

# Calibrated and Diverse News Coverage

Tianyi Zhou

KTH Royal Institute of Technology  
Stockholm, Sweden  
tzho@kth.se

Kiran Garimella

Rutgers University  
New Brunswick, USA  
kg766@comminfo.rutgers.edu

Stefan Neumann\*

TU Wien  
Vienna, Austria  
stefan.neumann@tuwien.ac.at

Aristides Gionis

KTH Royal Institute of Technology  
Stockholm, Sweden  
argioni@kth.se

## ABSTRACT

In recent years, there has been a debate about whether automated news aggregators, like Google News, lead readers to content that reinforces their existing beliefs and restricts their exposure to a biased subset of perspectives. To avoid bias, it has become common practice that news aggregators provide articles based on source diversity: for each story, they pick articles from news sources with different political leanings. In this paper, we ask whether this practice is sufficient. In particular, we study how well the diversity of viewpoints, in particular with respect to entities, is covered by articles picked using plain source diversity. We analyze a dataset fetched from Google News and find that, even though the top articles exhibit some diversity with respect to the leanings of the news outlets, many possible viewpoints towards the entities are missing. Based on this observation we design novel methods for selecting a small set of articles that cover all possible viewpoints; to ensure that our selections are useful we show how to incorporate the user preferences into our model. Our experiments on four real-world datasets show that our algorithms cover significantly more different viewpoints than previous baselines.

## CCS CONCEPTS

- Information systems → Document representation; Web searching and information discovery.

## KEYWORDS

News coverage; Diversity; Calibration; Data summarization

### ACM Reference Format:

Tianyi Zhou, Stefan Neumann, Kiran Garimella, and Aristides Gionis. 2025. Calibrated and Diverse News Coverage. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management (CIKM '25), November 10–14, 2025, Seoul, Republic of Korea*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3746252.3761149>

\*This work was done while the author was at KTH Royal Institute of Technology.



This work is licensed under a Creative Commons Attribution International 4.0 License.

## 1 INTRODUCTION

The consumption of news in the digital age has undergone a significant transformation. With the abundance of information available through various online platforms, readers are often overwhelmed by the sheer volume of content, leading to limited attention span and selective exposure to news sources. This phenomenon has led individuals to primarily consume news that align with their existing beliefs and opinions [12]. Consequently, there is a growing need for diversification in news consumption to promote a well-informed and balanced perspective on current events.

Achieving this goal, however, is not without its challenges. Given the limited attention span of readers, *striking a balance between coverage of various topics and catering of user interests is essential*. For any news-aggregation system to be effective and practical, it must take into account user preferences to ensure that the presented content is well received by users, especially for polarizing topics.

Previous attempts to address the problem of providing well-calibrated and diverse news coverage have fallen short in several aspects. For instance, many platforms focus on low-quality/clickbait content [28], or on highly-personalized topics [22], seeking to maximize user engagement while ignoring diversity [19]. On the other side, other platforms may account for diversity by selecting some contrarian content [16], but fail to consider user preferences, therefore not succeeding in keeping the users engaged [11, 38].

News aggregators, such as Google News, have been criticized for their role in fostering biased consumption [16]. Such platforms typically aggregate articles based on a theme or a story, frequently drawing information from a single news source with a particular ideological bias, and without considering alternative sources [31]. As a result, users are not exposed to a wide variety of viewpoints. In response to such criticism, platforms have adjusted their algorithms to increase the *source diversity*, e.g., offering articles from news sources with different political leanings. The premise is that by offering articles from sources with different leanings, all other viewpoints (such as stances towards entities) are covered as well.

In this paper, we ask whether source diversity is sufficient to cover the entire range of viewpoints in a given set of news articles. One of our main focal points is to evaluate whether the stances towards entities are well-represented in systems that pick articles based on source diversity. This question may also resonate with some politicians and celebrities who claim that the media coverage towards their persona is one-sided.

**Our contributions.** We find that source diversity is indeed not sufficient to provide a full coverage of the viewpoints in collections of news articles. On a Google News dataset covering two weeks of news stories, we find that the top articles it recommends leave out many different viewpoints towards the entities, indicating a bias in the set of selected articles.

To avoid such biased selection, we present a new approach for selecting a small set of news articles that provide a diverse coverage of news stories and balance different points of view. For this selection, we ensure that the picked articles align with given user interests in different topics, establishing the practicality of our approach. We further introduce realistic and practical diversity measures by extending the news viewpoints with stance detection, allowing us to detect bias in news stories' coverage and stances extracted from news articles. We present novel combinatorial algorithms for covering all viewpoints present in a set of news articles, while still calibrating to user preferences. Our approach is based on an annotation pipeline we developed, which extracts meta-information from given news articles; our pipeline allows us to assess the bias of groups of news articles.

We extensively evaluate our algorithms and compare them against interpretable baselines. We find that even if news aggregators satisfy source diversity and present articles for all stories from diverse news sources (e.g., conservative, as well as liberal news outlets), there is still bias regarding the entities (like politicians) mentioned in the articles. We show that our algorithms can mitigate this issue, presenting a less-biased coverage of the entities, while still covering all stories and satisfying the user preferences.

All of our code is available online in an anonymous repository [5]. The repository also contains an extended version of this paper with more implementation details and additional experiments.

## 2 RELATED WORK

Researchers have been exploring ways to diversify users' news diets for over a decade. The goal is to expose readers to multiple viewpoints while respecting their time constraints and ensuring they derive maximum value from their news consumption. Abbar et al. [1] address the problem of identifying diverse recommendations for a specific news article a user is currently reading, while Nikookar et al. [30] consider the diversification of recommendations across multiple browsing sessions. Reuver et al. [32] use distributional language models to place users and news articles in a multi-dimensional semantic space, where diversity can be modeled as distance and variance.

**Stance bias in news.** In addition to source bias, stance detection plays a crucial role in analyzing news bias, as emphasized by Alam et al. [3]. Various approaches have been developed for this task, including zero-shot learning [4], BERT-based models [25], and few-shot as well as cross-domain methods [17, 18]. Chuang [9] presents a practical framework that leverages pretrained large language models. In contrast to prior work that formulates stance detection as a classification task targeting predefined topics, our method dynamically identifies theme-relevant entities within news articles and uses them as stance targets. This entity-centric approach enables the aggregation of multiple bias dimensions and is well-suited for deployment in real-world news aggregators.

**Viewpoints in news.** Mulder et al. [27] emphasize the difference between source diversity and viewpoint diversity. They note that existing systems, such as Google News, primarily diversify based on news sources rather than viewpoints. Unlike our paper, they do not consider stances towards entities and focus on re-ranking articles. Tintarev et al. [36] develop a new distance measure for diversity within a topic. Their approach enables diversity while maintaining topic relevance, adapting the maximal marginal relevance (MMR) technique [8]. While most methods in this space focus on content-based diversification, Advani et al. [2] propose methods for detecting and diversifying news presentations by reranking of the recommendations. In contrast, our method broadens the notion of viewpoint diversity by jointly modeling source diversity, thematic coherence across related stories, and stances toward key entities. This integrated perspective allows us to capture more nuanced and dynamic forms of bias, making our method more suitable for real-world applications where diverse perspectives are essential.

**User preference.** A critical aspect of implementing diverse news recommendation systems is understanding and modeling user acceptance. Loecherbach et al. [24] develop a system to study how users make choices regarding diverse news content. Their study underlines the importance of incorporating user acceptance into recommendation models. This aligns with the findings of Mulder et al. [27], and Tintarev et al. [36] further emphasizing that user acceptance of diversification must be addressed in tandem with algorithmic solutions to enable a complete and effective approach.

In this context, our research aims to contribute to this important field by exploring the trade-off between covering all viewpoints and maintaining a manageable set of selected news articles while considering user preferences. Unlike classic recommender systems, our goal is to find a small (non-ranked) set of articles that contains all viewpoints. Our approach distinguishes itself through a broader notion of diversity, encompassing political leaning, stance towards specific entities, and perspectives on particular topics. Crucially, we operationalize this comprehensive view of diversity in a practical manner by integrating user interests into our model.

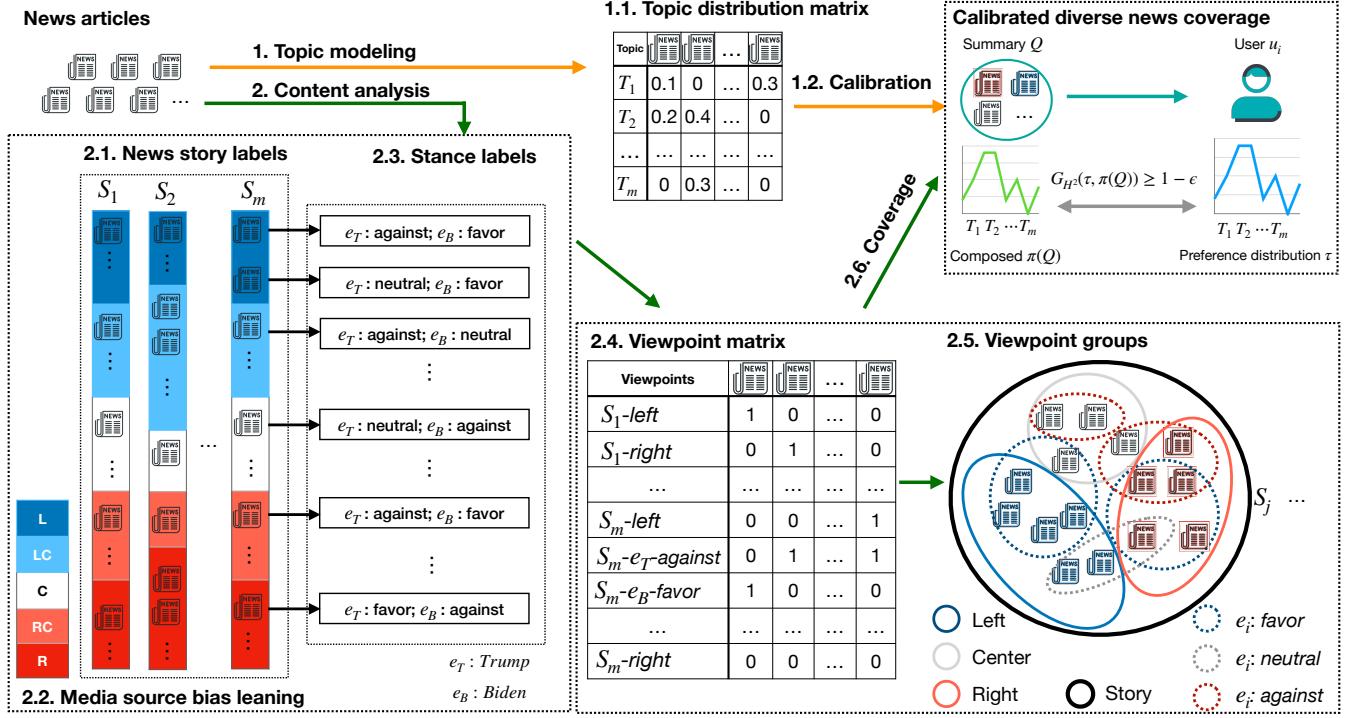
## 3 ANNOTATION PIPELINE

In this section we introduce our pipeline for processing news articles, which we refer to as DNA, for Diverse News Annotation. The pipeline utilizes semantic annotations to enhance news datasets by incorporating state of the art best practices from news-story discovery, topic modeling, entity extraction, and stance detection. An overview of our pipeline is given in Figure 1.

### 3.1 Input and output to the pipeline

**Input.** We start with a set of articles  $\mathcal{A} = \{a_1, \dots, a_n\}$ , each of which is given as plain text. For each article  $a_i$ , we also consider its time of publication  $t(a_i)$ , as well as its news source domain  $s(a_i)$  and we assume that for the news sources we know their political leaning. This is essentially a minimal input, and therefore we expect that our pipeline will be useful for other researchers in the future.

**Output.** The DNA pipeline enhances the input with various facets such as topics, story identification, entities, and stance labels.



**Figure 1:** Overview of the DNA pipeline, which analyzes news via topic modeling, story discovery, media bias labels, and stance detection. The orange path models user preferences by measuring the distance between selected articles  $Q$  and user profiles. The green path shows annotations: story labels (2.1), entities, and their stances (2.3). A bias matrix is then built from these labels (2.4), grouping articles by viewpoints (2.5). Finally, we cover all viewpoint groups while aligning with user preferences (2.6).

**Topic distributions.** First, we consider a set of  $m$  topics, e.g., politics, sports, etc. For each article  $a_i$ , we return a vector  $\mathbf{x}_i \in [0, 1]^m$ , which encodes the topic distribution of  $a_i$  over  $m$  topics, i.e.,  $\sum_{j=1}^m \mathbf{x}_i(j) = 1$ , where  $\mathbf{x}_i(j)$  is the fraction of content in article  $a_i$  about topic  $j$ . We let  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  denote the set of vectors of all articles. This step corresponds to Part 1.1 of Figure 1.

**News stories.** We return a grouping of the articles into a set of news stories  $\mathcal{S} = \{S_1, \dots, S_{|\mathcal{S}|}\}$ , where  $\mathcal{S}$  forms a partition of  $\mathcal{A}$ . Intuitively, each news story  $S_j$  consists of articles having a common theme. A story could be, e.g., a set of articles about the ceasefire in Gaza. This step corresponds to Part 2.1 of Figure 1. Note that, unlike for topics where each article may belong to multiple topics, each article is assigned to exactly one story.

**Entities.** For each article  $a_i$ , we return the  $entities E_i = \{e_1, e_2, \dots, e_{|E_i|}\}$  mentioned in the article. Entities can be, for instance, names of politicians or countries. We write  $E$  to denote the set of all entities.

**Bias labels.** We consider the *bias* of articles in different aspects. First, we consider the *political leaning of news sources*, e.g., whether they are left- or right-leaning. Second, we return the *stance toward the entities* mentioned in the articles, i.e., for each article we return whether its entities are discussed positively, negatively or in a neutral manner. This corresponds to Parts 2.2 and 2.3 of Figure 1.

### 3.2 Pipeline implementation

Next, we present an overview of the methods used to compute the outputs described in the previous section; more details are available in the online appendix [5]. We have employed state-of-the-art methods to ensure robustness and accuracy. Additionally, our implementation is designed for flexibility, allowing each subroutine to be easily replaced as advancements in algorithms occur.

**Topic modeling.** We approach topic modeling as a soft-clustering problem. We use BERTopic [14] and best practices reported therein to obtain article embeddings, article clusters, topic representations, and topic distributions for the article-topic vectors  $\mathbf{x}_i$ .

**News story discovery.** For news story discovery we use hard-clustering methods, i.e., each news article is assigned to only one story. Here, we employ the method proposed by Yoon et al. [39] to compute the stories  $S_j$ . The method considers the set of articles  $\mathcal{A}$  as a time-based stream and it clusters the articles based on their similarity, as computed in a sliding time window of a given size. This method achieves the best performance in our experiments, compared to other baselines [23, 26, 35], while it is robust and scalable.

**Entity detection.** We extract article entities using the BERT-based model NER<sup>1</sup> on a sentence level. We then postprocess the results to improve the data quality, where we take into account background

<sup>1</sup><https://huggingface.co/dslim/bert-base-NER>

information about politicians, political parties, and countries. This process gives us the set of entities  $E$ .

*Bias detection.* We obtain the biases for each article as follows. First, to obtain the *political leaning* of news sources, we use the data from Media Bias/Fact Check (MBFC), which is a widely used resource for evaluating bias and reliability of news sources [29, 33, 34]. The media source bias rating falls into the following five categories: {“left”, “left-center”, “center”, “right-center”, “right”}, and we label each article  $a_i$  with the leaning of its news source  $s(a_i)$ . Next, we consider the article’s *stances* towards entities with high frequency in each story, where we consider the labels {“in-favor”, “neutral-or-unclear”, “against”}. To obtain these labels, we use the zero-shot learning approach given by the FLAN-T5 model [10].

The code for our pipeline, which takes any set of news articles and enhances them with all the metadata is available online [5].

## 4 NOTATION AND PROBLEM DEFINITION

In this section, we formally introduce the problem that we study in this paper. Our goal is to select a subset of news articles from a collection of news articles, that takes into account the user interests (calibration), while simultaneously ensuring the diversity by coverage of different viewpoints.

*Calibration.* First we discuss how to measure calibration, that is, the alignment of a set of articles with the interests of a user. We consider an *ordered* set of articles  $Q \subseteq \mathcal{A}$  and a weight vector  $\mathbf{w} \in [0, 1]^{|Q|}$  with  $\|\mathbf{w}\|_1 = 1$  where vector  $\mathbf{w}$  represents the user’s attention. For simplicity, we assume selected items receive uniform attention and consider  $\mathbf{w}_i = \frac{1}{|Q|}$ , for all  $i$ .

The *composed topic distribution* of the articles  $Q$  is given by  $\pi(Q) = \sum_{i=1}^{|Q|} \mathbf{w}_i \mathbf{x}_i$ , where  $\mathbf{w}_i$  is the weight of the article  $i$ , and recall that  $\mathbf{x}_i$  represents the topic distribution of article  $i$ . Additionally, we let  $Q^{\leq j}$  be the subset of articles in  $Q$  assigned up to position  $j$ , and we set  $\pi(Q^{\leq j}) = \sum_{i=1}^j \mathbf{w}_i \mathbf{x}_i$  to the subdistribution composed by  $Q^{\leq j}$ . Note that  $\|\pi(Q^{\leq j})\|_1 \leq 1$  for all  $j \leq |Q|$ .

To measure how well a composed topic (sub-)distribution  $\pi$  approximates a target distribution  $\tau$ , we use the overlap measure

$$G_{H^2}(\tau, \pi) = \sum_{i=1}^m \sqrt{\tau_i \cdot \pi_i}, \quad (1)$$

which is the squared Hellinger distance and was proposed by Kleinberg et al. [20, Sec. 6] for its desirable properties: (1)  $G_{H^2}(\tau, \pi) \in [0, 1]$  for all  $\tau$  and  $\pi$ , (2) for fixed  $\tau$ ,  $G_{H^2}(\tau, \pi)$  is uniquely maximized at  $\tau = \pi$ , and (3)  $G_{H^2}(\tau, \pi)$  is a monotone and ordered-submodular function when adding articles to  $Q$  and when  $\tau$  and  $\mathbf{w}$  are fixed.

*Viewpoint groups.* We consider a *viewpoint grouping*  $C = \{C_1, \dots, C_{|C|}\}$ , where each *group*  $C_i$  is a subset of articles and we allow groups to overlap. Intuitively,  $C_i$  is a set of articles with the same viewpoint. For instance, a group  $C_i$  might contain all articles that have a negative stance towards a given entity, or all articles from left-leaning media outlets about a given story. See Part 2.5 in Figure 1.

We consider three concrete viewpoint groupings:

- The *story-viewpoint grouping* consists of groups  $C_{i,j}$ , for each story  $S_i$  and news outlet leaning  $j \in \{“left”, “left-center”, “center”, “right-center”, “right”\}$ . This is inspired by platforms like Google

News or Ground News<sup>2</sup> that aggregate and present information based on stories. However, they do not take into account stance bias within individual articles.

- The *entity-viewpoint grouping* consists of groups  $C_{e,j}$  containing all articles with entity  $e$  and news outlet leaning  $j \in \{“left”, “left-center”, “center”, “right-center”, “right”\}$ , as well as groups  $C_{e,k}$  containing all articles with entity  $e \in E$  and stance  $k \in \{“in-favor”, “neutral-or-unclear”, “against”\}$  towards  $e$ . Note that the entity-viewpoint grouping considers the viewpoints regarding entities, as well as the bias of media outlets, while disregarding specific story information.
- The *extensive-viewpoint grouping* is the most exhaustive version we consider. It consists of groups from the story-viewpoint grouping, as well as groups  $C_{i,e,j}$  containing all articles from story  $S_i$  with stance  $j$  towards entity  $e$ .

It is useful to encode the viewpoint grouping  $C$  in matrix form. Thus, we let  $\mathbf{d}_i \in \{0, 1\}^{|C|}$  denote the *viewpoint vector* of article  $a_i$  over the  $|C|$  groups, where  $\mathbf{d}_i(j) = 1$  indicates that article  $a_i \in C_j$ , i.e.,  $a_i$  expresses the viewpoint given by  $C_j$ , and  $\mathbf{d}_i(j) = 0$ , otherwise. We let  $\mathbf{D} \in \{0, 1\}^{|C| \times n}$  denote the *viewpoint matrix*, which contains each  $\mathbf{d}_i$  as its  $i$ -th column. This corresponds to 2.4 and 2.5 in Figure 1.

**Problem definition.** At a high level, we want to select a subset of articles that covers diverse viewpoints. Formally, our goal is to select a minimum number of articles  $Q$  such that *at least one* article from each viewpoint group  $C_i \in C$  is contained in  $Q$ , indicating that all viewpoint groups are covered. Since the groups  $C_i$  are overlapping, we can expect that  $|Q| \ll |C|$ . Further, observe that finding such a set  $Q$  of minimum cardinality is equivalent to solving the *hitting-set problem*, which is known to be NP-hard.

Besides taking into account the diversity of the articles, we are also interested in the user preferences. Thus, we introduce a utility constraint using the overlap measure from Equation (1) to guide the article-selection process to satisfy the user preferences. We consider a target distribution  $\tau$ , which represents the user preferences over the set of  $m$  topics. Given a selected subset of articles  $Q$  and the weight vector  $\mathbf{w}$ , we define the *user utility* as  $G_{H^2}(\tau, \pi(Q))$ . Naturally, we want to maximize the user utility.

Now, we formally present the calibrated diverse news coverage problem. The task is to select a minimum subset of articles that intersects all viewpoint groups, while covering topics that closely approximate the user preferences.

**Problem 1** (Calibrated diverse news coverage). *Given a set of articles  $\mathcal{A}$ , a viewpoint grouping  $C$ , the topic distribution vectors in  $\mathbf{X}$ , a threshold parameter  $\epsilon \in [0, 1]$ , and a target distribution  $\tau$ , the goal is to find a subset  $Q \subseteq \mathcal{A}$  of minimum size*

$$\text{such that } Q \cap C_i \neq \emptyset \quad \text{for all } C_i \in C, \text{ and} \\ G_{H^2}(\tau, \pi(Q)) \geq 1 - \epsilon. \quad (2)$$

The first constraint in Problem 1 is a *hitting-set constraint*, stating that all viewpoints must be covered by the selected articles. The second constraint is a *calibration constraint* given by the overlap measure, and a threshold parameter  $\epsilon$ .

Since the overlap measure is nonnegative, for  $\epsilon = 1$ , Problem 1 generalizes the hitting-set problem. Thus, we obtain the following:

<sup>2</sup><https://ground.news/>

**Lemma 1.** Problem 1 is NP-hard.

We also consider a budget-constrained version of Problem 1. This variant addresses the fact that the optimal size of  $Q$  might be larger than a user's attention. Thus, we introduce a budget parameter  $k$  to model the user's maximum attention, and we seek to maximize the coverage rate of subsets in the viewpoint group when picking at most  $k$  articles. The formal definition is as follows.

**Problem 2** (Calibrated diverse news coverage with a budget). Given a set of articles  $\mathcal{A}$ , a viewpoint group  $C$ , the topic distribution vectors in  $\mathbf{X}$ , a threshold parameter  $\epsilon \in [0, 1]$ , a target distribution  $\tau$ , and a budget parameter  $k$ , the goal is to find a subset  $Q \subseteq \mathcal{A}$  to satisfy

$$\begin{aligned} \max_{Q \subseteq \mathcal{A}} \quad & \sum_{C_i \in C} \mathbb{I}(Q \cap C_i \neq \emptyset) \\ \text{such that} \quad & G_{H^2}(\tau, \pi(Q)) \geq 1 - \epsilon \quad \text{and} \quad |Q| \leq k, \end{aligned} \quad (3)$$

where  $\mathbb{I}(\cdot)$  is the binary indicator function with  $\mathbb{I}(Q \cap C_i \neq \emptyset) = 1$  if  $Q \cap C_i \neq \emptyset$  and 0 otherwise.

Note that while both Problems 1 and 2 guarantee that the user's preferences are satisfied, Problem 2 takes the ordering of the selected articles in  $Q$  into account: it prioritizes articles that cover many unseen viewpoints due to the limited budget.

## 5 PROPOSED METHOD

We now describe our main algorithm, which combines a *greedy multi-objective optimization process* with *local-search postprocessing*. Our main challenge is that articles that are useful for satisfying the hitting-set constraint may not be helpful to improve the overlap measure from the calibration constraint, and vice versa. We consider two different strategies to trade-off between these competing goals.

Let us first define the gain obtained by adding an article  $a_i$  to a set of articles  $Q^{\leq j}$ . We let  $H(Q^{\leq j}, C) = \sum_{C_i \in C} \mathbb{I}(Q^{\leq j} \cap C_i \neq \emptyset)$  denote the number of subsets  $C_i$  intersecting (hitting) the first  $j$  items in  $Q$ .

The *coverage gain* of adding element  $a_i$  to  $Q$  is

$$F_H(a_i, Q^{\leq j}) = H(Q^{\leq j} \cup \{a_i\}, C) - H(Q^{\leq j}, C).$$

Observe that  $F_H(a_i, Q^{\leq j}) \in [0, |C|]$ , for all  $a_i$ .

Similarly, we define the *overlap-measure gain* as

$$F_{G_{H^2}}(a_i, Q^{\leq j}) = G_{H^2}(\tau, \pi(Q^{\leq j} \cup \{a_i\})) - G_{H^2}(\tau, \pi(Q^{\leq j})).$$

Observe that  $F_{G_{H^2}} \in [0, 1]$ .

To address Problems 1 and 2, we apply a greedy algorithm that iteratively selects the article with the highest gain. Since rankings by overlap and coverage gains may conflict and are on different scales, we propose two strategies to choose elements that improve both simultaneously.

**Strategy 1: Index minimization (Greedy-IM).** In our first strategy, we compute  $F_H(a_i, Q^{\leq j})$  and  $F_{G_{H^2}}(a_i, Q^{\leq j})$  for all  $a_i$  and we sort them by descending gains. Based on this ordering, we define the ordered article lists  $L_H$  and  $L_{G_{H^2}}$ . We let  $I(a_i, L_H)$  denote the index of  $a_i$  in list  $L_H$  and  $I(a_i, L_{G_{H^2}})$  the index of  $a_i$  in list  $L_{G_{H^2}}$ . Now the *index score* of article  $a_i$  is given by  $I(a_i) = I(a_i, L_H) + I(a_i, L_{G_{H^2}})$ . Note that low index scores correspond to elements that have high coverage gain and high overlap gain. Our strategy returns the article  $a_i$  with the smallest index score  $I(a_i)$ .

Note that Greedy-IM treats coverage and overlap gains independently, selecting the article ranked highest in both lists. This avoids issues from their differing scales. However, since only the order is considered, gains may be unevenly distributed, and the method cannot prioritize one aspect over the other.

**Strategy 2: Balanced objectives (Greedy-BO).** Our second strategy alleviates the problem mentioned above and allows us to trade-off between coverage and overlap gain.

More concretely, we consider the *normalized coverage rate gain*

$$\bar{F}_H(a_i, Q^{\leq j}) = \frac{F_H(a_i, Q^{\leq j})}{F_H^*(Q^{\leq j})},$$

where  $F_H^*(Q^{\leq j}) = \max_{a_\ell} F_H(a_\ell, Q^{\leq j})$  is the maximum coverage gain achieved by any article. Note that  $F_H^*(a_i, Q^{\leq j})$  could be 0 (if all subset in  $C$  are hit) and in this case we set  $\bar{F}_H(a_i, Q^{\leq j}) = 0$ . Similarly, we consider the *normalized overlap-measure gain*

$$\bar{F}_{G_{H^2}}(a_i, Q^{\leq j}) = \frac{F_{G_{H^2}}(a_i, Q^{\leq j})}{\max_{a_\ell} F_{G_{H^2}}(a_\ell, Q^{\leq j})}.$$

Due to the normalization, both scores are in the interval  $[0, 1]$ .

To trade-off between the normalized gains, we introduce a weight parameter  $\beta \in [0, 1]$ , which allows us to prioritize either calibration or coverage. We consider the following score:

$$F(a_i, Q^{\leq j}) = \beta \cdot \bar{F}_{G_{H^2}}(a_i, Q^{\leq j}) + (1 - \beta) \cdot \bar{F}_H(a_i, Q^{\leq j}).$$

Now the strategy chooses the article  $a_i$  with the largest  $F(a_i, Q^{\leq j})$ .

**Greedy solution.** To obtain a solution of  $k$  articles, we start by initializing an empty set of articles  $Q \leftarrow \emptyset$ . Then we greedily add articles to  $Q$  based on Greedy-IM or Greedy-BO until  $Q$  contains  $k$  articles. The time complexity is  $O(knm + kn \log n)$  where  $m$  is the size of the viewpoint group.

**Local search.** Optionally, we apply a post-processing step with local search to prune  $Q$  while covering all viewpoint sets  $C_j$  and meeting the calibration constraint. The search iterates until a constraint in Problem 1 is violated. In each iteration, it explores neighbors of  $Q$  by removing one article and adopts the solution that least increases the objective. When  $Q$  shrinks, the weight vector  $w$  used in  $\pi(Q)$  is updated by redistributing the final position's weight equally across the remaining ones.

### 5.1 Baselines

For our empirical evaluation we consider the following baselines.

**BL1: Hitting set.** As mentioned before, without the calibration constraint, Problem 1 is equivalent to the hitting-set problem. Therefore, BL1 is a greedy hitting-set algorithm, which we obtain by running Greedy-BO with  $\beta = 0$ . This baseline allows us to understand how “coverage without calibration” may differ from user preferences.

**BL2: Calibration.** BL2 selects a small set of articles aligned with the user's target distribution, ignoring coverage constraints. As in BL1, we set  $\beta = 1$  in Greedy-BO, replacing the objective with the calibration constraint and removing threshold  $\epsilon$ . This baseline shows how “calibration without coverage” can miss important topics.

**BL3: Maximal marginal relevance.** BL3 adapts the diversity-promoting *maximum marginal relevance (MMR)* method [6], which

**Table 1: Summary of our datasets. #Sources refers to the number of news sources among the articles.**

Dataset	#Articles	#Stories	#Groups	#Sources	#Entities
COVID-Jan	4 933	88	216	197	40
GoogleNews	5 713	110	971	889	157
NELA-Jan	9 094	97	1 055	110	105
WCEP-2018	11 591	225	1 005	389	146

iteratively ranks items by a trade-off between relevance and diversity. Bourgeois et al. [6] applied MMR to mitigate news selection bias; we adapt it for article ranking using

$$MMR(a_i) := \beta \cdot relevance(a_i, \tau) - (1 - \beta) \cdot \max_{a_j \in Q^{\leq k}} sim(a_i, a_j), \quad (4)$$

where  $\beta$  controls diversification and  $Q^{\leq k}$  is the set already chosen. Relevance is defined as  $F_{G_{H^2}}(a_i, Q^{\leq k})$ , and similarity as cosine similarity of bias vectors  $d_i$  and  $d_j$ . Thus, relevance favors high overlap gain, while diversification prefers articles differing in viewpoints from the selected set.

**BL4: Nonnegative orthogonal matching pursuit (NOMP).** Lapras et al. [21] used the NOMP algorithm [7] for selecting a representative subset of elements. Formally, given a target vector  $\tau_D$  and an integer  $k$ , NOMP aims to find a binary vector  $z_Q \in \{0, 1\}^n$ , which contains at most  $|Q|$  1-entries, such that  $\|\tau - Dz_Q\|_2^2$  is minimized. We adopt NOMP to find subsets of articles to approximate the target distributions based on  $X$  and  $D$ . In particular, we use the average vector of  $D$  as the target vector  $\tau_D$  and apply NOMP to compute  $z_Q$  with budget  $k$ . We pick the articles corresponding to non-zero entries in  $z_Q$  to compose the selected articles.

## 6 EXPERIMENTAL EVALUATION

We empirically evaluate our method and compare with the baselines. For a detailed evaluation please refer to the online appendix [5]. Our implementation is also available online [5].

### 6.1 Datasets and experimental setup

**Datasets.** We use data from Google News and three public datasets. The GoogleNews dataset covers Nov 25–Dec 8, 2024, including all articles from the full coverage page of the top 10 US stories each day ( $\approx 50$  articles per story). The COVID dataset, provided by AYLIEN,<sup>3</sup> contains English articles from 440 sources between Nov 2019–Jul 2020, all related to COVID-19. The NELA [15] dataset includes 1.78M articles from 361 sources spanning 2022. Finally, the WCEP [13] dataset, built from the *Wikimedia Current Event Portal* and *Common Crawl Archive*, provides ground-truth story labels.

We analyze different time spans with our DNA pipeline for topic modeling, story discovery, entity detection, and stance annotation: two weeks for GoogleNews, one month for COVID and NELA, and one year for WCEP. Table 1 summarizes the annotated datasets. The COVID-Jan and NELA-Jan subsets use articles from Jan 2020 and Jan 2022, respectively, while WCEP-2018 contains 2018 news stories with at least 50 articles.

<sup>3</sup><https://aylien.com/>

**Experimental environment.** We conduct our experiments on a Linux server with 2 AMD Epyc 7742 CPUs, 1 TB of RAM and 1 NVIDIA DGX-A100 GPU. Our code is written in Python v3.11.7.

**Target distributions.** To generate realistic user target distributions  $\tau$ , we analyzed the news consumption of topics by users in the Microsoft News dataset [37]. We found that users’ interests show an exponential decay when sorting the topics based on how often a user clicks on them. Inspired by this observation, we generate our target distributions as follows. First, we pick a number of seed topics uniformly at random from the  $m$  topics. We then compute the top 10 most similar topics to each seed topic by their keyword representation in BERTopic [14]. We obtain  $\tau$  by assigning exponentially decaying weights to the seed topic and its similar topics.

**Algorithms.** We write Greedy-IM and Greedy-BO to denote the two greedy strategies in Section 5; we write Greedy-IM-LS when combining Greedy-IM with local-search post-processing. The baselines are denoted by  $BL_i$ , for  $i \in \{1, 2, 3, 4\}$ . We solve Problem 2 and refer to the parameter  $k$  as *budget*, i.e., the number of selected articles.

For the number of articles,  $k$ , picked by our algorithms, we do *not* report absolute numbers. Instead, to ensure that the results are interpretable across different datasets, we report the number of articles *per story in the dataset*. Here, our reasoning is that if there are more stories, the algorithms need more budget and that the number of entities should grow with the number of stories.

**Evaluation measures.** We evaluate our algorithms using the coverage objective function  $\sum_{C_i \in C} \mathbb{I}(Q \cap C_i \neq \emptyset)$  from Problem 2 and the overlap measure  $G_{H^2}$  from Equation (1).

We also consider the *entity balance score*, which measures the balance of news articles with respect to stances towards the entities. Formally, given a set of articles  $Q$ , we consider the balance vector  $b \in [0, 1]^{3|E|}$ , which contains three entries for each entity  $e \in E$ , corresponding to the fraction of articles in  $Q$  that have stance positive/neutral/negative towards  $e$ . Ideally, each fraction equals  $\frac{1}{3}$ . We measure balance by overlap with the uniform distribution, which we prefer over calibration, since flooding users with many items of the same stance—even if representative—should be avoided. In particular, we define the *entity balance score* as  $B(Q) = G_{H^2}\left(\frac{1}{|E|}b, \frac{1}{3|E|}\mathbf{1}\right)$ , where  $\mathbf{1}$  is the all-ones vector and the normalization ensures that both vectors are distributions.

### 6.2 Experiments

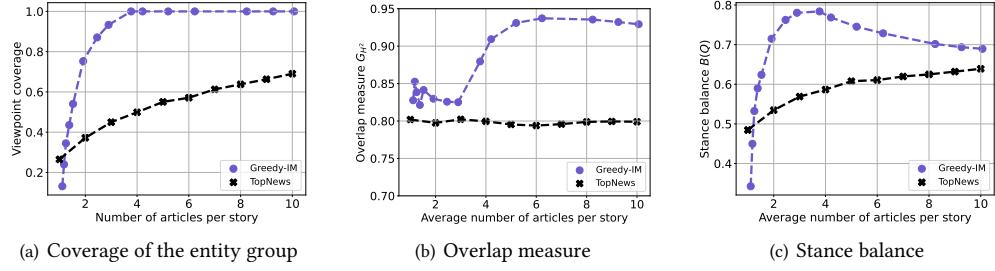
Our evaluation aims to answer the following questions:

**Q1:** Does the full coverage feature on news aggregators like Google News lead to biased news selection?

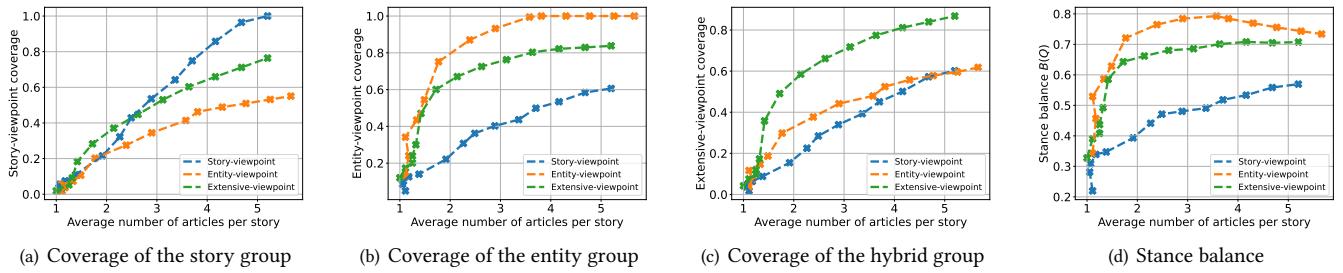
**Q2:** What is the interplay between different coverage objectives? In particular, if we seek to cover story-based viewpoints, how well are entity-based viewpoints covered?

**Q3:** What are the best-performing methods and best settings to achieve diverse viewpoints coverage while considering user preferences? Can we achieve an effective trade-off between coverage and user preferences?

**Q1: News aggregator limitations.** To study the limitations of existing news aggregators, we consider the GoogleNews dataset and build a baseline that is based on Google News’ “Top Stories” feature, denoted as TopNews. We are interested in how well the



**Figure 2: Performance of Greedy-IM and TopNews for covering the entity-viewpoint grouping on GoogleNews.** The  $y$ -axis in (a) shows the coverage rate of the entity-viewpoint grouping, in (b) the overlap measure  $G_{H^2}$ , and in (c) the stance balance  $B(Q)$ . The  $x$ -axis shows the budget in an average number of articles per story. We report the best result for each budget.



**Figure 3: Performance of Greedy-IM on GoogleNews for covering story-viewpoint, entity-viewpoint, and extensive-viewpoint groupings.** The  $y$ -axis shows coverage of each grouping in (a)–(c), and the stance balance score  $B(Q)$  in (d). The  $x$ -axis shows the number of selected articles, normalized by story count.

TopNews covers all viewpoints. For each day, TopNews selects the 10 highest ranked stories on Google News and for each story it picks articles among the top 10 articles suggested in the ranking by Google News for this story. The articles are picked such that the coverage of the viewpoints is maximized; note that due to this strategy, the coverage we obtain will be an *upper bound* on the coverage that a user may obtain who reads the same number of articles for each story (since the user might pick suboptimally).

We present the results of TopNews applied to the GoogleNews dataset in Figure 2 and compare it with Greedy-IM. Both algorithms select  $k$  articles per story (10 stories total), with  $k$  on the  $x$ -axis. Figures 2(a)–2(c) demonstrate that Greedy-IM outperforms TopNews in viewpoint coverage, calibration, and balance. With about 4 articles per story, Greedy-IM reaches full coverage, strong calibration, and a high stance balance score. By contrast, TopNews covers only 48% of viewpoints, yields an overlap of 0.8, and a stance balance score  $\leq 0.6$ . Moreover, TopNews shows slow gains in coverage and balance as  $k$  grows, while overlap remains nearly unchanged.

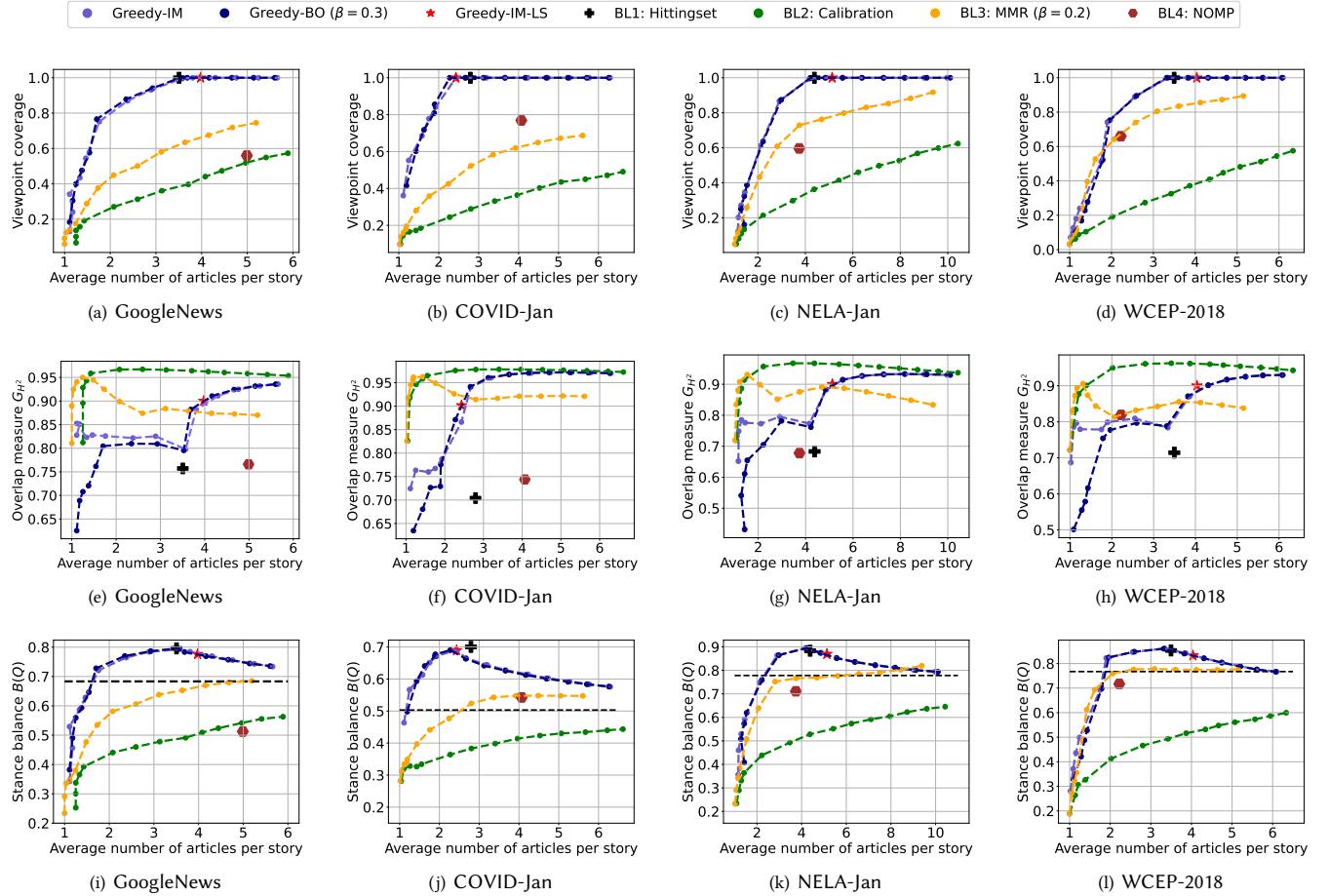
In summary, we show that by following Google News’ recommendations, one faces challenges in facilitating unbiased news consumption, struggling to ensure comprehensive viewpoint coverage while adequately aligning with user preferences. Greedy-IM shows the ability to achieve full viewpoint coverage along with strong calibration to user preference with a small budget. Notably, the Google News website displays four articles as thumbnails for

**Table 2: Performance of Greedy-IM-LS across viewpoint groupings and datasets.** The second column specifies the objective used; the budget is the number of articles per story. For example, the first row shows that with objective story-viewpoint, Greedy-IM-LS needs 5.2 articles per story to fully cover story-viewpoint, while covering 60% of both entity-viewpoint and extensive-viewpoint.

Dataset	Coverage objective	Budget	$B(Q)$	Viewpoint coverage (%)		
				entity	story	extensive
GoogleNews	story	5.2	0.57	60	100	60
	entity	3.9	0.77	100	47	53
	extensive	8.7	0.71	90	100	100
COVID-Jan	story	3.9	0.46	71	100	76
	entity	2.4	0.69	100	26	33
	extensive	5.1	0.58	92	100	100
NELA-Jan	story	5.0	0.63	55	100	67
	entity	5.1	0.86	100	55	60
	extensive	6.9	0.85	91	100	100
WCEP-2018	story	5.7	0.72	88	100	85
	entity	4.0	0.82	100	48	52
	extensive	6.4	0.78	97	100	100

each story, which matches the number of articles our algorithm picked to achieve full viewpoint coverage.

**Q2: Interaction of different viewpoint coverage objectives.** We study how the algorithm covers different viewpoint groupings when optimizing for specific objectives, analyzing the interaction



**Figure 4: Performance of all algorithms and baselines on different datasets when optimizing for the entity-viewpoint grouping. The  $y$ -axis in (a)–(d) shows coverage rates; in (e)–(h), the overlap measure  $G_{H^2}$ ; and in (i)–(l), the stance balance  $B(Q)$ , with the entity balance score for the full set shown as a dashed line. We vary the budget  $k$  and average results over five randomly generated user preference vectors.**

between coverage objectives and groupings. Figure 3 shows results of Greedy-IM optimizing different objectives and evaluating coverage across three groups. In Figures 3(a)–3(c), we observe that (unsurprisingly) optimizing for the objective aligned with its corresponding viewpoint grouping achieves the best performance.

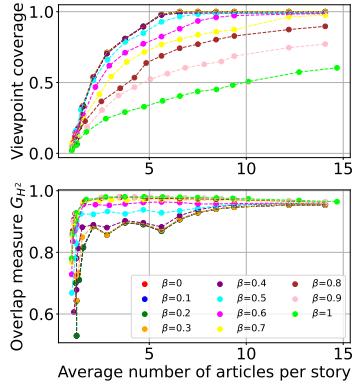
We observe that optimizing for the story-viewpoint objectives results in a low coverage for the entity-viewpoint grouping, and similarly optimizing for the entity-viewpoint objectives performs poorly for covering the story-viewpoint grouping. This indicates that when only considering the coverage of a single objective (either stories or entities), this will still lead to a biased set of articles with respect to entity stances. This result highlights that enforcing only news source diversity in news aggregators can lead to biased recommendations. Notably, optimizing for the extensive-viewpoint objective strikes an excellent balance, achieving a strong trade-off in coverage across different objectives and viewpoints.

Next, in Table 2 we report the performance of Greedy-IM-LS for covering the three viewpoint groupings on all datasets. We ensure

the grouping tied to the objective is fully covered and the overlap score satisfies  $G_{H^2} \geq 0.9$ . Reported are the minimum budget, stance balance  $B(Q)$ , and coverage of the entity-viewpoint, story-viewpoint, and extensive-viewpoint groupings. Covering entity-viewpoint requires the smallest budget for its objective and achieves the highest balance score across datasets.

By contrast, covering the story-viewpoint grouping requires a larger budget to achieve the desired goal. This can be attributed to the fact that each article belongs to one story, but it may contain multiple entities. However, we observe that it obtains the smallest entity balance score. Again, this result highlights that by solely enforcing source diversity in news aggregators, one might end up with biased stances towards the entities.

On the other hand, covering the extensive-viewpoint grouping results in missing about 3–10% in coverage of the entity-viewpoint grouping; except on GoogleNews, its budget requirement is significantly less than summing the budgets for covering both story-viewpoint and entity-viewpoint groupings. Except on COVID, it



**Figure 5: Performance of Greedy-BO with varying parameter  $\beta \in [0, 1]$  on NELA-Jan.** The  $y$ -axis on the first plot shows the coverage rate of the extensive-viewpoint group, and the  $y$ -axis in the second plot shows the overlap measure score  $G_{H^2}$ .

achieves balance scores comparable to those obtained when covering the entity-viewpoint grouping.

In summary, covering story-viewpoint risks biased content by ignoring entity stances. When all stories must be included, extensive-viewpoint yields more balanced results than story-viewpoint and also covers entities, though at the cost of more articles. When full story coverage is unnecessary, entity-viewpoint is effective for calibrated, diverse coverage: it balances stances, covers many stories from sources with different biases, satisfies user preferences, and uses only a moderate budget.

**Q3: Comparison of methods and varying budget.** Next, we compare Greedy-IM and Greedy-BO against the baselines when varying the budget parameter  $k$ . We consider covering the entity-viewpoint grouping on different datasets. The results are reported in Figure 4.

First, we report the result of the coverage rates in Figures 4(a)–(d). As expected, BL1 achieves full coverage with the smallest budget (2–4 articles per story), while BL2 achieves the lowest coverage. Greedy-IM and Greedy-BO perform similarly to BL1, achieving full coverage with the same budget in all experiments, and they outperform BL2, BL3 and BL4 constantly by a large margin. We observe that initially the coverage increases sharply with  $k$ , but then flattens off.

Next, we report the result for the overlap measure  $G_{H^2}$  in Figures 4(e)–(h). A higher overlap score indicates that the selected articles are well-calibrated to the user’s target distribution  $\tau$ . As expected, BL2 achieves the highest overlap score (around 0.97 for 2 articles per story), while BL1 achieves the lowest (around 0.7). Greedy-IM and Greedy-BO perform similarly after selecting 2 articles per story and show an overall upward trend as the budget increases. For small budget, Greedy-IM and Greedy-BO behave differently, with Greedy-IM achieving  $G_{H^2} \geq 0.75$  early on.

Next, in Figures 4(i)–(l) we report the entity bias score of the selected articles  $Q$ . Notice that although the coverage constraint guarantees to cover all possible stances towards entities, it is possible that the distribution of stances is skewed, e.g., for some entities there might be more articles in favor of them than against. We find that Greedy-IM and Greedy-BO have less bias than the bias of the entire set of articles (given by the dashed line) after covering about 50% of the group. As the budget increases, their bias converges to

the average bias (as expected). BL1 and BL4 also achieve good results here, whereas BL2 and BL3 perform worse than our algorithms.

In summary, the results for BL1 and BL2 show a large gap between the goal of unbiased coverage and calibration: unbiased articles may fail to satisfy users, while well-calibrated articles may lead to biased content. The results of BL2 particularly highlight that when overfitting to user preferences, the coverage of the viewpoints is low and the stance balance is more biased than the average of the article collection and much worse than our methods. Greedy-IM and Greedy-BO can effectively balance the goals of providing unbiased coverage and calibration. With the local-search improvement, a budget of just 2–4 articles per story is sufficient to create an unbiased news summary tailored to user preferences ( $G_{H^2} \geq 0.9$ ).

Next, we explore the trade off between coverage and calibration. In Figure 5 we run Greedy-BO with different choices of  $\beta$ . Recall that large values of  $\beta$  emphasize user preferences, whereas small values of  $\beta$  emphasize coverage. As expected, the algorithm’s behavior aligns well with the choice of  $\beta$ . This therefore provides a meaningful trade off between coverage and calibration.

## 7 CONCLUSION, LIMITATIONS, AND FUTURE WORK

We study the problem of calibrated and diverse news coverage. Our motivation is to test whether source diversity alone ensures full viewpoint coverage. We show it does not, as many entity stances remain underrepresented when only source diversity is enforced, including in Google News recommendations. To address this, we develop algorithms that jointly cover stories and entity stances, while respecting user interests and reducing bias.

While our study advances methods for relevant and diverse news coverage, it also has limitations and points to future work. First, we rely on synthetic user preference distributions inferred from real histories, which enable controlled experiments but may not capture real behaviors; user studies could improve robustness. Second, our model assumes a fixed time window, whereas temporal or streaming settings would better reflect evolving news and preferences. Third, we focus on entity-based stance detection, leaving other targets such as events or abstract concepts unexplored.

Designing recommender systems that satisfy user preferences while meeting coverage objectives is a promising direction. While we focus on aggregating a small set of diverse articles in static settings, future work could incorporate users’ reading histories as new articles appear. Moreover, assessing user-centered outcomes such as satisfaction requires real-world deployment, which introduces scalability and implementation challenges. We leave these questions for future investigation.

## ACKNOWLEDGMENTS

This research was supported by the ERC Advanced Grant RE-BOUND [834862], the Vienna Science and Technology Fund (WWTF) [Grant ID:10.47379/VRG23013], the Swedish Research Council (VR) [2024-05603], the European Commission MSCA DN [101168951], the US National Science Foundation (NSF) [2318843], and the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

## GENERATIVE AI USAGE DISCLOSURE

We acknowledge the use of generative AI tools during the development of this work. Specifically, LLMs, like ChatGPT, were employed to assist with surveying related literature during the early research phase, helping to identify and summarize relevant papers. Generative models were also used to write scripts for data collection and preprocessing, such as formatting and visualizing data; these scripts were used only during preliminary analysis and are not part of the final experimental pipeline or released codebase. Furthermore, our proposed method integrates a generative model directly into the stance detection component of the pipeline. In the preliminary stages of the project, we also evaluated LLM-based stance detection using standard zero-shot and few-shot prompting approaches, following best practices in the field. On the other hand, no generative AI was used in the writing of the present manuscript.

## REFERENCES

- [1] Sofiane Abbar, Sihem Amer-Yahia, Piotr Indyk, and Sepideh Mahabadi. Real-time recommendation of diverse related articles. In *WWW*, pages 1–12, 2013.
- [2] Rishi Advani, Paolo Papotti, and Abolfazl Asudeh. Maximizing neutrality in news ordering. In *KDD*, pages 11–24, 2023.
- [3] Mehwish Alam, Andreea Iana, Alexander Grote, Katharina Ludwig, Philipp Müller, and Heiko Paulheim. Towards Analyzing the Bias of News Recommender Systems Using Sentiment and Stance Detection. In *WWW (Companion Volume)*, pages 448–457, Virtual Event, Lyon France, April 2022.
- [4] Emily Allaway and Kathleen McKeown. Zero-Shot Stance Detection: A Dataset and Model using Generalized Topic Representations. In *EMNLP*, pages 8913–8931, Online, 2020.
- [5] Anonymous authors. Appendix, code, and datasets. <https://anonymous.4open.science/r/DNA-pipeline-7868>.
- [6] Dylan Bourgeois, Jérémie Rappaz, and Karl Aberer. Selection bias in news coverage: Learning it, fighting it. In *WWW*, pages 535–543, 2018.
- [7] Alfred M Bruckstein, Michael Elad, and Michael Zibulevsky. On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations. *IEEE Transactions on Information Theory*, 54(11):4813–4820, 2008.
- [8] Jaime Carbonell and Jade Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *SIGIR*, pages 335–336, 1998.
- [9] Yun-Shiuan Chuang. Tutorials on Stance Detection using Pre-trained Language Models: Fine-tuning BERT and Prompting Large Language Models, July 2023. URL <http://arxiv.org/abs/2307.15331>. arXiv:2307.15331 [cs].
- [10] Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Eric Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, Albert Webson, Shixiang Shane Gu, Zhuyun Dai, Mirac Suzgun, Xinyun Chen, Aakanksha Chowdhery, Sharan Narang, Gaurav Mishra, Adams Yu, Vincent Zhao, Yanping Huang, Andrew Dai, Hongkun Yu, Slav Petrov, Ed H. Chi, Jeff Dean, Jacob Devlin, Adam Roberts, Denny Zhou, Quoc V. Le, and Jason Wei. Scaling instruction-finetuned language models, 2022. URL <https://arxiv.org/abs/2210.11416>.
- [11] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Reducing controversy by connecting opposing views. In *WSDM*, pages 81–90, 2017.
- [12] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In *WWW*, pages 913–922, 2018.
- [13] Demian Gholipour Ghalandari, Chris Hokamp, Nghia The Pham, John Glover, and Georgiana Ifrim. A large-scale multi-document summarization dataset from the wikipedia current events portal. In *ACL*, pages 1302–1308, 2020.
- [14] Maarten Grootendorst. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*, 2022.
- [15] Maurício Gruppi, Benjamin D. Horne, and Sibel Adali. NELA-GT-2021: A large multi-labelled news dataset for the study of misinformation in news articles. *CoRR*, abs/2203.05659, 2022. doi: 10.48550/ARXIV.2203.05659. URL <https://doi.org/10.48550/arXiv.2203.05659>.
- [16] Mario Haim, Andreas Graefe, and Hans-Bernd Brosius. Burst of the filter bubble? effects of personalization on the diversity of google news. *Digital journalism*, 6(3):330–343, 2018.
- [17] Momchil Hardalov, Arnav Arora, Preslav Nakov, and Isabelle Augenstein. Cross-Domain Label-Adaptive Stance Detection, September 2021. URL <http://arxiv.org/abs/2104.07467>. arXiv:2104.07467 [cs].
- [18] Yan Jiang, Jinhua Gao, Huawei Shen, and Xueqi Cheng. Few-Shot Stance Detection via Target-Aware Prompt Distillation. In *SIGIR*, pages 837–847, July 2022.
- [19] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. The challenge of understanding what users want: Inconsistent preferences and engagement optimization. *Management science*, 2023.
- [20] Jon M. Kleinberg, Emily Ryu, and Éva Tardos. Calibrated recommendations for users with decaying attention. In *SAGT*, volume 15156, pages 443–460, 2024.
- [21] Theodoros Lappas, Mark Crovella, and Evmaria Terzi. Selecting a characteristic set of reviews. In Qiang Yang, Deepak Agarwal, and Jian Pei, editors, *KDD*, pages 832–840, 2012.
- [22] Huyen Le, Raven Maragh, Brian Ekdale, Andrew High, Timothy Havens, and Zubair Shafiq. Measuring political personalization of google news search. In *WWW*, pages 2957–2963, 2019.
- [23] Bang Liu, Di Niú, Kunfeng Lai, Linglong Kong, and Yu Xu. Growing story forest online from massive breaking news. In *CIKM*, pages 777–785, 2017.
- [24] Felicia Loecherbach, Kasper Welbers, Judith Moeller, Damian Trilling, and Wouter Van Atteveldt. Is this a click towards diversity? Explaining when and why news users make diverse choices. In *WebSci*, pages 282–290, 2021.
- [25] Laura Mascarell, Tatyana Ruzsics, Christian Schneebeli, Philippe Schlattner, Luca Campanella, Severin Klingler, and Cristina Kadar. Stance Detection in German News Articles. In *FEVER*, pages 66–77, 2021.
- [26] Sebastião Miranda, Arturs Znotins, Shay B. Cohen, and Guntis Barzdins. Multi-lingual clustering of streaming news. In *EMNLP*, pages 4535–4544, 2018.
- [27] Mats Mulder, Oana Inel, Jasper Oosterman, and Nava Tintarev. Operationalizing Framing to Support Multiperspective Recommendations of Opinion Pieces. In *EAccT*, pages 478–488, Virtual Event Canada, 2021.
- [28] Kevin Munger. All the news that's fit to click: The economics of clickbait media. *Political Communication*, 37(3):376–397, 2020.
- [29] Preslav Nakov, Giovanni Da San Martino, et al. Fact-checking, fake news, propaganda, and media bias: Truth seeking in the post-truth era. In *EMNLP*, pages 7–19, 2020.
- [30] Sepideh Nikookar, Mohammadreza Esfandiari, Ria Mae Borromeo, Paras Sakharak, Sihem Amer-Yahia, and Senjuti Basu Roy. Diversifying recommendations on sequences of sets. *The VLDB Journal*, 32(2):283–304, 2023.
- [31] Cornelius Puschmann. Beyond the bubble: Assessing the diversity of political search results. *Digital Journalism*, 7(6):824–843, 2019.
- [32] Myrthe Reuver, Nicolas Mattis, Marijn Sax, Suzan Verberne, Nava Tintarev, Natali Helberger, Judith Moeller, Sanne Vrijenhoek, Antske Fokkens, and Wouter Van Atteveldt. Are we human, or are we users? The role of natural language processing in human-centric news recommenders that nudge users to diverse content. In *Proceedings of the 1st Workshop on NLP for Positive Impact*, pages 47–59, Online, 2021.
- [33] Ronald E Robertson, Jon Green, Damian J Ruck, Katherine Ognyanova, Christo Wilson, and David Lazer. Users choose to engage with more partisan news than they are exposed to on google search. *Nature*, 618(7964):342–348, 2023.
- [34] Ivan Srba, Branislav Pecher, Matus Tomlein, Robert Moro, Elena Stefancova, Jakub Simko, and Maria Bielikova. Monant medical misinformation dataset: Mapping articles to fact-checked claims. In *SIGIR*, 2022.
- [35] Todor Staykovski, Alberto Barrón-Cedeño, Giovanni Da San Martino, and Preslav Nakov. Dense vs. sparse representations for news stream clustering. In Alíprio Mário Jorge, Ricardo Campos, Adam Jatowt, and Sumit Bhatia, editors, *Text2Story@ECIR*, volume 2342 of *CEUR Workshop Proceedings*, pages 47–52. CEUR-WS.org, 2019.
- [36] Nava Tintarev, Emily Sullivan, Dror Guldin, Sihang Qiu, and Daan Odijk. Same, Same, but Different: Algorithmic Diversification of Viewpoints in News. In *UMAP*, pages 7–13, 2018.
- [37] Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhuan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, et al. Mind: A large-scale dataset for news recommendation. In *ACL*, pages 3597–3606, 2020.
- [38] De-Nian Yang, Hui-Ju Hung, Wang-Chien Lee, and Wei Chen. Maximizing acceptance probability for active friending in online social networks. In *KDD*, pages 713–721, 2013.
- [39] Susik Yoon, Dongha Lee, Yunyi Zhang, and Jiawei Han. Unsupervised story discovery from continuous news streams via scalable thematic embedding. In *SIGIR*, pages 802–811, 2023.