Proceedings of the Nineteenth International AAAI Conference on Web and Social Media (ICWSM 2025)

What Do You Learn About Political Hard News from Soft News Outlets - A Case Study from People Magazine Online

Jason Yan[1], **Kerri Milita**[2], **Yanna Krupnikov**[3], **Sabina Tomkins**[1]
[1]School of Information, University of Michigan, Ann Arbor, MI, USA
[2]Department of Politics and Government, Illinois State University, Normal, IL, USA
[3]Department of Communications and Media, University of Michigan, Ann Arbor, MI, USA
{jasonyan, yanna, stomkins}@umich.edu, kmilita@ilstu.edu

Abstract

People are increasingly avoiding the news in a phenomenon termed *news avoidance*. Prior work has found that many people are not following political news, with the median American consuming zero articles per year from traditional news outlets. This behavior has troubling implications, as a well-informed public is typically considered essential to a functioning democracy. However, it is possible that people are accidentally picking up political information through entertainment or "soft news" outlets. The subject of this paper is *People Magazine*, a soft news outlet selected for its popularity (with an average of 187.1 million monthly visits). To understand how political news is covered by entertainment and soft news outlets, we make the following contributions: (1) We propose two potential and complementary frameworks for differentiating hard news content from soft news content; (2) We collect a large dataset of articles published on *People Magazine*'s website; (3) We apply our frameworks to this dataset and contribute an analysis of the political content offered by a soft news outlet; (4) We evaluate our framework on two additional platforms. Together, these contributions additively result in the first exploration of the political content of individual news stories - all prior work has labeled entire outlets as hard or soft. As people increasingly avoid outlets with hard news, our work shows that they may still encounter political information through soft news outlets, a finding that has critical implications for political communication.

Introduction

A well-educated electorate is generally considered to be normatively good for democracy (Prior 2003; Kuklinski et al. 2000). Ostensibly, as modern technology – and an increase in news outlets – makes information more accessible than at any time in the past, the public could become increasingly better informed about politics. Yet, rather than leading to more political knowledge, a growth of outlets and greater accessibility to information has instead correlated with the public increasingly turning away from "hard news" about politics (Prior 2003; Wojcieszak et al. 2024; Baum 2003). In turn, for many people exposure to political content is coming not from "hard" news outlets – places whose sole purpose is to cover political news – but "soft news" and entertainment outlets (Wojcieszak et al. 2024).

Soft news typically refers to coverage and news stories that are, potentially, more entertaining – but, in contrast to hard news, are less focused on important, timely events (Boukes and Boomgaarden 2015). The implications of this growing consumption of soft news sources – relative to hard news sources – are still unknown. While it may be that soft news sources can increase general political knowledge, there is limited evidence about its broader effects (Prior 2003; Boukes and Boomgaarden 2015). In contrast, hard news sources are understood to be essential for a well-informed electorate (Reinemann et al. 2012; Patterson 2000; Shoemaker and Cohen 2012). A key challenge, however, is understanding how (and whether) soft news outlets cover consequential current *events* (Boukes and Boomgaarden 2015).

The media generally plays an influential role in shaping how the public understands events – and even which events are elevated to the status of "consequential", "important" or "historic". In covering events which are regular occurrences (like elections) or events which are entirely unexpected (like the Covid-19 pandemic or the January 6th, 2021 attack on the U.S. Capitol), the media have power to set the public agenda by determining how much attention to devote to a given event and how long to maintain event coverage (McCombs 2005; Boydstun, Hardy, and Walgrave 2014). Given the media's disproportionate power to elevate, shape narratives and direct public attention to events, researchers have often turned their attention to patterns in the amount (and content) of media coverage awarded to various events (Boydstun 2013; Costley White et al. 2024). Events which are accompanied with an increase in coverage have been termed "media waves" and "media storms" by those who study them (Boydstun, Hardy, and Walgrave 2014).

Although these types of waves and storms, in theory, transcend outlets and can engulf media as a whole, this type of research has often focused on mainstream and "legacy" news media – outlets like network television channels (e.g. ABC), cable news (e.g. CNN) and large newspapers (e.g. *The New York Times, LA Times*) (Zulli, Coe, and Isaacs 2023; Boydstun 2013; Boydstun, Hardy, and Walgrave 2014; Hart, Chinn, and Soroka 2020). In turn, scholars suggest that the decisions made by editors and journalists at these legacy outlets carry implications for the way ordinary people understand political events. For example, media coverage of the events of January 6 shaped how people view American democracy more

Concept	Question ID	Questions
RECENT	Q1	Does the article mention something that suggests that it was written about a recent event?
MENTIONSPFIGURE	Q2	Does this article mention political figure(s)?
MENTIONSCELEB	Q3	Does this article mention celebrity(s)?
CENTRALPFIGURE	Q4	Are political figure(s) the central subject of the article?
CENTRALCELEB	Q5	Are celebrity(s) the central subject of the article?
CENTRALPEVENT	Q6	Is this an article that mainly focuses on a political event(s)?
CELEBIMPLICATIONS	Q7	Does this article explicitly discuss how a political figure, event, or trend affects ordinary people?
PIMPLICATIONS	Q8	Does this article explicitly discuss how a political figure, event, or trend affects politician(s) or political party(s)?
PISSUE	Q9	Does the article discuss political issues? (e.g., Economic Issues, Immigration, Abortion, Crime, Health, etc.)

Table 1: Characteristics of article content assessed in our annotation study.

broadly (Kreiss 2024).

This focus on "legacy" or mainstream media – outlets which "reach sizeable national audiences online" (Zulli, Coe, and Isaacs 2023) – is an important approach to considering the way media cover, frame and produce narratives. At the same time, a focus on these outlets leaves an important omission: many people do not follow these outlets. As Wojcieszak et al. (2023) find, "news content is nearly unnoticeable in the context of overall information and communication ecology of most individuals". Rather, many individuals are receiving news through outlets whose main mission is not the transmission of political news (Kim 2024).

Related Work

Researchers across disciplines have long posed questions about the role of soft news outlets – and often *People Magazine* (*People*) in particular – in informing people about politics (Maddox and Robins 1981; Niven and Zilber 1998; Prior 2003; Baum 2003; Baum and Jamison 2006). Beginning with work focusing on the way soft news addresses politics and political figures (e.g. Diamond 1978; Maddox and Robins 1981) and shifting to the strategic use of soft news by political candidates (e.g. Niven and Zilber 1998; Brewer and Cao 2006) researchers have grappled with the intersection of entertainment and news that often forms soft news outlets (Baum 2003; Boukes and Boomgaarden 2015; Serazio 2018).

This research, however, has led to conflicting findings about the ability of soft news outlets to inform the public (Prior 2003; Brewer and Cao 2006; Baum 2003). It is possible that outlets which primarily focus on soft news can offer people useful information about politics (e.g. Baum 2003), but it is also possible that soft news offers *little* by way of political information (e.g. Otto, Glogger, and Boukes 2017). The difference in viewpoints about whether soft news outlets provide political information may be driven by research approaches. Work on soft news, for example, has at times depended on surveys asking people whether they watch or read soft news outlets; this research cannot track the type of content people encounter through these outlets. Other work has categorized soft news at the *outlet* level – rather than at the news story level (as more recent theories of soft news suggest might be important (e.g., Otto, Glogger, and Boukes 2017)); this again means that we are working off of assumptions about the content of soft news outlets. As a result, this work leaves a number of open questions fundamental to understanding the power of soft news to inform the public. Can outlets which are, primarily, soft news outlets ever provide people with hard news about politics? We answer this question by addressing central gaps in the existing literature:

- **Classifying content on the article level -** While previous research has underscored the importance of distinguishing between exposure to hard and soft news (Prior 2003; Baum 2003; Baum and Jamison 2006; Boukes and Boomgaarden 2015), these earlier studies classified news at the outlet level, without addressing soft news at the level of an individual article. Our work contributes to this literature by taking a data-driven approach to classify content at an article level.
- **Analyzing soft coverage of political events -** While research in communication argues that "media waves" are important in generating coverage of political news, and ultimately, shaping how people interpret political events (Boydstun, Hardy, and Walgrave 2014), there is comparatively little investigation into what soft news outlets bring to these media waves. Such an analysis, however, would be difficult without a computational approach which is able to differentiate hard and soft news content across outlets. Our approach makes such an analysis possible.
- **Considering soft news in the computational social sciences -** We bring questions around soft news content to the CSS community. Existing work has used digital trace data to understand the types of news that people consume (Houidi et al. 2019), how sharing behavior differs from viewing behavior (Trilling et al. 2022), and the political implications of how news is consumed and spread online (Eady et al. 2020; Kümpel, Karnowski, and Keyling 2015). Other work using large datasets of online news consumption has revealed that digital outlets cover celebrities more

frequently than print media (Burggraaff and Trilling 2020). We are the first to raise and answer questions around political information on soft outlets through a data-driven approach to classify content on an article-level at scale, contributing to this larger body of work and to the interdisciplinary study of political information environments (Theocharis and Jungherr 2021; Hemphill and Schöpke-Gonzalez 2020).

Our approach builds on methods across disciplines which begin with large corpora of text. In this way, for example, our work shares similar "components" to the digital humanities – like the use of text and "datafication" of text (e.g. Drucker 2021, p. 2-3). While the digital humanities focus on "humanities materials" and bring together "computational methods and humanistic research" (Drucker 2021, p. 7), our focus is on articles produced by news outlets. Our contributions are within political communication research and computational social science. That is, we employ computational methods to answer critical questions about how people gain political information through soft news outlets.

Our Contributions

We make five contributions in this work:

- We create a novel conceptualization and operationalization of what constitutes political hard news.
- We apply a machine learning framework for classifying news content as hard or soft according to both a classic, and our novel definition.
- We collect a large-scale dataset of political coverage on *People*.
- We use this dataset to answer novel research questions, e.g. how does *People* cover *political events*.
- We expand our analysis to two other outlets besides *People*, one which is considered a soft news outlet (*New York Daily News*) and one which is considered a hard news outlet (*The New York Times*).

These contributions allow us to answer critical open questions in the existing literature:

- To what extent does *People* offer hard news content in general?
- Does *People* cover *political events* and does this coverage contain hard news?
- How does *People* cover *political events* through soft news?
- Do other soft news outlets cover hard news as well?

Classifying News Content

Our goal is to classify news articles as being either "hard news" or "soft news". Generally, hard news refers to articles that center around political figures or events. In contrast, soft news may focus on a celebrity, an entertainment event or human interest features.

Defining Hard News

Previous research on hard news has often been limited by two main issues: (1) the scope of the definition and (2) the operationalization of the concept. Different researchers have, for example, suggested different components of the definition (Reinemann et al. 2012). Synthesizing these different components, Reinemann et al. (2012) suggests three ideas that differentiate hard news from soft news in the context of politics. First, hard news focuses on recent developments and has a time-limited nature; it has "to be reported right away" (Shoemaker and Cohen 2006). Second, hard news is more likely to focus on political events. Finally, hard news focuses on the behavior of political actors related to those events. By this definition, for example, an article highlighting how a couple of ordinary people or entertainment celebrities are feeling about an upcoming election would not be considered hard news. On the other hand, an article tracking the latest electoral polls and which candidate is winning or losing an election could be considered a hard news article.

The example of an article covering polling (often termed the "game frame" approach to elections) highlights another scope challenge to the definition of hard news. Scholars, for example, have long argued that the public most benefits from news coverage that addresses political issues and the broader implications of political events – rather than polling or electoral strategy (Lawrence 2000). While an assumption is that hard news will be superior in preparing people to participate in politics, the definition of hard news does not address these more fine-grained ideas of content and substance (Lawrence 2000).

Aside from issues with the definition, a second challenge has been the operationalization of hard news as a measure. Researchers who have tracked political hard news have often conducted measurements at the level of the *outlet* – classifying an outlet as entirely hard news or entirely soft news – rather than measuring content by news story (Baum 2003; Prior 2003). This is an important omission, as outlets often vary in the content they offer. The closest work, to date, took a more fine-grained approach that considered hard vs. soft news at the level of a *program*, yet they still did not consider the specific program content (Boukes and Boomgaarden 2015).

Our approach overcomes both challenges. Addressing the definitional challenge, we propose two distinct definitions of hard news. The first, which we term the *classic definition*, addresses the three overarching components of hard news found in the literature (Reinemann et al. 2012): (1) recency, (2) focus on political events and (3) behavior of political actors related to those events.

The second, which we term the *novel definition*, addresses the three components along with the more fine-grained content that scholarship suggests is most beneficial (Lawrence 2000): the inclusion of issues in the news and the discussion of broader implications. Through these definitions, we also address the second challenge by classifying news at story level, rather than at outlet or program level.

Let a_i be an article and y_i^c be a binary variable which is 1 if a_i is a hard news article according to the classic definition and let y_i^n be a binary variable which is 1 if a_i is a hard news article according to the novel definition. We define y^c and y^n below (dropping index i for simplicity).

In the first definition (classic) y^c, we are interested in whether an article is discussing something which happened recently and focuses on a political figure or a political event. These points address the three over-arching characteristics of hard news in existing scholarship (Reinemann et al. 2012).

This definition requires the introduction of several logical predicates. First, we introduce the variable $\textsc{Recent}(a)$, which is True if article a discusses a recent event. Next, we introduce the variable $\textsc{CentralPFigure}(a)$, which is True if article a discusses a political figure as a central subject of the article. For example, an article about Mick Jagger which mentions all of the political figures whom the artist has met would not qualify as true, as Mick Jagger is the central subject of the article and not a political figure. Likewise, we introduce the variable $\textsc{CentralPEvent}(a)$, which equals True if article a discusses a political event as a central event of the article.

$$y^c = \begin{cases} 1, & if\ \textsc{Recent}(a) \wedge \\ & (\textsc{CentralPFigure}(a) \vee \textsc{CentralPEvent}(a)). \\ 0, & \text{Otherwise.} \end{cases} \tag{1}$$

In the second definition (novel) y^n, we are also interested in articles which are published in a timely fashion and concerns politicians or political events. Here, however, we extend the previous definition to specify articles which discuss one of the following: how recent events connect to broader political issues, an explicitly political event experienced by a political figure, or political implications for political actors. The first extra condition allows us to discard articles that may discuss recent events on a politically superficial level. This may include traditional "game frame" articles which only report polling numbers, for example. The next condition allows us to discard articles that may mention political figures in non-political situations (e.g., President Obama on a golfing trip). The final condition allows us to differentiate between articles which mention political actors in non-political situations (e.g., President Trump on a golfing trip) and those that mention political actors in non-political situations but which still mention political implications (like President Obama on a golfing trip with political implications for foreign relations).

The goal of the novel definition is to consider a more fine-grained approach in order to speak to research about news content which, scholars suggest, is most beneficial to political knowledge in a democracy (Lawrence 2000). In other words, while the definition of hard news can include a variety of political information, we consider the idea that some hard news may be especially beneficial for people's ability to understand political events and their implications.

Here, we introduce the variable $\textsc{PIssues}(a)$, which is True if article a discusses any of the following political issues (Economic issues, Democracy and representation, Social issues and discrimination, Corporate regulation, Crime, Defense and foreign policy, Education, Environment, Gun control, Health, Immigration, Abortion and birth control, Social safety net, Other). Next, we introduce the variable $\textsc{PImplications}(a)$, which is true if an article discusses the implications of an event for society in general (such as the implications of a politician's endorsement of abortion on the conversation about access to birth control), or the implications for a political party (such as how a court ruling might shift voter support in upcoming elections).

$$y^n = \begin{cases} 1, & \text{if } \textsc{Recent}(a) \wedge \\ & (\textsc{CentralPEvent}(a) \wedge \textsc{PIssues}(a)) \vee \\ & (\textsc{CentralPFigure}(a) \wedge \textsc{CentralPEvent}(a)) \vee \\ & (\textsc{CentralPFigure}(a) \wedge \textsc{PImplications}(a)). \\ 0, & \text{Otherwise.} \end{cases} \tag{2}$$

Problem Definition

Let a be a text-based article. In the classic definition, a can be described as classic hard news (y^c) if it is recent and centers on a political figure. In the novel definition, a can be described as novel hard news (y^n) if it is recent, centers on a political figure, and discusses either political issues or the implications of an event or issue for either politicians or society at large.

Consider a dataset $\mathbb{D} = \{(x_1, y_1), \ldots, (x_N, y_N)\}$. In this dataset, each y_i is represented by the binary label y_i^c in the classic setting and by the binary label y_i^n in the novel setting. Each x_i belongs to a multimodal feature space $\mathcal{X}$, which is determined by the available data. In this paper, we work exclusively with news articles in text form. Our objective is twofold: in the classic setting, we aim to learn a function $f : \mathcal{X} \rightarrow y^c$; in the novel setting, we seek to learn a function $f : \mathcal{X} \rightarrow y^n$.

Our Approach

We propose a supervised learning framework to classify news articles as classic or novel hard news. This framework requires labeled data for both the classic and novel definitions. Thus, our approach consists of the following steps: (1) We collect a large dataset of news articles; (2) We design an annotation task to label these articles according to our definitions of hard news; (3) We compare different machine learning models and choose an appropriate method to label all articles in our dataset.

In the annotation task, we ask annotators to describe articles in terms of a variety of concepts. We then use those concepts to assign an article a label. These labels are then used to train a classification model. This approach allows us to effectively combine information into a single composite measure of hard news and allows us to exploit existing large pre-trained Transformer-based language models, which are effective at classification tasks with text data inputs.

Annotation Design

We employed a wisdom-of-the-crowds framework (Simoiu et al. 2019) to generate annotations of articles published on *People*. This involved creating two resources: (1) A TRAINING CURRICULUM and (2) An ANNOTATION INTERFACE. We developed these resources iteratively. In particular, we refined the training curriculum until we obtained a sufficiently large number of annotators who achieved sufficient performance on the training task (as described below).

Our TRAINING CURRICULUM consists of the following components: (1) instructions, (2) examples with feedback, (3) an attention check, and (4) an assessment. First, participants are shown *instructions* detailing each of the concepts which contribute to the operationalizations of hard and soft news (see Table 1). Next, for each of these concepts, participants are asked to answer questions about *example* articles. They are then given *feedback* on their answers. Then, before being directed to a final training quiz, each participant must pass an *attention check*. Finally, participants take an *assessment* where they are required to score greater than 80% to pass.

Our ANNOTATION INTERFACE consists of a set of questions for 10 articles, and an attention check. Participants begin by consenting to engage in the annotation task. Next, they are asked to provide their annotator ID. This ID triggers a request to a database which returns 10 articles for them to annotate. Participants are asked to answer up to 9 questions for each article. Halfway through these questions, they are asked to pass an attention check.

We asked 9 questions for each article to assess the existence and intensity of political information in the article. These questions are described in Table 1. Together, these questions provide labels for the concepts needed for both the classic and novel definitions of hard news.

Empirical Evaluation

Our empirical analysis consists of several components. We begin by introducing our dataset of articles from *People* and other comparative outlets. Then, we describe our process for collecting annotations and investigating the quality of the annotations through multiple validation methods. Next, we discuss our machine learning framework for classifying hard news according to both classic and novel definitions, and present the quantitative results of this classification. Finally, we end with a qualitative analysis of the articles we've classified as being hard news, as well as those classified as soft news, to better understand the different ways political content is presented across these categories. We further validate our approach by extending our analysis to additional news outlets (*The New York Times* and *New York Daily News*) to demonstrate the generalizability of our framework.

Dataset

We originally collected approximately 8.4 million articles published between October 2016 and November 2023 in *People* through web scraping. Out of these 8.4 million articles, there were 275,000 unique articles, which we refer to as ALLPEOPLE. The data collection process was conducted during January 2024, utilizing three laptops running the scraper continuously for one month. All articles in *People* that appeared on its search page were scraped, rather than a random sample. The scraper was programmed to access the Internet Archive's captures of *People*'s website and retrieve the first available snapshot from each day in the specified time period.

To focus on politically relevant content, we applied keyword filtering, yielding approximately 25,000 articles (POLITICALKEYWORDS). The keywords used are: Supreme Court, Politics, Election, President, Democrat, Republican, Obama, Clinton, Romney, Trump, Biden, and Congress.

Our analysis also incorporates additional datasets that are subsets of POLITICALKEYWORDS. Detailed descriptions of these datasets can be found in Table 2.

Our analysis centers on what readers might encounter when they see political content in a source largely focused on entertainment and human interest. Indeed, *People* – and other entertainment-focused outlets – may be a place which people follow and visit to avoid political spaces (e.g Prior 2007; Wojcieszak et al. 2024). As a result, exposure through this outlet may be highly important as it may be the *only* exposure to political news for some.

This approach allows us to assess the quantity and quality of political information available in this particular soft news context, where readers may be unintentionally engaging with political content.

Annotating the Data

We counducted our annotation study through Prolific with 22 annotators. A total of 1,026 articles were randomly sampled from POLITICALKEYWORDS, the filtered set of 25,000 articles, and then annotated. This annotated dataset of articles is denoted as ANNOTATEDPOLITICALKEYWORDS. Each article receives an odd number of annotations, with a minimum of 3. We conducted this process in five waves from May to August 2024.

We curated a gold dataset of articles that were annotated by a domain expert on the team (GOLDDATASETPOLITICALKEYWORDS). Treating this gold dataset as the ground truth, we computed the Macro F1 score between the aggregated label from the annotations and the expert-labeled gold set. This metric is shown in Table 3. We utilize three label aggregation methods: majority vote, Dawid-Skene (Passonneau and Carpenter 2014), and MACE (Hovy et al. 2013).

We see that the novel definition of hard news is a more difficult annotation task. For example, one concept employed in this definition is whether an article is discussing implications for a politician or society at large. This task is more subjective than the others in the annotation design, and so we are not surprised that it proves to be more difficult.

Overall, we find that MACE label aggregation performs the best, achieving a F1-score of 0.715 for the classic definition and 0.640 for the novel definition. Therefore, we use the labels from MACE for the remainder of the paper. We also include an analysis with Majority Vote in the Appendix (see *Results for Majority Vote Label Aggregation*).

Data From Other Outlets We also assessed whether our models, trained solely on *People*, can generalize to unseen articles from other outlets. We selected *The New York Times* (*NYT*) and the *New York Daily News* (*DailyNews*) for this purpose, as the *The New York Times* is widely known as a hard news outlet. The *Daily News* is another legacy publication known for being a softer outlet (Pressman 2021) and will be used to assess the generalizability of our models to other soft news outlets besides *People*. Detailed descriptions of datasets based on *NYT* and the *DailyNews* can be found in Table 2.

Name	Description	Time Frame	# Articles	Filters	Figures/Tables
AllPeople	All articles in *People* during time frame.	Oct 2016 – Nov 2023	275,000	None	None
Political-Keywords	All articles which may discuss politics or political events.	Oct 2016 – Nov 2023	25,000	The article body must contain a keyword in **PoliticalKeywords**.[1]	Figure 1
Annotated-Political-Keywords	Random sample of 1026 articles from PoliticalKeywords that were annotated to train and evaluate models as part of the classification experiments.	Oct 2016 – Nov 2023	1026	Random sampling of 1026 articles from PoliticalKeywords.	Table 4
GoldDataset-Political-Keywords	A gold dataset that consists of a random sample of 41 articles from AnnotatedPoliticalKeywords. These articles were further annotated by our domain expert to evaluate the quality of labels procured by the annotators and label aggregation methods.	Oct 2016 – Nov 2023	41	Random sampling of 41 articles from AnnotatedPolitical-Keywords.	Table 3
Political-Events	All articles published around the period of a certain political event. For example, one event we study is the January 6th insurrection.	The two weeks following the event.[2]	621	Must be published in a time frame surrounding a political event and the article body must contain a keyword in one of the sets of keywords in **PoliticalEventsKeywords**.[3]	Figure 2
NYT-7-19	Articles in *NYT* during time frame.	July 21 – July 28, 2019	1228	News articles only (excludes opinions, obituaries, newsletters). Text needs to have at least 100 words.	None
DailyNews-7-19	Articles in *DailyNews* during time frame.	July 21 – July 28, 2019	124	Any article	None
NYT-20	Articles in *NYT* during time frame that include the keyword "election".	Nov 3 – Nov 17, 2020	662	The article body must contain 'election' keyword; news articles only (excludes opinions, obituaries, newsletters). Text needs to have at least 100 words.	None
DailyNews-20	Articles in *DailyNews* during time frame that include the keyword "election".	Nov 03 – Nov 17, 2020	99	The article body must contain 'election' keyword.	None
Annotated-NYT	Random sample of 41 articles that includes the keyword "election" from *NYT-20* that were annotated by our domain expert to evaluate the generalizability of the classifiers trained on AnnotatedPoliticalKeywords.[4]	Oct 20 – Nov 17, 2020	41	Random Sampling of 41 articles.	Table 5
Annotated-NYDaily	Random sample of 41 articles that includes the keyword "election" from *DailyNews-20* that were annotated by our domain expert to evaluate the generalizability of the classifiers trained on Annotated-PoliticalKeywords.[4]	Oct 20 – Nov 17, 2020	41	Random Sampling of 41 articles.	Table 5

[1] **PoliticalKeywords** = {Supreme Court, Politics, Election, President, Democrat, Republican, Obama, Clinton, Romney, Trump, Biden, Congress}
[2] For example, the timeframe for the January 6th insurrection would be January 6, 2021 to January 20, 2021.
[3] **PoliticalEventKeywords** = { "Trump Impeachment Vote": [impeachment, impeach, inquiry], "Biden becomes Presumptive Democratic Nominee": [biden, democratic, nominee], "George Floyd Murder and Protests": [Black Lives Matter, George Floyd, protest, BLM, police, violence, george, floyd], "Election Day": [election, turnout, poll, election day], "Capitol Insurrection": [Capitol, insurrection, protest, jan 6th, 6th, storm, stolen, violence]}
[4] Note that for validation, articles were sampled using the same filters as *NYT-20* and *DailyNews-20*, but with an expanded time frame covering 2 weeks prior, in addition to the 2 weeks following Election day.

Table 2: Overview of datasets used in analysis. Datasets include: complete article collection (Oct 2016-Nov 2023); political keyword-filtered articles containing keywords related to elections, parties, and political figures; event-specific articles spanning two weeks before/after major political events; and comparative data from *NYT* and *DailyNews*. Full description of datasets, political keywords, and event-specific criteria available in the Appendix (see *Overview of the Datasets Used in Our Analysis*).

Dataset	Annotation Method	Classic			Novel		
		F1	Precision	Recall	F1	Precision	Recall
People	Majority Vote	0.663	0.661	0.673	0.468	0.450	0.486
	Dawid-Skene	0.670	0.702	0.690	0.610	0.631	0.598
	MACE	**0.715**	**0.711**	**0.728**	**0.640**	**0.712**	**0.611**

Table 3: Comparison of Annotation Methods against the Gold Dataset for *People* (N=41). Labels compared to the gold dataset were aggregated from multiple annotators per article. This figure is generated with GOLDDATASETPOLITICALKEYWORDS.

Classifying Hard News

To analyze how an outlet, such as *People*, discusses the news, we must first be able to distinguish between hard and soft news at scale. Towards this end, we develop a suite of classification models and evaluate them with respect to their ability to classify news content.

Classification Models We used Logistic regression with TF-IDF embeddings as the baseline model. Next, our proposed method uses Transformer-based classifiers (Vaswani 2017), such as RoBERTa (Liu 2019), DeBERTa (He et al. 2020), and DistilBERT (Sanh et al. 2019). An MLP classification head with one hidden layer, ReLU, and dropout was used for the Transformer-based models. ANNOTATEDPOLITICALKEYWORDS, consisting of 1026 annotated articles, was split into 6 folds, with the 6th fold held out as the test set. The remaining 5 folds formed a stratified 5-fold cross-validation to perform hyperparameter tuning. The experiments were run across 3 random seeds (1, 2, and 3). Further details can be found in the Appendix (see *Implementation Details*).

Quantitative Results Table 4 presents the results of the classification experiments for various models using the POLITICALKEYWORDS dataset. The F1, precision, and recall scores are all evaluated on the test set with the best model as determined by the best validation F1 score. For both classic and novel hard news, the BERT* models had the best F1 scores. DistilBERT had the highest F1 score for classic hard news at 0.802 (with a corresponding macro precision of 0.802 and recall of 0.810). For novel hard news, DistilBERT and DeBERTa tied for the highest F1 score at 0.695, with DistilBERT having precision of 0.692 and recall of 0.704, while DeBERTa had precision of 0.705 and recall of 0.700.

We observe a substantial gap in performance between logistic regression with TF-IDF and the BERT* models. This disparity likely stems from the fact that determining hard news requires a contextual understanding of how words interact to generate meaning - a nuance that cannot be captured simply by considering just the individual words themselves (an additive model). Indeed, hard news as a construct is often tacitly understood by most, making it a challenge to define explicitly. We see that all of the BERT* are able to reliably predict whether an article can be considered hard news according to the classic definition. Similarly, although the novel definition proves to be more difficult, the F1-score is reasonably high for the BERT* models.

As an extra validation of our predictive models, we assessed whether the formality of an article varies with respect to hard news categorization. While the formality of a given text is not necessarily tied to its content, we do expect that the average formality of hard news articles would be greater than that of soft news articles. To test this hypothesis, we derived the formality score of each article using a RoBERTa-based classifier[1] that was fine-tuned on GYAFC (Rao and Tetreault 2018) and the Online Formality Corpus (Pavlick and Tetreault 2016). We found that novel hard news had an average formality score of 0.803 (95% CI: [0.799, 0.806]) and classic hard news had an average formality score of 0.757 (95% CI: [0.754, 0.759]). In comparison, soft news had an average formality score of 0.688 (95% CI: [0.686, 0.690]). Next, we inspect the articles classified as hard and soft news to understand how readers would learn about political events through their encounters with *People*.

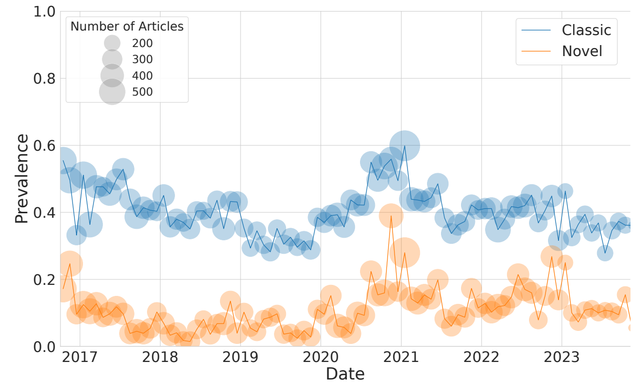

Figure 1: Each circle shows the fraction of articles classified as classic or novel hard news (labels were derived from MACE) over a four week time period. The circle corresponding to January 2021 has 633 articles. On average, there were 237 articles per month. This figure is generated with POLITICALKEYWORDS.

Qualitative Results

We began this work with the research questions described above, briefly summarized as: (1) To what extent does *People* offer hard news content?, (2) Does *People* cover *political events*?, and (3) How does *People* cover *political events* through soft news? To investigate these questions, we classify all 25,000 articles in our dataset according to whether they can be considered as either classic or novel hard news.

Overall, we found that 42% of articles are classic hard news and 12% of the articles are novel hard news. In Figure 1, we show how these percentages vary over time. Thus, with respect to our first research question, we see that yes, *People* does offer hard news content. Moreover, our analysis shows that the amount of hard news reporting changed significantly during the 2020 U.S. presidential election. Interestingly, while we might expect this increase in hard news to return to normal after the election, it remained at higher levels compared to what we saw before the election period.

[1]https://huggingface.co/s-nlp/roberta-base-formality-ranker

Model	Classic			Novel		
	F1	Precision	Recall	F1	Precision	Recall
Logistic Regression	$0.741_{0.000}$	$0.741_{0.000}$	$0.742_{0.000}$	$0.637_{0.000}$	$0.641_{0.000}$	$0.634_{0.000}$
DistilBERT	$\mathbf{0.802}_{0.001}$	$\mathbf{0.802}_{0.002}$	$\mathbf{0.810}_{0.004}$	$\mathbf{0.695}_{0.011}$	$0.692_{0.016}$	$0.704_{0.012}$
RoBERTa	$0.759_{0.006}$	$0.760_{0.005}$	$0.762_{0.007}$	$0.681_{0.018}$	$0.675_{0.007}$	$\mathbf{0.705}_{0.036}$
DeBERTa	$0.776_{0.006}$	$0.776_{0.006}$	$0.782_{0.008}$	$\mathbf{0.695}_{0.018}$	$\mathbf{0.705}_{0.026}$	$0.700_{0.039}$

Table 4: Comparison of Model Performance for Classic and Novel Hard News (labels were derived from MACE). Subscripts denote standard deviation across 3 seeds (1, 2, and 3). Best results for each column are bolded. This figure generated with ANNOTATEDPOLITICALKEYWORDS.

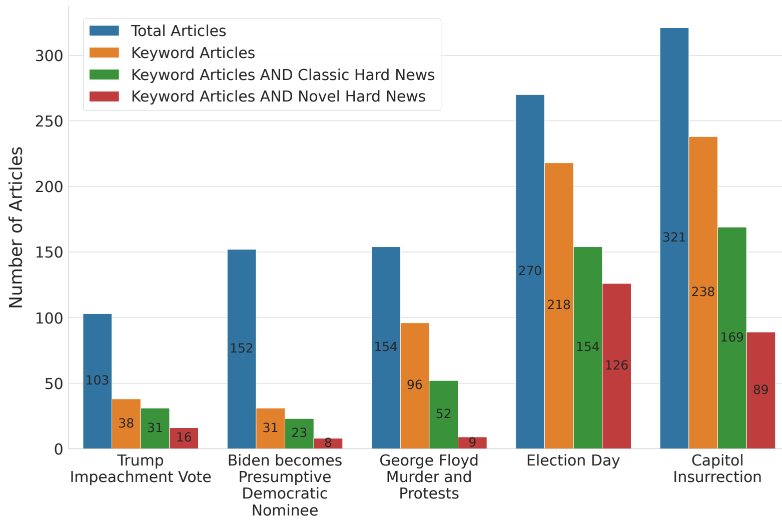

Figure 2: Distribution of articles for 2 weeks following key events. Labels were derived from MACE. Bars show total articles (all articles published on the day of, and the two weeks following the event), articles containing event-specific keywords, and keyword articles classified as classic or novel hard news. Number inside each bar denotes the number of articles. For comparison, we observe 12% of all articles in the corpus (POLITICALKEYWORDS) to be novel hard news and 42% to be classic hard news. This figure is generated with POLITICALEVENTS.

This sustained change suggests that *People* may have fundamentally altered their approach to news coverage, rather than simply responding temporarily to election-related events.

Next, towards the second research question, we investigate *political events*. Here, we are interested in the following events: the impeachment vote of former President Trump on December 18th, 2019; the presumptive nomination of President Biden on April 18th, 2020; the murder of George Floyd on May 25th, 2020; the election of President Biden on November 3, 2020; and the insurrection at the US Capitol on January 6th, 2021. We focus on these dates as previous research has identified these particular points as especially consequential political events (Druckman et al. 2024). As a result, tracking the news which people may have encountered about these events suggests important implications about the information environment.

For each of these events, we form a dataset of all articles published on the day of, and the two weeks following the event. We calculate the percentage of articles containing event-related keywords, followed by estimating the percentage that qualify as hard news under each definition. These results are presented in Figure 2. Our analysis reveals two key findings regarding political event coverage. First, within two weeks following major political events, a substantial portion of published political news focuses on these events. Second, among articles covering these events, the majority were classified as classic hard news, with a smaller but significant portion classified as novel hard news. For example, of the 218 articles published about Election Day within the subsequent two-week period, 154 (70.6%) met classic hard news criteria, while 126 (57.8%) qualified as novel hard news. These proportions significantly exceed the general novel hard news rate (12%) found in *People*'s *political* coverage, challenging the common perception of *People* as primarily a soft news outlet.

Soft coverage of *political events* We see that *People* provided both hard and soft news coverage of each event. This spurs the third question, how is this coverage different? How would *People* cover an event like the murder of George Floyd in a non-politically substantive way?

We explore this question by analyzing those articles that contain keywords relevant to *political events* published within

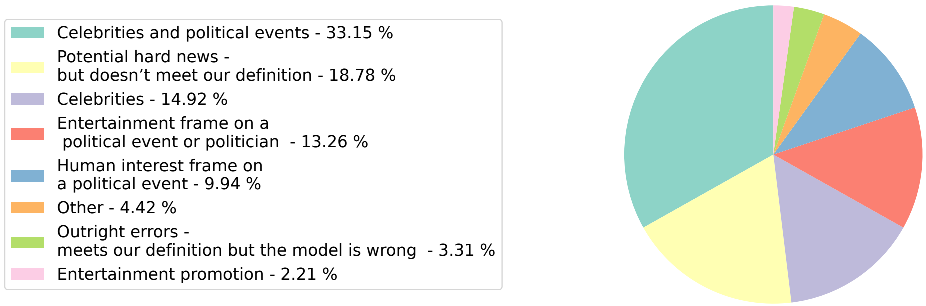

Figure 3: Categorization of articles from POLITICALKEYWORDS that were classified as soft news (N=182). Celebrity reactions and participation (33%) represent the largest category, followed by potential hard news not meeting definition criteria (19%).

a 2-week period, but which are classified as soft news. Here, soft news refers to articles that our classifier predicts do not meet the criteria for either classic or novel hard news. This forms a dataset of 182 articles. That is, 29% of articles that reference the *political events* are classified as soft news. Within this dataset, we manually categorize each article into one of eight categories:

- *Celebrities and political events:* this category describes articles that feature interviews with celebrities, discuss the statements or posts celebrities have made about *political events*, as well as articles about celebrities participating in events (such as joining protests).
- *Potential hard news - but doesn't meet our definition:* this category describes articles which upon closer inspection may be hard news, but which do not meet our politics-focused definition. For example, this would describe an article that is about police brutality in general, but does not center on particular recent *political events*. It would also describe an article about crime.
- *Entertainment frame on a political event or politician:* this category describes content that is not hard news, but which is still framed around political events. For example, an article about how a TV show writer was fired due to racist remarks following George Floyd's murder would fit this category.
- *Celebrities:* this category describes articles that center on a celebrity and address politics only tangentially. For example, a story about a celebrity taking part in a non-political show that briefly mentions that the show was taking place during a political event.
- *Human interest frame on a political event:* this category describes articles that focus on a political event, but which employ a general or human interest frame. For example, an article about how airlines are responding to increased aggression in the wake of the capitol insurrection would fall under this category.
- *Outright errors - meets our definition but the model is wrong:* this category describes articles that should have been classified as hard news but were not.
- *Entertainment promotion:* this category describes articles which are about the promotion of a single instance of entertainment, such as a book or TV show.
- *Other:* this category describes all articles that can't be described by the other categories.

We see in Figure 6 that the largest category of articles describe how celebrities respond to and participate in *political events*. These articles demonstrate a unique way in which *People* informs readers about current events. These articles may serve to engage readers who avoid hard news around current events through a softer celebrity-focused lens.

Another informative category is that of *Potential hard news - but doesn't meet our definition*. This category has clear implications for future work. It contains articles that appear qualitatively more similar to hard news than soft news but which challenge our current definition. Typically, these articles are either not recent, or not focused on national political actors. This suggests that in future work we may want to expand our definition to better encompass articles with different time horizons and which centers on potentially politically-relevent, if not political, figures.

The category of *Entertainment frame on a political event or politician* is, to the best of our knowledge, a new way of describing soft news content. An article in this category may provide a list of movies to watch if one is interested in learning more about social justice in the wake of the George Floyd murder. This category demonstrates how outlets such as *People* may capitalize on attention surrounding *political events* to deliver content that leverages their specialized style.

Validation With Other Outlets

We analyze the content on different outlets as a means of validating our method. Here, we consider two outlets during two time periods. We consider the outlet *DailyNews* as a similar soft news outlet to *People*. In contrast, the *NYT* is considered the most influential hard news outlet (Benoit, Stein, and Hansen 2005). We consider two periods which are useful for validation. The first period, July 21-28, 2019, allows us to inspect content during a politically calm time. The second period, November 3-17, 2020, allows us to inspect content during a politically eventful time.

Here, we perform three tests. First, we evaluate the ability to use our classifiers to inspect these outlets without collecting platform-specific labeled data. Second, we analyze content during the July 2019 period. Here, we test the hypothesis that, by examining all platform content without keyword filtering, we will see more hard news content on *NYT*. Third, we analyze content during the 2020 election period. Given that we saw 71% of the election related content on *People* during this time period was classified as classic hard news, we expect to see similar amounts on the other platforms. That is, if our operationalization of hard news is relevant beyond *People*, we should see hard news on *DailyNews* as well, especially around political events. We now evaluate these three tests.

Validation 1: Method Generalization To assess our classifiers' generalizability beyond *People*, we evaluated their performance on ANNOTATEDNYT and ANNOTATEDNYDAILY. Both datasets comprise of 41 articles each, randomly sampled using the same filter as *NYT-20* and *DailyNews-20* (see Table 2), except with an expanded time frame covering 2 weeks prior, in addition to the 2 weeks after Election day. We focused specifically on articles pertaining to Election day as this is of particular interest and importance when understanding political news.

Table 5 presents the Macro F1, recall, and precision scores comparing classifier predictions against expert-generated labels. The models exhibited robust performance despite having no prior exposure to content from these outlets during training.

Dataset	Classic			Novel		
	F1	Precision	Recall	F1	Precision	Recall
NYT	0.559	0.565	0.560	0.581	0.581	0.582
DailyNews	0.603	0.598	0.629	0.707	0.704	0.710

Table 5: Validation of classifiers trained on MACE-aggregated labels from *People* based on the expert-annotated datasets for *NYT* and *DailyNews* (N=41). This table is generated with ANNOTATEDNYT and ANNOTATEDNYDAILY.

Validation 2: Prevalence of Hard News Across Content During a Politically Calm Period We find that *NYT-7-19* comprised 23% novel and 14% classic hard news, while *DailyNews-7-19* contained lower proportions at 16% novel and 9% classic hard news. During the same time period, we inspect a subset of ALLPEOPLE for the same week of the 21-28 of July, 2019. For this week's worth of data, we observe that 3% of *People*'s content was classic hard news, and 0.3% was novel hard news.

We further validated these results by inspecting the section headings for hard and soft for *NYT-7-19*. Doing so we find that articles classified as hard news typically appear in the US, Foreign Desk, and National Desk sections, while articles classified as soft news are much less concentrated, but tend to appear in the Arts, Opinion, and Sports Desk sections. Together, these result show that our method, although trained on *People*, is making meaningful distinctions between hard and soft content on other outlets.

Validation 3: Coverage During a Political Event Period – 2020 Election The analysis revealed that *NYT-20* comprised 70% classic and 59% novel hard news in 2020, while *DailyNews-20* contained higher proportions at 76% classic and 64% novel hard news in 2020. During the same period, the political content filtered to include the same "election" keyword on *People* was 71% classic and 58% novel hard news. We see that all outlets provide similar proportions of hard news content in their election coverage. While it may seem surprising that the *NYT* has a lower proportion of hard news articles than *DailyNews* or *People* during the election, we propose two explanations. One is that the *NYT* published many more articles about the election (662 vs. 99 on *DailyNews* and 218 on *People*). Given the space for political stories allocated by the *NYT*, they can publish a variety of articles. For example, features which interview potential voters or long in-depth articles about social issues, which add content about the election but would not qualify as hard news. Additionally, the performance of the classifier would likely improve if trained on *NYT* exclusively, and we may be missing stories that would qualify as hard news with the current model.

Validation Summary We find that our framework for operationalizing hard news generalizes well. This is evidenced by several findings. First, when examining all content across the three outlets (*People*, *DailyNews*, and *NYT*) during the politically calm period of July 2019, we found that *NYT* contained the highest proportion of hard news. We further validate this finding by inspecting the section headings of the hard and soft news articles on *NYT*, and find that they follow our expectations. For example, hard news falls under political section headings, while soft news falls under headings like *Art*. Next, we inspect coverage of the 2020 election and find that *DailyNews* news also offers coverage of this political event, at a rate similar to that of *People*. Thus, we find that our operationalization is not limited to *People* only, but generalizes to other outlets as well.

Discussion

We propose two definitions of hard news designed to capture the nuances of covering politics in a modern media environment. We then apply these definitions to a large dataset of articles on *People*, an outlet often considered to focus on soft news and entertainment. To do so, we develop a training program to collect high-quality annotations of hard news and use these annotations to train a scalable machine learning pipeline. We then utilize this final machine-labeled dataset to answer our research questions.

We find that *People* does provide hard news coverage, with 42% of *People*'s political coverage being estimated as classic hard news. Our novel definition of hard news introduces increased constraints over the classic definition. Yet, even according to this strict definition, we estimate that roughly 12% of *People*'s political coverage is novel hard news.

We also find that *People* covers *political events*. From the Trump Impeachment Vote to the Capitol Insurrection, *People* offered both classic and novel hard news coverage. Furthermore, we find that *People* offers unique soft news coverage around these events as well.

When directly related to the events, this soft news coverage typically focuses on how celebrities respond to and participate in political events. For example, George Clooney's thoughts on the Capitol Insurrection, and Tyler Perry reaching out to George Floyd's family after his death.

Future Work This work suggests several avenues for further research. One potential direction is to develop a more expansive definition of hard news that aligns with the stylistic and contextual nuances found in soft news environments. For instance, this extended definition could incorporate political information framed through human interest narratives or celebrity-focused storytelling. While *People* included numerous articles with a hard news frame and a formal news style, other entertainment-oriented outlets may blur the boundaries between political and non-political content more thoroughly.

Another critical direction involves examining audience composition and information consumption patterns. For example, how many readers engage exclusively with celebrity-centered coverage of *political events*, and how does their understanding of these *political events* differ from those who also consume hard news across a range of outlets?

Finally, an important but still unexplored question concerns the practical implications of applying different hard news definitions—classic versus novel—for informing and engaging readers. Future work could investigate whether exposure to content classified under our "novel" definition of hard news results in more knowledgeable or politically engaged readers, compared to those encountering "classic" hard news. For example, this research could consider more subtle aspects of engagement, such as immediate reader perceptions and emotional responses, which can influence how individuals process political information over time.

We focus on news published predominantly by *People*, but also by two other legacy outlets: *NYT* and *DailyNews*. These outlets are incredibly important to understand, given their vast readership. Even on social media, users are often engaging with content produced by such outlets (Kümpel, Karnowski, and Keyling 2015; Majó-Vázquez et al. 2017; Rosengard, Tucker-McLaughlin, and Brown 2014). Future work could investigate hard and soft content on social media or inspect how people respond to different types of stories.

Limitations

We see three main limitations with this work. The first is in our annotation design. To overcome the potential noise from recruiting from an online platform, we went through many rounds of evaluating annotators and removing those who failed to meet a performance threshold on a held-out dataset. As demonstrated in Table 3, this ultimately led to a high-quality set of annotated data, especially for the definition of classic hard news. However, there is still noise in the annotations and this may affect downstream results, especially for the definition of novel hard news. Similarly, although we see strong performance from the machine learning models, future work may improve this performance.

Second, while we know what *People* is publishing, we do not know what is being consumed. Understanding who is consuming what kinds of content would greatly add to our study and enable us to understand if hard news is less popular on the platform than other kinds of content.

Third, we treat our labeled data as ground truth. However, we would like to point out that this is an oversimplification. In reality, there is no ground truth label as to whether an article is hard news or not. Instead, it may be more realistic to imagine the label of hard news as a latent variable. From this perspective, we may propose a latent-variable model where we would like to infer the probability that an article is hard news given its observable characteristics. Alternatively, one could pose this learning task as a weakly supervised problem. Here, we would view each composite concept (e.g. recency) as contributing to the overall classification of hard news. We leave these formulations for future work.

References

Baum, M. A. 2003. Soft news and political knowledge: Evidence of absence or absence of evidence? *Political communication*.

Baum, M. A.; and Jamison, A. S. 2006. The Oprah effect: How soft news helps inattentive citizens vote consistently. *The Journal of Politics*.

Benoit, W. L.; Stein, K. A.; and Hansen, G. J. 2005. New York Times coverage of presidential campaigns. *Journalism & Mass Communication Quarterly*.

Boukes, M.; and Boomgaarden, H. G. 2015. Soft News With Hard Consequences? Introducing a Nuanced Measure of Soft Versus Hard News Exposure and Its Relationship With Political Cynicism. *Communication Research*.

Boydstun, A. 2013. *Making the news: Politics, the media, and agenda setting*. University of Chicago Press.

Boydstun, A. E.; Hardy, A.; and Walgrave, S. 2014. Two Faces of Media Attention: Media Storm Versus Non-Storm Coverage. *Political Communication*.

Brewer, P. R.; and Cao, X. 2006. Candidate Appearances on Soft News Shows and Public Knowledge About Primary Campaigns. *Journal of Broadcasting & Electronic Media*.

Burggraaff, C.; and Trilling, D. 2020. Through a different gate: An automated content analysis of how online news and print news differ. *Journalism*.

Costley White, K.; Kreiss, D.; McGregor, S. C.; and Tromble, R. 2024. *Media and January 6th*, chapter Understanding Media's Role in January 6, 2021. Oxford University Press.

Diamond, E. 1978. *Good News, Bad News*. MIT Press.

Druckman, J.; Klar, S.; Krupnikov, Y.; Levendusky, M.; and Ryan, J. B. 2024. *Partisan Hostility and American Democracy: Explaining Political Divisions and When They Matter*. University of Chicago Press.

Eady, G.; Bonneau, R.; Tucker, J. A.; and Nagler, J. 2020. News Sharing on Social Media: Mapping the Ideology of News Media Content. *Citizens, and Politicians*.

Hart, P.; Chinn, S.; and Soroka, S. 2020. Politicization and Polarization in COVID-19 News Coverage. *Science Communication*.

He, P.; Liu, X.; Gao, J.; and Chen, W. 2020. Deberta: Decoding-enhanced bert with disentangled attention. *arXiv preprint arXiv:2006.03654*.

Hemphill, L.; and Schöpke-Gonzalez, A. M. 2020. Two computational models for analyzing political attention in social media. In *Proceedings of the International AAAI Conference on Web and Social Media*.

Houidi, Z. B.; Scavo, G.; Traverso, S.; Teixeira, R.; Mellia, M.; and Ganguly, S. 2019. The news we like are not the news we visit: News categories popularity in usage data. In *Proceedings of the International AAAI Conference on Web and Social Media*.

Hovy, D.; Berg-Kirkpatrick, T.; Vaswani, A.; and Hovy, E. 2013. Learning whom to trust with MACE. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Kim, E. 2024. *The American Mirage: How Reality TV Upholds the Myth of Meritocracy*. Princeton University Press.

Kreiss, D. 2024. *Media and January 6th*, chapter Political Communication Research at a Time of Democratic Crises. Oxford University Press.

Kuklinski, J. H.; Quirk, P. J.; Jerit, J.; Schwieder, D.; and Rich, R. F. 2000. Misinformation and the currency of democratic citizenship. *The Journal of Politics*.

Kümpel, A. S.; Karnowski, V.; and Keyling, T. 2015. News sharing in social media: A review of current research on news sharing users, content, and networks. *Social media+ society*.

Lawrence, R. 2000. Game-Framing the Issues: Tracking the Strategy Frame in Public Policy News. *Political Communication*.

Liu, Y. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint arXiv:1907.11692*.

Maddox, W. S.; and Robins, R. 1981. How People Magazine Covers Political Figures. *Journalism & Mass Communication Quarterly*.

Majó-Vázquez, S.; Zhao, J.; Nurse, J.; and Nielsen, R. 2017. Digital-born and legacy news media on Twitter during the UK general election. *Reuters Institute for the Study of Journalism*.

McCombs, M. 2005. A Look at Agenda-setting: past, present and future. *Journalism Studies*.

Niven, D.; and Zilber, J. 1998. "What's Newt Doing in People Magazine?" The Changing Effect of National Prominence in Congressional Elections. *Political Behavior*.

Otto, L.; Glogger, I.; and Boukes, M. 2017. The Softening of Journalistic Political Communication: A Comprehensive Framework Model of Sensationalism, Soft News, Infotainment, and Tabloidization. *Communication Theory*.

Passonneau, R. J.; and Carpenter, B. 2014. The benefits of a model of annotation. *Transactions of the Association for Computational Linguistics*.

Patterson, T. E. 2000. Doing well and doing good. *Available at SSRN 257395*.

Pavlick, E.; and Tetreault, J. 2016. An empirical analysis of formality in online communication. *Transactions of the association for computational linguistics*.

Pressman, M. 2021. The New York Daily News and the History of Conservative Media. *Modern American History*, 4.

Prior, M. 2003. Any good news in soft news? The impact of soft news preference on political knowledge. *Political communication*.

Prior, M. 2007. *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections*. Cambridge University Press.

Rao, S.; and Tetreault, J. 2018. Dear sir or madam, may I introduce the GYAFC dataset: Corpus, benchmarks and metrics for formality style transfer. *arXiv preprint arXiv:1803.06535*.

Reinemann, C.; Stanyer, J.; Scherr, S.; and Legnante, G. 2012. Hard and soft news: A review of concepts, operationalizations and key findings. *Journalism*.

Rosengard, D.; Tucker-McLaughlin, M.; and Brown, T. 2014. Students and social news: How college students share news through social media. *Electronic news*.

Sanh, V.; Debut, L.; Chaumond, J.; and Wolf, T. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Serazio, M. 2018. Producing Popular Politics: The Infotainment Strategies of American Campaign Consultants. *Journal of Broadcasting & Electronic Media*.

Shoemaker, P. J.; and Cohen, A. A. 2006. *News around the World. Content, Practitioners, and the Public*.

Shoemaker, P. J.; and Cohen, A. A. 2012. *News around the world: Content, practitioners, and the public*.

Simoiu, C.; Sumanth, C.; Mysore, A.; and Goel, S. 2019. Studying the "wisdom of crowds" at scale. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*.

Theocharis, Y.; and Jungherr, A. 2021. Computational social science and the study of political communication. *Political Communication*.

Trilling, D.; Kulshrestha, J.; De Vreese, C.; Halagiera, D.; Jakubowski, J.; Möller, J.; Puschmann, C.; Stepińska, A.; Stier, S.; and Vaccari, C. 2022. Is sharing just a function of viewing? The sharing of political and non-political news on Facebook. *Journal of Quantitative Description: digital media*.

Vaswani, A. 2017. Attention is all you need. *arXiv preprint arXiv:1706.03762*.

Wojcieszak, M.; de Leeuw, S.; Menchen-Trevino, E.; Lee, S.; Huang-Isherwood, K. M.; and Weeks, B. 2023. No polarization from partisan news: Over-time evidence from trace data. *The International Journal of Press/Politics*.

Wojcieszak, M.; Menchen-Trevino, E.; Clemm von Hohenberg, B.; de Leeuw, S.; Gonçalves, J.; Davidson, S.; and Gonçalves, A. 2024. Non-news websites expose people to more political content than news websites: Evidence from browsing data in three countries. *Political Communication*.

Zulli, D.; Coe, K.; and Isaacs, Z. 2023. News Framing in the Aftermath of the January 6 Attacks on the U.S. Capitol: An Analysis of Labels, Definitional Uncertainty, and Contextualization. *American Behavioral Scientist*.

Ethics Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, see Introduction
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, see Introduction, Annotation Design, and Dataset
   (e) Did you describe the limitations of your work? Yes, see Limitations
   (f) Did you discuss any potential negative societal impacts of your work? Yes, see Introduction, Limitations, and Contributions to Related Work
   (g) Did you discuss any potential misuse of your work? Yes, see Limitations
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, see Ethics Statement
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? NA
   (b) Have you provided justifications for all theoretical results? NA
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA
   (e) Did you address potential biases or limitations in your theoretical framework? NA
   (f) Have you related your theoretical results to the existing literature in social science? NA
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? See Annotation Design and Empirical Evaluation
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes, see Implementation Details
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? Yes, see Table 4
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? Yes
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? Yes, see Limitations.
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? Yes
   (b) Did you mention the license of the assets? NA
   (c) Did you include any new assets in the supplemental material or as a URL? NA
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes, see Annotation Design
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes, see Annotation Design
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? NA
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? NA
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? Yes, see Table 1
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? Yes, see Ethics Statement
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? Yes, see Ethics Statement
   (d) Did you discuss how data is stored, shared, and deidentified? Yes, see Ethics Statement

Ethics Statement

We will follow the guidelines of our university. We follow the ethical guidelines of the Internal Review Board of the university. As we do not intend to collect, analyze, or publish any information that could be used to identify individuals, the study presents no risk to individuals.

We will store data in our university's data center. The data center is protected through multiple layers of physical security and layered network protection firewalls. The hardware is monitored by a professional technical team running continuous monitoring software to detect malicious access, running automatically scheduled updates, and cooperating with the campus computing teams to keep abreast of potential and active security threats.

We recruited participants through Prolific and paid them an hourly rate of \$30-\$40.

Implementation Details

The dataset consisting of 1026 annotated articles was split into 6 folds, with the 6th fold held out as the test set. The remaining 5 folds formed a stratified 5-fold cross-validation to perform hyperparameter tuning. The experiments were run across 3 random seeds (1, 2, and 3). Logistic Regression with TF-IDF was implemented with scikit-learn. The Transformer-based models were implemented in full precision (FP32) with PyTorch and Hugging Face. Cross-entropy loss with an AdamW optimizer was used to train all Transformer-based models.

In our hyper-parameter search we iterated over different values of the learning rate, weight decay (for AdamW), hidden layer size, dropout, and number of epochs. A batch size of 16 was used for all experiments. Each model was trained for 20 epochs. Our approach combined a grid search for epochs and hidden layer size with random sampling for the remaining parameters. Specifically, we performed a grid search across all hidden layer sizes. For each grid point, we randomly sampled 20 combinations of learning rate, weight decay, and dropout. These sampled parameters were then combined with the corresponding grid search values, resulting in a set of hyperparameters defining a unique model configuration. Hyperparameter ranges can be found in the Appendix (see Table 7). All experiments were run with 8 A5000 GPUs with 24 GB of memory. Macro F1 scores were used to evaluate model performance. Hyperparameters were selected based on the highest validation F1 score averaged across all folds and seeds; the corresponding test F1, precision, and recall scores were reported.

Inter-Annotator Agreement and Simulations

Table 8 lists the Krippendorff's Alpha and the simulated agreement levels for each concept. Simulated Krippendorff's Alpha was derived by simulating labels for each concept. Labels are drawn from a Bernoulli(p) distribution, where p is the prevalence of the True class for the particular concept based on unaggregated annotations. Since the labels are randomly drawn, the Krippendorff's Alpha, which adjusts for chance agreements, for each concept is expected to be 0. We observe that the simulated Krippendorff's Alpha scores for all concepts are near 0, consistent with expectation.

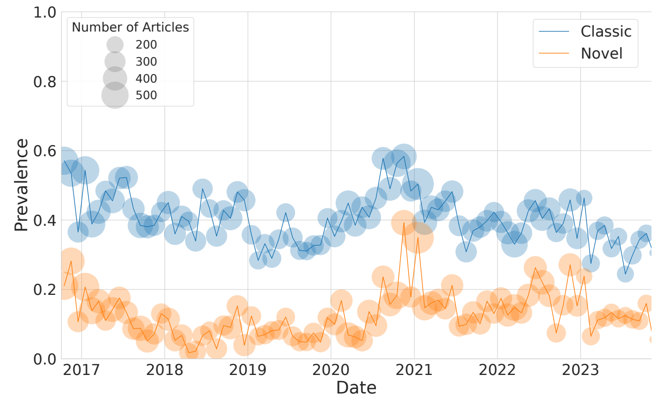

Figure 4: Each circle shows the fraction of articles classified as classic or novel hard news (labels were derived from Majority Vote) over a four week time period. The circle corresponding to January 2021 has 633 articles. On average, there were 237 articles per month. This figure is generated using POLITICALKEYWORDS.

Results for Majority Vote Label Aggregation

Experiments and analyses with labels generated by majority vote are presented in this section. Table 6 shows the classification experiments for classic and novel hard news with the majority vote labels. Figure 4 and Figure 5 report the updated prevalence of hard news over time and publication behaviors surrounding key events based on the MACE labels.

Annotator Demographics

Our study involved 22 annotators aged 22–52 years (mean and median age: 36 years). The group comprised of 10 males and 12 females. Ethnic composition was predominantly White (15 participants), followed by Black (5) and Mixed ethnicity (2). All participants were U.S. nationals and native English speakers.

Regarding employment status, most were employed full-time (12 participants), others part-time (6), one was due to start a new job soon, one was unemployed and seeking work, and employment data was unavailable for two participants. Educational status included 5 students, 12 non-students, with data unavailable for five participants.

Overview of the Datasets Used in Our Analysis

- **ALLPEOPLE:** Includes all scraped articles from October 2016 to November 2023 without any filtering.
- **POLITICALKEYWORDS:** Comprises articles published between October 2016 and November 2023 that contain specified political keywords. Political keywords include: Supreme Court, Politics, Election, President, Democrat, Republican, Obama, Clinton, Romney, Trump, Biden, and Congress.
- **POLITICALEVENTS:** Consists of articles related to specific political events, published within two weeks before and after each event. The keywords for each event are:

Model	Classic F1	Classic Precision	Classic Recall	Novel F1	Novel Precision	Novel Recall
Logistic Regression	$0.790_{0.000}$	$0.791_{0.000}$	$0.790_{0.000}$	$0.689_{0.000}$	$0.687_{0.000}$	$0.692_{0.000}$
DistilBERT	$0.819_{0.005}$	$0.823_{0.005}$	$0.825_{0.006}$	$0.741_{0.001}$	$0.729_{0.003}$	$0.763_{0.010}$
RoBERTa	$0.803_{0.008}$	$0.804_{0.008}$	$0.804_{0.009}$	$\mathbf{0.750_{0.009}}$	$0.749_{0.003}$	$\mathbf{0.767_{0.023}}$
DeBERTa	$\mathbf{0.827_{0.008}}$	$\mathbf{0.828_{0.008}}$	$\mathbf{0.830_{0.008}}$	$0.735_{0.001}$	$\mathbf{0.751_{0.020}}$	$0.735_{0.012}$

Table 6: Comparison of Model Performance for Classic and Novel Hard News. Labels were derived from Majority Vote. Subscripts denote standard deviation across 3 seeds (1, 2, and 3). This figure is generated with ANNOTATEDPOLITICALKEYWORDS.

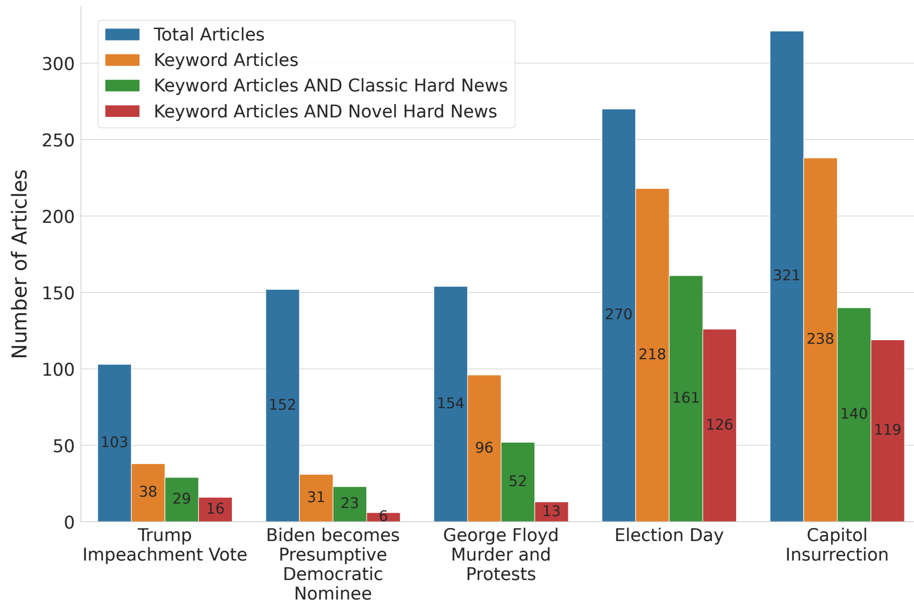

Figure 5: Article breakdown for 2 weeks following key events. Labels were derived from Majority Vote. Bars show total articles (all articles published on the day of, and the two weeks subsequent to the event), articles containing event-specific keywords, and keyword articles classified as classic or novel hard news. Number inside each bar denotes the number of articles within that particular category. On average, we observe 46% of all articles in the corpus to be novel hard news and 38% to be classic hard news. This figure is generated using POLITICALEVENTS.

Hyperparameters	Range
Epoch	{1,2,3,4,...,20}
Batch size	{16}
Classification Head Hidden Dimension	{512, 1024}
Learning rate	{loguniform(1e-6, 1e-4)}
Weight Decay	{loguniform(1e-4, 1e-2)}
Dropout	{uniform(0.1, 0.5)}

Table 7: Hyperparameter tuning was conducted using these ranges to determine the optimal settings.

  - **Trump Impeachment Vote:** impeachment, impeach, inquiry
  - **Biden Becomes Presumptive Democratic Nominee:** biden, democratic, nominee
  - **George Floyd Murder and Protests:** Black Lives Matter, George Floyd, protest, BLM, police, violence, george, floyd
  - **Election Day:** election, turnout, poll, election day
  - **Capitol Insurrection:** Capitol, insurrection, protest, jan 6th, 6th, storm, stolen, violence

- **Comparative Datasets:** Includes two datasets from other news sources:
  - Articles containing the keyword *election* from the *NYT* (1239 articles) and *DailyNews* (124 articles) during October-November 2020.
  - Annotated subsets of these articles: ANNOTATEDNYTand ANNOTATEDNYDAILY, each consisting of 41 articles, used to evaluate classifier generalizability.
  - Articles published during the last week of July 2019 (7/21/19 - 7/28/19) for both *NYT* and *DailyNews*, resulting in *NYT-7-19* and *DailyNews-7-19*. Non-news articles - such as opinions, obituaries, and newsletters - in *NYT*are removed. No keyword filtering is applied.

Concept	Krippendorff's Alpha	Simulated Krippendorff's Alpha
RECENT	0.24	0.02
MENTIONSPFIGURE	0.62	-0.01
MENTIONSCELEB	0.59	-0.01
CENTRALPFIGURE	0.63	0.00
CENTRALCELEB	0.59	0.03
CENTRALPEVENT	0.40	-0.01
CELEBIMPLICATIONS	0.36	-0.01
PIMPLICATIONS	0.40	0.03
NOVEL	0.38	0.00
CLASSIC	0.48	-0.00

Table 8: Simulated Krippendorff's Alpha was derived by simulating labels for each concept. Labels are drawn from a Bernoulli(p) distribution, where p is the prevalence of the *True* class for the particular concept based on unaggregated annotations.

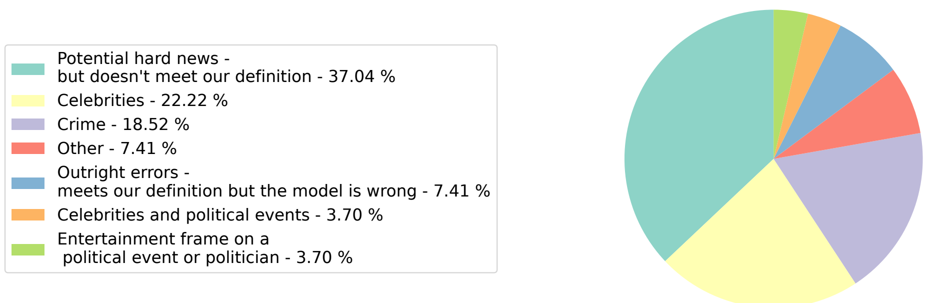

Figure 6: Categorization of soft news articles covering political events in *DailyNews* (N=27). Potential hard news not meeting definition criteria (37%) represent the largest category, followed by celebrities (22%).

Inspecting Soft Election Coverage on *DailyNews*

Generalization to other outlets We additionally inspect how other outlets cover events with soft news frames. For example, we inspect coverage of the same set of events as those included in the *People* analysis in the *DailyNews* articles. We follow the same approach where we manually inspect each article which is tagged with event key words and which our classifier predicts as being soft news.

Doing so we find that the majority of articles (81%) can be categorized using the taxonomy developed on *People*.

We find that the largest proportion of articles (37%) can be described with the category of potential hard news. This is to be expected, as our definition focuses on *political* hard news. The next most common category is described with the category of celebrities. However, we find that *DailyNews* also offers a category of *Crime* articles, which we define as having a focus on a crime that has occurred or police activity. While *People* also offers *Crime* articles, they are currently grouped under *Other* as they rarely occur in that outlet.

Although we find that that soft news outlets do contain some coverage of political events that reflects a hard news approach, we also find that these outlets cover political events in different ways than hard news outlets. Given their focus, however, this coverage is likely to be supplemented via articles connecting current political events to celebrity activity, the entertainment industry or framed through a more human-interest lens. While the exact way that each outlet employs this strategy likely varies, we expect that readers of soft news outlets are likely to be made aware of these events even if they only read these types of traditional soft news articles.

Robustness Check

We performed a robustness check to ensure that our results are robust to the idiosyncracies of individual annotators. That is, we selected the top-15 annotators (out of 22) and trained a supervised machine learning model using the aggregated labels from these 15 annotators, where each article in this training set still must have at least 3 annotations and an odd count. To determine the best annotators, we computed each annotators' Macro F1 score for the novel hard news classification task using the Dawid-Skene model. This resulted in all 15 annotators having a Macro F1 score higher than 0.70 and a training set of 280 articles (compared to 1026 in ANNOTATEDPOLITICALKEYWORDS). Due to time and resource constraints, we trained a logistic regression with TF-IDF embeddings to predict whether each article is hard news (both classic and novel). We find that the predicted rates of hard news in POLITICALKEYWORDSis robust: there were 38% classic and 16% novel hard news based on a model trained with the top annotators. In contrast, using all annotations from the full set of 22 annotators, we find 42% classic and 12% novel hard news in *People*'s political coverage. This assures us that the full dataset is of reasonable quality as to produce reliable results.