Note: The name of the paper, dataset, and MAW are currently referred to as DocGen because I couldn't think of one. Suggestions welcome 😃

1. Introduction

The task of legal document generation using large language models (LLMs) has not been extensively explored in the legal context. To address this gap, we propose to create a new academic paper titled "Legal Document Generation Using LLMs". This paper will introduce a novel dataset, the DocGen Dataset, and a model-agnostic wrapper (MAW) called DocGen, specifically designed for the task of legal document generation.

2. Dataset: DocGen Dataset

The DocGen Dataset is a question-answering (QA) dataset prepared for instruction tuning LLMs. It contains 500 legal documents that were human-created, along with prompts generated by using larger LLMs (Gemini Pro) to reverse-create the prompts, resulting in QA pairs.

3. Model -Agnostic Wrapper (MAW): DocGen

The DocGen MAW is a model-agnostic wrapper designed to augment the size of documents and improve the performance of LLMs for the task of legal document generation. Although context windows in LLMs are increasing, the size of coherent output that these models can generate has not been proportionally increasing, making them more suited for short text generation rather than long text generation required for legal document generation.

4. Evaluation

We will conduct four types of evaluation testing to assess the effectiveness of the proposed approach:

* Vanilla LLM
* Vanilla LLM + DocGen MAW
* Instruction Tuned Open Source LLM
* Instruction Tuned Open Source LLM + DocGen MAW

We can create multiple MAWs and test them to determine which one performs the best.

5. Human Evaluation Set

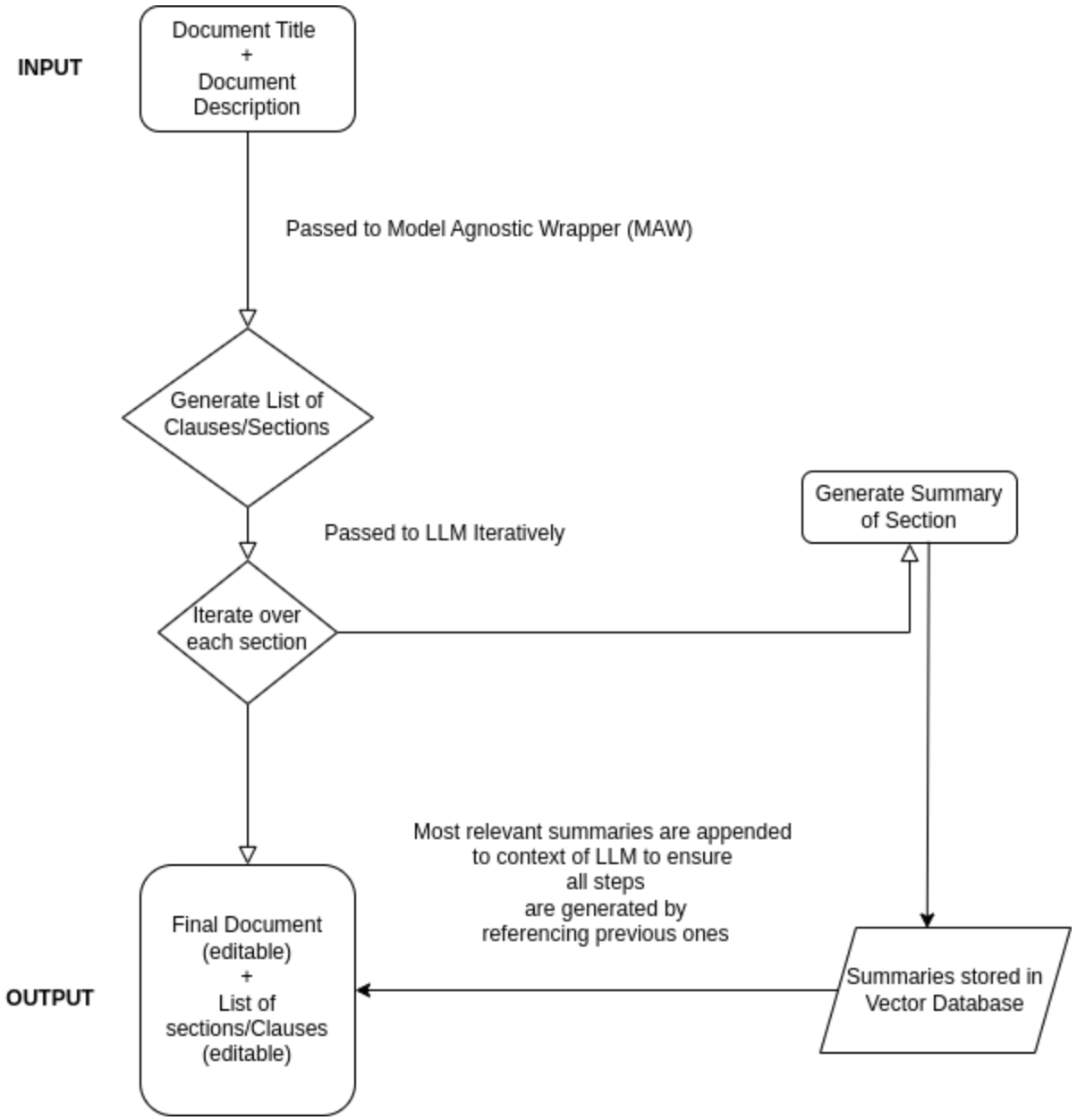
A human evaluation set of 20 questions will be prepared to gauge the LLMs' ability to generate long, coherent texts based on prompts.

6. Conclusion

The proposed paper aims to fill the gap in the legal document generation task using LLMs. By introducing the DocGen Dataset and the DocGen MAW, we hope to provide a valuable resource for researchers and practitioners in the legal field. The four types of evaluation testing will help to assess the effectiveness of the proposed approach and provide insights into the strengths and limitations of LLMs for legal document generation.

Model Agnostic Wrapper

1. User is asked for the document title and the description of the document that is to be generated.
2. Both title and description is passed to the LLM to generate a list of sections that are parsed out and stored in a list datatype.
3. This list is iterated through using a for loop wherein based on the current section title passed and the document title and document description, the LLM generates the section and as well as a short summary of the section.
4. This summary is then parsed and stored in a vector database (minilm-v6).
5. During subsequent iterations of the for loop, the vector database is queried and the summaries of the two most relevant sections are added into the context of the LLM.
6. Final Document is cleaned and displayed, along with the list of sections generated initially.



Related Works

1. LEGAL-BERT: The Muppets straight out of Law School
2. TST_GAN Legal Document Generation based on text style Transfer
3. Contract Understanding Atticus Dataset (CUAD)
4. Kleister: Key Information Extraction Datasets Involving Long Documents with Complex Layouts
5. RECURRENTGPT: Interactive Generation of (Arbitrarily) Long Text

Links to Paper Dataset

1. Spreadsheet containing the Test set which includes the prompts and the associated human-created document (20 Documents)
2. GDrive File containing all the human-created legal documents for Testing.
3. QA Dataset created using human-created legal documents and using the document as input to generate a suitable prompt (494 Documents).

✅ Checklist

S No.	Task	Status
1.	Discussion on motivation, contributions of Paper.	✅
2.	Perusal of Related Works.	✅
3.	Finalising MAW architecture.	✅
4.	Procuring and Cleaning and Processing DocGen Training Dataset.	✅
5.	Generating the Prompts for the DocGen Training	✅
6.	Dataset based on Reverse-engineering using the procured Documents.	✅
7.	Procuring and Cleaning and Processing DocGen Test Dataset.	✅
8.	Evaluation of Test Dataset on Vanilla LLMs.	✅
9.	Creation and testing of MAW v1.0.	✅
10.	Fine-tuning of LLMs (Llama2 7B, Phi-3 mini, Llama3 8B) on DocGen Training Dataset.	✅
11.	Evaluation of vanilla LLMs + MAW on test dataset.	❌

12.	Evaluation of fine tuned LLMs on test dataset.	✅
13.	Evaluation of fine tuned LLMs + MAW on test dataset.	❌
14.	Evaluation of CPT + SFT + MAW LLMs on test dataset.	✅
15.	Human Evaluation of Results.	❌