USING LARGE LANGUAGE MODELS TO EXPLORE THE IMPACTS OF PFAS CONTAMINATION AND RENEWABLE ENERGY POLICIES

AUTHORS

Samuel Maley, Aryaan Verma, Yuan Yao, Molly Allen

ADVISORS

Dr. Xinde "James" Ji

AFFILIATIONS

Department of Computer & Information Science & Engineering, University of Florida

Department of Statistics, University of Florida

Department of Sustainability Studies, University of Florida

ABSTRACT

This research introduces a cutting-edge model leveraging large language models (LLMs) to interrogate legal documents, focusing on renewable energy commitments and the socio-economic effects of PFAS contamination. Our model systematically parses state legislation and case files, offering an unprecedented lens into the dynamics of environmental law and public health impacts. In doing so, we demonstrate the potential of advanced LLMs to revolutionize the extraction of knowledge from legal texts, promising significant contributions to policy formulation, legal practice, and environmental advocacy.

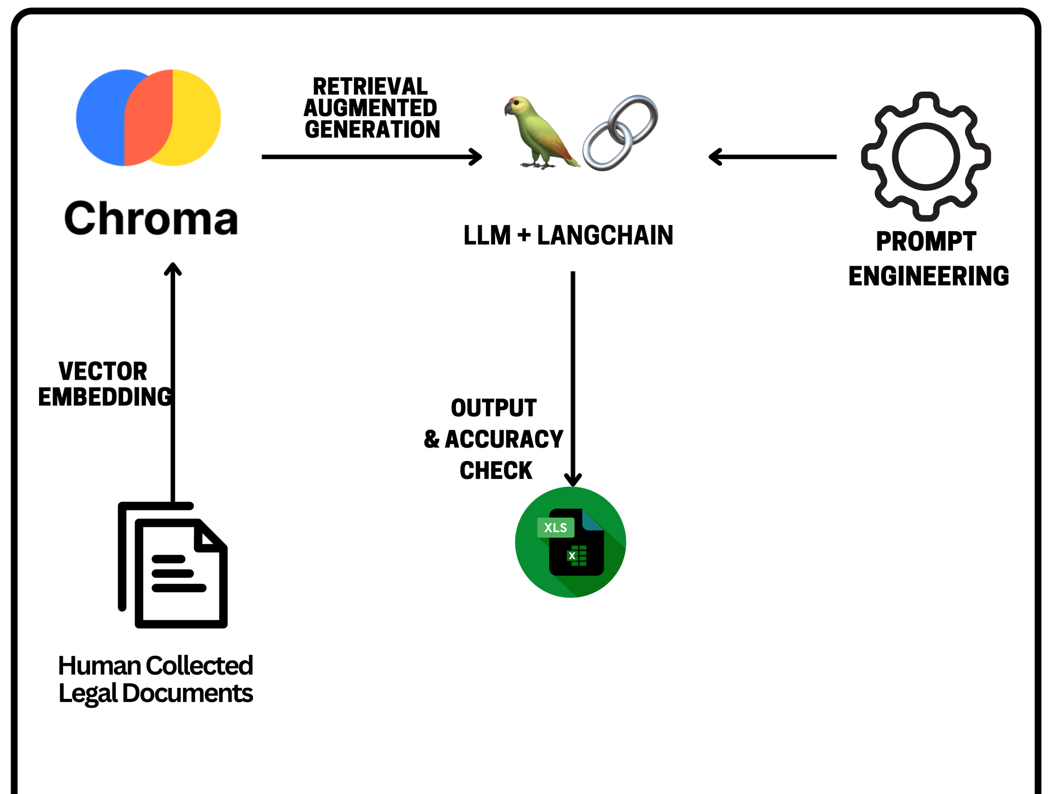

Figure 1 - Data Pipeline Architecture

INTRODUCTION

The intersection of environmental sustainability, public health, and legal frameworks presents complex challenges and opportunities for policy and legal practice. As AI technologies have improved with Natural Language Processing and Analysis, their use in legal and political scenarios as an aid is starting to be researched. This research leverages the capabilities of large language models (LLMs) to navigate this landscape, focusing on renewable energy legislation and the socio-economic implications of PFAS contamination. They can be utilized to automatically analyze documents to improve efficiency and allow humans to focus on more personal and subjective aspects of their work. Our work underscores the importance of innovative legal analysis tools in addressing contemporary environmental and legal issues.

OBJECTIVE

Our primary objective is to develop a novel model using LLMs for efficient and comprehensive analysis of legal documents related to renewable energy commitments and PFAS contamination impacts. We aim to uncover insights into state-level legislative trends, identify barriers to sustainable energy adoption, and assess the socio-economic fallout from environmental contaminants. Additionally, we seek to explore the model's potential applicability in other legal and policy analysis domains.

METHODOLOGY

- **Model Architecture (Figure 1)**: We designed a data pipeline using the Langchain framework that takes pdf/query input and outputs responses. First, we manually collected legal documents pertaining to the context of PFAS exposure and RPS. We pass those documents into an embedding model and store vector representations of each document's text. We then query the documents with our chosen questions and utilize LangChain's "retriever" functionality to obtain sources for each response returned.
- **Retrieval Augmented Generation:** A key functionality of LangChain is Retrieval Augmented Generation (RAG). RAG limits the context of the model to just the inputted documents, reducing model hallucination and increasing accuracy.
- **Checking Accuracy**: We passed in nine questions to model for each document inputted. For this task, we used only the RPS documents. Human annotators manually parsed each document and recorded the correct response for each question in each document. Four questions were categorical, meaning that there was a text output. For these questions we compared the semantic similarity of the model response to the ground truth and returned a value from zero to one that reflected how semantically similar the two were responses were. The other five questions returned a numerical value, which was directly compared to the ground truth value. A value of zero was returned if the responses differed and one if the responses were an exact match. The results of our tests can be found in Figure 2 and 3.
- **Model Selection:** We tested this system on OpenAI's gpt-3.5-turbo and Google's Gemini. We tested 61 files on gpt-3.5, however, we were only able to test Gemini on 10 files. This was due mainly to rate limitations on the LLM API's.

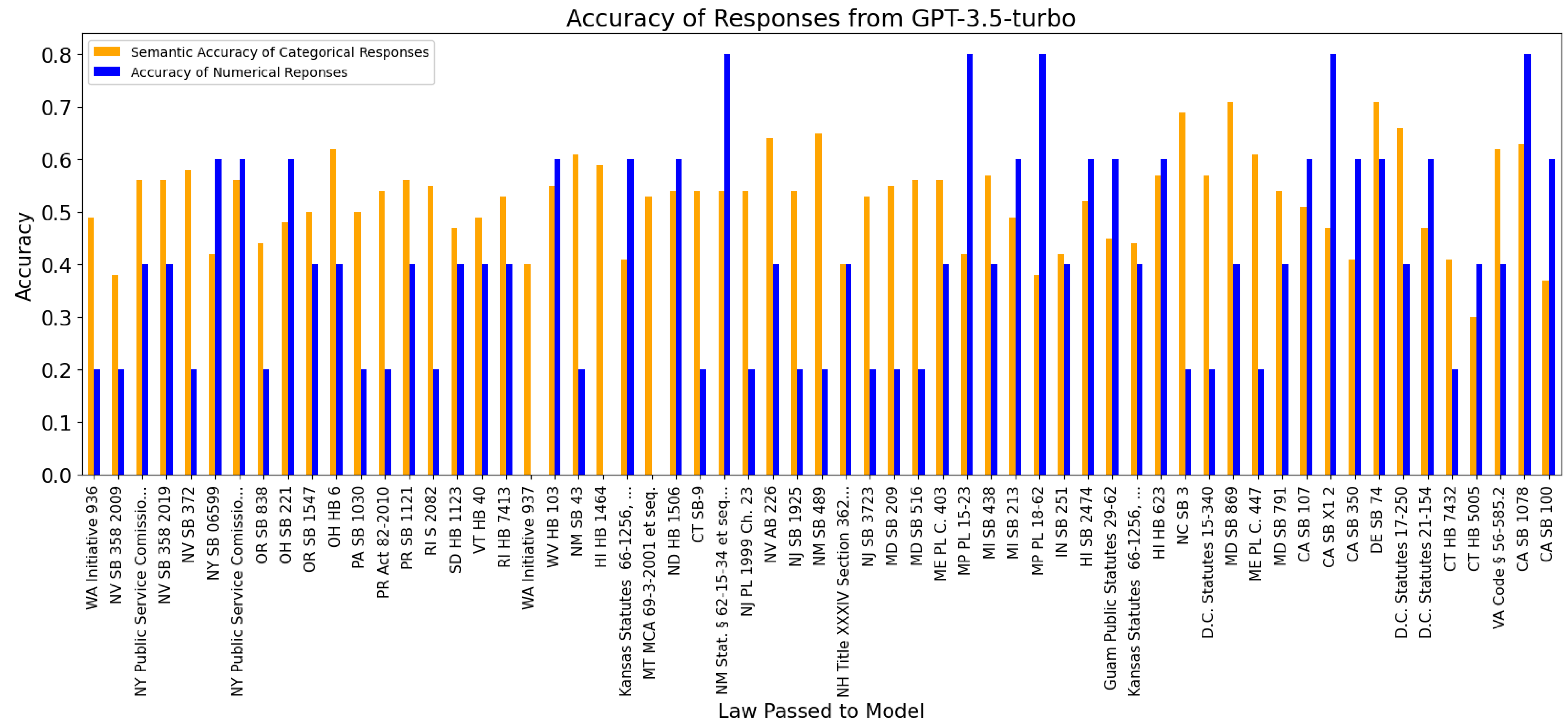

Figure 2 - Chatgpt-3.5-turbo Response Accuracy on RPS Laws

	Which state, area, or territory is the reference law regarding to?	What is the legal commitment name?	List energy sources that meet the criteria of RPS.	List energy sources that receive multifold credit.	What is the date at which the commitment passed into law	What's the RPS/CES commitment in the policy?	What's the year by which the commitment must be met?	What is the RPS commitment for the year the policy was introduced?	What's the percentage of the RPS/CES commitment that is voluntary?
gpt-3.5-turbo OpenAI	98%	80%	28%	3%	10%	46%	54%	8%	82%
Gemini	0%	0%	3%	0%	0%	20%	20%	10%	20%

Figure 4 - LLM Average Response Accuracy by Question

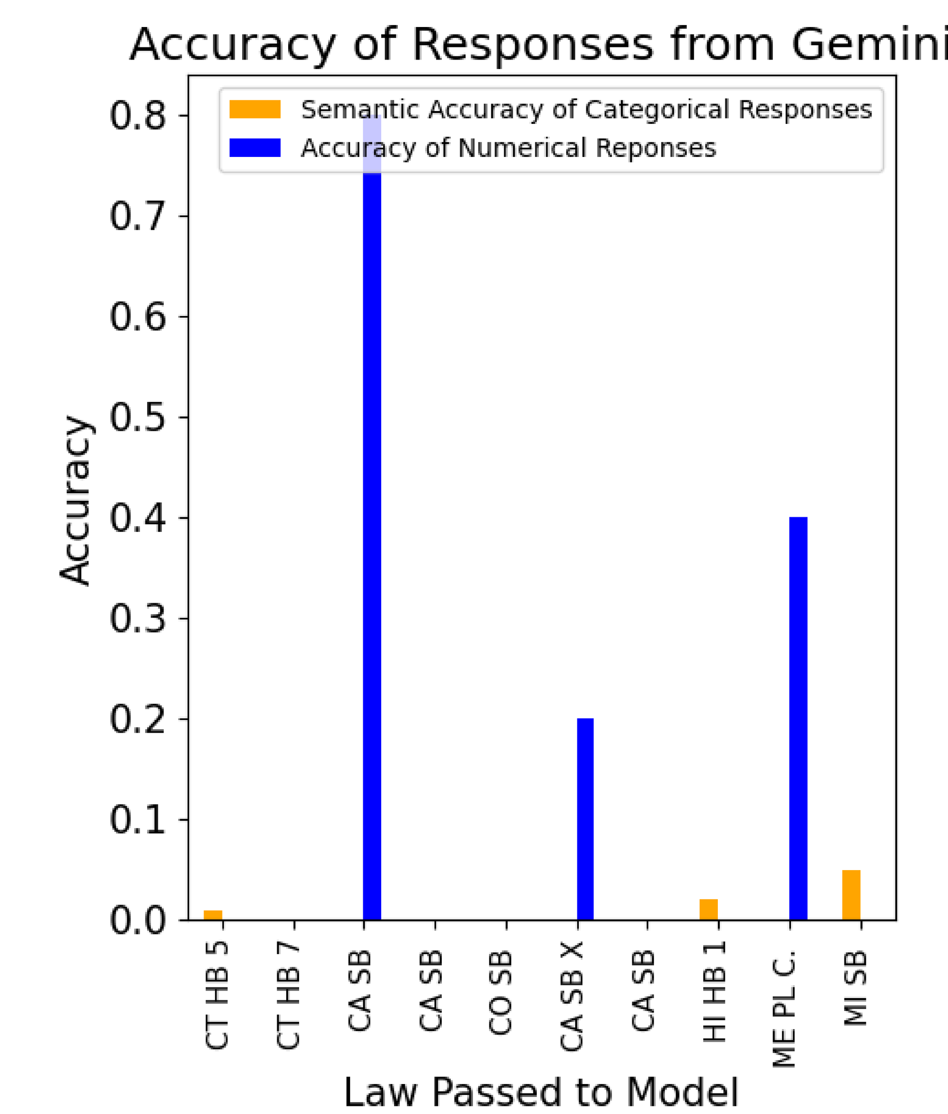

Figure 3 - Gemini Response Accuracy on limited RPS Laws

RESULTS

Based on our findings, ChatGPT 3.5 is currently more apt than Google Gemini Pro for Retrieval Augmented Generation. GPT-3.5 had an average accuracy of 52% for categorical tasks and 40% for numerical tasks. This performance is greater than Gemini, which achieved corresponding accuracies of 1% and 14%. We acknowledge the limitations in our study, as Gemini was not tested on a dataset as expansive as the one used on GPT-3.5. However, our results lead us to believe that the capabilities of GPT-3.5 heavily outweigh Gemini in RAG tasks. Also, based on the outputs of GPT-3.5, there is a greater fluctuation in accuracy for numerical questions compared to categorical questions, which potentially indicates a variance in how the LLM processes different types of data.

CONCLUSION

In this research, we have systematically compared multiple LLM's to evaluate their effectiveness in processing legal documents regarding PFA's and renewable energy. It can be observed that LLMs have considerable utility in extracting relevant information from legal documents and conducting preliminary analyses. However, due to the uncontrollable aspects of LLMs, the model currently can only provide limited valid information. At this stage, we are unable to rely solely on this model for a comprehensive and accurate analysis. Nevertheless, the outputs offer valuable information and potential research directions for subsequent analyses.With the increasing amount of LLM's becoming available to the public, it is likely some will be better than others for NLP. We plan on comparing other LLM's available to the public to determine the best LLM's for this legal processing use case. The implications of these findings are profound for both the development of AI in legal practices and the ethical considerations it entails. As LLMs become more integrated into legal workflows, it is essential to maintain rigorous standards of accuracy and fairness, ensuring these tools do not perpetuate existing biases or introduce new ones.

REFERENCES

Bartik, A., Gupta, A., & Milo, D. (2023). The Costs of Housing Regulation: Evidence from Generative Regulatory Measurement. Social Science Research Network. https://doi.org/10.2139/ssrn.4627587

Ji, X., Li, Q., Mavrothalassitis, F. & Luke, M. (2023). Cleaner bills, cleaner air: The environmental consequence of Renewable Portfolio Standards policy design. Working paper.

Gemini. https://gemini.google.com

OpenAI. https://openai.com