeline can yield better results. This research facilitates the creation of websites like Justia and Oyez to extract important details of district and circuit cases without much human effort.

CCS CONCEPTS

• Computing methodologies → Information extraction

• Computing methodologies → Natual Language Processing

KEYWORDS

Case Holding, Deep Learning, Natural Language Processing, Summarization, Entity and Information Extraction

1 Introduction

Legal summarization refers to the process of condensing complex legal documents into concise and understandable summaries, making it easier for lawyers, legal experts, and the general public to understand and navigate the law. Building a website like Oyez [1] and Justia [2] involves creating a platform that provides access to legal information and resources, such as court cases, legal codes, and articles, in a user-friendly manner. These types of websites serve as a valuable tool for anyone looking to research and understand the law, and they also provide a comprehensive resource for legal professionals looking to stay up-to-date on the latest legal developments. With the increasing importance of technology in the legal industry, a website that provides comprehensive and easy-to-use legal information and resources is a valuable resource for anyone involved in the legal field. However, extracting key legal entities such as primary holding, key questions etc. from case judgment can be a complex and challenging task. It requires a deep understanding of the legal system, laws, and regulations and finally the ability to collect, categorize, and present information in a clear and accessible manner.

Over the past few decades, there has been growing interest in developing automated methods for information retrieval and legal summarization to aid lawyers, legal researchers, and other professionals in the field. The advent of Deep Learning and technological capabilities has facilitated extensive research for its potential applications in legal summarization in recent years. Galgani et al. (2012) [3] developed a rule-based approach to summarization that uses a knowledge base, statistical information, and other handcrafted features like POS tags, specific legal terms, and citations. Case Summarizer (Polsley et al., 2016) [4], an automated text summarization tool, uses word frequency augmented with additional domain-specific knowledge to score the sentences in the case document. Recently Abhisek et. al.(2022)[5] proposed a multi-task learning framework for extractive summarization that combines sentence classification and information extraction. In the growing body of literature, this study focuses on identifying the optimal model architecture for summarization and information extraction from US Legal judgment text.

2 Dataset

The primary source of the data was from Justia and Oyez. Justia is a platform that provides access to legal information and resources, including U.S. Supreme Court decisions, federal and state laws, and legal articles written by lawyers and legal professionals. The website offers a wide range of information, including summaries of court cases, legal opinions, and laws, as well as links to legal organizations, lawyers, and other relevant resources. Similarly, Oyez is a free, non-profit, multimedia law library. It aims to be a complete and authoritative source of information about the Supreme Court of the United States. Oyez offers a vast collection of information about the Court, including audio recordings of oral arguments and opinions, biographies of the Justices, a timeline of the Court's history, and summaries of major cases.

Justia contains the legal experts annotated “Primary Holding” and “Summary” and Oyez contains “Facts, Question and Conclusion” of U.S. Supreme Court Cases. We scrape the data from these websites. Overall, our dataset contains 8,236 case opinion texts. The cases in concern date back from 1966 to 2021 where each row of the dataset has information about each of the cases including the name, the docket number, the term (or year), and a short description of the case.

As a part of data processing, we inspect our dataset for any anomalies. We first remove HTML tags from the documents to extract clean tokens. Then, we look at the length of the input documents and the length of each task. We generate a distribution of the length of tokens for each task to identify any outliers, such as unusually short or long tokens. For example, we find two cases with token length 2 that contained the phrases “Currently available" and “Currently unknown". These rows are interpreted as having no useful information, so we drop them from the dataset. Additionally, we remove any rows that are null from the training set prior to fine-tuning our model for each task.

For each task, there are varying proportions of null values in our training set and the rows containing nulls are all dropped from our dataset prior to fine-tuning our model for each task. After removal of null or insufficient cases, we are left with the following numbers of document-summary pairs.

Task	**Number of Data Points**
Justia Summary	1278
Justia Holding	781
Facts of the Case	3357
Questions	3356
Conclusion	3356

Table 1: Description of Data

The input document of our model has token length 95 to 158,087 and the average token length is 14,594. The main challenge of our project is to design a model that can handle the varying input document lengths and generate a unique output for each task by identifying relevant information. A table showing the mean, minimum, and maximum length of the outputs is shown in Table 2.

3 Literature Survey

In general, text summarizations can be divided into two main types: extractive summarization and abstractive summarization. Extractive summarization is a method of creating a summary by selecting a few key sentences or phrases from the original text and using them to form the summary. As a result, the sentences or phrases in an extractive summary are all taken directly from the original text. One well-known algorithm for extractive summarization is TextRank (Mihalcea and Tarau [6]). TextRank is a technique for summarizing documents by modifying Google’s PageRank algorithm.

	Mean	Min	Max
Total Input	14,594	95	158,087
Summary Input	13,090	148	83,690
Summary Output	351	60	692
Holdings Input	16,438	148	139,484
Holdings Output	40	10	224
Facts Input	14,478	59	189,086
Facts Output	210	14	1,203
Questions Input	14,485	59	189,086
Questions Output	37	6	198
Conclusion Input	14,479	59	189,086
Conclusion Output	192	9	1,032

Table 2: Summary of the Data

PageRank represents documents as a graph structure and continuously updates the importance of the nodes to generate rankings. However, a disadvantage of this approach is that the model’s language generation ability is limited because it can only use existing sentences and phrases.

To improve on this, abstractive summarization is a method of summarizing the original text by generating new sentences or phrases that reflect the main context of the original text, even if they were not present in the original text. Abstractive summarization is like a human summarizing a text, where they can use their own words and sentences to convey the main points of the original text. Abstractive summarization algorithms often rely on seq2seq models, and different algorithms can be used in combination to produce the most accurate summaries. For example, Seq2seq models with attention mechanisms (Chopra et al. [7]) and pointer-generator methods (See et al. [8]) allow the model to focus on specific parts of the input text, enabling it to generate summaries that accurately reflect the content of the original text.

Extractive summarization and abstractive summarization can be performed in more advanced ways using transformer based architectures like BERT (Devlin et al. [9]), GPT-2(Radford et al. [10]), BART (Lewis et al. [11]), or T5 (Raffel et al. [12]), respectively. BERT can be used for extractive summarization by identifying important sentences in the original document, and GPT-2, BART, or T5 can be used for abstractive summarization by generating a summary of the document in a new, coherent form. Models like Longformer (Beltagy et al., 2020) [13] introduce transformer architectures with more efficient attention mechanisms that enables them to summarize long documents (up to 16×1024 input tokens). Bajaj et al. (2021)[14] developed a two-step extractive-abstractive approach for long document summarization – they use a pre-trained BART model over compressed documents generated by identifying salient sentences.

Although these methods have shown good performance on summarization datasets that contain small documents, they may not be as effective on longer documents or in low-resource environments (Bajaj et al. [14]).

Several domain-specific models and approaches have been specifically designed for summarizing legal case documents. Khan et.al. (2017) [15] proposed a method for extractive summarization of legal decisions based on multi-task learning and maximal marginal relevance. Galgani et al.(2015) [16] presented a summarization method that leverages bi-directional citation analysis to identify the most relevant documents for a given summary task. The proposed method combines information from both backward and forward citations to evaluate the importance of individual documents in a corpus. Anand and Wagh (2019) [17] used recurrent neural networks (RNNs) and long-short-term memory (LSTM) networks for the summarization of legal texts.

4 Methodology

In this project, our primary task was to build a model that automates the case summarization process and extracts relevant information from extremely long legal documents. Using the Supreme Court's public case opinion texts as input, our model is expected to identify Holdings, Facts, Questions, and Conclusions as a part of information extraction and produce the summary of the case text. The below table shows a sample case from the actual case opinions

Input (truncated) NOTE:Where it is feasible, a syllabus (headnote) will be released, as is being done in connection with this case, at the time the opinion is issued. The syllabus constitutes no part of the opinion of the Court but has been prepared by the Reporter of Decisions for the convenience of the reader. See United States v. Detroit Timber & Lumber Co.,200 U.S. 321, 337. SUPREME COURT OF THE UNITED STATES Syllabus Oklahoma v. Castro-Huerta certiorari to the court of criminal appeals of oklahoma No. 21429. Argued April 27, 2022 Decided June 29, 2022 In 2015, respondent Victor Manuel Castro-Huerta was charged by the State of Oklahoma for child neglect. Castro-Huerta was convicted in state court and sentenced to 35 years of imprisonment. While Castro-Huertas state-court appeal was pending, this Court decided ...
Output - Summary (truncated) Castro-Huerta was convicted of child neglect in Oklahoma state court. The Supreme Court subsequently held that the Creek Nation's eastern Oklahoma reservation was never properly disestablished and remained "Indian country." Castro-Huerta then argued that the federal government had exclusive jurisdiction to prosecute him (a non-Indian) for a crime committed against his stepdaughter (Cherokee Indian) in Tulsa (Indian country). The Oklahoma Court of Criminal Appeals vacated his ...
Output - Holdings The federal government and the state have concurrent jurisdiction to prosecute crimes committed by non-Indians against Indians in Indian country.
Output - Facts Victor Manuel Castro-Huerta, a non-Native, was convicted in Oklahoma state court of child neglect, and he was sentenced to 35 years. The victim, his stepdaughter, is Native American, and the crime was committed within the Cherokee Reservation.Castro-Huerta challenged his conviction, arguing that under the Supreme Court's 2020 decision in McGirt v. Oklahoma, which held that states cannot prosecute crimes committed on Native American lands without federal approval. Oklahoma argued that McGirt involved a Native defendant, whereas Castro-Huerta is non-Native, so McGirt does not bar his prosecution by the state.
Output - Question Do states have the authority to prosecute non-Natives who commit crimes against Natives on Native American lands?
Output - Conclusion (truncated) The federal government and the state have concurrent jurisdiction to prosecutecrimes committed by non-Natives against Natives on Native American land. Justice Brett Kavanaugh authored the majority opinion of the Court.The Court has held that States have jurisdiction to prosecute crimes committed by non-Natives against non-Natives on Native American lands. Native American land is not separate from state territory. And States have jurisdiction to prosecute crimes committed on ...

1) Summary: Summarization of the entire case opinion.
2) Holding: Court's decision of the case.
3) Facts of the case: Detail information of the event that are legally relevant to the court's decision including history of the dispute, legal claims, and defenses.
4) Question: Statement of the question of law that the court must answer to make a decision.
5) Conclusion: Decision made by a judge regarding a question of law.

We experiment with two different approaches.

Approach 1 - Building a pipeline that uses current SOTA transformer model, BART.

We first create a pipeline that takes case opinion text as input and outputs a distinctive outcome for each of the five tasks. We use several SOTA transformer models that are well-known for their performance on summarization tasks. Specifically, we experiment with BART, T5, and PEGASUS. For evaluating the performance of our models, we use the Rouge-L score as a metric for comparison. Rouge-L is based on the longest common subsequence (LCS) shared between the model output and the reference. A longer shared sequence indicates a higher level of similarity between the two sequences. We use an 80-20 split on our dataset and experiment with 10 different random seeds for each model to assess the robustness of the model by checking for consistency in the Rouge-L score.

Baseline Rouge-L scores of our models are used as our starting point and we apply different methods found in other related works in order to track any improvements made to our models. In other words, using Rouge-L as our metric, we experiment with various methods to see if they help us achieve higher scores for any of our tasks.

Approach 2 - Using a transformer model (LED) that handles longer documents

The main summarization task models that we mentioned under Step 1 have a limitation on maximum input token lengths. For example, Bart can only take in as its input maximum of 1,024 tokens and truncates the rest of the documents. This means that our model only looks at only 1/10 of the entire document length and generate a summary from only the beginning of the document. Therefore, we applied an LED (Longformer-Encoder-Decoder, Beltagy et al. [3]), which is designed for longer documents, with the capacity to handle at most 16,384 tokens.

Experimental Setup

The below diagram shows the pictorial representation of the experimentation design.

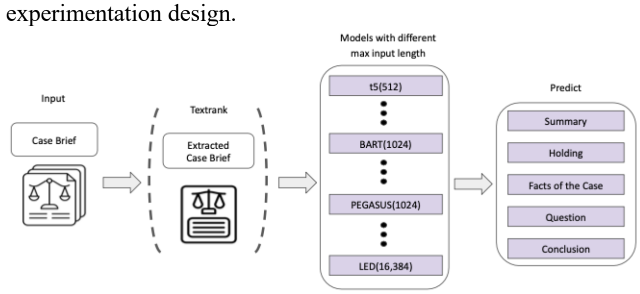

Figure 1: Experimental Design

First, we split the train and test set with an 8:2 ratio using a random seed of 0. Then, we define the maximum input token length based on the maximum input length of the model. However, for the LED model, we change the input token length from 1,024 to 16,384 to compare the performance of LED with the other models when the input token length is set to the same value. The output length is fixed to 512, as the longest token length in the output is less than 512. We use a batch size of 8, a learning rate of 5e-05, a weight decay of 0.02, and 10 epochs as our default hyperparameters setting. We also set an early stopping point based on the Rouge-L score. We encode the input and output, train the model, and generate predictions for the summary on the test set. We decode the prediction and evaluate the result based on the Rouge-L score. We expand this architecture to 5 different tasks.

Models/architectures

We apply various encoder-decoder transformer architectures. As a baseline approach, we use T5, BART, and PEGASUS pre-trained models, as those are widely used for summarization tasks. Additionally, we apply the LED model as it is capable of handling at most 16,384 input tokens. We use LED model by changing its maximum input tokens from 1,024 to 16,384 and comparing the performance of the models.

5 Model Results and interpretation

The table below shows the Rouge-L scores from different experiments

Model (Input Length)	Rouge L Scores				
	Summary	Holding	Facts	Question	Conclusion
t5-small(512)	0.3280	0.2527	0.2819	0.2696	0.2558
t5-base(512)	0.3892	0.2611	0.2731	0.2736	0.2599
t5-large(512)	0.1171	0.1187	0.1196	0.1220	0.1242
bart-base(1024)	0.4375	**0.2972**	0.2861	0.2988	0.2856
bart-large(1024)	0.4631	0.2942	0.2987	0.2966	**0.2895**
bart-large-cnn(1024)	0.4698	0.2559	**0.3054**	0.2406	0.2808
Pegasus-xsum(512)	0.4619	0.2909	0.3014	**0.3061**	0.2556
Pegasus-large(1024)	**0.4753**	0.2806	0.3024	0.2996	0.2591
LED(1024)	0.4232	0.2875	0.2697	0.2768	0.2489
LED(2048)	0.4360	0.2804	0.2795	0.2835	0.2682
LED(4096)	0.4273	0.2872	0.2893	0.2951	0.2835
LED(8192)	0.4125	0.2922	0.2884	0.3002	0.2749
LED(16384)	0.4269	0.2862	0.2718	0.3023	0.2831

We trained each task by using T5 (small, base, large), BART (base, large, large-CNN), and PEGASUS (xsum, large) models. The performance of each model varied depending on the task. No single model was able to excel at all tasks, and for each task, the model that performed best was different. For “Summary", PEGASUS-large performed the best showing Rouge-L of 0.4753. For “Holding", BART-base performed the best showing Rouge-L of 0.2972. For “Facts", BART-large-CNN performed the best showing Rouge-L of 0.3054. For “Question", PEGASUS-xsum performed the best showing Rouge-L of 0.3061. For “Conclusion", BART-large performed the best showing Rouge-L of 0.2895. Overall, our experiment shows that the general performance of these models in all tasks is approximately Rogue-L of 0.3.

We extended our experiments to larger models that can have more maximum token length than the baseline models. LED is capable of at most 16,384 input tokens, and we experimented with different sizes of inputs from 1,024 tokens to 16,384 tokens. Unfortunately, LED models showed similar or lower Rouge-L scores compared to the baseline results.

Example Outputs

<u>Example 1 : Roe v. Wade (1973)</u>

Roe v. Wade is a landmark case in the United States, decided by the Supreme Court on January 22, 1973. The case involved a Texas law that made it a crime to perform an abortion except on medical advice to save the life of the mother. Below is the outcome from the best models.

	Actual	Predicted
Holding	A person may choose to have an abortion until a fetus becomes viable, based on the right to privacy contained in the Due Process Clause of the Fourteenth Amendment. Viability means the ability to live outside the womb, which usually happens between 24 and 28 weeks after conception.	The Due Process Clause does not permit a state to criminalize an abortion on the mother's behalf without regard to the stage of her pregnancy.
Conclusion	"Inherent in the Due Process Clause of the Fourteenth Amendment is a fundamental “right to privacy” that protects a pregnant woman’s choice whether to have an abortion. However, this right is balanced against the government’s interests in protecting women's health and protecting “the potentiality of human life.” The Texas law challenged in this case violated this right.Justice Harry Blackmun delivered the opinion for the 7-2 majority of the Court.First, the Court considered whether the case was moot, concluding that it was not. When the subject of litigation is “capable of repetition yet evading review,” a case need not be dismissed as moot. Pregnancy is a “classic justification for a conclusion of nonmootness.”The Due Process Clause of the Fourteenth Amendment protects against state action the right to privacy, and a woman’s right to choose to have an abortion falls within that right to privacy. A state law that broadly prohibits abortion without respect to the stage of pregnancy or other interests violates that right. Although the state has legitimate interests in protecting the health of pregnant women and the “potentiality of human life,” the relative weight of each of these interests varies over the course of pregnancy, and the law must account for this variability.In the first trimester of pregnancy, the state may not regulate the abortion decision; only the pregnant woman and her attending physician can make that decision. In the second trimester, the state may impose regulations on abortion that are reasonably related to maternal health. In the third trimester, once the fetus reaches the point of “viability,” a state may regulate abortions or prohibit them entirely, so long as the laws contain exceptions for cases when abortion is necessary to save the life or health of the mother.	Justice William J. Brennan, Jr. delivered the opinion of the 6-3 majority.The Court held that the state had standing to sue because it had legitimate interests in protecting both the pregnant woman’s health and the potentiality of human life, and not simply when the action was initiated.In this case, the Court also held that there was no evidence that the Texas law violated the Due Process Clause of the Fourteenth Amendment.Justice John M. Harlan wrote a concurring opinion in which he argued that the Court should not have granted injunctive relief to Roe and Hallford because the case was not before the Court on appeal.Justice Harlan also argued that, while the state may regulate the abortion procedure in ways that are reasonably related to maternal health, it may not regulate the procedure in such a way as to protect the mother.Justice Thurgood Marshall joined in the concurrence.
Facts of the Case	In 1970, Jane Roe (a fictional name used in court documents to protect the plaintiff’s identity) filed a lawsuit against Henry Wade, the district attorney of Dallas County, Texas, where she resided, challenging a Texas law making abortion illegal except by a doctor’s orders to save a woman’s life. In her lawsuit, Roe alleged that the state laws were unconstitutionally vague and abridged her right of personal privacy, protected by the First, Fourth, Fifth, Ninth, and Fourteenth Amendments.	A pregnant single woman brought a class action challenging the constitutionality of Texas criminal abortion laws, which proscribe procuring or attempting an abortion except on medical advice for the purpose of saving the mother's life.A licensed physician who had two state abortion prosecutions pending against him was permitted to intervene.A childless married couple (the Does), the wife not being pregnant, separately attacked the laws, basing alleged injury on the future possibilities of contraceptive failure, pregnancy, unpreparedness for parenthood, and impairment of the wife's health.A three-judge District Court, which consolidated the actions, held that Roe and Hallford, and members of their classes, had standing to sue and presented justiciable controversies.Ruling that declaratory, though not injunctive, relief was warranted, the court declared the abortion statutes void as vague and overbroadly infringing those plaintiffs' Ninth and Fourteenth Amendment rights.The court ruled the Does' complaint not alsoiciable.
Question	Does the Constitution recognize a woman's right to terminate her pregnancy by abortion?	Did the Texas law violate the Due Process Clause of the Fourteenth Amendment?

<u>Example 2: Moose Lodge No. 107 v. Irvis (1972)</u>

Moose Lodge No. 107 v. Irvis was a United States Supreme Court case that was decided in 1972. The case involved a black man named Thomas Irvis who was denied access to a private club, Moose Lodge No. 107, in Pennsylvania because of his race.

	Actual	Predicted
Holding	Giving a liquor license to a private club that engages in discriminatory practices does not rise to the level of state action.	The Fourteenth Amendment does not allow a private club to deny guests access to its dining room and bar solely because of their race.
Conclusion	In a 6–to–3 decision, the Court held that the Moose Lodge\'s refusal to serve food and beverages to Irvis because he was black did not violate the Fourteenth Amendment. The Court noted that the state action doctrine did not necessarily apply to all private entities that received benefits or services from the government; otherwise, the Court reasoned, all private associations that received electricity, water, and fire protection would be subject to state regulation. The Court found that the Moose Lodge "a private social club in a private building," and thus not subject to the Equal Protection Clause.	Justice William H. Rehnquist delivered the opinion of the 6-3 majority.The Court held that the operation of the Pennsylvania liquor board’s regulatory scheme, except as noted below, did not sufficiently implicate the state in Irvis’ discriminatory guest practices so as to make those practices “state action” within the purview of the Equal Protection Clause.In his opinion concurring in the judgment, Justice William O. Douglas wrote that there was no evidence that the liquor board intended to restrict the availability of liquor at Moose Lodge.Justice William J. Brennan, Jr. wrote a dissenting opinion in which he argued that the Court should not have granted injunctive relief to Irvis because he had not applied for or been denied membership in Moose Lodge and was not aware of any state action.Justice Thurgood Marshall joined in the dissent.
Facts of the Case	K. Leroy Irvis, a black man who was a guest of a white member of the Moose Lodge No. 107, was refused service at the club\'s dining room because of his race. The bylaws of the Lodge limited membership to white male Caucasians. Irvis challenged the club\'s refusal to serve him, arguing that the action of the Pennsylvania liquor board issuing the Lodge a license made the club\'s discrimination "state action.	A member of Moose Lodge, a private club in Philadelphia, was refused service at the club's dining room and bar solely because of his race.\nThe Pennsylvania liquor board issued Moose Lodge a private liquor license.\nIrvis, a black guest of the club, sued in federal district court and argued that the state action was state action, and thus a violation of the Equal Protection Clause of the Fourteenth Amendment.\nThe district court found Moose Lodge's membership and guest practices discriminatory and declared the liquor license invalid as long as Moose Lodge continued its discriminatory practices.\nThe court denied the motion.\nFollowing the District Court's decision, the applicable bylaws were amended to exclude as guests those who would be excluded as members.
Question	Did the discriminatory practices violate the Equal Protection Clause of the Fourteenth Amendment?	Did the amendment to the liquor board’s bylaws violate the Equal Protection Clause of the Fourteenth Amendment?

Summaries were not shown due to its length. In the above examples we can clearly see how well the model is able to predict/retrieve the relevant information from the case text.

6 Conclusion

In conclusion, the research on legal summarization and information retrieval has demonstrated the effectiveness of deep learning models, specifically the BART and Pegasus model, in generating quality output for different tasks from legal texts. The results highlight the potential of deep learning models in improving the efficiency and accuracy of legal research and information retrieval. The findings of this research can be leveraged to improve the functionality and usability of websites like Justia for district and circuit courts, providing a more accessible and efficient platform for legal information retrieval.

The development of such websites is crucial for the general public, legal practitioners, and researchers to stay updated on the latest legal developments and decisions.

ACKNOWLEDGMENTS

We would like to thank Yoobin Cheong, Yoon Tae Park and Yeong Koh from New York University for their contribution in this research.

REFERENCES

[1]. Oyez - https://www.oyez.org/

[2]. Justia - https://www.justia.com/

[3]. Galgani, Filippo, Paul Compton, and Achim Hoffmann. "Combining different summarization techniques for legal text." Proceedings of the workshop on innovative hybrid approaches to the processing of textual data. 2012.

[4]. Polsley, Seth, Pooja Jhunjhunwala, and Ruihong Huang. "Casesummarizer: A system for automated summarization of legal texts." Proceedings of COLING 2016, the 26th international conference on Computational Linguistics: System Demonstrations. 2016.

[5]. Agarwal, Abhishek, Shanshan Xu, and Matthias Grabmair. "Extractive Summarization of Legal Decisions using Multi-task Learning and Maximal Marginal Relevance." arXiv preprint arXiv:2210.12437 (2022).

[6]. Rada Mihalcea and Paul Tarau. TextRank: Bringing order into text. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, pages 404– 411, Barcelona, Spain, July 2004. Association for Computational Linguistics. URL https: //aclanthology.org/W04-3252.

[7]. Sumit Chopra, Michael Auli, and Alexander M. Rush. Abstractive sentence summarization with attentive recurrent neural networks. In Proceedings of the 2016 Conference of the North Ameri- can Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 93–98, San Diego, California, June 2016. Association for Computational Linguistics. doi: 10.18653/v1/N16-1012. URL https://aclanthology.org/N16-1012.

[8]. Abigail See, Peter J. Liu, and Christopher D. Manning. Get to the point: Summarization with pointer-generator networks, 2017. URL https://arxiv.org/abs/1704.04368.

[9]. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2018. URL https://arxiv.org/ abs/1810.04805.

[10]. Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.

[11]. Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, 2019. URL https://arxiv.org/abs/1910.13461.

[12]. Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer, 2019. URL https://arxiv.org/abs/1910.10683.

[13]. Beltagy, Iz, Matthew E. Peters, and Arman Cohan. "Longformer: The long-document transformer." arXiv preprint arXiv:2004.05150 (2020).

[14]. Ahsaas Bajaj, Pavitra Dangati, Kalpesh Krishna, Pradhiksha Ashok Kumar, Rheeya Uppaal, Bradford Windsor, Eliot Brenner, Dominic Dotterrer, Rajarshi Das, and Andrew McCallum. Long document summarization in a low resource setting using pretrained language models, 2021. URL https://arxiv.org/abs/2103.00751.

[15]. Agarwal, Abhishek, Shanshan Xu, and Matthias Grabmair. "Extractive Summarization of Legal Decisions using Multi-task Learning and Maximal Marginal Relevance." arXiv preprint arXiv:2210.12437 (2022).

[16]. Filippo Galgani, Paul Compton, and Achim Hoffmann. 2015. Summarization based on bi-directional citation analysis. Information processing & management, 51(1):1–24.

[17]. Deepa Anand and Rupali Wagh. 2019. Effective deep learning approaches for summarization of legal texts. Journal of King Saud University-Computer and Information Sciences.