CivilSum

A dataset for abstractive summarization of Indian court decisions

Citation for published version:
Malik, M, Zhao, Z, Fonseca, M, Rao, S & Cohen, SB 2024, CivilSum: A dataset for abstractive summarization of Indian court decisions. in *SIGIR 2024 - Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval.* ACM, pp. 2241-2250, 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, Washington D. C., Washington, United States, 14/07/24. https://doi.org/10.1145/3626772.3657859

Digital Object Identifier (DOI):
10.1145/3626772.3657859

Published In:
SIGIR 2024 - Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval

CivilSum: A Dataset for Abstractive Summarization of Indian Court Decisions

Manuj Malik*
IIIT Bangalore
Bangalore, India
manuj.malik@iiitb.org

Zheng Zhao*
The University of Edinburgh
Edinburgh, United Kingdom
zheng.zhao@ed.ac.uk

Marcio Fonseca*
The University of Edinburgh
Edinburgh, United Kingdom
m.fonseca@ed.ac.uk

Shrisha Rao
IIIT Bangalore
Bangalore, India
shrao@ieee.org

Shay B. Cohen
The University of Edinburgh
Edinburgh, United Kingdom
scohen@inf.ed.ac.uk

ABSTRACT

Extracting relevant information from legal documents is a challenging task due to the technical complexity and volume of their content. These factors also increase the costs of annotating large datasets, which are required to train state-of-the-art summarization systems. To address these challenges, we introduce CivilSum, a collection of 23,350 legal case decisions from the Supreme Court of India and other Indian High Courts paired with human-written summaries. Compared to previous datasets such as IN-Abs, CivilSum not only has more legal decisions but also provides shorter and more abstractive summaries, thus offering a challenging benchmark for legal summarization. Unlike other domains such as news articles, our analysis shows the most important content tends to appear at the end of the documents. We measure the effect of this *tail bias* on summarization performance using strong architectures for long-document abstractive summarization, and the results highlight the importance of long sequence modeling for the proposed task. CivilSum and related code are publicly available to the research community to advance text summarization in the legal domain.[1]

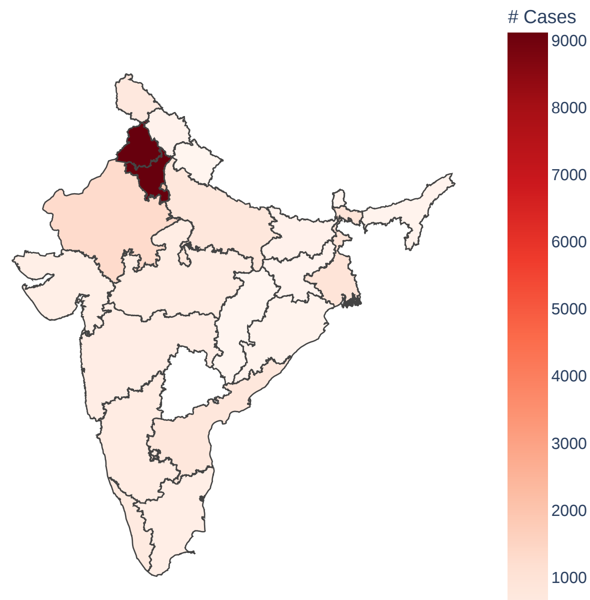

Figure 1: Distribution of CivilSum legal cases across Indian states. The majority of samples originate from the Supreme Court of India (4,499; not on the map) and the High Courts of Punjab and Haryana (9,111), and Delhi (1,790).

CCS CONCEPTS

• **Computing methodologies** → **Natural language processing**; • **Applied computing** → Law; • **Information systems** → **Summarization**.

KEYWORDS

Abstractive text summarization; Dataset; Legal IR; Legal document summarization

ACM Reference Format:
Manuj Malik, Zheng Zhao, Marcio Fonseca, Shrisha Rao, and Shay B. Cohen. 2024. CivilSum: A Dataset for Abstractive Summarization of Indian Court Decisions. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24), July 14–18, 2024, Washington, DC, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3626772.3657859

1 INTRODUCTION

With the growing demand for automation of legal systems, the development of natural language processing (NLP) techniques for analyzing legal documents has become a critical area of research [6, 8, 28, 40]. In particular, summarizing legal documents is an important problem as effective summaries can benefit stakeholders across various sectors, including legal professionals, corporate entities, government agencies, academia, and the general public. It streamlines legal research, aids in decision-making processes, enhances access to justice, facilitates academic studies, and promotes public understanding of legal matters. However, summarizing legal

*Equal contribution.
[1] https://github.com/ra-MANUJ-an/CivilSum. We release our corpus under the CC BY-NC-SA 4.0 license.

SIGIR '24, July 14–18, 2024, Washington, DC, USA

ACM ISBN 979-8-4007-0431-4/24/07
https://doi.org/10.1145/3626772.3657859

documents is challenging due to their length and technical complexity, which increases the difficulty and cost of the collection of high-quality reference summaries required by state-of-the-art supervised summarization approaches.

To address these challenges, we introduce CivilSum, a dataset for abstractive summarization of legal documents. CivilSum comprises a collection of 23,350 legal case decisions from the Supreme Court of India and other Indian High Courts, each paired with a summary written by a legal professional. The dataset provides a rich source of information for training and evaluating NLP models for legal summarization tasks.

In this work, we describe the process of constructing the CivilSum dataset and compare its quantitative and qualitative characteristics with previous work in this domain. Our analysis reveals an interesting observation that the most important summarizable content tends to appear at the end of the legal documents, which is opposite from the lead bias observed in other domains such as news articles [30, 31]. We also evaluate our dataset using two architectures for long-document abstractive summarization, namely Longformer [2] and FactorSum [14]. Our results reveal that abstractive approaches outperform paragraph-based extractive methods, emphasizing the need for fine-grained, intra-paragraph abstractive processing to generate high-quality summaries on the CivilSum dataset. Our findings also suggest that the end of documents contains more informative content, as observed by comparing the results with lead and tail content guidance. Given recent advances in large language models (LLMs) and their effectiveness in news summarization [38], we also assess our dataset using Llama 2 [36], an open-source LLM with the capacity to model lengthy text. Although ROUGE performance is inferior to the other two methods, human evaluation of overall summary quality shows Llama 2 summaries were more favored.

2 RELATED WORK

This section provides an overview of previous research on legal document summarization, specifically examining the existing datasets. Since legal document summarization falls under the broader category of long document summarization, we are exposed to similar challenges such as summary coherence and scaling computation for longer contexts [22]. Various datasets have been released for summarizing long texts in the academic domain [5, 7, 27, 32] and government reports [4, 18, 33]. In the legal domain, an early investigation was conducted in the area of legislative text summarization, with a focus on US congressional proceedings and California state bills [23]. Further research on legal documents spanned diverse areas such as legal acts [1], legal cases [3, 12, 15, 29, 34], debate dialogue [10], and regulatory documents [21].

While most of the legal NLP work focuses on US datasets, other jurisdictions are also studied, including the summarization of 4,595 curated European regulatory documents [EUR-LexSum; 21], and topic modeling applied to multi-document summarization in the Brazilian lawmaking process [35]. The LEGAL-BERT family of models [6] and JuriBERT [9] have shown competitive performance in legal NLP tasks through pre-training on large arrays of English and French legal texts, respectively. An argument mining approach is used to improve abstractive summarization of Canadian legal cases [12], proposing a dataset consisting of 1,262 legal cases obtained through an agreement with the Canadian Legal Information Institute[2].

Table 1: A list of the most frequent topic keywords in CivilSum training split.

Frequency	Topic Keywords
367	engineer, engineers, promotion, executive
353	medical, admission, seats, college
337	compassionate, family, deceased
297	compensation, accident, workmens
234	promotion, post, assistant, promoted
227	admission, examination, university, course
216	candidates, selection, commission, posts
196	police, constables, constable, inspector
192	seniority, posts, candidates, recruitment
175	disability, disease, pension, medical
167	workmen, labour, industrial, act

In the Indian context specifically, similar summarization problems have been explored [3, 15, 33, 34]. Bhattacharya et al. [3] provided a dataset and performed extractive summarization operations on the dataset. Shukla et al. [34] developed three datasets, primarily "IN-Abs" consisting of 7,130 document-summary pairs obtained from the website of the Legal Information Institute of India[3], "IN-Ext" consisting of 50 manually annotated summaries of judgments, and "UK-Abs" from the website of UK Supreme court[4] having 693 cases. For those datasets, they perform and evaluate both extractive and abstractive summarization models. Parikh et al. [33] has introduced a large dataset of 10,764 Indian legal documents and employed a weakly supervised approach to tackle the challenge of summarizing legal documents. Our work aims to increase the scale of summarization datasets in the Indian legal domain.

3 CIVILSUM: DESIGN OF BENCHMARK DATASET

The focus of the CivilSum dataset is on civil cases heard by the Supreme Court of India and Indian High Courts from the country's independence (1947) up until the 2010–2011 calendar year. In comparison to previous work, CivilSum is significantly larger in dataset size, including 3 times more samples than IN-Abs [34]. The distribution of cases per state is illustrated in Figure 1.[5]

In addition, our human-written summaries have a higher compression ratio, providing more concise and informative summaries.[6] Previous datasets have a compression ratio of around 5-10. In contrast, the summaries in CivilSum have a higher compression ratio of around 16, making the task of summarization more challenging. Following Narayan et al. [31], we also compute the fraction of n-grams in the summary that are not present in the original

[2]https://www.canlii.org/en/

[3]http://www.liiofindia.org/in/cases/cen/INSC/

[4]https://www.supremecourt.uk/decided-cases/

[5]This map was generated using the Plotly tool and may not include disputed regions or other areas of contention. The boundaries and names shown on the map do not imply official endorsement or acceptance by the authors or their affiliations.

[6]The compression ratio is calculated as the ratio of the number of words in the original document to the number of words in the summary

Table 2: Statistics for legal summarization datasets, including the number of documents, average length in words/sentences, and summary abstractiveness (measured as the percentage of novel n-grams).

Dataset	# docs (train/val/test)	Avg. document length		Avg. summary length		% novel n-grams in summary			
		words	sentences	words	sentences	1-gram	2-gram	3-gram	4-gram
IN-Abs [34]	7,030/-/100	4,378	-	1,051	-	18.95	34.71	47.19	56.12
EUR-LexSum [21]	3,447/689/459	11,864	340	1,011	32	43.70	71.00	84.62	90.29
CivilSum	21,015/1,168/1,167	2,123	90	104	4.5	62.60	91.52	98.87	99.77

document. A summary of the main dataset statistics compared to existing legal summarization datasets is provided in Table 2.

To get a better insight into the content covered by CivilSum, we generated topics for the training split (21,015 samples) using the BERTopic topic modeling framework with its default settings [17]. After filtering out non-informative topics, we found the most frequent topics cover labor, professional, and admission issues. A sample of the most frequent topics is provided in Table 1. In the next sections, we present details about the collection, cleaning, and post-processing of legal cases and summaries.

3.1 Data Collection

The legal cases and corresponding summaries were obtained from the Patiala Law House, Patiala, India, through an agreement.[7] The data consists of judgments from various judicial courts situated in different parts of India immediately after the independence of India till the calendar year 2010-11. For each case judgment, we obtain the following information (in DOCX format): 1) A document identifier derived from the position on the hit list returned by the system; 2) The petitioner and respondent's and contesting parties' names. It is important to notice that the 'versus' clause may contain numerous petitioners' and respondents' names; 3) The name of the court that rendered this ruling; 4) The judgment's summary, usually referred to as a headnote (in legalese); 5) References to previous relevant cases. These references pertain to the legal cases cited by the adjudicating authority based on earlier judgments referred to by the parties in the case; 6) The text conveying the judge's decision; 7) A summary of the judgment. The summaries are condensed from the judgments and are manually written without automation.

3.2 Data Cleaning

The data is preprocessed and cleaned, starting from the original DOCX files containing the judgments. We developed a matching algorithm to recognize six different types of information: ranks, names of the contesting parties, name of the court where the judgment appeared, references from previous judgments (from which the current judgment draws support for its claims), judgment text, and summaries (headnotes in legalese). The written documents contain errors, thus several edge cases are addressed. For instance, documents are divided by words with spelling variations like `ORDER`, `COMMON ORDER`, `JUDGMENT`, `JUEDGMENT`, `JUDGEMENT`, and `JUDGMEN2T` to separate judgment content from the remaining text. The names of the contesting parties receive a similar level of attention, and the strings `cases referred` and `case referred` are used to separate reference cases in the judgment document as this is a common pattern found after carefully assessing a subset of decisions. Our matcher misses certain faults as the documents were prepared manually by human experts and are therefore prone to human error.

[7]Copyright details are provided in Appendix A.

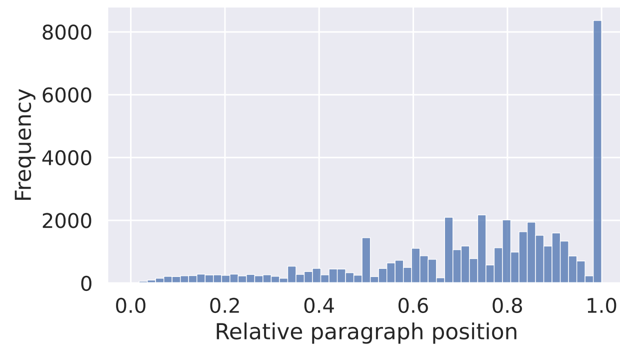

Figure 2: Distribution of relevant paragraph positions in the documents (training split) exhibiting tail bias.

3.3 Paragraph Reference Extraction

A salient stylistic feature of the dataset is that most of the paragraphs in the summaries include textual references to the relevant paragraphs in the judgments, which we hypothesize is an important signal for summarization modeling. To leverage this data, we devised a pattern-matching algorithm to extract paragraph references of the form `[Paras 10, 15, 17]` (refer to Appendix B for an example). By applying this heuristic, we create a dataset where each paragraph in a judgment is labeled as 1 if mentioned in the summary, and 0 otherwise. Out of 23,350 documents in the dataset, 22,682 ($\approx$ 97%) contain at least one referenced paragraph in the reference summaries.

This paragraph reference information reveals an interesting insight about the information distribution in the dataset: most of the relevant content is located towards the end of the documents, a characteristic we refer as to *tail bias* (Figure 2). A consequence of this finding is that summarization systems that are biased towards leading information, as commonly seen in news summarization [16, 41], should not perform well on our benchmark. We explore this tail bias in various settings in Section 4.

4 SUMMARIZATION EXPERIMENTS

We describe summarization experiments with various types of architectures designed to process long documents. Our objective is to provide a baseline performance assessment for future work and to measure how the distribution of relevant information in the documents affects summarization performance. The models are detailed as follows:

- **Random extractive baseline.** To get an estimate for the task difficulty, we randomly sample paragraphs from the documents up to 7% of the total words, subject to a minimum of 100 words. If the document has 100 or fewer words, the entire document is used as the summary.
- **Extractive oracle paragraphs.** We also obtain oracle extractive summaries that include only paragraphs mentioned in the reference summaries (refer to Section 3.3 for details). The budget constraints are the same as the random extractive baseline described above. In contrast to the random baseline, the oracle paragraphs are added in the same order as they appear in the input document.
- **FactorSum** [13, 14], an abstractive summarization model that employs a sampling mechanism to generate several summary snippets (*summary views*), which are then combined into a final summary following a guidance optimization objective. We leverage guidance to bias the resulting summary to focus on the start of the document (*lead guidance*) and the end of the document (*tail guidance*). Additionally, we measure the performance using extractive oracle paragraphs as guidance, that is, we encourage the final summary to be similar (using ROUGE-1 as the similarity metric; see below) to oracle paragraphs. We choose FactorSum because it can handle long documents by relying on a relatively small sequence-to-sequence backbone [BART-base; 24] and a short input context (1,024 tokens).
- **Longformer** [2], a transformer-based model that implements an attention mechanism that scales linearly with the input length, which makes it suited for the processing of the long documents from our dataset. We experiment with various input configurations, including 1) the first 4,096 input tokens, 2) truncated documents of the first 1,024 tokens (lead), and 3) truncated documents of the last 1,024 tokens (tail). Additionally, we test the performance of the model when using only oracle paragraphs as inputs.
- **Llama 2** [36], a transformer-based large language model. With up to 70 billion parameters, Llama 2 can process lengthy texts. We leverage the finetuned chat version of Llama 2 and provide it with 4,096 tokens as input. In this work, "Llama 2" or "Llama 2-chat" refers to the chat version.

Data pre-processing. For FactorSum, we augment the document-summary pairs by creating pairs of *document views* and *summary views* that capture different perspectives of the original documents. To this end, we first perform sentence tokenization on both documents and summaries. Then, we uniformly sample 20% of the sentences in the documents to serve as document views for each one of the 21,013 documents in the training set, resulting in 420,260 shorter training samples. Each document view is paired with a corresponding subset of the original summary, which we refer to as a summary view. This pairing of sentences is determined by the minimization of the sum of ROUGE-1 and ROUGE-2 scores (refer to Section 2.1 from Fonseca et al. [14] for more details). Using the same approach, we obtain 23,360 and 23,340 document-summary view pairs for the validation and test sets respectively. Apart from the usual input truncation in transformer models, no further pre-processing is performed for Longformer and Llama 2.

Training and inference details. We use a BART-base [24] checkpoint from HuggingFace[8] as starting point to train FactorSum summary views generator. The maximum length for generation per view is set to 128 tokens, the effective batch size is 64, and we use the Adam optimizer [20] with a learning rate of 5×10^{-5}, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. The training is performed for 50,000 steps on 4 GeForce GTX 1080 Ti GPUs, and we choose the checkpoint with the highest ROUGE-1 F1 score on the validation split. We employ a pre-trained LED-base checkpoint from HuggingFace[9] for Longformer and fine-tuned the model using a learning rate of 1×10^{-4} on 4 NVIDIA A100 GPUs with 128 effective batch size. The maximum length for summary is set to 256 tokens. All other training details align with those used for FactorSum. During inference, FactorSum performs the greedy optimization described by Fonseca et al. [14] using the same sampling hyperparameters as the training phase (20 document views per sample, each with 20% of the original sentences), with a budget constraint of 190 words per summary. Longformer uses a beam size of 3. For Llama 2-chat, we query the model with the prompt template: "`${document}.\n Write a summary of the text above in 4 sentences.`", and parse the model's completion as the candidate summary. For sampling hyperparameters, we use a value of 0.6 for temperature, and 0.9 for top-p filtering.

Evaluation metrics. In all experiments, we report performance measured by ROUGE-1/2/L F1 score [25] and BERTScore [37], following previous work in the summarization literature. ROUGE metrics measure the word overlap, bigram overlap, and longest common sequence between system-generated and reference summaries. BERTScore, a neural metric with a range between 0 and 1, evaluates the similarity between generated summaries and reference summaries.

5 RESULTS

5.1 Automatic Evaluation

The results in Table 3 show a large gap in performance between a paragraph-based extractive summarizer and the abstractive approaches. This result suggests that summarizing more fine-grained, intra-paragraph abstractive processing is required to generate high-quality summaries. Still, we can verify that paragraph references are highly informative, improving the scores in ≈ 14 R-1 over the random extractive summarizer, and ≈ 6 R-1 over FactorSum with lead guidance. Moreover, Longformer's performance with only oracle paragraphs as input is approximately ≈ 2.4 R-1 points higher than when using the first 1,024 tokens (lead). It is comparable to the performance achieved with 4,096 tokens, providing additional evidence for the informativeness of the referenced paragraphs. We

[8]https://huggingface.co/facebook/bart-base
[9]https://huggingface.co/allenai/led-base-16384

Table 3: ROUGE and BERTScore F-1 scores for the summarization task. *lead* and *tail* refer to summaries focusing on the start and end of documents, respectively. The *paras.* leverage oracle paragraph information as described in Section 4. Underlined results are statistically the best scores ($p < 0.05$, using bootstrap methods [11]).

Model	Input Tokens	R-1	R-2	R-L	BERTScore
Extractive (random)	-	31.81	8.38	20.92	0.835
Extractive (paras.)	-	33.12	10.11	22.01	0.840
Llama 2-chat-7B	4096	37.12	12.55	25.43	0.851
Llama 2-chat-13B	4096	36.73	11.63	25.61	0.847
Llama 2-chat-70B	4096	37.39	12.61	25.74	0.852
FactorSum (lead)	1024	40.33	15.74	31.98	0.863
FactorSum (tail)	1024	41.80	16.53	33.30	0.867
FactorSum (paras.)	1024	**46.51**	**20.67**	**37.07**	**0.874**
Longformer	4096	44.80	18.37	36.85	0.868
Longformer (lead)	1024	41.77	16.15	34.37	0.864
Longformer (tail)	1024	44.35	17.65	36.32	0.867
Longformer (paras.)	1024	44.10	18.53	36.77	0.867

observe similar trends in BERTScore results. Despite a high Spearman correlation (ρ=0.998) between ROUGE (R-1) and BERTScore, the latter appears to be less sensitive to different systems.

Another salient pattern is the higher informativeness towards the end of the documents, which can be verified by comparing the results of FactorSum with lead and tail guidance. Similarly, we observe a strong loss in Longformer performance by truncating the documents to the first 1,024 tokens (lead) compared to using 4,096 tokens, but the loss in performance is much smaller when using the last 1,024 tokens (tail). Finally, we observe that Llama 2 exhibits superior performance to extractive summarization approaches, yet remains inferior to other abstractive methods. We posit this stems from evaluating Llama 2 in a zero-shot setting without fine-tuning. As LLMs show promise in legal summarization, we leave finetuning Llama 2 with in-domain data for future work. Additionally, we observe that scaling Llama 2 parameters does not further improve performance. Nonetheless, the zero-shot results demonstrate Llama 2's capability to generate reasonable abstractive summaries without training. Further tuning could likely adapt the model to the target summaries' style and content.

5.2 Human Evaluation

In addition to automated measures like ROUGE, we designed a human evaluation to collect preference annotations. For each given document, annotators were presented with summaries from all three summarization systems (FactorSum, Longformer, and Llama 2). They were first instructed to select their most preferred summary to replace a technical judgment abstract. Subsequently, they were prompted to choose the best summary accounting for criteria such as informativeness and fluency. Refer to Appendix C for further details. Our evaluators comprised two trained Indian lawyers familiar with the cases, who examined 25 randomly selected samples.

Regarding the results for the first question, the first annotator preferred the FactorSum, Longformer, and Llama 2 summaries 10, 9, and 6 times, respectively. The second annotator preferred them 8, 13, and 4 times. The inter-annotator agreement as measured by Cohen's kappa was 0.44, indicating moderate agreement. For the second question, the preferences were 6, 6, 13 and 6, 7, 12, respectively. The inter-annotator agreement was 0.52, again suggesting moderate alignment. These results imply that for technical adequacy, summaries from supervised models like FactorSum and Longformer were preferred. However, considering overall summary quality, the Llama 2 summaries were favored.

In addition, the annotators observed that although the summaries were generally adequate and captured key points successfully, there were deficiencies in sentence construction ambiguity, erroneous interpretations of interest payment, and sporadic incompleteness. Furthermore, concerns were raised about the lack of conciseness and occasional omission of conclusions, which are crucial elements in summarizing legal judgments. In the sample examples (see Table 5 and Table 6), the evaluator highlighted specific issues, including the overuse of articles 14 and 16 of the Constitution of India without proper contextual relevance, a tendency to refer to party names instead of directly addressing the real issue, and the use of personal pronouns instead of maintaining an objective tone. Moreover, there was a lack of sufficient attention to the legal aspects of the issue, resulting in an incomplete and inadequate portrayal of the real issue from a legal standpoint. Overall, the evaluation suggests that our dataset is challenging, and current summarization systems struggle to produce satisfactory summaries.

6 CONCLUSIONS AND FUTURE WORK

In this work, we introduce CivilSum, a novel dataset for legal summarization containing 23,350 court decisions paired with human-written summaries. We describe the steps for dataset construction and provide extractive and abstractive summarization baselines to serve as a benchmark for further investigation. We explore stylistic features of the documents such as paragraph references and measure how information tail bias affects the summarization performance in diverse settings. A promising direction for future work would be to assess the factuality of generated summaries, including identifying hallucination [19, 26, 39], in terms of relevant entities such as legislation references, which is a crucial aspect of court decisions.

ACKNOWLEDGMENTS

We thank the reviewers and Matt Grenander for their useful feedback. Zheng Zhao is supported by the UKRI Centre for Doctoral Training in Natural Language Processing (EP/S022481/1). We also gratefully acknowledge the use of computational resources through Baskerville Tier 2 HPC service at the University of Birmingham.

REFERENCES

[1] Dennis Aumiller, Ashish Chouhan, and Michael Gertz. 2022. EUR-Lex-Sum: A Multi- and Cross-lingual Dataset for Long-form Summarization in the Legal Domain. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 7626–7639. https://doi.org/10.18653/v1/2022.emnlp-main.519

[2] Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. Longformer: The Long-Document Transformer. arXiv:2004.05150 [cs.CL]

[3] Paheli Bhattacharya, Soham Poddar, Koustav Rudra, Kripabandhu Ghosh, and Saptarshi Ghosh. 2021. Incorporating domain knowledge for extractive summarization of legal case documents. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law* (São Paulo, Brazil) *(ICAIL '21)*. Association for Computing Machinery, New York, NY, USA, 22–31. https://doi.org/10.1145/3462757.3466092

[4] Shuyang Cao and Lu Wang. 2022. HIBRIDS: Attention with Hierarchical Biases for Structure-aware Long Document Summarization. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Dublin, Ireland, 786–807. https://doi.org/10.18653/v1/2022.acl-long.58

[5] Ronald Cardenas, Matthias Gallé, and Shay B. Cohen. 2024. On the Trade-off between Redundancy and Cohesion in Extractive Summarization. *Journal of Artificial Intelligence Research* (2024).

[6] Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. LEGAL-BERT: The Muppets straight out of Law School. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics, Online, 2898–2904. https://doi.org/10.18653/v1/2020.findings-emnlp.261

[7] Arman Cohan, Franck Dernoncourt, Doo Soon Kim, Trung Bui, Seokhwan Kim, Walter Chang, and Nazli Goharian. 2018. A Discourse-Aware Attention Model for Abstractive Summarization of Long Documents. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*. Association for Computational Linguistics, New Orleans, Louisiana, 615–621. https://doi.org/10.18653/v1/N18-2097

[8] Robert Dale. 2019. Law and word order: NLP in legal tech. *Natural Language Engineering* 25, 1 (2019), 211–217.

[9] Stella Douka, Hadi Abdine, Michalis Vazirgiannis, Rajaa El Hamdani, and David Restrepo Amariles. 2021. JuriBERT: A Masked-Language Model Adaptation for French Legal Text. In *Proceedings of the Natural Legal Language Processing Workshop 2021*. Association for Computational Linguistics, Punta Cana, Dominican Republic, 95–101. https://doi.org/10.18653/v1/2021.nllp-1.9

[10] Xinyu Duan, Yating Zhang, Lin Yuan, Xin Zhou, Xiaozhong Liu, Tianyi Wang, Ruocheng Wang, Qiong Zhang, Changlong Sun, and Fei Wu. 2019. Legal Summarization for Multi-role Debate Dialogue via Controversy Focus Mining and Multi-task Learning. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (Beijing, China) *(CIKM '19)*. Association for Computing Machinery, New York, NY, USA, 1361–1370. https://doi.org/10.1145/3357384.3357940

[11] Bradley Efron. 1992. *Bootstrap Methods: Another Look at the Jackknife*. Springer New York, New York, NY, 569–593. https://doi.org/10.1007/978-1-4612-4380-9_41

[12] Mohamed Elaraby and Diane Litman. 2022. ArgLegalSumm: Improving Abstractive Summarization of Legal Documents with Argument Mining. In *Proceedings of the 29th International Conference on Computational Linguistics*. International Committee on Computational Linguistics, Gyeongju, Republic of Korea, 6187–6194. https://aclanthology.org/2022.coling-1.540

[13] Marcio Fonseca and Shay B. Cohen. 2024. Can Large Language Model Summarizers Adapt to Diverse Scientific Communication Goals? arXiv:2401.10415 [cs.CL]

[14] Marcio Fonseca, Yftah Ziser, and Shay B. Cohen. 2022. Factorizing Content and Budget Decisions in Abstractive Summarization of Long Documents. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 6341–6364. https://doi.org/10.18653/v1/2022.emnlp-main.426

[15] Satyajit Ghosh, Mousumi Dutta, and Tanaya Das. 2022. Indian Legal Text Summarization: A Text Normalization-based Approach. In *2022 IEEE 19th India Council International Conference (INDICON)*. Institute of Electrical and Electronics Engineers, Kochi, India, 1–4. https://doi.org/10.1109/INDICON56171.2022.10039891

[16] Matt Grenander, Yue Dong, Jackie Chi Kit Cheung, and Annie Louis. 2019. Countering the Effects of Lead Bias in News Summarization via Multi-Stage Training and Auxiliary Losses. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, 6019–6024. https://doi.org/10.18653/v1/D19-1620

[17] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv:2203.05794 [cs.CL]

[18] Luyang Huang, Shuyang Cao, Nikolaus Parulian, Heng Ji, and Lu Wang. 2021. Efficient Attentions for Long Document Summarization. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Online, 1419–1436. https://doi.org/10.18653/v1/2021.naacl-main.112

[19] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of Hallucination in Natural Language Generation. *ACM Comput. Surv.* 55, 12, Article 248 (mar 2023), 38 pages. https://doi.org/10.1145/3571730

[20] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. http://arxiv.org/abs/1412.6980 cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.

[21] Svea Klaus, Ria Van Hecke, Kaweh Djafari Naini, Ismail Sengor Altingovde, Juan Bernabé-Moreno, and Enrique Herrera-Viedma. 2022. Summarizing Legal Regulatory Documents using Transformers. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (<conf-loc>, <city>Madrid</city>, <country>Spain</country>, </conf-loc>) *(SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 2426–2430. https://doi.org/10.1145/3477495.3531872

[22] 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.

[23] Anastassia Kornilova and Vladimir Eidelman. 2019. BillSum: A Corpus for Automatic Summarization of US Legislation. In *Proceedings of the 2nd Workshop on New Frontiers in Summarization*. Association for Computational Linguistics, Hong Kong, China, 48–56. https://doi.org/10.18653/v1/D19-5406

[24] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 7871–7880. https://doi.org/10.18653/v1/2020.acl-main.703

[25] Chin-Yew Lin. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out*. Association for Computational Linguistics, Barcelona, Spain, 74–81. https://aclanthology.org/W04-1013

[26] Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. 2020. On Faithfulness and Factuality in Abstractive Summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 1906–1919. https://doi.org/10.18653/v1/2020.acl-main.173

[27] Rui Meng, Khushboo Thaker, Lei Zhang, Yue Dong, Xingdi Yuan, Tong Wang, and Daqing He. 2021. Bringing Structure into Summaries: a Faceted Summarization Dataset for Long Scientific Documents. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Association for Computational Linguistics, Online, 1080–1089. https://doi.org/10.18653/v1/2021.acl-short.137

[28] Julian Moreno-Schneider, Georg Rehm, Elena Montiel-Ponsoda, Víctor Rodriguez-Doncel, Artem Revenko, Sotirios Karampatakis, Maria Khvalchik, Christian Sageder, Jorge Gracia, and Filippo Maganza. 2020. Orchestrating NLP Services for the Legal Domain. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, 2332–2340. https://aclanthology.org/2020.lrec-1.284

[29] Gianluca Moro and Luca Ragazzi. 2022. Semantic Self-Segmentation for Abstractive Summarization of Long Documents in Low-Resource Regimes. *Proceedings of the AAAI Conference on Artificial Intelligence* 36, 10 (Jun. 2022), 11085–11093. https://doi.org/10.1609/aaai.v36i10.21357

[30] Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, Caglar Gulcehre, and Bing Xiang. 2016. Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond. In *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*, Stefan Riezler and Yoav Goldberg (Eds.). Association for Computational Linguistics, Berlin, Germany, 280–290. https://doi.org/10.18653/v1/K16-1028

[31] Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018. Don't Give Me the Details, Just the Summary! Topic-Aware Convolutional Neural Networks for Extreme Summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Brussels, Belgium, 1797–1807. https://doi.org/10.18653/v1/D18-1206

[32] Laura Nguyen, Thomas Scialom, Benjamin Piwowarski, and Jacopo Staiano. 2023. LoRaLay: A Multilingual and Multimodal Dataset for Long Range and Layout-Aware Summarization. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, Dubrovnik, Croatia, 636–651. https://doi.org/10.18653/v1/2023.eacl-main.46

[33] Vedant Parikh, Vidit Mathur, Parth Mehta, Namita Mittal, and Prasenjit Majumder. 2021. Lawsum: A weakly supervised approach for indian legal document summarization. *arXiv preprint arXiv:2110.01188* (2021).

[34] Abhay Shukla, Paheli Bhattacharya, Soham Poddar, Rajdeep Mukherjee, Kripabandhu Ghosh, Pawan Goyal, and Saptarshi Ghosh. 2022. Legal Case Document Summarization: Extractive and Abstractive Methods and their Evaluation. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Online only, 1048–1064. https://aclanthology.org/2022.aacl-main.77

[35] Nádia F. F. da Silva, Marília Costa R. Silva, Fabíola S. F. Pereira, João Pedro M. Tarrega, João Vitor P. Beinotti, Márcio Fonseca, Francisco Edmundo de Andrade, and André C. P. de L. F. de Carvalho. 2021. Evaluating Topic Models in Portuguese Political Comments About Bills from Brazil's Chamber of Deputies. In *Intelligent*

Systems: 10th Brazilian Conference, BRACIS 2021, Virtual Event, November 29 – December 3, 2021, Proceedings, Part II. Springer-Verlag, Berlin, Heidelberg, 104–120. https://doi.org/10.1007/978-3-030-91699-2_8

[36] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv:2307.09288 [cs.CL]

[37] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. BERTScore: Evaluating Text Generation with BERT. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net. https://openreview.net/forum?id=SkeHuCVFDr

[38] Tianyi Zhang, Faisal Ladhak, Esin Durmus, Percy Liang, Kathleen McKeown, and Tatsunori B. Hashimoto. 2024. Benchmarking Large Language Models for News Summarization. *Transactions of the Association for Computational Linguistics* 12 (01 2024), 39–57. https://doi.org/10.1162/tacl_a_00632 arXiv:https://direct.mit.edu/tacl/article-pdf/doi/10.1162/tacl_a_00632/2325685/tacl_a_00632.pdf

[39] Zheng Zhao, Shay B. Cohen, and Bonnie Webber. 2020. Reducing Quantity Hallucinations in Abstractive Summarization. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics, Online, 2237–2249. https://doi.org/10.18653/v1/2020.findings-emnlp.203

[40] Haoxi Zhong, Chaojun Xiao, Cunchao Tu, Tianyang Zhang, Zhiyuan Liu, and Maosong Sun. 2020. How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 5218–5230. https://doi.org/10.18653/v1/2020.acl-main.466

[41] Chenguang Zhu, Ziyi Yang, Robert Gmyr, Michael Zeng, and Xuedong Huang. 2021. Leveraging Lead Bias for Zero-shot Abstractive News Summarization. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, Canada) *(SIGIR '21)*. Association for Computing Machinery, New York, NY, USA, 1462–1471. https://doi.org/10.1145/3404835.3462846

A COPYRIGHT INFORMATION

In accordance with the provisions of the Indian Copyright Act, 1957, it is affirmed that the judicial pronouncements are readily accessible through the website https://indiankanoon.org/ by conducting a search using the name of the specific case. It should be noted that the headnotes (called in legalese) or summaries of these judicial pronouncements are protected under the Indian Copyright Act, 1957, with copyright belonging to Copyright © 2016 Patiala Law House (https://patialalawhouse.blogspot.com/).

Furthermore, this dataset's license is restricted to specific purposes such as conducting academic or educational research or study. It should be duly acknowledged that the utilization of the judicial pronouncements from the aforementioned website is carried out within the confines of the license provided, and thus does not infringe upon the provisions set forth by the copyright act.

B DATA AND SUMMARY SAMPLES

We provide a sample of our dataset in Table 4. Each summary paragraph starts with a supporting legislation and usually ends with references to relevant paragraphs from the source documents (shown in blue color). We also provide samples of generated summary from Longformer, Llama 2, and FactorSum in Table 5 and Table 6.

C HUMAN EVALUATION GUIDELINES

In order to assess the quality of summaries written for legal judgments, we conducted a human evaluation study. We use system generated summaries from FactorSum (with paragraph guidance), Longformer (with 4096 input tokens), and Llama 2-chat (70B) for the study. The purpose of this study was to gather subjective assessments from human evaluators based on specific guidelines. The guidelines were designed to evaluate the relevance, consistency, fluency, and coherence of the output summaries.

We asked human evaluators to find the answers to two questions for each summary pair which included summaries generated by FactorSum, Longformer, and Llama 2: Initially, we tasked human evaluators with selecting the superior summary to replace a technical judgment abstract. Subsequently, the second question pertained to identifying the best summary overall, taking into account factors such as informativeness, fluency, and more. We defined the following evaluation criteria for human evaluators for the second question and instructed them to select the best summary based on these definitions:

(1) **Relevance:** The rating measures how well the output summary captures the key points of the Judgment. Consider whether all and only the important aspects are contained in the output summary.
(2) **Consistency:** The rating measures whether the facts in the output summary are consistent with the facts in the original Judgment. Consider whether the output summary reproduces all facts accurately and does not include untrue information.
(3) **Fluency:** This rating measures the quality of individual sentences, whether they are well-written and grammatically correct. Consider the quality of individual sentences.
(4) **Coherence:** The rating measures the quality of all sentences collectively, and how well they fit together and sound natural. Consider the quality of the output summary as a whole.
(5) **Informativeness:** The rating measures whether the summary abstract encompasses all the essential details contained within the judgment.

We provided the following instructions to the human evaluators:

(1) Carefully read the Judgment and be aware of the information it contains.
(2) Read the three provided generated summaries.
(3) Pick the best replacement for the reference legal summary.
(4) Pick the best output summary on the five dimensions (Relevance, Consistency, Fluency, Coherence, Informativeness).
(5) Consider the definitions provided for each criterion while rating the output summary.

Table 4: Sample abstract and Judgment from the CivilSum test set (ID = 716).

For Respondent No. 3. :- R.K. Malik, Advocate. A. Haryana Labour Department (Group A and Group B) Rules, 1987, Rules 9 and 7 - It is noted that the existing rules have been repealed and the Draft Service Rules framed and approved by Public Service Commission, but the draft rules have not been notified in Gazette and thus, cannot be considered as executive instructions. 1985(1) SLR 41, relied upon. **[Paras 7 and 8]**

B. Haryana Labour Department (Group A and Group B) Rules, 1987, Rules 7 and 9 - In relation to the constitutional validity of Article 16, seniority and acting promotion granted to the petitioner, it was established that the petitioner's promotion was regularised from 6.10.1986, but with no back salary. However, the respondent was appointed to the post with effect from 24.2.1984 and appointment regularised by Public Service Commission with effect from 11.1.1986, thereby proving that the respondent was senior to the petitioner. **[Paras 7 and 8]**

N.K. Kapoor, J. - The petitioner has sought issuance of writ of certiorari quashing promotion order Annexure P which Mange Ram stated to be junior to the petitioner has been promoted without considering the claim of the petitioner.

2. The petitioner joined the Labour Department as a Clerk in the year 1961 and after getting few promotions is presently working as Statistical Officer which is Class-II post. [...] Provided that their inter se seniority for purpose of consideration for promotion shall be on the basis of continuous service on the post or (ii) by direct, or (iii) by transfer or deputation of an officer already in the service of any State Government, or the Government of India."

3. It is according to Rule 9 that the post of Deputy Labour Commissioner is to be filled up from amongst the Labour Officer-cum-Conciliation Officer, statistical Officer, and Welfare Officer (Women). [...] In any case, even if the Rules have not been notified, the same can be taken as executive instructions.

⋮

6. The matter was heard on 4.10.1993. In view of the submissions made, it was though appropriate that a direction be given to the State to file a detailed reply specifically indicating whether Draft Rules have been approved by the Public Service Commission and given effect thereto or it is the stand of the State that the post of Statistical Officers are not at all to be considered for the purpose of promotion to the posts of Deputy Labour Commissioners. [...] It is a settled law that Draft Rules are no Rules in the presence of notified Rules. It is also clarified that Class I and II are redesignated as Group A and B respectively."

7. We have heard learned counsel for the parties and perused the relevant material referred to during the course of their submissions. The petitioner has challenged the appointment of respondent No. 3 to the post of Deputy Labour Commissioner primarily on two counts : (i) that while promoting the respondent No. 3, claim of the petitioner has not been considered; (ii) that Punjab Labour Service (Class I and II) Rules, 1955, having already been repealed and the Draft Rules approved by the Haryana Public Service Commission, entitle the petitioner for consideration in terms of Rule 9; and (iii) the Draft Rules having been approved by the Public Service Commission even if no notified can be taken as executive instructions and if construed so, the petitioner's case has merit.

8. Admittedly, under 1955 Rules there was no provision with the Labour Department for promotion to the post of Statistical Officer. It is on the advice of the Haryana Public Service Commission that avenue was created subject to the persons possessing the requisite qualifications and experience. [...] Even otherwise, as per decision in Jitender Kumar Verma's case (supra), Draft Rules cannot be equated with administrative instructions. Thus, finding no merit in any of the contentions raised by the counsel for the petitioner, the civil writ petition is dismissed. Parties to hear their own costs.

Table 5: Sample abstract and generated summaries from the CivilSum test set (ID = 2). FactorSum summaries includes the same sentences in different order. References to paragraphs in the source document are shown in blue color.

Reference abstract (Tokens: 52)
Constitution of India, Article 300-A - Retirement Pension - Denial - The petitioner was previously employed before entering into the University and was receiving pension from their previous employer, therefore, the petitioner is eligible for pension from the University. **[Paras 4 and 12]**
Longformer summary Tokens: 142; ROUGE-1: 34.73; ROUGE-2: 15.76; ROUGE-L: 28.74
constitution of india, articles 14 and 300-a - university pension regulations, 1990, regulations 2 and 3(2) - pension - re-employment - discrimination - petitioner, who was appointed as a security guard in the university on 8.1.80 and had his services regularised and regular pay scale granted, sought pensionary benefits. however, the respondent university argued that the petitioner was not covered under the definition of re-employed as he was already receiving pension from his previous employer. the court held that the respondent's action was violative of articles 14, as the judgment of the single judge affirmed the entitlement of the petitioner to pension by the university. consequently, the university was directed to grant all retiral benefits from the date of superannuation. **[paras 9 to 14]**
Llama 2-chat summary Tokens: 112; ROUGE-1: 20.65; ROUGE-2: 5.23; ROUGE-L: 16.77
The five connected civil writ petitions were disposed of by the court, with the judgment in favor of the petitioner. The petitioner, who was a security guard at the University of Rajasthan, had been denied pensionary benefits on the ground that he was a re-employed pensioner. The court held that the petitioner was entitled to all retiral benefits as per the provisions of the 1990 Regulations, and the arrears of the retiral benefits were to be paid to the petitioner within 60 days. The court also directed that in case of non-payment of arrears, the petitioner would be entitled to interest on the due amount at the rate of 12% per annum.
FactorSum - Tail content guidance Tokens: 154; ROUGE-1: 31.77; ROUGE-2: 13.10; ROUGE-L: 23.53
Constitution of India, Articles 14 and 16 - Pension Regulations for University of Rajasthan, 1990, Regulation 3(b) - Retiral benefits - Re-employment - The petitioner, who was a regular employee of the respondent University, retired on attaining the age of superannuation and was governed by the Pension Regulations which were applicable to University Employees as per Register No. 2 subject to the provisions of proviso (a) to (d) and that the petitioner was not falling under any of the categories of the proviso - Therefore, the petitioner is entitled to all retiral benefits as per the 1990 Regulations. Constitution of India, Articles 14, 16 and 226 - Pension - Re-employment - The petitioner, who retired as a Security Guard in the University, is entitled to all retiral benefits from the date of his superannuation.
FactorSum - Lead content guidance Tokens: 127; ROUGE-1: 34.90; ROUGE-2: 10.88; ROUGE-L: 22.82
Constitution of India, Articles 14, 16 and 226 - Pension - Re-employment - The petitioner, who retired as a Security Guard in the University, is entitled to all retiral benefits from the date of his superannuation. Constitution of India, Articles 14 and 16 - Rajasthan Civil Services (Pension) Rules, 1990, Rule 3 - Retiral benefits - Re-employment - The petitioner, who was appointed as a Security Guard on 8.1.80, retired on 30.4.94 and was governed by the Pension Regulations applicable to University Employees as per Register No. 2 subject to the provisions of proviso (a) to (d) and was not falling under any of the categories of the proviso.

Table 6: Sample abstract and generated summaries from the CivilSum test set (ID = 8). References to paragraphs in the source document are shown in blue color.

Reference abstract (Tokens: 94)
Under the Uttar Pradesh Public Services (Reservation for Scheduled Castes, Scheduled Tribes and other Backward Classes) Act, 1994, Rule 3(4) for Recruitment Eligibility, the State Engineering Services (Special Recruitment) Examination is conducted for the recruitment of SC and ST candidates.', 'In this process, it is reasonable to fix a minimum aggregate of 40% marks as the standard for evaluating candidate suitability.', 'It was established that the petitioner did not obtain the required minimum marks and was subsequently not called for an interview, therefore, there is no justification for intervention. **[Paras 7, 8, and 9]**
Longformer summary Tokens: 98; ROUGE-1: 58.95; ROUGE-2: 31.91; ROUGE-L: 43.16
constitution of india, article 16 - u.p. public services (reservation for scheduled castes, scheduled tribes and other backward classes) act, 1994, section 3(4) - interview - suitability - petitioners who appeared in state engineering services (special recruitment) examination, 1995 conducted by respondent no. 2 and were not called for interview - it was held that the petitioners were unsuitable for interview as they could not secure the 40% minimum marks required for judging their suitability. however, the criterion was not deemed arbitrary and did not violate any laws. hence, the petition was dismissed. **[paras 7, 8 and 9]**
Llama 2-chat summary Tokens: 109; ROUGE-1: 39.81; ROUGE-2: 16.67; ROUGE-L: 26.21
The petitioners, who were candidates in a special recruitment examination for S.C. and S.T. candidates, filed a writ petition against the respondents, claiming that they were not called for an interview due to the respondents' illegal fixing of 40% minimum marks in aggregate. The Court found that the requirement of securing 40% minimum marks in aggregate was not arbitrary and did not suffer from any legal infirmity. Therefore, the Court dismissed the writ petition, finding that the respondents did not commit any error in not calling the petitioners for an interview. The petitioners did not obtain the required 40% minimum marks in aggregate, making them ineligible for the interview.
FactorSum - Tail content guidance Tokens: 167; ROUGE-1: 47.49; ROUGE-2: 17.51; ROUGE-L: 21.0
T. candidates - The requirement of 40% marks in aggregate for assessing the suitability of a candidate for being called for interview is not illegal . **[Para 5]** Constitution of India, Articles 14, 16 and 226 - Appointment - Selection - Criteria of selection - The minimum standard for assessing the suitability of candidates is a minimum of 40% marks in aggregate - The fixing of the 40% minimum marks by respondent No. 2 does not suffer from any legal infirmity warranting interference by the Court . **[Para 3]** Constitution of India, Articles 14, 16 and 226 - Selection - Interview - Selection for the post of S . C. and S .T. candidates - Petitioners appeared in State Engineering Services (Special Recruitment) Examination, 1995 conducted by respondent No. 2 for recruitment of SC and ST candidates. T. candidates requires a minimum of 40% marks in aggregate.
FactorSum - Lead content guidance Tokens: 155; ROUGE-1: 54.88; ROUGE-2: 27.23; ROUGE-L: 26.05
Constitution of India, Articles 14, 16 and 226 - Selection - Interview - Selection for the post of S . C. and S .T. candidates - Petitioners appeared in State Engineering Services (Special Recruitment) Examination, 1995 conducted by respondent No. 2 for recruitment of SC and ST candidates. T. candidates - The requirement of 40% marks in aggregate for assessing the suitability of a candidate for being called for interview is not illegal . **[Para 5]** P. Public Services (Reservation for Scheduled Castes, Scheduled Tribes and other Backward Classes) Act, 1994, Section 4 - Recruitment - Interview - Post of Lecturer - Petitioners, who were appointed as Lecturers, challenged the appointment of Respondent No. 2 as Lecturer after obtaining 40% marks in aggregate. Constitution of India, Articles 14 and 16 - U .