NIT-Surat@L-Summ: A Semantic Retrieval-Based Framework for Summarizing Indian Judicial Documents

Nita Jadav[1]   Ashok Urlana[2]   Pruthwik Mishra[1]
[1]SVNIT Surat, India
[2]TCS Research, Hyderabad, India
ds24ai001@aid.svnit.ac.in, ashok.urlana@tcs.com
pruthwikmishra@aid.svnit.ac.in

Abstract

The shared task of Legal Summarization (L-Summ) focuses on generating abstractive summaries for the Indian court judgments in English. This task presents unique challenges in producing fluent, relevant, and legally appropriate summaries given voluminous judgment texts. We experiment with different sequence-to-sequence models and present a comprehensive comparative study of their performance. We also evaluate various Large Language Models (LLM) with zero-shot settings for testing their summarization capabilities. Our best performing model is fine-tuned on a pre-trained legal summarization model where relevant passages are identified using the maximum marginal relevance(MMR) technique. Our findings highlight that retrieval-augmented fine-tuning is an effective approach for generating precise and concise legal summaries. We obtained a rank of 5th overall[1].

1 Introduction

One of the main impediments to India's growth is its pending cases, in which cases remain unsolved due to the shortage of competent manpower (Katju, 2019). According to the national judicial data grid[2], 82 percent of criminal cases are pending for more than 9 years in various district courts of India. The facts show that the Indian legal system alone has more than 44 million pending cases. Most of the processes in the legal system still remain manual with very few examples of transition to fully digital court rooms.

To assist the legal community, several works such as the series of shared tasks on Artificial Intelligence for Legal Assistance (AILA), semantic segmentation of judicial text using rhetorical roles

(Malik et al., 2022), summarization (Parikha et al., 2021; Datta et al., 2023), machine translation (Kocmi et al., 2025), language understanding models specific to the Indian legal system (Paul et al., 2023), and legal reasoning models (Joshi et al., 2024) have been proposed. In a similar vein, the Legal Summarization shared task under the JustNLP workshop[3] attempts to develop legal summarization systems capable of handling domain-specific legal knowledge, legal understanding, legal reasoning, and coherent abstraction for Indian judgments in English.

Reasoning over large specialized legal judgments and generating abstractive summaries make this task even more challenging. As we are dealing with large documents, semantic segmentation of the document into passages becomes extremely important. Similarly, locating the relevant passages corresponding to a summary is also equally crucial for the overall performance of any summarization model especially in this domain.

Taking all these complexities into account, our contributions in this paper are threefold.

1. Semantic Segmentation of documents into meaningful passages using different LLMs

2. Relevant passage retrieval corresponding to a summary using different techniques with dense embeddings

3. Fine-tuning domain-specific summarization models with comprehensive comparison

2 Related Work

Although document summarization remains a very active and competitive research area in general and other technical domains, it has very sparse representation in the legal domain. Publicly available

---

[1]The demonstration of the approaches used in the paper can be found in https://github.com/nitajadav8/Legal_TextSumm

[2]https://njdg.ecourts.gov.in/njdg_v3/

[3]https://exploration-lab.github.io/JUST-NLP/

Metrics	Train		Validation		Test	
	Judgment	Summary	Judgment	Summary	Judgment	Summary
Number of Samples	1200	1200	200	200	400	400
(Min, Max) Words	(169, 149087)	(29, 2404)	(336, 92322)	(106, 1552)	(183, 155282)	(44, 1934)
Average Words	8294	606	6815	603	7920	621
(Min, Max) Sentences	(1, 3129)	(1, 18)	(2, 813)	(1, 13)	(1, 2228)	(1, 18)
Average Sentences	79	3	61	4	78	3
Compression Ratio	0.74		0.70		0.70	
Density	34.03		34.74		32.25	
Abstractivity	0.93		0.93		0.94	

Table 1: Dataset Statistics for Train, Validation, Test variants of L-Summ.

legal datasets are very few even for highly resource rich languages such as English. The majority of the available data span through sub-domains such as patents (Sharma et al., 2019) and contracts (Manor and Li, 2019), whereas actual court rulings (de Vargas Feijó and Moreira, 2018) are scarce.

To overcome the lack of rich resources, there have been few attempts in the Indian context to generate high-quality summarization datasets (Parikh et al., 2021). In-Abs (Shukla et al., 2022) consists of 7000 abstractive summaries or headnotes of the Indian Supreme Court judgments using reliable legal resources. IL-TUR (Joshi et al., 2024) presents a comprehensive benchmark and strong baseline models for the understanding and reasoning over the Indian legal texts. Their best performing summarization model is fine-tuned on the legal Pegasus (Sairam, 2021) model. The legal Pegasus model is a fine-tuned version of the Pegasus (Zhang et al., 2020) model on the US securities litigation dataset.

Most of the research in the domain of legal summarization is limited to only English. MILDSum (Datta et al., 2023) is an attempt to provide understandable summaries in native languages through cross-lingual summarization for complex court proceedings written in English. With the advent of LLMs and improved performance in the majority of tasks, the research in the direction of building domain-specific LLMs has gained traction. Domain adaptive pre-training and converting raw corpora into reading comprehension tasks (Cheng et al., 2023) enhance performance across various tasks in different domains, including law.

3 L-Sum Dataset Description

The dataset used in the task is shared by the organizers. Each entry in the dataset consists of a judgment

and a corresponding concise summary in English. To perform in-depth data analysis, we adopted several metrics from Urlana et al. (2022, 2023b,a) and reported the number of words, the size of the vocabulary, compression-ratio and density of judgment to summary. The detailed data analysis statistics of the dataset are listed in Table 1.

4 Data processing

4.1 Preprocessing

The given documents in the L-Summ dataset do not have any annotations related to rhetorical roles (Malik et al., 2022) that could be leveraged for their semantic segmentation. Moreover, the document has additional noise of broken or incomplete sentences, misspelled words, non-ASCII characters, and title statements without proper structure. Considering this complexity, we incorporate different pre-processing techniques to remove the noise and divide each judgment into semantically coherent passages or paragraphs to harness the effectiveness of domain-specific summarization models that are explained in the following sections. All non-ASCII characters, special symbols, and redundant spaces are removed to ensure text uniformity.

4.2 Normalization

A judgment can have a wide variety of abbreviations from the legal domain as well as short sentences about sections, articles, writ numbers, and other legal information. It also contains unstructured sentences with improper spacing. Normalizing these sentences is essential for preparing the data for the models. We automate the normalization process using GPT-4 (OpenAI et al., 2024) and the corresponding prompt mentioned in Appendix A.1.

1. *Expansion of Abbreviations:* Common abbre-

	Model	R-2	R-L	BLEU
Zero-shot	Qwen-32b (Yang et al., 2025)	14.7	18.7	6.7
	gpt-oss-120b (Agarwal et al., 2025)	17.8	21.8	9.1
	Llama-4 17b (AI, 2025)	22.0	24.0	9.8
	Legal-Pegasus (Sairam, 2021)	17.0	20.7	6.2
	LongFormer (Beltagy et al., 2020)	21.6	21.9	15.4
	Llama 3.1 8b (Dubey et al., 2024)	21.1	23.8	10.5
	Llama 3 70b (Dubey et al., 2024)	21.4	23.6	7.9
Fine-tuning	Legal-Pegasus-DPR (Karpukhin et al., 2020)	19.4	20.2	9.9
	LongFormer-MCS (Shukla et al., 2022)	22.3	23.2	14.7
	Legal-Pegasus-MCS (Shukla et al., 2022)	22.9	23.0	9.6
	LongFormer-MMR (Xie and Liu, 2008)	21.7	21.9	15.6
	Legal-Pegasus-MMR (Xie and Liu, 2008)	**24.4**	**24.7**	**16.8**

Table 2: Various models' performance comparison on Validation Dataset.

viations such as *dept.*, *Addl.*, *secy.*, and *Lko* are expanded into their full forms—*department*, *Additional*, *Secretary* and *Lucknow*—to enhance text clarity and uniformity.

2. Coreference resolution of legal entities: Legal judgments often contain implicit or missing references to entities such as courts. In this step, such references are made explicit.

3. Completion of Incomplete Sentences: Fragmented or unstructured sentences are transformed into grammatically complete forms for better readability.

4.3 Passage Retrieval

Passage retrieval is very important in summarization tasks when we are dealing with large documents. It helps identify and extract relevant content that is semantically aligned with the summary, improving summary coherence and informativeness. We also perform experiments without considering the semantic relevance. In both experimental setups, the judgment document is divided into passages of a maximum of 1024 tokens, preserving the long sentence boundaries. The passage length is increased (maximum 2048 tokens) for models such as Longformer (Beltagy et al., 2020), as they are capable of handling large documents. For fine-tuning, three approaches are used to retrieve passages from a long judgment and its corresponding summary.

4.3.1 Dense Passage Retrieval (DPR)

The DPR approach (Karpukhin et al., 2020; Faizullah et al., 2024) is based on the idea of covering all summary sentences to identify their contributing passages from the judgment. The judgment is divided into passages with a maximum of 1024 tokens (to align with the model's capacity), covering whole sentences only. The cosine similarities between the embeddings of summary sentences and the embeddings of passages are calculated to

Approach	R-2	R-L	BLEU
Legal-Pegasus-MCS	25.56	23.82	9.24
Legal-Pegasus-MMR	25.38	24.75	11.53
Legal-Pegasus-MMR-epoch:8	25.9	24.95	13.06

Table 3: Various methods performance comparison on Test Dataset.

find the most relevant passage for the summary sentence. We perform the experiment by identifying the three most relevant passages for each summary sentence. The training data is formed by combining all three relevant passages as the input and the corresponding summary sentence as the target.

4.3.2 Maximum Cosine Similarity (MCS)

In the MCS approach (Shukla et al., 2022), passages are ranked using maximum cosine similarity (MCS). Both documents and summaries are sentence-segmented and represented with SBERT embeddings (Reimers and Gurevych, 2019). For each document sentence, the most similar summary sentence is selected, and the process continues until reaching a 1024-token limit. The final training set thus contains passages capped by length and paired with their most relevant summary chunks.

4.3.3 Maximum Marginal Relevance (MMR)

To capture both semantic similarity and relevance between a judgment and its summary, we utilize the MMR (Xie and Liu, 2008; Zhong et al., 2019) approach. MMR balances two competing objectives: selecting passages that are *highly relevant* to the summary sentence while also maintaining *diversity* among the selected passages.

Let s denote the embedding of a summary sentence, and let $C = \{c_1, c_2, \ldots, c_n\}$ denote the set of embeddings of n candidate passages obtained from a judgment document. The relevance of each passage to the summary sentence is computed using the cosine similarity as follows:

$$\text{Rel}_i = \text{cosine\_sim}(s, c_i), \quad \forall i \in \{1, 2, \ldots, n\} \tag{1}$$

The pairwise similarity between two passages c_i and c_j is given by:

$$\text{Sim}_{ij} = \text{cosine\_sim}(c_i, c_j) \tag{2}$$

Step 1: Initial Selection. The first passage c_i to be selected is the one that is the most relevant to the summary sentence:

$$S_1 = \arg\max_i(\text{Rel}_i) \tag{3}$$

Step 2: Iterative Selection using MMR. For the remaining candidates, the MMR score is computed as:

$$\text{MMR\_score}(c_i) = \lambda \cdot \text{Rel}_i - (1 - \lambda) \cdot \max_{c_j \in S} \text{Sim}_{ij} \quad (4)$$

where S is the set of passages already selected, and $\lambda \in (0, 1)$ controls the trade-off between relevance and diversity. A higher λ value emphasizes *relevance* (passages similar to the summary sentence). A lower λ value emphasizes *diversity* (passages dissimilar to previously selected ones).

Step 3: Passage Update. The next passage to include is the one with the maximum MMR score:

$$S_i = \arg\max_{c_i \notin S}[\text{MMR\_score}(c_i)] \quad (5)$$

$$S \leftarrow S \cup \{S_i\} \quad (6)$$

Implementation Details. In our experiments, a maximum of two semantically relevant passages were selected for each summary sentence using $\lambda = 0.7$. A higher value of λ (> 0.5) ensures that retrieved passages are highly relevant to the summary sentence while maintaining moderate diversity.

Each selected passage is associated with its corresponding summary sentence in a mapping dictionary, since a passage may be relevant to multiple summary sentences. To retrieve semantically coherent passages, the judgment document is segmented into passages with a maximum token limit of 1024, ensuring complete sentences and a cosine similarity threshold above 0.2. For each retrieved passage, the corresponding summary chunk is then obtained using the constructed mapping dictionary.

5 Model Description

The summary generation from the preprocessed judgment is carried out using several popular and open-source models. All the models have demonstrated better efficiency in the general summarization task. The most accurate domain-specific models, like Legal-pegasus and Longformer, were also used to infer the result. The experiments were performed with zero-shot and fine-tuning models.

5.1 Zero-Shot Experiments

We utilize Qwen3 (Yang et al., 2025), gpt-oss (Agarwal et al., 2025), Llama 3 (Dubey et al.,

2024), Llama 4 (AI, 2025), Longformer (Beltagy et al., 2020), and Legal-Pegasus (Sairam, 2021) to perform zero-shot experiments and obtain a summary from a preprocessed judgment. For LLM-based summary generation, Groq API[4] is used, which costs us a total of 26.32 dollars (USD) including normalization of raw judgments.

5.2 Fine-tuning Encoder-Decoder Models

We fine-tune two encoder-decoder (Sutskever et al., 2014) based models. The selection of models is based on the type of data on which those models were trained and the context length. The Longformer model is chosen because it can handle long documents similar to legal judgments. We select Legal-Pegasus as a base model, as it is fine-tuned on legal texts. The details of the hyper-parameters are given in Table 7. Fine-tuning is performed on the datasets created after passing the original documents through the pre-processing and normalization techniques, followed by the three passage retrieval techniques defined in Section 4.3.

6 Results and Evaluation

Each generated summary is evaluated on the basis of ROUGE-2, ROUGE-L (Lin, 2004) and BLEU (Papineni et al., 2002) scores. Table 2 and Table 3 present the results on the validation data and the test data, respectively. The approach of retrieving passages based on the MMR technique with the fine-tuned Legal-Pegasus performs the best. The efficiency of the MMR technique can be attributed to its ability to retrieve passages relevant to the summary with diversity among them. To improve the performance of the model, we also try different number of epochs for fine-tuning and the best performing model is fine-tuned for 8 epochs. The performance of LLMs to generate summaries from legal documents with zero-shot prompting is also compared. We also compare the results of different passage retrieval approaches. Table 9 shows the generated summaries for a sample document in Table 8 by models using different passage retrieval techniques.

Conclusion

As a part of the L-Summ shared task, we fine-tune domain-specific models using the semantic passage retrieval framework, considering the reference summary. Semantically relevant passage retrieval from

---

large documents with domain knowledge plays a crucial role in improving the performance of summarization models. As a successor to this work, we are exploring techniques to fine-tune Legal LLMs.

7 Limitation

We analyze our results using the evaluation metrics recommended by the shared task organizers; however, the scope of this evaluation is limited and can be further expanded. Currently, our experiments are restricted to LLM-based zero-shot settings, and we have not performed any fine-tuning of LLMs due to computational constraints.

References

Sandhini Agarwal, Lama Ahmad, Jason Ai, Sam Altman, Andy Applebaum, Edwin Arbus, Rahul K Arora, Yu Bai, Bowen Baker, Haiming Bao, and 1 others. 2025. gpt-oss-120b & gpt-oss-20b model card. *arXiv preprint arXiv:2508.10925*.

Meta AI. 2025. Llama 4 models.

Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. Longformer: The long-document transformer. *Preprint*, arXiv:2004.05150.

Daixuan Cheng, Shaohan Huang, and Furu Wei. 2023. Adapting large language models via reading comprehension. In *The Twelfth International Conference on Learning Representations*.

Debtanu Datta, Shubham Soni, Rajdeep Mukherjee, and Saptarshi Ghosh. 2023. MILDSum: A novel benchmark dataset for multilingual summarization of Indian legal case judgments. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 5291–5302, Singapore. Association for Computational Linguistics.

Diego de Vargas Feijó and Viviane Pereira Moreira. 2018. Rulingbr: A summarization dataset for legal texts. In *International Conference on Computational Processing of the Portuguese Language*, pages 255–264. Springer.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, and 1 others. 2024. The llama 3 herd of models. *arXiv e-prints*, pages arXiv–2407.

Abdur Rahman Bin Mohammed Faizullah, Ashok Urlana, and Rahul Mishra. 2024. Limgen: Probing the llms for generating suggestive limitations of research papers. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 106–124. Springer.

Abhinav Joshi, Shounak Paul, Akshat Sharma, Pawan Goyal, Saptarshi Ghosh, and Ashutosh Modi. 2024. Il-

tur: Benchmark for indian legal text understanding and reasoning. *arXiv preprint arXiv:2407.05399*.

Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. 2020. Dense passage retrieval for open-domain question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6769–6781, Online. Association for Computational Linguistics.

Justice Markandey Katju. 2019. Backlog of cases crippling judiciary.

Tom Kocmi, Ekaterina Artemova, Eleftherios Avramidis, Rachel Bawden, Ondrej Bojar, Konstantin Dranch, Anton Dvorkovich, Sergey Dukanov, Mark Fishel, Markus Freitag, and 1 others. 2025. Findings of the wmt25 general machine translation shared task: Time to stop evaluating on easy test sets. In *Proceedings of the Tenth Conference on Machine Translation, China. Association for Computational Linguistics*.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

Vijit Malik, Rishabh Sanjay, Shouvik Kumar Guha, Angshuman Hazarika, Shubham Nigam, Arnab Bhattacharya, and Ashutosh Modi. 2022. Semantic segmentation of legal documents via rhetorical roles. In *Proceedings of the Natural Legal Language Processing Workshop*, volume 2022, pages 153–171.

Laura Manor and Junyi Jessy Li. 2019. Plain English summarization of contracts. In *Proceedings of the Natural Legal Language Processing Workshop 2019*, pages 1–11, Minneapolis, Minnesota. Association for Computational Linguistics.

OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, and 262 others. 2024. Gpt-4 technical report. *Preprint*, arXiv:2303.08774.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Vedant Parikh, Upal Bhattacharya, Parth Mehta, Ayan Bandyopadhyay, Paheli Bhattacharya, Kripabandhu Ghosh, Saptarshi Ghosh, Arindam Pal, Arnab Bhattacharya, and Prasenjit Majumder. 2021. Overview of the third shared task on artificial intelligence for legal assistance at fire 2021. In *Fire (working notes)*, pages 517–526.

Vedant Parikha, Vidit Mathurc, Parth Mehtab, Namita Mittalc, and Prasenjit Majumdera. 2021. Lawsum: A

weakly supervised approach for indian legal document summarization. *arXiv preprint arXiv:2110.01188*.

Shounak Paul, Arpan Mandal, Pawan Goyal, and Saptarshi Ghosh. 2023. Pre-trained language models for the legal domain: A case study on indian law. In *Proceedings of 19th International Conference on Artificial Intelligence and Law - ICAIL 2023*.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Naren Sairam. 2021. Legal pegasus model.

Eva Sharma, Chen Li, and Lu Wang. 2019. BIG-PATENT: A large-scale dataset for abstractive and coherent summarization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2204–2213, Florence, Italy. Association for Computational Linguistics.

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)*, pages 1048–1064, Online only. Association for Computational Linguistics.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27.

Ashok Urlana, Sahil Manoj Bhatt, Nirmal Surange, and Manish Shrivastava. 2023a. Indian language summarization using pretrained sequence-to-sequence models. In *Fire*.

Ashok Urlana, Pinzhen Chen, Zheng Zhao, Shay Cohen, Manish Shrivastava, and Barry Haddow. 2023b. PMIndiaSum: Multilingual and cross-lingual headline summarization for languages in India. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 11606–11628, Singapore. Association for Computational Linguistics.

Ashok Urlana, Nirmal Surange, Pavan Baswani, Priyanka Ravva, and Manish Shrivastava. 2022. TeSum: Human-generated abstractive summarization corpus for Telugu. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5712–5722, Marseille, France. European Language Resources Association.

Shasha Xie and Yang Liu. 2008. Using corpus and knowledge-based similarity measure in maximum marginal relevance for meeting summarization. In *2008 IEEE international conference on acoustics, speech and signal processing*, pages 4985–4988. IEEE.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, and 1 others. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter Liu. 2020. PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 11328–11339. PMLR.

Linwu Zhong, Ziyi Zhong, Zinian Zhao, Siyuan Wang, Kevin D. Ashley, and Matthias Grabmair. 2019. Automatic summarization of legal decisions using iterative masking of predictive sentences. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, ICAIL '19, page 163–172, New York, NY, USA. Association for Computing Machinery.

A Appendix

A.1 Examples of Prompt

For the data normalization task, the prompt consists of instructions to expand abbreviations, standardize entity names, structuring broken sentences. We design prompt in such a way that it covers all requirements for normalizing judgment which can be used later for semantic passage retrieval. Here, we instruct the LLM to preserve the original meaning of sentences without paraphrasing and also avoid summarizing at this phase. Table 6 gives the prompt sample, we used for performing the normalization and generation parameter values in our experiment. The sample outcome of normalization is given in Table 4.

We design summarization prompts to generate an abstractive summary with clear instruction to use domain specific and fluent legal terms. We also constraint the summary length maximum of 500 tokens as per the rules of shared task. Our prompt example is given in Table 5 with generation parameters.

The LLM based experiments are performed using the Groq API [5] for high speed inference. The Groq API offers a free-tier usage limit of tokens per day per api call for selected large language models (LLM). To adhere to this constraint, the judgment texts were divided into smaller passages, ensuring that each segment remained within the maximum token limit which in our case was 50000 tokens. All the Groq inferences conducted by setting the temperature=0 to encourage the reproducibility. Considering the shared task deadline and the need for higher inference capacity, the Groq API key was subsequently upgraded to the developer tier to support extended model utilization.

Raw text:
Case :- WRIT - C No. - 11383 of 2023 Petitioner :- Syed Hamidul Bari Respondent :- State Of U.P. Thru. Addl. Chief/Prin. Secy. Housing And Urban Planning Deptt. Lko. And 4 Others
Normalized text:
Case: Writ Petition No. 11383 of 2023, Petitioner: Syed Hamidul Bari, Respondent: State of Uttar Pradesh through Additional Chief/Principal Secretary, Housing and Urban Planning Department, Lucknow, and four others.

Table 4: Example of the Raw text Vs. normalized text sample.

---

[5] https://console.groq.com

You are an abstractive text summarizer for legal document in the Indian legal domain. Clean the document by removing unwanted noisy data and frame sentences in clear legal-style English. Produce a single fluent abstractive summary containing Context and background of the case with the court name, Legal reasoning and Final decision/outcome made by court. Return the result as a summarized paragraph in human-readable form with a maximum of 500 words.
Document:**{doc}**

End-of-Prompt

Table 5: Prompt Template for zero-shot prompting.

A.2 Fine Tuning Parameters

In the fine-tuning phase, several key hyperparameters were configured to optimize model performance while maintaining computational efficiency. To address GPU memory constraints encountered during the fine-tuning of long-former models, the gradient accumulation steps were set to 4, allowing effective larger batch sizes without exceeding memory limits. For effective training, we consider 20% of train data as validation data. To increase our score on the best performing model, we finally fine-tune the legal pegasus model with eight epochs to infer the test results. All experiments are carried out on a single NVIDIA H-100 GPU with 94GB RAM. The complete set of selected hyperparameter values is summarized in Table 7.

Parameter Name	Legal-Pegasus	LongFormer
Epochs	4, 8	4
Train Batch	1	1
Evaluation Batch	1	1
Warmup Step	200	750
Gradient accumulation steps	1	4
Evaluation accumulation steps	1	1
Learning Rate	1e-4	1e-4

Table 7: Hyperparameters for Fine-tuning

Normalizing Prompt:

You are a text cleaning expert for Indian legal case documents. Take the following legal document text and split it into clean paragraphs based on following conditions.

1. Remove meaningless statements, 2. Expand abbreviations according to the Indian legal context, 3. Use full court names according to the Indian legal system instead of short forms, 4. Add full court name if not present, 5. Remove empty lines, 6. Remove numbering like "1.", "(i)", "(a)" from the beginning of lines, 7. Merge broken lines into proper sentences, 8. Remove Unicode or non-standard characters, 9. Do **not** rephrase, 10. Do **not** summarize, 11. Correct spelling of misspelled or noisy English words containing digits or non-alphabetic characters, 12. Form paragraphs containing at most 1000 tokens, keeping complete sentences without breaking them, 13. Return the output **only** as a JSON array of paragraphs.

Document: {doc}

End-of-Prompt

Table 6: Prompt template for GPT-4-based normalization of Indian legal documents.

Judgment: Petitioner :- Suresh Devi And Another Respondent :- State Of U P And 13 Others Counsel for Petitioner :- Jamil Ahamad Azmi,Anand Swaroop Gautam,Ashutosh Kumar Tiwari,Dharmendra Singh Counsel for Respondent :- G A Hon'ble Ashwani Kumar Mishra,J Hon'ble Deepak Verma,J This petition has been filed invoking jurisdiction of this Court under Article 226 of the Constitution of India with the allegation that petitioner's son has been murdered in police custody in the night intervening 11/12th December, 2020 while he was in police custody in Case Crime No 1181 of 2020, under Section 366 I P C , Police Station Khurja Nagar, District Bulandshahar A prayer has also been made to protect the life and liberty of the petitioners as their son was subjected to torture and murder as he had contracted inter caste marriage out of his own free will It is contended that the authorities have been most unfair in dealing with the grievance raised by the petitioners Attention of the Court has been invited to the provision contained in Section 176 Cr P C , sub Section (1)A whereof contemplates holding of a judicial enquiry where death is caused in the custody of the police It is also stated that neither any post mortem has been carried out nor the body has been burried and instead the police personnels have cremated the body contrary to all settled norms Learned AGA does not dispute the fact that a judicial enquiry was initiated with a request made to District Judge on 06 01 2021 and states that a report is still awaited In matters, where the allegation is with regard to custodial death, judicial enquiry under Section 176 (1)A cannot be allowed to drag for so long These are instances which have to be viewed with greatest sensitivity and concern We, therefore, direct the Registry to enquire from the District Judge, Bulandshahar as to when the enquiry report has been submitted in the matter and, in the event, such a report is not submitted, the explanation of the Judicial Officer in that regard shall be placed before us by the next date fixed as a period of more than one year has expired We hasten to add that in the event, such enquiry has not been concluded so far, the same shall be concluded most expeditiously by following the procedure in law List this matter, once again, on 27 01 2022 at 02:00 PM

Reference Summary: The Allahabad High Court recently observed that judicial inquiry in custodial death cases cannot be dragged on for long [Suresh Devi and Another v. State of Uttar Pradesh and Others]. A Division Bench of Justices Ashwani Kumar Mishra and Deepak Verma held, "In matters, where the allegation is with regard to custodial death, judicial enquiry under Section 176 (1)A cannot be allowed to drag for so long. These are instances which have to be viewed with greatest sensitivity and concern." The Court was hearing a plea alleging that the petitioner's son was murdered in police custody in December 2020. It was claimed that the son was tortured and killed while in police custody after he had contracted inter-caste marriage out of his own free will. The petitioner had also prayed for a direction to protect the life and liberty of the parents of the deceased. It was the case of the petitioner that the concerned authorities have been most unfair in dealing with their grievances, as no post-mortem was carried out on the body of the deceased, which was cremated contrary to all settled norms. Moreover, the Court's attention was drawn to the judicial enquiry initiated by the District Judge, Bulandshahr, on January 6, 2021. A report on the same was still awaited. While noting that more than one year has expired since the judicial enquiry was ordered, the Bench directed the High Court Registry to enquire from the District Judge, Bulandshahr as to when the enquiry report was submitted. In the event such a report is not submitted, the explanation of the judicial officer was directed to be placed before the Court. "We hasten to add that in the event, such enquiry has not been concluded so far, the same shall be concluded most expeditiously by following the procedure in law," the Court added. The matter was listed on January 27 for further hearing. Advocates Jamil Ahamad Azmi, Anand Swaroop Gautam, Ashutosh Kumar Tiwari and Dharmendra Singh appeared for the petitioners.

Table 8: Document-Summary Sample

Legal-Pegasus-MCS:
In matters, where the allegation is with regard to custodial death, judicial enquiry under Section 176 (1)A cannot be allowed to drag for so long.These are instances which have to be viewed with greatest sensitivity and concern.We, therefore, direct the Registry to enquire from the District Judge, Bulandshahar as to when the enquiry report has been submitted in the matter and, in the event, such a report is not submitted, the explanation of the Judicial Officer in that regard shall be placed before us by the next date fixed as a period of more than one year has expired.We hasten to add that in the event, such enquiry has not been concluded so far, the same shall be concluded most expeditiously by following the procedure in law.List this matter, once again, on 27.01.2022 at 02:00 PM.We, therefore, direct the Registry to enquire from the District Judge, Bulandshahar as to when the enquiry report has been submitted in the matter and, in the event, such a report is not submitted, the explanation of the Judicial Officer in that regard shall be placed before us.

Legal-Pegasus-DPR:
A prayer has also been made to protect the life and liberty of the petitioners as their son was subjected to torture and murder as he had contracted inter-caste marriage out of his own free will. We hasten to add that in the event, such enquiry has not been concluded so far, the same shall be concluded most expeditiously by following the procedure in law.

Legal-Pegasus-MMR:
Advocates Jamil Ahamad Azmi, Anand Swaroop Gautam, Ashutosh Kumar Tiwari, and Dhirajendra Singh appeared for the petitioner. Advocate Ashwani Kumar Mishra and Justice Deepak Verma represented the State of UP. The Court was hearing a petition filed by the father of a man who was found dead in police custody on December 11, 2020 while he was in custody in a case related to his alleged inter-caste marriage out of his own free will. Ä prayer has also been made to protect the life and liberty of the petitioners as their son was subjected to torture and murder as he had contracted inter-caste marriage out of his own free will. It is contended that the authorities have been most unfair in dealing with the grievance raised by the petitioners,̈the Court said. The Allahabad High Court recently observed that judicial inquiry cannot be allowed to drag on for so long when it comes to custodial death [Suresh Devi v. State of UP]. In matters where the allegation is with regard to custodial death, judicial enquiry under Section 176 (1)A cannot be allowed to drag for so long. These are instances which have to be viewed with greatest sensitivity and concern. Ẅe, therefore, direct the Registry to enquire from the District Judge, Bulandshahar as to when the enquiry report has been submitted in the matter and, in the event, such a report is not submitted, the explanation of the Judicial Officer in that regard shall be placed before us by the next date fixed as a period of more than one year has expired,̈the Court ordered. The Allahabad High Court on Wednesday directed the registry to enquire from the District Judge, Bulandshahar as to when the enquiry report has been submitted in the matter and, in the event, such a report is not submitted, the explanation of the Judicial Officer in that regard shall be placed before us by the next date fixed as a period of more than one year has expired.

Table 9: Generated Summaries for the Judgment in Table 8