

Senior Research Project Blogs

Blog Entry 1

This week I have been focusing on 2 main things: understanding the ProPublica Congress API and reading papers on gun control data pipelines. The ProPublica API is essential to my project to get the actual congressional legislation. The API itself seems pretty easy to use, but it does have a problem giving the text from legislation that failed to pass. While failed legislation isn't particularly important since it has no real-world impact, I am looking at the framing of issues from the legislators, so even failed legislation could be useful as it gives me more writing from legislators. My current idea of a workaround that I have been implementing is using the Playwright library to automate the task of copying the text. As for the gun control data pipelines, I am reading papers that propose pipelines that involve scraping (getting text from online sources) online news articles relating to the gun control controversy. This data could then be used for in-context learning, which means training pre-trained models on context-specific text to improve accuracy. However, all of these pipelines have only been proposed, meaning I would need to spend time seeing if these pipelines are actually feasible for me.

Regarding my internship at the ASU Data Mining and Reinforcement Learning Lab, I decided to work with Ph.D. student Longchao Da on a project relating to the paper "Safe Reinforcement Learning with Natural Language Constraints" [1]. We will be having weekly meetings, first trying to understand the previous paper and methodology, and then going from there and constructing our research question. This project will help me understand the natural language constraints in machine learning models and how to improve accuracy without domain expertise, which may be required in my project if the idea for pipelines is not feasible. The overall lab weekly meetings are pretty simple, as each student just presents a paper they've read in the previous week relating to their projects.

Next week, I plan to continue working on the data collection. I plan to make a simple system using the ProPublica Congress API to get the legislation in JSON format for easy access to the models. I also plan to start working on the pipelines and hopefully make a rough draft of one starting out, and then building off from there. As I continue these steps, I will be writing documentation to help readers replicate my research process and for me to remember the steps I take to further improve the methodology section in my paper.

[1] T.-Y. Yang, M. Hu, Y. Chow, P. Ramadge, and K. R. Narasimhan, "Safe Reinforcement Learning with Natural Language Constraints," *Neural Information Processing Systems*, vol. 34, May 2021.