Summarization of Legal & Medical Documents

Team 26 Gatri Reddy Nishal Karamsetty Deepak Reddy

Contents

1 Introduction **2**

2 Motivation **2**

3 Problem Statement **2**

4 Research Gap **3**

5 Objectives **3**

6 Scope and Limitations **3**

7 Datasets **4**
7.1 BillSum (Legal Domain) 4
7.2 PubMed Summarization (Medical Domain) 4
7.3 Combined Dataset 4

8 Methodology **4**

9 System Flow **5**

10 Experiments **6**

11 Results **6**

12 Discussion **6**

13 Conclusion **7**

14 Future Work **7**

Abstract

Legal and medical documents are often long, technical, and difficult to read efficiently. This project focuses on building a summarization system using a T5 model fine-tuned with Low-Rank Adaptation (LoRA) to produce concise and meaningful summaries for legal bills and biomedical research articles. We combine two domain-specific datasets, BillSum and PubMed, and apply parameter-efficient fine-tuning to adapt T5 to both domains. Our model is evaluated using ROUGE metrics and compared against classical baselines such as Lead-3 and TextRank. Results show that our LoRA-based T5 summarizer generates clearer and more accurate summaries while requiring minimal computational resources.

1 Introduction

Legal and medical texts are essential sources of information in policy, research, and clinical decision-making. However, these documents are long, dense, and difficult to read quickly. Manual summarization requires significant time and domain knowledge. As document volume and complexity increase, automated text summarization becomes increasingly important.

Modern NLP models, particularly transformer-based architectures, have dramatically improved summarization quality. However, most are trained on general datasets and do not perform well on specialized domains. In this project, we explore whether a single summarization model can be adapted to two very different domains: legal bills and biomedical research articles. Using LoRA fine-tuning, we efficiently adapt the T5 model to both types of documents.

2 Motivation

Professionals working with legal and medical documents often need to extract key information quickly. Long and complex text slows down decision-making, increases cognitive load, and can lead to missing essential details. Manual summarization is not scalable.

Automating this process can save time and make domain-specific content more accessible. Since legal and medical texts use highly specialized vocabulary and structure, a general summarization tool will not perform well. This motivated us to explore domain-aware summarization using efficient fine-tuning methods like LoRA.

3 Problem Statement

Traditional summarization models often fail when applied to domain-specific documents due to differences in writing style, terminology, and structure. They may miss important details, misunderstand context, or produce inaccurate summaries.

The main problem addressed in this project is: *How can we build a single summarization model that generates reliable, concise, and domain-aware summaries for both legal and medical documents using limited computational resources?*

4 Research Gap

While extensive research exists on general text summarization, studies focusing on legal and biomedical domains are more limited. Gaps identified include:

- Most existing models are domain-specific and not capable of cross-domain summarization.
- Few works attempt to summarize both legal and medical documents using a unified model.
- Parameter-efficient fine-tuning methods such as LoRA are not widely explored for multi-domain summarization.
- Limited comparison with classical baselines on domain-specific datasets.

This project aims to address these gaps by applying LoRA to T5 in a multi-domain setup.

5 Objectives

Primary Objective

To develop a transformer-based summarization model that generates accurate and concise summaries for both legal and medical documents.

Secondary Objectives

- Fine-tune the T5 model using LoRA for efficient domain adaptation.
- Combine legal (BillSum) and medical (PubMed) datasets to train a unified model.
- Compare performance against baseline summarization techniques.
- Evaluate using ROUGE metrics and qualitative analysis.

6 Scope and Limitations

Scope

- Summarization of legal bills from the BillSum dataset.
- Summarization of biomedical research abstracts from the PubMed dataset.
- Evaluation using ROUGE metrics and sample analysis.

Limitations

- Input sequence limited to 512 tokens due to model constraints.
- Occasional factual inaccuracies in highly technical biomedical summaries.
- Limited GPU resources restricted training epochs.
- Not optimized for extremely long documents.

7 Datasets

7.1 BillSum (Legal Domain)

The BillSum dataset contains U.S. Congressional and California State legislative bills along with expert-written summaries. The language is formal, structured, and domain-specific. Documents are long and often exceed typical model limits.

7.2 PubMed Summarization (Medical Domain)

The PubMed dataset includes biomedical research articles paired with abstracts. These texts contain scientific terminology and detailed factual information, making them difficult to summarize accurately.

7.3 Combined Dataset

We sampled balanced subsets from both datasets and mixed them during training. This ensures that the model learns patterns from both domains, improving generalization.

8 Methodology

Preprocessing

Text was cleaned by removing extra spaces, broken lines, and formatting artifacts. The T5 tokenizer was used to convert text into input token sequences. Inputs were truncated to 512 tokens and outputs to 128 tokens.

Model Selection

We selected T5-Small due to its balance of performance and efficiency. It treats all tasks as text-to-text, making it suitable for summarization.

LoRA Fine-Tuning

Low-Rank Adaptation (LoRA) was applied to the attention layers of T5. Key parameters:

- Rank (r): 8
- Alpha: 32
- Dropout: 0.1

LoRA reduces training cost while maintaining strong performance.

Training Setup

Training details:

- GPU: Google Colab T4
- Optimizer: AdamW
- Epochs: 1–3
- Batch size: 2–4

Baselines

We compared our model against:

- Lead-3 (first three sentences)
- TextRank (extractive summarization)

Evaluation Metrics

ROUGE-1, ROUGE-2, and ROUGE-L were used to measure summary overlap with reference summaries.

9 System Flow

The summarization pipeline includes:

1. Input document (legal or medical)
2. Preprocessing and tokenization
3. T5 model with LoRA adapters
4. Summary generation
5. Evaluation (ROUGE)
6. Output final summary

10 Experiments

Experiments were conducted on mixed-domain data with varying epochs. The goal was to observe how well the model generalizes to both legal and medical styles. Resource limitations required efficient fine-tuning, which LoRA provided.

11 Results

The performance of our model was compared against two baseline approaches: Lead-3 and TextRank. ROUGE-1, ROUGE-2, and ROUGE-L metrics were used to evaluate summary quality. The final results are shown in Table 1.

Model	**ROUGE-1**	**ROUGE-2**	**ROUGE-L**
Lead-3	27.44	9.62	17.71
TextRank	31.51	10.08	18.46
T5 + LoRA (ours)	**28.04**	**10.86**	**18.86**

Table 1: ROUGE evaluation results for baseline models vs. our LoRA-tuned T5 model.

12 Discussion

The ROUGE scores present an interesting observation. TextRank achieves the highest ROUGE-1 score (31.51), outperforming our model's 28.04. This is expected because TextRank is an *extractive* method—it directly selects sentences from the source document, leading to strong keyword overlap.

However, our T5 + LoRA model outperforms TextRank in both ROUGE-2 (10.86 vs. 10.08) and ROUGE-L (18.86 vs. 18.46). These scores are more indicative of meaningful summarization quality because they measure sentence structure, fluency, and logical coherence.

The higher ROUGE-2 score suggests that our model constructs smoother, more natural sentence transitions. Similarly, the higher ROUGE-L score shows better overall summary structure compared to the graph-based extraction of TextRank.

Therefore, although TextRank wins in simple keyword matching, our model demonstrates stronger *abstractive* capability—rephrasing content, organizing ideas better, and generating summaries that read more naturally. This supports the value of using transformer-based fine-tuning, even with lightweight approaches like LoRA.

Our results show that LoRA is effective for adapting summarization models to specialized domains with limited hardware. The combined dataset helped create a more generalizable model. Limitations included handling very technical biomedical terminology and input length restrictions.

13 Conclusion

We successfully built a unified summarizer for legal and medical documents using LoRA fine-tuning. The model generated more concise and coherent summaries than traditional baselines while being resource-efficient. This approach can be extended to other domains and applications.

14 Future Work

Future work can explore:

- Long-context models like LongT5 or LED.
- Knowledge-enhanced summarization for biomedical factual accuracy.
- Summarization across more specialized domains.
- Deployment as a web application or API.

Team Contributions

- **Gatri Reddy:** Dataset preparation, preprocessing, evaluation.
- **Nishal Karamsetty:** Model training, LoRA implementation, debugging.
- **Deepak Reddy:** Baseline models, ROUGE analysis, report and presentation creation.

References

- Raffel, C. et al. (2020). *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*. Journal of Machine Learning Research.
- Beltagy, I., Peters, M. E., & Cohan, A. (2020). *Longformer: The Long-Document Transformer*. arXiv:2004.05150.
- Wolf, T. et al. (2020). *Transformers: State-of-the-Art Natural Language Processing*. EMNLP.
- Kingma, D. P., & Ba, J. (2015). *Adam: A Method for Stochastic Optimization*. ICLR.
- Hu, E. et al. (2021). *LoRA: Low-Rank Adaptation of Large Language Models*. arXiv:2106.09685.
- Kornilova, A., & Eidelman, V. (2019). *BillSum: A Corpus for Automatic Summarization of U.S. Legislation*. ACL.
- Cohan, A. et al. (2018). *A Discourse-Aware Attention Model for Abstractive Summarization of Long Documents*. NAACL.