1. Motivation:

What question(s) do you seek to answer and why?

1. Background:

Who has worked on this before? How does this project relate to the content covered in the lectures and tutorials?

1. Data:

Describe the data and any resources used to collect it. Explain how you filtered data, normalized values, computed additional variables, etc.

1. Methods:

Detail the data analysis and statistical methods. Justify your choice.

1. Results:

Show the results of statistical analyses and visualizations that assess the question(s).

1. Discussion:

Evaluate answers to the question and their reliability. Assess limitations and critique. Relate to results from other work or replicated paper

1. **Motivation**

*RQ1: How well do LLMs estimate general public opinion?*

* Analyze how accurately an LLM can predict the results of surveys. Is this answer skewed towards its own stance on the topic?

Check if a better performance can be obtained by inputting comments from social media. *RQ2: Are comments a faithful reflection of society (as measured by the surveys)? When fed into an LLM, does it then yield more polarized estimations of public opinion?*

1. **Background**

I think this is the most related paper:

<https://osf.io/5ecfa/download>

**Artificially Precise Extremism: How Internet-trained LLMs Exaggerate Our Differences**

Large Language Models (LLMs) offer new research possibilities for social scientists, but their potential as “synthetic data” is still largely unknown. In this paper, we investigate how accurately the popular closed-source LLM ChatGPT can recover public opinion, prompting the LLM to adopt different “personas” and then provide feeling thermometer scores for 11 sociopolitical groups. The average scores generated by ChatGPT correspond closely to the averages in our baseline survey, the 2016–2020 American National Election Study. Nevertheless, sampling by ChatGPT is not reliable for statistical inference: there is less variation in responses than in the real surveys, and regression coefficients often differ significantly from equivalent estimates obtained using ANES data. We also document how the distribution of synthetic responses varies with minor changes in prompt wording, and we show how the same model yields significantly different results over a three-month period. Altogether, our findings raise serious concerns about the quality, reliability, and reproducibility of synthetic survey data generated by LLMs.

* Papers that work with ANES – other applications not really this “social census”
* Papers that work with extended context-window. We explored SelfExtent blabla

1. **Data**

Reddit data via PRAW.

* Filter for 2020 (ANES years)
* Filter for topics according to key terms; selected submissions according to number of comments/ upvotes threshold
* Manually select submissions and extract comments

ANES Survey

* Topics chosen: something about the possible responses range? Not yes/no questions, or different formats
* Abortion, climate change, immigration, gun control, healthcare: why these topics? More clear inclination between progressive and conservatives?

1. **Methods**

We explore in-context learning, using OS fine-tuned models that extend the context window. For GPT we take advantage of the “open session” memory.

* Use Flash Attention

Use ~3 models: GPT, Mistral, Codellama

* Zero shot
* Few shots:
  + Top level comments
  + Comment forest
* Several runs (3?)

3\*5\*3 = 45 runs for each model

1. **Results**

* Explore nice measures/visualizations for comparing distributions

1. Discussions

* Bias of Reddit population - also interesting if we show it as a measure of polarization on Reddit though, but without having a non-biased group or exact description of Reddit population we cannot do comparisons.
* Extracted from r/news, which we assume as with no political leaning but probably has a more democrat leaning (maybe we could do a quick exploration of which newspapers are mostly shared there and see the classification). Are we leaving conservatives out? Could explore
* Further applications: parliamentary debates, newspaper articles, other social media.