ABSTRACT

Our aim is to analyze the terms and conditions document of an organization and return a summary of any potential clauses, or sections in the T&Cs which may cause a breach in the client's privacy.

Terms and conditions document is a huge document that mentions the terms and conditions for using the service provided by the organization, which is very complex and filled with legal jargon that only tends to confuse the client when the client goes through it. Our goal is to make a tool that will simplify the terms and conditions document of any organization and inform the client about any potential privacy concerns such that the client will not have to go through the document.

TABLE OF CONTENTS

Chapter No.	Title	Page No.
1.	**INTRODUCTION**	3
2.	**PROBLEM STATEMENT**	6
3.	**LITERATURE REVIEW**	8
	3.1 Deep Learning-Based Approaches	8
	3.2 Domain-Specific Approaches	10
	3.3 Innovative Techniques	11
4.	**DATA**	12
	4.1 Overview	12
	4.2 Dataset Selection and Initial Challenges	12
	4.3 Steps for Dataset Preparation	13
	4.4 Rationale for the Chosen Methodology	14
	4.5 Final Dataset for Fine-Tuning	14
5.	**METHODOLOGY**	15
	5.1 Overview	15
	5.2 Dataset Preparation	17
	5.3 Choice of Large Language Model (LLM)	18
	5.4 Fine-Tuning Process	19
6.	**RESULTS AND DISCUSSION**	20
	6.1 Qualitative Results	20
	6.2 Quantitative Results	21
	6.3 Discussion	21
7.	**CONCLUSION AND FUTURE WORK**	22
REFERENCE/ BIBLIOGRAPHY		23

CHAPTER 1

INTRODUCTION

In today's increasingly digital world, software applications, whether accessed via mobile devices or through web platforms, have become indispensable in virtually every aspect of our daily lives. From social networking and online shopping to banking and healthcare, these applications serve as gateways to essential services and conveniences. However, with the ease of access comes a significant challenge: the necessity for users to agree to terms and conditions (T&Cs) before utilizing these services. These T&Cs, while legally binding, are often lengthy documents filled with complex legal jargon and vague provisions that can be difficult for the average user to understand.

This complexity is not merely an inconvenience; it poses a significant risk to user privacy and autonomy. The intricacy of legal language in T&Cs often overwhelms users, leading them to accept the terms without fully grasping the implications of their consent. As a result, users may unknowingly agree to conditions that compromise their privacy, grant excessive permissions, or expose them to potential legal vulnerabilities. Studies have consistently shown that a significant majority of users tend to bypass reading these documents due to their cumbersome nature, leaving them unaware of what they have agreed to. This widespread disregard for T&Cs highlights a critical need for more accessible and user-friendly methods of conveying this information, ensuring that users can make informed decisions about the services they use.

The rapid advancements in natural language processing (NLP) and artificial neural networks offer a promising avenue for addressing this challenge. NLP, a branch of artificial intelligence, focuses on the interaction between computers and human language, enabling machines to process and analyze large amounts of natural language data. Recent innovations in NLP have been driven by the development of self-supervised or pretrained language models, which have significantly enhanced the ability of machines to understand and interpret human language in a more nuanced and context-aware manner.

One of the most notable models in this field is Google's BERT (Bidirectional Encoder Representations from Transformers), which has revolutionized NLP by introducing a new paradigm in language understanding. Unlike traditional models that process text in a unidirectional manner (either left-to-right or right-to-left), BERT processes text bidirectionally, allowing it to capture the context of a word based on both its preceding and succeeding words. This bidirectional approach enables a more profound understanding of language nuances, making BERT particularly effective for a range of natural language understanding (NLU) tasks, including sentiment analysis, question answering, and text classification.

BERT's ability to grasp the context of words within a sentence has set a new standard for NLP applications. By pretraining on vast collections of general texts, such as Google Books and Wikipedia, BERT has developed a robust understanding of the structure and meaning of natural language. This extensive pretraining allows BERT to be fine-tuned with domain-specific datasets, enabling it to excel in specialized tasks, such as analyzing legal documents or summarizing complex texts like T&Cs.

However, while BERT and similar models have demonstrated remarkable success in general NLP tasks, their performance in highly specialized domains, such as legal text analysis, still poses challenges. The language used in T&Cs is often analogous to that found in legal documents, which are characterized by their formal tone, precise terminology, and complex sentence structures. To fully leverage the capabilities of NLP in this context, domain-specific pretraining and fine-tuning are essential. These processes involve training the models on legal corpora, allowing them to better understand and process the unique characteristics of legal language, thereby improving their ability to summarize and interpret T&Cs effectively.

Moreover, the development of automatic text summarization (ATS) systems, powered by large language models (LLMs), has the potential to transform the way users interact with T&Cs. ATS systems can automatically generate concise and informative summaries of lengthy documents, making the essential information more accessible to users. By focusing on the most relevant clauses and highlighting potential privacy risks, these systems can help users make more informed decisions about the services they engage with, without needing to wade through pages of dense legal text.

In this context, self-supervised learning and fine-tuning models like BERT on domain-specific data are critical for achieving accurate and reliable summaries. Traditional evaluation metrics for summarization, such as ROUGE, which rely on n-gram matching, have limitations in capturing the true quality of a summary, particularly in legal contexts where the accuracy and completeness of the summary are paramount. Advanced evaluation metrics like BERTScore, which compares contextual embeddings rather than surface-level word matches, offer a more sophisticated approach to assessing the quality of text summarization. BERTScore's ability to evaluate semantic similarity aligns more closely with human judgment, making it a valuable tool in the development and assessment of ATS systems.

As we continue to push the boundaries of NLP, integrating these advanced models and techniques into the analysis of T&Cs can significantly enhance user comprehension and awareness. By automating the summarization of these complex documents, we can bridge the gap between legal expertise and user understanding, ensuring that individuals are better equipped to navigate the digital landscape with confidence and informed consent.

CHAPTER 2

PROBLEM STATEMENT

The increasing complexity of terms and conditions (T&Cs) documents has become a significant barrier to user comprehension and privacy protection. These documents are often dense, filled with legal jargon, and difficult for the average user to navigate, leading to a widespread tendency among users to agree to terms without fully understanding their implications. This behavior can result in unintentional consent to privacy-compromising clauses and other unfavorable terms.

While existing solutions attempt to address this issue through manual simplification or generic text summarization, they fall short in capturing the nuances of legal language, especially in documents as varied and complex as T&Cs. Moreover, the vast quantity of T&Cs documents across different services makes manual review and summarization impractical, necessitating an automated and scalable approach.

To overcome these challenges, our paper proposes a novel approach that leverages Low-Rank Adaptation (LoRA) for fine-tuning large language models (LLMs), specifically tailored to the task of summarizing T&Cs documents. By utilizing LoRA, we can efficiently adapt pre-trained models to focus on the specific characteristics of legal text, enabling our tool to generate concise, accurate, and contextually relevant summaries.

Our approach is particularly suitable and feasible for several reasons:

1. **Efficiency and Scalability:** LoRA enables fine-tuning of large models with a significantly reduced computational cost, making our approach more scalable for widespread application across various T&Cs documents.
2. **Domain-Specific Adaptation:** Unlike general NLP models, which may struggle with the intricacies of legal language, our approach fine-tunes models on a carefully curated dataset of T&Cs, ensuring that the generated summaries are not only accurate but also sensitive to the legal nuances that are critical for user comprehension.
3. **Focused Privacy Analysis:** Our methodology is designed to highlight potential privacy risks

within T&Cs, providing users with clear and actionable insights into the clauses that may affect their privacy. This targeted approach addresses the core issue of user vulnerability in the digital landscape.

4. **Automatic and Real-Time Summarization:** The automation of T&Cs summarization allows for real-time analysis and presentation of summaries to users as they interact with services, thereby enhancing informed consent without requiring users to manually sift through lengthy documents.

By integrating these elements, our approach presents a practical solution to the ongoing challenge of making T&Cs more accessible and comprehensible to users. We believe that this method not only advances the current state of T&Cs analysis but also sets the stage for broader applications in legal document summarization and user privacy protection.

CHAPTER 3

LITERATURE SURVEY

Automated summarization of data privacy clauses in terms and conditions has become increasingly important due to the growing concerns over data privacy and the complexity of legal documents. In this literature review, we categorize and discuss existing approaches to summarizing terms of service clauses, focusing on methodologies such as automatic text summarization, leveraging large language models (LLMs), and innovations in natural language processing (NLP) techniques. The review highlights the effectiveness of these approaches in making legal content more accessible and comprehensible for users.

3.1 Deep Learning-Based Approaches

In recent advancements within legal NLP, the identification of relevant holdings in cited cases has seen significant improvements through the development of specialized datasets and pretraining techniques. A notable contribution is the CaseHOLD dataset, which presents a challenging task for legal professionals by requiring the identification of the pertinent holding of a cited case through multiple choice questions. The baseline BiLSTM model achieved an F1 score of 0.4, demonstrating the inherent difficulty of this task. However, leveraging a Transformer architecture like BERT, pretrained on a general corpus such as Google Books and Wikipedia, yielded improved performance. More strikingly, domain-specific pretraining with a custom legal vocabulary resulted in substantial gains, enhancing the F1 score by 7.2%, which translates to a 12% improvement over BERT. This finding underscores the effectiveness of domain-specific pretraining when the task closely aligns with the pretraining corpus, and it was corroborated by consistent performance improvements across two additional legal tasks.

Another relevant study addressed the challenge of detecting unfair clauses in online terms of service . This research involved a comprehensive comparison of various machine learning systems, including advanced deep learning architectures for text categorization and structured SVMs for collective classification. The study employed several approaches for sentence-wide classification, such as Support Vector Machines

(SVMs), Long Short-Term Memory Networks (LSTMs), and collective classification techniques combining SVMs with Hidden Markov Models (SVM-HMMs). Additionally, the research explored different feature sets for text categorization, utilizing both bag-of-words (BoW) models and word embeddings. These methodologies highlight the diverse strategies and advancements in the field of legal text analysis.

3.2 Domain-Specific Approaches

Recent advancements in legal reasoning tasks have explored the effectiveness of prompt engineering, particularly utilizing datasets from the COLIEE 2021-22 competition. These studies emphasize reason-based prompting mechanisms and leverage zero to few-shot learning approaches with pretrained large language models (LLMs). A notable framework in this context is the IRAC (Issue, Rule, Application, Conclusion) method, which facilitates the structuring of legal reasoning processes.

In addition, research has applied clustering techniques to COLIEE training data for entailment tasks, specifically focusing on bar exam questions. This method is more oriented towards classification and may not directly address the broader requirements for training LLMs in comprehensive legal reasoning tasks.

Another innovative approach involves ensuring the privacy policy compliance of IoT wearables with regulations by analyzing privacy policies from major companies such as Apple, Samsung, Fitbit, Garmin, Withings, and Hexoskin. This research constructs an ontology graph to identify key regulatory terms, aiding in the identification of privacy risk areas. While this approach is tailored to IoT devices and wearables, its generalizability to other contexts and regulatory requirements may be limited.

Furthermore, recent research on annotating and classifying relevant clauses in terms-and-conditions contracts has utilized the Contract Understanding Atticus Dataset. This study employs few-shot prompting and achieved a high inter-coder agreement with a Cohen's Kappa score of 0.92 on 20% of the data. The annotation guidelines were refined through iterative processes, incorporating clearer definitions, edge-cases, and examples. The study uses Google's FLAN-T5 model to categorize clauses into general categories, focusing specifically on the Italian language aspect of legal corpus clause categorization and leveraging the dataset for training and fine-tuning the LLM.

3.3 Innovative Techniques

Recent research has introduced several innovative techniques for analyzing privacy policies and enhancing user understanding. One such technique is Polisis [polisis], which employs the OPP-115 dataset and a two-stage approach. In the unsupervised stage, domain-specific word embeddings are generated, while the supervised stage involves training privacy text classifiers. The process includes tokenizing segments using PENN Treebank tokenization and optimizing classifier hyper-parameters through a randomized grid-search. The classifiers are evaluated using metrics such as top-1 precision, F1 score, recall, and precision. Polisis detects privacy clauses by identifying specific labels within the clauses and compares model-generated labels with human annotations using Cohen's Kappa. Additionally, PriBot extends the practical application of this system by providing real-time answers to free-form user questions about previously unseen privacy policies, thus enhancing user comprehension.

In contrast, PrivacyCheck takes a different approach by focusing on automatic summarization through data mining. Using a corpus of 400 privacy policies, PrivacyCheck classifies clauses into three risk categories: red for high risk, yellow for medium risk, and green for low risk. This tool is implemented as a browser extension, offering users immediate risk assessments of the clauses within privacy policies.

Another innovative study, "Demystifying Legalese", utilizes the ToSDR dataset alongside models such as RoBERTa, PrivBERT, Linear SVM, and Random Forest. Case classification is handled by Linear SVM, with the F1 score addressing the dataset's imbalance. The study measures annotation consistency and agreement using Cohen’s Kappa. Rather than a summarization approach, this research employs a simplification strategy by labeling data based on annotations from the ToSDR website. The study highlights the preference for BERT-family transformers over LLMs due to their superior operability in end-to-end pipelines with user interfaces.

CHAPTER 4

DATA

This chapter serves to describe the data under consideration. Understanding how all the data works is vital in the process of creating a good solution to the problem at hand.

4.1 Overview

The fine-tuning of our Large Language Model (LLM) for summarizing data privacy clauses in terms and conditions documents required meticulous preparation of the dataset. We decided to use the *Terms of Service; Didn't Read (TOS;DR)* dataset, which, despite its extensive coverage, required significant modification to be suitable for our specific task.

4.2 Dataset Selection and Initial Challenges

The TOS;DR dataset contains approximately 10,000 documents including terms and conditions, cookie policies, privacy policies, refund policies, etc. However, the dataset lacked pre-existing summaries, which posed a challenge as there were no reference summaries to compare our results against. This absence of ground truth summaries hindered our ability to compute the loss and fine-tune the LLM effectively.

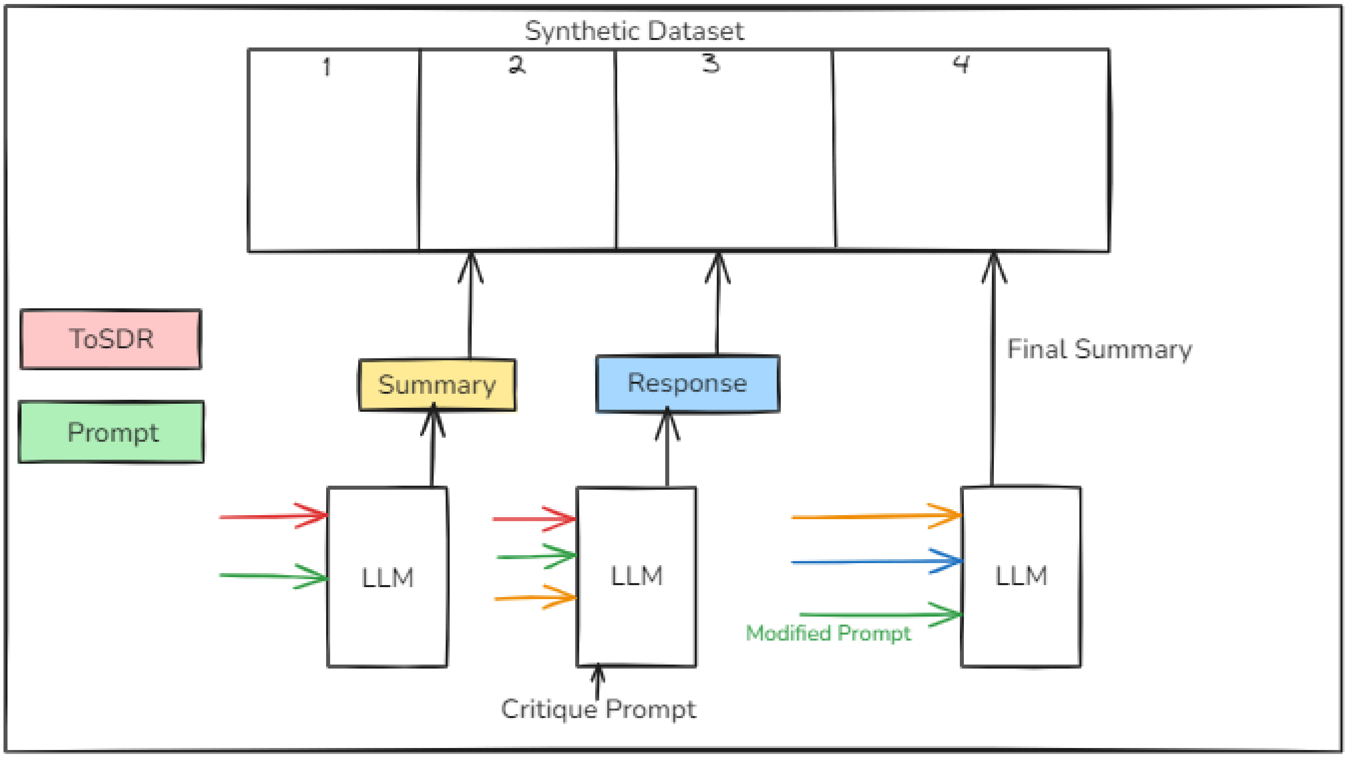

Figure 1: Creation of the Synthetic Dataset

4.3 Steps for Dataset Preparation

Summary Generation for Each Document:

- **Objective**: To generate a summarized version for each document in the TOS;DR dataset that adheres to the rules outlined in the *Digital Private Data Protection Bill* passed in India in 2023.
- **Tool Used**: We used the Gemini model, a large-scale language model, to generate these summaries.
- **Prompt Design**: The prompt used for generating summaries was carefully crafted based on an in-depth understanding of the Digital Private Data Protection Bill. The prompt ensured that the generated summaries would cover all the critical points mandated by the bill.

Ensuring Adherence and Accuracy:

- **Verification**: After generating the summaries, we employed another round of processing with Gemini to ensure that the summaries strictly followed the prompt's rules and were free from hallucinations. This step involved creating a new prompt specifically for checking the accuracy and adherence of the generated summaries to the bill.

- **Feedback Mechanism**: The original document, the new prompt, and the generated summary were input into Gemini to receive feedback, which was then saved for further processing.

Final Summary Creation:

- **Refinement**: The initial summary and the feedback provided by Gemini were used to create a final version of the summary. This final summary incorporated the necessary changes suggested during the verification step, ensuring it was both accurate and aligned with the requirements.
- **Outcome**: The refined summary from this step was then used as the final input for fine-tuning our model, ensuring that all outputs adhered to the rules and minimized hallucination risks.

4.4 Rationale for the Chosen Methodology

Why Gemini?:

- Gemini was chosen for its ability to handle complex legal language and produce coherent, concise summaries that align with legal standards. Its capacity for understanding nuanced instructions made it an ideal tool for our dataset preparation.

Importance of Multi-Step Validation:

- The iterative process of summary generation and validation was crucial to ensure that the final dataset was not only accurate but also legally compliant. This rigorous approach helped avoid data quality issues that could negatively impact the LLM's performance during fine-tuning.

4.5 Final Dataset for Fine-Tuning

Prepared Dataset:

- The final dataset used for fine-tuning consisted of original documents from the TOS;DR dataset, the corresponding generated summaries, and the verification feedback. This structured approach provided a reliable dataset for training the LLM, ensuring its output would be both accurate and aligned with the intended legal standards.

CHAPTER 5

METHODOLOGY

This chapter comprehensively details the methodology used to solve the problem at hand.

5.1 Overview

In this section, we describe the methodology employed to summarize data privacy clauses in terms and conditions documents. Our approach is structured into three key components: dataset preparation, choice of Large Language Model (LLM), and the fine-tuning process using Low-Rank Adaptation (LoRA).

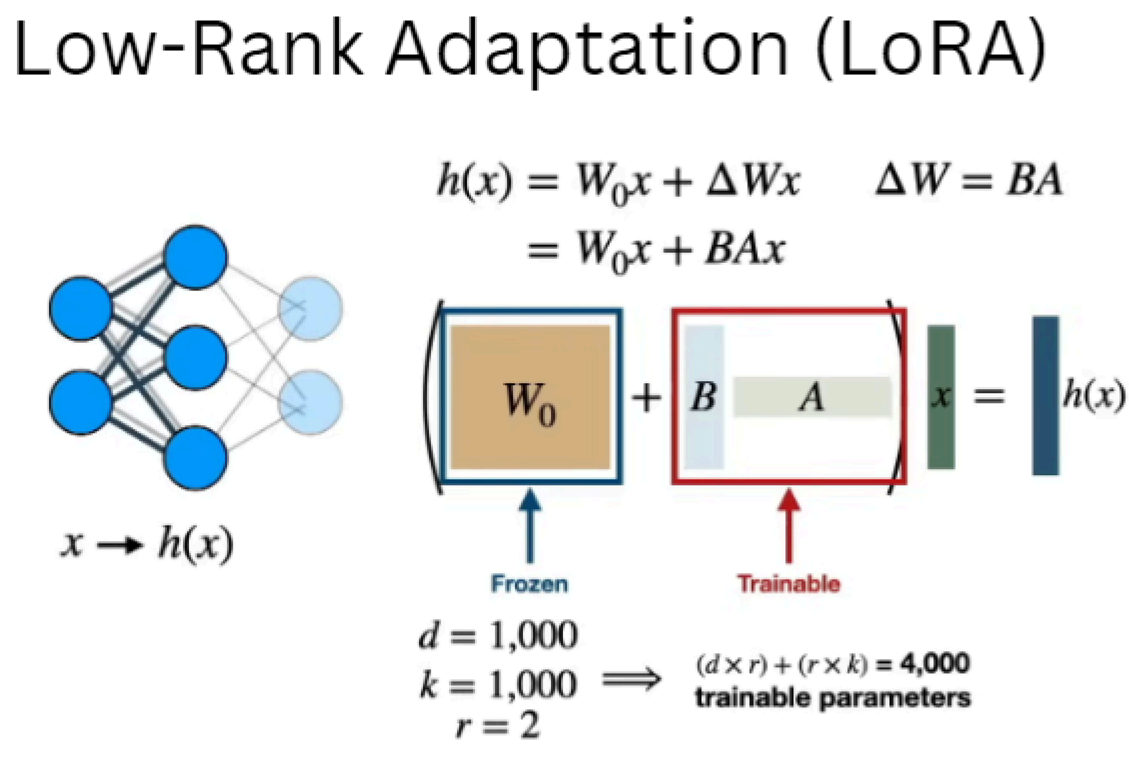

Figure 2: Low-Rank Adaptation

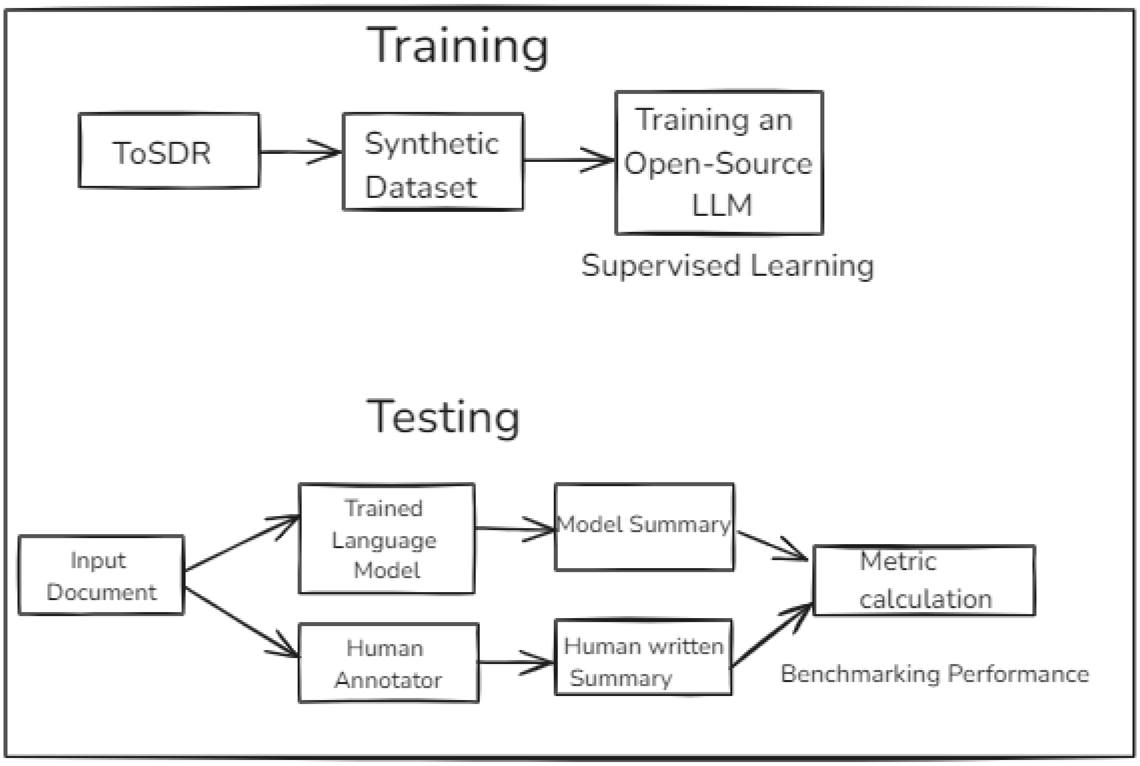

Figure 3: Abstract Architecture Overview

5.2 Dataset Preparation

To fine-tune our LLM effectively, we meticulously prepared a dataset specifically tailored for the task of summarizing data privacy clauses. We selected the *Terms of Service; Didn't Read (TOS;DR)* dataset, which comprises approximately 9,000 documents, including terms and conditions, cookie policies, privacy policies, and refund policies. However, the dataset initially lacked summaries, which are crucial for training purposes as they provide ground truth for evaluating and improving the LLM.

To address this, we undertook the following steps:

1. **Summary Generation**:
   - **Objective**: Generate a summary for each document in the TOS;DR dataset that aligns with the *Digital Private Data Protection Bill* passed in India in 2023.
   - **Tool Used**: We utilized Gemini, a large-scale language model, to generate the summaries.
   - **Prompt Design**: A carefully crafted prompt, based on a thorough understanding of the Digital Private Data Protection Bill, was used to ensure that the generated summaries covered all necessary points.
2. **Verification and Validation**:
   - **Objective**: Ensure that the summaries generated adhere to the rules specified in the prompt and are free from hallucinations.
   - **Process**: A secondary prompt was created to check the accuracy of the generated summaries. This involved inputting the original document, the new prompt, and the generated summary back into Gemini for feedback.
   - **Result**: Feedback was saved and used for further refinement.
3. **Final Summary Refinement**:
   - **Objective**: Incorporate the feedback from Gemini to produce a final, refined summary.
   - **Process**: The initial summary was adjusted based on the feedback to ensure compliance with the intended guidelines.
   - **Outcome**: The final summaries were used to fine-tune our model, ensuring accuracy and minimizing the risk of hallucinations.

5.3 Choice of Large Language Model (LLM)

For the summarization task, we chose the Mistral-7B-Instruct-v0.2-GPTQ, an open-source LLM with seven billion parameters. The Mistral model is based on transformer architecture, which underpins modern language models. Key features of the chosen model include:

- **Efficiency**: The GPTQ model is quantized, enhancing efficiency in terms of memory and computational resources without significantly sacrificing performance.
- **Instruction Tuning**: Fine-tuned on a diverse dataset for instruction-following tasks, such as summarization and question-answering.
- **Contextual Handling**: The model's context length of 8192 tokens was deemed sufficient for processing lengthy documents like terms and conditions.

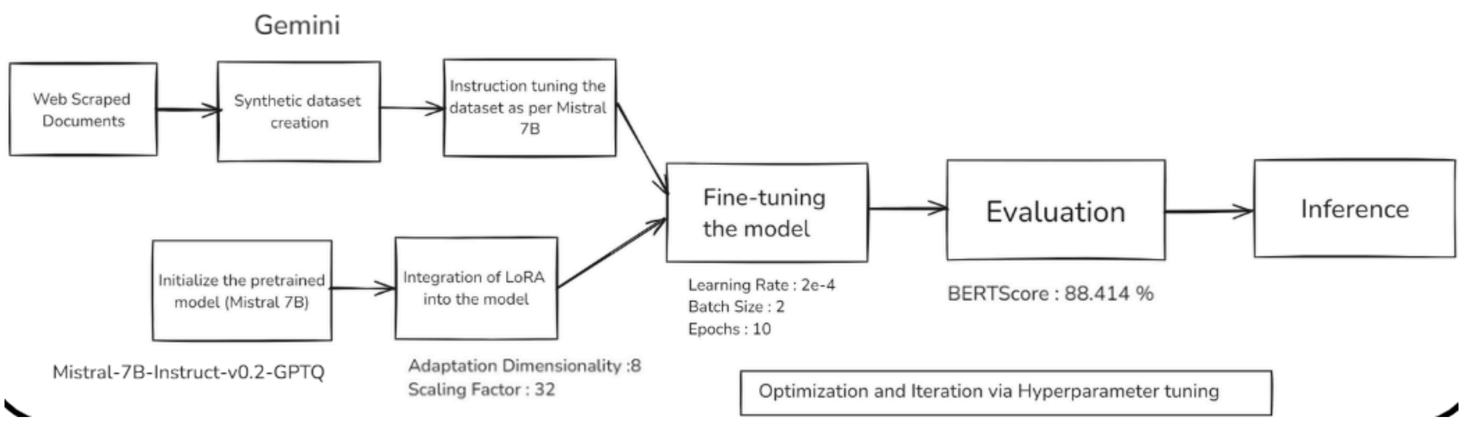

Figure 4: System Architecture Implementing Low-Rank Adaptation (LoRA)

5.4 Fine-Tuning Process

The fine-tuning of our LLM was carried out using Low-Rank Adaptation (LoRA), a parameter-efficient fine-tuning technique. LoRA optimizes the fine-tuning process by focusing on a small subset of the model's weights, specifically those with the most significant impact on the task.

The fine-tuning process involved the following steps:

1. **Model Setup**:
   - **Initial Setup**: The model was put into training mode with gradient checkpointing enabled to reduce memory usage during backpropagation. Quantized training was also prepared to further improve efficiency.
   - **LoRA Configuration**: The rank of the low-rank matrix was set to 8, with a scaling factor of 32 to control the adaptation extent. A dropout rate was defined for the LoRA layers to introduce regularization and prevent overfitting.
2. **Managing Large Text Inputs**:
   - **Challenge**: The vast size of terms and conditions documents required an approach to manage large text inputs efficiently.
   - **Sliding Window Approach**: A sliding window approach with a chunk size of 512 tokens was implemented. This technique preserved contextual integrity across text chunks while efficiently managing system resources.

CHAPTER 6

RESULTS AND DISCUSSION

In this section, we present and discuss the results obtained from our experiments. The performance of our LoRA-based approach to summarizing data privacy clauses within terms and conditions documents is evaluated through both qualitative and quantitative analyses. These results highlight the effectiveness of our methodology and its potential applications in legal document analysis, compliance checking, and consumer awareness.

6.1 Qualitative Results

Through a detailed qualitative analysis, we assessed the human-like summarization capabilities of our LoRA-based model. This evaluation was guided by the use of BERTScore, a state-of-the-art metric that compares the semantic similarity between generated summaries and reference texts using contextual embeddings from pre-trained transformer models like BERT.

- **Human-like Summarization**:
    - Our analysis revealed that BERTScore successfully captured the essence and semantic meaning of the text, providing a nuanced and context-aware evaluation of the summaries. This ability to reflect the quality of summaries in a manner akin to human evaluation validates the efficacy of our approach.
- **Dataset Quality Control**:
    - To ensure high-quality outputs, we manually inspected the summaries generated by Gemini to identify and eliminate any hallucinations. This meticulous quality control process prevented any degradation in the performance of our model, thereby maintaining the integrity of the synthetic dataset used for fine-tuning.

The qualitative results demonstrate that our approach effectively preserves critical information while summarizing, ensuring that the generated summaries are both accurate and contextually relevant. This is particularly important in real-world applications where semantic correctness is paramount.

6.2 Quantitative Results

The quantitative analysis further substantiates the performance of our model, measured using BERTScore. This metric assesses the semantic similarity between generated summaries and reference texts, ranging from 0 to 1, with higher values indicating a closer alignment.

- **Efficiency Gains**:
  - Utilizing the quantized version of the Mistral-7B-Instruct-v0.2-GPTQ model, we achieved a 50% reduction in training time per epoch. This significant improvement in computational efficiency did not compromise the model's performance but instead balanced resource utilization with output quality.
- **Training and Validation**:
  - After training the model for 10 epochs, we calculated the BERTScores for both the training and validation datasets. The training BERTScore reflects the model's ability to learn from the synthetic dataset, while the validation BERTScore measures its generalization ability on a held-out dataset.
- **Accuracy**:
  - Upon testing our model on 20 unseen documents, it achieved an average accuracy of 88.414%. This high level of accuracy demonstrates the effectiveness of our approach in generating concise and semantically coherent summaries that preserve the essential meaning and context of the original documents.

6.3 Discussion

The results from our qualitative and quantitative analyses highlight the success of our LoRA-based approach in summarizing legal texts. The use of BERTScore as both a qualitative and quantitative metric allowed for a comprehensive evaluation, demonstrating that our method is capable of producing summaries that are not only accurate but also contextually and semantically rich.

The efficiency gains achieved through model quantization and the high accuracy obtained during testing underscore the practicality of our approach for real-world applications. By maintaining the semantic integrity of the summarized content, our model offers a robust solution for legal document analysis, potentially aiding in compliance checking and enhancing consumer awareness.

CHAPTER 7

CONCLUSION AND FUTURE WORK

In conclusion, this report has presented a novel approach to summarizing data privacy clauses within terms and conditions by employing Low-Rank Adaptation (LoRA) techniques. Utilizing a LoRA-enhanced version of the Mistral-7B-Instruct-v0.2-GPTQ model, we have effectively generated concise and semantically coherent summaries that preserve the essential information from the original clauses. This includes providing users with a clear overview of the types of personal data collected and highlighting any clauses that may potentially violate the Indian Privacy Bill. The framework we developed is designed to be adaptable for analyzing various legal documents, offering users summaries that are both accurate and contextually relevant, thereby simplifying the understanding of legal jargon.

Our experimental results have demonstrated the effectiveness and robustness of this approach across a range of terms and conditions and related privacy policy documents. This underscores its potential for real-world applications in legal document analysis, compliance checking, and enhancing consumer awareness. Future research directions include exploring the integration of this architecture into a tool that could be configured as a browser extension for seamless and automatic summarization. We also aim to further refine summarization strategies, optimize them for real-time use, enhance domain-specific adaptability, and incorporate user feedback mechanisms. By pursuing these developments, we seek to advance the field of legal text summarization and improve information transparency and accessibility for users worldwide.

BIBLIOGRAPHY

[1] Terms of Service; Didn't Read. 2012. Project Website for 'Terms of Service; Didn't Read'. https://tosdr.org/

[2] Zheng, Lucia, et al. "When does pretraining help? assessing self-supervised learning for law and the casehold dataset of 53,000+ legal holdings." Proceedings of the eighteenth international conference on artificial intelligence and law. 2021.

[3] Lippi, Marco, et al. "CLAUDETTE: an automated detector of potentially unfair clauses in online terms of service." Artificial Intelligence and Law 27 (2019): 117-139.

[4] Alex, Neel, et al. "RAFT: A real-world few-shot text classification benchmark." arXiv preprint arXiv:2109.14076 (2021).

[5] Bizzaro, Pietro Giovanni, et al. "Annotation and Classification of Relevant Clauses in Terms-and-Conditions Contracts." arXiv preprint arXiv:2402.14457 (2024).

[6] Yu, Fangyi, Lee Quartey, and Frank Schilder. "Exploring the effective-ness of prompt engineering for legal reasoning tasks." Findings of the Association for Computational Linguistics: ACL 2023.

[7] Wei, Jason, et al. "Finetuned language models are zero-shot learners."arXiv preprint arXiv:2109.01652 (2021).

[8] Webson, Albert, and Ellie Pavlick. "Do prompt-based models really understand the meaning of their prompts?." arXiv preprint arXiv:2109.01247 (2021).

[9] Echenim, Kelvin Uzoma, Lavanya Elluri, and Karuna Pande Joshi."Ensuring privacy policy compliance of wearables with iot regulations."2023 5th IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA). IEEE, 2023.

[10] Galassi, Andrea, et al. "Unfair clause detection in terms of service across multiple languages." Artificial Intelligence and Law (2024): 1-49.

[11] Soneji, Shikha, et al. "Demystifying Legalese: An Automated Approach for Summarizing and Analyzing Overlaps in Privacy Policies and Terms of Service." arXiv preprint arXiv:2404.13087 (2024)..

[12] Harkous, Hamza, et al. "Polisis: Automated analysis and presentation of privacy policies using deep learning." 27th USENIX Security Symposium (USENIX Security 18). 2018.

[13] Zaeem, Razieh Nokhbeh, Rachel L. German, and K. Suzanne Barber.”Privacycheck: Automatic summarization of privacy policies using data mining.” ACM Transactions on Internet Technology (TOIT) 18.4 (2018):1-18.

[14] Zeng, Xianlong, Fanghao Song, and Ang Liu. ”Similar Data Points Identification with LLM: A Human-in-the-loop Strategy Using Summarization and Hidden State Insights.” arXiv preprint arXiv:2404.04281 (2024).

[15] Jin, Hanlei, et al. ”A comprehensive survey on process-oriented automatic text summarization with exploration of llm-based methods.” arXiv preprint arXiv:2403.02901 (2024).

[16] Gao, Mingqi, et al. ”Human-like summarization evaluation with chatgpt.” arXiv preprint arXiv:2304.02554 (2023).

[17] Zhang, Tianyi, et al. ”Bertscore: Evaluating text generation with bert.”arXiv preprint arXiv:1904.09675 (2019).

[18] Devlin, Jacob, et al. ”Bert: Pre-training of deep bidirectional transformers for language understanding.” arXiv preprint arXiv:1810.04805 (2018).

[19] Obar, Jonathan A., and Anne Oeldorf-Hirsch. ”The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services.” Information, Communication & Society 23.1 (2020): 128-147.

[20] Lin, Chin-Yew. ”Rouge: A package for automatic evaluation of summaries.” Text summarization branches out. 2004.

[21] Brown, Tom, et al. ”Language models are few-shot learners.” Advances in neural information processing systems 33 (2020): 1877-1901

[22] Ouyang, Long, et al. ”Training language models to follow instructions with human feedback.” Advances in neural information processing systems 35 (2022): 27730-27744.