AI-Powered Summarization for Civil Rights Case Data

Community Partner description:

The the University of Michigan in collaboration with the Civil Rights Clearinghouse for Justice (https://clearinghouse.net/) provides comprehensive summaries of civil rights litigation and related resources.

Community Partner contact information:

Main Contact:

Margo Schlanger, mschlan@umich.edu

Advisory Team:

Maryam Berijanian, berijani@msu.edu

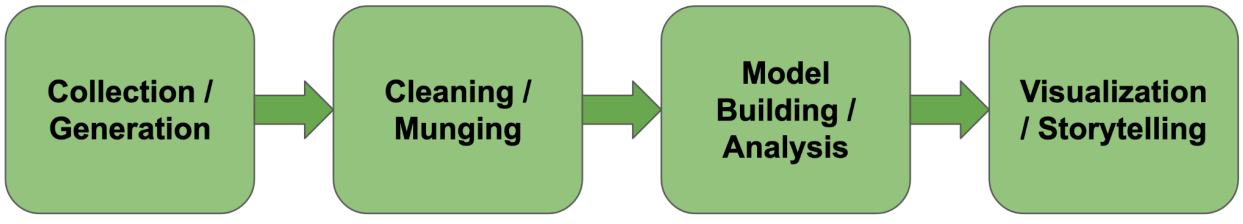
Project overview:

This project aims to develop a “human-in-the-loop” tool to shrink long litigation summaries into a readable format suitable for multiple and different audiences. The tool will use Large Language Models (LLM) and build off prior work that can automatically build summaries. Students will adapt this research into a practical tool that can be added to the Clearinghouse website. A stretch goal is to incorporate mechanisms for human feedback to refine and update summaries over time.

Research questions:

- How can LLMs be applied to rewrite summaries appropriate for different audiences while maintaining accuracy and clarity?
- Can tools like “PersonaMatrix” be adapted into a web-based summarization tool?
- What strategies can enable human-in-the-loop refinement of summaries?

Project Details:



Source of data:

Student's will be provided to hundreds of existing texts and summaries from Clearinghouse.net. These are all in the public domain and will act as validation dataset.

Required cleaning of data:

Summaries may need preprocessing, such as removing legal jargon, normalizing text, and segmenting content for training models.

Data models:

Potential use of NLP models for summarization (e.g., LLM-based models). Adaptation of "PersonaMatrix" methodology for multi-summary synthesis.

Visualizations:

Visual representation of summary length reduction and readability metrics. Possibly dashboards for human-in-the-loop review and feedback.

Project risks and backup plan:

Key to success of the project will be to isolate example summaries that can be used to validate results. Success metrics in tasks with unstructured output is particularly difficult to evaluate and students should work closely with the community partner team to make sure that the results are usable.

Deliverables:

In addition to the project deliverables students should build a system prototype that is intended to integrate with the clearinghouse.net website.

Programming language and software:

Likely Python for NLP and web integration. Possible use of frameworks like Flask/Django for web deployment and Hugging Face for NLP models

NDA and IP agreement:

The data and project is open to the public. NO NDA or IP agreement is required.