

# Estimating Disparate News Media Stance on US Policing

*US policing, text entailment, separability tests, active learning, inconsistency sampling*

## Extended Abstract

This paper analyzes the responses of three major US cable news networks to three seminal policing events in the US spanning a thirteen month period – the murder of George Floyd by police officer Derek Chauvin, the Capitol riot, Chauvin’s conviction, and his sentencing. Although these events happened in rather a quick succession, the Capitol role cast the police in the role of valiant saviors of democracy while Floyd’s murder triggered fierce debate on Police use of racially motivated excessive force. Using sophisticated natural language processing methods, this paper tracks the US policing portrayal over three major US cable news networks – Fox News, CNN, and MSNBC – and seeks to investigate the following two research questions: **RQ1:** *How to quantify differences in police discourse across these three cable news networks?* **RQ2:** *How did the support for police evolve in these three US cable news networks across these momentous policing events?*

**Data:** We collect news transcripts from the official websites of the three news networks from May 25, 2019 to July 24, 2021. We divide this time duration to four non-overlapping slices:

- $\mathcal{T}_{pre-Floyd}$  (*Pre-George Floyd*, starting on May 25 2019 and ending on May 24 2020).
- $\mathcal{T}_{Floyd}$  (*George Floyd’s murder, protest, and civil unrest*, starting on May 25 2020 and ending on Dec 31 2020).
- $\mathcal{T}_{capitol}$  (*Capitol riot*, starting on Jan 1 2021 and ending on March 7 2021).
- $\mathcal{T}_{trial}$  (*Chauvin’s conviction and sentencing*, starting on Mar 8 2021 and ending on July 24 2021).

Filtering based on general keywords is a high-recall approach to obtain relevant discussions [2]. To zero in on the discourse about police, from 34,710 news transcripts, we extract 119,579 sentences with the token `police` in them.

**(RQ1) Reimagining Text Classification as a Separability Test:** Many computational social science research problems eventually often boil down to quantifying differences (or similarities) between two (or multiple) sets of text entities. For a single pair of short documents, sentence similarity is a well-studied problem with several solutions [4]. When the sets contain millions of tokens worth of sentences that makes learning rich distributional representations of words feasible, machine translation based frameworks can compute cross-corpora dissimilarities [3]. When dealing with a few hundred thousand sentences, methods suited for neither of the two extremes work. This paper casts the problem of quantifying corpus similarities as a text classification problem. The intuition is if two classes are easily separable (i.e., highly dissimilar), a classifier would have little or no trouble in distinguishing between them. Hence, classification accuracy can act as a proxy for corpus similarity – a low classification accuracy implying that the two corpora are not easy to differentiate from each other. In our specific problem context, the intuition is if the linguistic signals present in two networks’ reporting on policing are distinguishable, the classifiers will be able to predict the source news network with high accuracy. The control setting is defined by how well a text classifier can classify any other

	CNN	Fox News	MSNBC
CNN	-	74.8 $\pm$ 0.4% 60.4 $\pm$ 2.0%	62.9 $\pm$ 1.2% 59.8 $\pm$ 0.9%
Fox News	74.8 $\pm$ 0.4% 60.4 $\pm$ 2.0%	-	73.5 $\pm$ 0.5% 62.2 $\pm$ 0.7%
MSNBC	62.9 $\pm$ 1.2% 59.8 $\pm$ 0.9%	73.5 $\pm$ 0.5% 62.2 $\pm$ 0.7%	-

Table 1: Text classification accuracy as a proxy for text separability. A cell  $\langle i, j \rangle$  presents the accuracy on the binary classification task of predicting the news network given a sentence from a news transcript with label choices network  $i$  and  $j$ . Each cell,  $\langle i, j \rangle$ , summarizes the classification accuracy as  $a / b$  where where  $a$  (top) is the classification accuracy when models are trained and tested on sentences containing the token *police*;  $b$  (bottom) is the classification accuracy when models are trained and tested on any sentence from the news transcripts of network  $i$  and  $j$ . All models are trained on a data set of 11,000 sentences each from two relevant corpora (5,500 each) with an 80/20 train/test split. We fine-tune BERT [1], a well-known pre-trained language model, for this classification task.

sentences from these news networks. Table 1 shows that Fox News can be easily distinguished from CNN and MSNBC on police-related sentences.

**(RQ2) Text Entailment to Estimate Support for Police:** Table 1 establishes that the reporting on policing on Fox News is considerably different from that on CNN and MSNBC. We estimate the support for police casting the problem as a natural language inference (NLI) task. Given a premise  $\mathcal{P}$  and a hypothesis  $\mathcal{H}$ , the natural language inference (NLI) task involves predicting entailment, contradiction, or semantic irrelevance. Textual entailment is much more relaxed than pure logical entailment and can be viewed as a human reading  $\mathcal{P}$  would infer most likely  $\mathcal{H}$  is true. We consider *Police protect us* and *Police make us safe* as our hypotheses and compute the fraction of sentences from each news networks that entail these hypotheses. We develop a supervised solution that leverages a rich ensemble of active learning methods, and novel sampling technique that exploits logical inconsistencies (e.g., the same premise cannot both entail or contradict a hypothesis and its negation). Our substantive finding (see. Figure 1 is concerning. We observe that across cable networks coverage of politically salient events responds quickly and dramatically to the partisan preferences of their viewership.

## References

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *NAACL:HLT*, pages 4171–4186, 2019.
- [2] Andrew Halterman, Katherine A. Keith, Sheikh Muhammad Sarwar, and Brendan O’Connor. Corpus-level evaluation for event QA: the indiapoliceevents corpus covering the 2002 gujarat violence. In *ACL/IJCNLP (Findings) 2021*, pages 4240–4253, 2021.
- [3] Ashiqur R. KhudaBukhsh, Rupak Sarkar, Mark S. Kamlet, and Tom M. Mitchell. We don’t speak the same language: Interpreting polarization through machine translation. In *AAAI 2021*, pages 14893–14901, 2021.
- [4] Xiaofei Sun, Yuxian Meng, Xiang Ao, Fei Wu, Tianwei Zhang, Jiwei Li, and Chun Fan. Sentence Similarity Based on Contexts. *TACL*, 10:573–588, 2022.

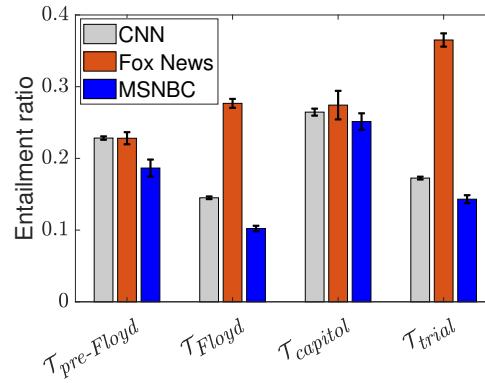


Figure 1: Temporal trend of news network support for  $\mathcal{H}_{protect}$  (*Police protect us*) in terms of entailment ratio. For a corpus  $\mathcal{D}_i$  and a hypothesis  $\mathcal{H}$ , we computed the entailment ratio as the overall fraction of the individual sentences present in  $\mathcal{D}_i$  that entails  $\mathcal{H}$ .