

---

Article

Not peer-reviewed version

---

# Perspective-based Microblog Summarization

---

[Chih-Yuan Li](#) <sup>\*</sup> , [Soon Ae Chun](#) <sup>\*</sup> , [James Geller](#)

Posted Date: 24 February 2025

doi: [10.20944/preprints202502.1889.v1](https://doi.org/10.20944/preprints202502.1889.v1)

Keywords: User perspective-based Summarization; social media; COVID-19 Vaccine Tweets; Semantic triples; microblog summarization; entity; event and social features for summarization



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Perspective-Based Microblog Summarization<sup>†</sup>

Chih-Yuan Li <sup>1,\*</sup>, Soon Ae Chun <sup>2,\*</sup> and James Geller <sup>3</sup>

<sup>1</sup> St. Francis College, Brooklyn, NY, USA, cli@sfc.edu

<sup>2</sup> City University of New York College of Staten Island, New York, NY, USA, soon.chun@csi.cuny.edu

<sup>3</sup> New Jersey Institute of Technology, Newark, NJ, USA, james.geller@njit.edu

\* Correspondence: cli@sfc.edu (C.-Y. L.) and soon.chun@csi.cuny.edu (S.A.C.)

<sup>†</sup> This article is a revised and expanded version of a paper entitled Multiple View Summarization Framework for Social Media, which was presented at FLAIRS-36: 36th International Florida Artificial Intelligence Research Society Conference, May 14-17, 2023.

**Abstract:** Social media allows people to express and share a variety of users' experiences, opinions, beliefs, interpretations, or viewpoints on a single topic. Summarizing a collection of social media textposts (microblogs) on one topic may be challenging and can result in an incoherent summary due to these multiple perspectives by different users. We introduce an approach of microblog summarization based on user perspectives, called Multiple View Summarization Framework (MVSF), a novel approach designed to efficiently generate multiple summaries from the same social media dataset depending on a chosen perspective, delivering personalized and fine-grained summaries. The MVSF leverages component-of-perspective computing that can recognize the perspectives expressed in microblogs, such as sentiments, critical views, political orientations, or unreliable opinions (fake news), etc. The perspective computing can filter social media data to summarize them according to a specific user-selected perspective. For the summarization methods, our framework implements three extractive summarization methods: Entity-based, Social Signal-based, and Triple-based. We conduct comparative evaluations of MVSF summarizations against state-of-the-art summarization models, including BertSum, SBert, T5, and Bart-Large-CNN, by using a gold standard BBC news dataset and Rouge scores. Furthermore, we utilized a dataset of 18,047 tweets about COVID-19 vaccines to demonstrate the applications of MVSF. Our contributions include the innovative approach of using user perspectives in summarization methods as a unified framework, capable of generating multiple summaries that reflect different perspectives, in contrast to prior approaches of one summary for one dataset. The practical implication of MVSF is that it offers end-users diverse perspectives from social media data. Our prototype web application is also implemented using ChatGPT to show the feasibility of our approach.

**Keywords:** Multiple-View Summarization; COVID-19 Vaccine Tweet Summarization; Microblog Summarization; Social feature-focused summarization; Entity-based Summarization; Distance-centered Summarization

## 1. Introduction

In 2023, Twitter (now "X") had around 450 million monthly active users, and 6,000 tweets were posted on average every second [1]. If we extrapolate the tweet counts, there are 360,000 tweets posted every minute, 518 million tweets a day, and 189 billion tweets a year. Due to the large volume, it is infeasible for humans to review the tweets to investigate a certain topic or event. A number of Machine Learning and Deep Learning algorithms for social media summarization have been proposed, e.g., ([26–33]) using reinforcement learning along with attention layers and a deep learning-based model, by using a Recurrent Neural Network (RNN) and a Bi-LSTM (Bidirectional Long Short Term Memory) network. Others have used K-means clustering, and a Twitter Online Word Graph Summarizer for a set of related tweets.

However, social media posts can talk about the same topic or content, but they reflect different user perspectives, such as negative or positive sentiments, emotions, biases, political views, or distorted opinions (so-called fake vs. real news), etc. Thus, providing one single summary of social posts may be misleading, or not representing different perspectives of the users, such as their emotions, opinions and their attitudes toward content. Besides, despite the potential that social media platforms have to democratize access to diverse social and political perspectives [63], the meteoric rise of these platforms has further fueled the creation and ossification of "echo chambers" [64–66]. Echo chambers are environments where the opinions, political leanings, or beliefs of users about a topic get reinforced due to repetitive interactions with peers having similar tendencies and attitudes [67]. The echo chamber effect usually arises concerning controversial topics, e.g., gun control, vaccinations, abortion, school prayer, etc. [67]. Spending time in communities of like-minded individuals not only makes individuals become more exposed to pro-attitudinal messages, but also decreases the exposure to counter-attitudinal information. This leads to the issue of echo chambers, where citizens do not see or hear different topics or ideas. This issue also limits their capacity to reach common ground on political issues [68].

A number of methods have been researched to overcome echo chamber effects, for example, increasing the diversity of information sources [70]. Receiving information from different sources would help users to better understand a topic from different perspectives, develop more in-depth opinions, and make better decisions [71,72]. However, users often restrict their personal exposure to only like-minded individuals or platforms [73]. According to [74], the primary cause of polarization and extremism is the loss of truly public platforms/forums. Therefore, recreating a new environment where people are more likely to encounter opposing perspectives and opinions could counteract this tendency, as having trust in "the media" is likely to increase confidence in seeking and receiving information from diverse sources [75]. It is plausible that the more media trust someone has, the more likely they are to explore different views from these sources, thereby becoming less susceptible to echo chambers [73].

In this paper, we propose the *perspective-based summarization approach* to address the challenges posed by the overwhelming volume and echo chamber issues in social media. This approach is designed to summarize and compare the subtle nuances between different viewpoints, e.g., the dissemination and reception of fake and real news. By employing this comparative analysis, our framework aims to highlight the discrepancies between authentic and manipulated content, which are important in our big data world. The proposed summarization framework, *Multiple View Summarization Framework (MVSF)* generates different perspective-based summaries from the same set of social media posts based on user-centered views, opinions, or emotions that can flexibly combine their desired perspectives expressed in social media posts. Providing summaries with different perspectives of the same social media content would help individuals to understand the contrary opinions and diverse perspectives on the same topic.

With MVSF, users can obtain summaries reflecting specific interests and viewpoints expressed in social media. For example, a positive summary on Pfizer vaccines in COVID-19 related posts, or a summary with only negative sentiment regarding the same vaccines. This can help policy makers realize how the vaccines are perceived by different groups. The summaries can also focus on further content analyses for understanding their focus of positivity vs. negativity, or who may be the influential figures in these different perspective summaries.

Our MVSF has two components, the summarization component and the perspective computing component. The summarization component summarizes the social media data using one of three summarization methods: Entity-based (EbS), Triple based (TbS), and Social Signal-based (SbS). The perspectives computing component is intended to detect specific user perspectives expressed in social media posts, such as negative sentiments. The combination of two components enables a wide range of fine-grained summaries, such as a summary of only a negative sentiment views whose focus is the topical entity. As another example, it can combine a triple-based summary with positive sentiments.

The summarization methods use the semantic analysis of each social media post in the form of  $\langle \text{subject}, \text{predicate}, \text{object} \rangle$  triples 2. We show that the semantic analysis-based summarization

outperforms state-of-the-art extractive and abstractive summarization models from the literature, as indicated by better Rouge F-1 scores [3] of around 14%.

The perspectives can be composed together to convey summaries with multiple perspectives. When a journalist writes a story on social media posts on a politician's policy, (s)he can focus on one perspective, such as (s)he can get a summary of negative opinions about the policy, or (s)he can ask for a summary of user posts that are negative *and* fake opinions about the policy. In this way (s)he can convey summaries that contain one or multiple perspectives. By employing MVSF's various methods and perspectives, users can navigate through widespread social media content, gaining deeper insights and enhancing information extraction. The framework's adaptability achieves its significance in providing comprehensive and user-centric summarization solutions.

The work in this paper contributes in the following ways:

1. We have developed three summarization methods for social media posts: Entity-based, Triple-based and Social Feature-focused summarization.
2. We developed the Multiple View Summarization Framework (MVSF) that is capable of generating summaries based on multiple *perspectives*, such as political orientations, sentiments, or fake opinions, combining them with any of the three summarization methods. The framework provides greater flexibility of choosing summarization methods as well as combining different perspectives, generating fine-grained and personalized summaries tailored to the end-user's preferences.
3. Theoretical implications are the integration of multiple perspectives and extractive methods within a unified framework, in contrast to one-size-fits-all summarization. This enables comprehensive information extraction from social media content, and constitutes an extended paradigm for summarization research. On the practical side, we have developed summarization algorithms for the MVSF to allow users to tailor summaries to their needs.
4. Through extensive performance experiments, we compared the summaries generated by our Entity-based and Triple-based methods, both independently and in conjunction with a view. When benchmarking the results against prominent summarization models from the literature, such as Bart-large-CNN [5], Text-To-Text-Transfer-Transformer (T5) [6], BertSum [13], and SBert [14], our summaries achieved a performance improvement of 14% in terms of Rouge Scores.
5. Our framework was effectively applied to a X/Twitter dataset of 18,047 COVID-19 vaccine-related tweets, demonstrating the flexibility of presenting different summaries of this topic.
6. We leverage our MVSF to present a comparative analysis of fake and real news to enhance its utility in discerning truth and misleading information distributed across social media platforms, which is increasingly important in today's world.
7. We have developed a user-friendly web prototype (<http://ai4sg.njit.edu/ai4sg/Summarize>) leveraging the power of OpenAI ChatGPT [11] for text summarization. This application works within the MVSF framework, but does not suffer from the hardware-intensive resource use of our algorithms. The application allows users to input text and customize summarization preferences based on MVSF's methods and perspectives, serving as an accessible perspective-based summarization tool.

These contributions collectively underscore the significance and practical applicability of MVSF in addressing the challenges of efficiently summarizing extensive social media content, while catering to diverse user needs. Our summarization approach is novel as opposed to a one-size-fits-all summarization algorithm.

This paper is organized as follows. In Section 2, we provide a synopsis of related work. We present our summarization framework in Section 3. In Sections 4 to 6, we detail each components of MVSF, including knowledge triple extraction and processing, perspective analysis and detection models, and our summarization methods, respectively. The evaluation of our summarization methods is in Section 7. We present the application of our summarization framework to a large dataset, the results, and our findings in Section 8. We compare and present the findings of summaries of fake news and real news in Section 9. In Section 10, we present our web prototype application that leverages the power of OpenAI ChatGPT [11] to perform text summarization within our MVSF framework. Sections 11 contain Discussion, Conclusions and Future Work.

## 2. Related Work

In this section we introduce general social media summarization approaches along with different applications, as well as echo chamber effects in social media.

### 2.1. General Summarization Approaches

Researchers distinguish between abstractive and extractive summarization methods [4]. Abstractive summarization generates a summary by capturing the prominent ideas of the source text. The summaries contain new sentences that do not exist in the original text. The abstractive summarization methods in the literature include: Bart-large-CNN [5], Text-To-Text Transfer Transformer (T5) [6], FactSumm [7], FAIRSEQ [8], PEGASUS [9], XNLG [10], ChatGPT [11], GPT-2 [12], etc. On the other hand, extractive summarization selects a subset of the sentences that are able to best represent the original document. The extractive summarization models in the literature include: BertSum [13], SBert [14], RankSum [15], HAHSum [16], NeRoBERTa [17], DebateSum [18], MemSum [19], Gensim [20], etc. While abstractive summarization is beneficial in situations where high rates of compression are required [76], microblogs are the antithesis to long documents. Abstractive systems usually perform best in limited domains, since they require outside knowledge sources. The abstractive approaches might not work so well with microblog posts, as they are unstructured and diverse in their subject matter. Furthermore, the summaries generated by abstractive models usually face factual inconsistency problems [77]. There is also a huge inference speed gap between the abstractive and extractive summarization methods [78]. Extractive techniques are known to scale better in highly diverse domains [61]. It is also beneficial to work with tweet fragments rather than entire tweets [25]. Therefore, in this paper, we work on extractive summarization along with this tweet post fragment approach.

Joshi et al. [15] proposed a method based on LDA (Latent Dirichlet Allocation) [21] topic modeling and word embeddings [22] for the extractive summarization of single documents. Their work, based on CNN Daily Mail datasets [23], achieved state-of-the-art performance at the time. LDA and word embeddings are effective when sentences are coherent, however, when sentences are not coherent, which often happens in social media, extra contextual information is needed [24]. One of our summary methods identifies contextual information and bridges this gap.

### 2.2. Microblogging Summarization Approaches

Many algorithms for social media summarization have been proposed, e.g., ([26–33]). Yadav et al. [26] used reinforcement learning along with attention layers and a deep learning-based model, by using a Recurrent Neural Network (RNN) and a Bi-LSTM (Bidirectional Long Short Term Memory) network for summarization tasks. Their work achieved state-of-the-art performance when evaluated with BLEU and ROUGE scores. Modhe et al. [27] made use of TF-IDF encoding to obtain both single-post and multi-post summaries of Twitter activity, based on the rankings of words and a user-defined threshold. Geng et al. [28] proposed a microblogging cluster stream (Microblog Cluster Vectors) and a ranking method, by using K-means clustering [34] and query-lexical-rank to generate a query-focused extractive summary of Twitter data. Dutta et al. [30] performed the first systematic analysis of approaches towards summarizing Twitter posts during disasters. They found that different algorithms applied to the same input would yield summaries with significant differences, which is superficially similar to our results. However, we achieve the different summaries *intentionally* and as part of a unified framework. Olariu et al. [31] introduced a Twitter Online Word Graph Summarizer, which was the first online abstractive summarization algorithm for tweets. In their experiments, for a set of related tweets, they generated a high quality summary. However, when applied to unrelated tweets, the initially generated summary lacks meaning. This happened because event-related signals (in their case, bigrams) stand out when analyzing similar tweets. As a better solution, they built a word graph from trigrams to solve the issue.

Sharifi et al. [32] chose an extractive approach, since it is more appropriate for the structure and diversity of microblogs. They first applied Phrase Reinforcement [62], which generates summaries by looking for the most commonly occurring phrases. Then they processed the results of the Phrase Reinforcement approach by their proposed Hybrid TF-IDF algorithm where the TF (term frequency)

component is computed from the entire collection of posts, whereas the IDF (inverse document frequency) component is computed from a single post. Inouye [33] proposed a clustering-based algorithm and a threshold-based Hybrid TF-IDF algorithm. The first step performs clustering by combining the K-Means++ algorithm [59] with the bisecting K-means algorithm [34] to cluster posts into subtopics and then each cluster is summarized individually. The second step uses the modified Hybrid TF-IDF summarization algorithm [32], so that it can produce multiple post summaries. Originally [32], the algorithm only selected the best summarizing topic sentence, but later Inouye [33] modified the method to select the top four highest weighted posts.

### 2.3. Other Summarization Approaches

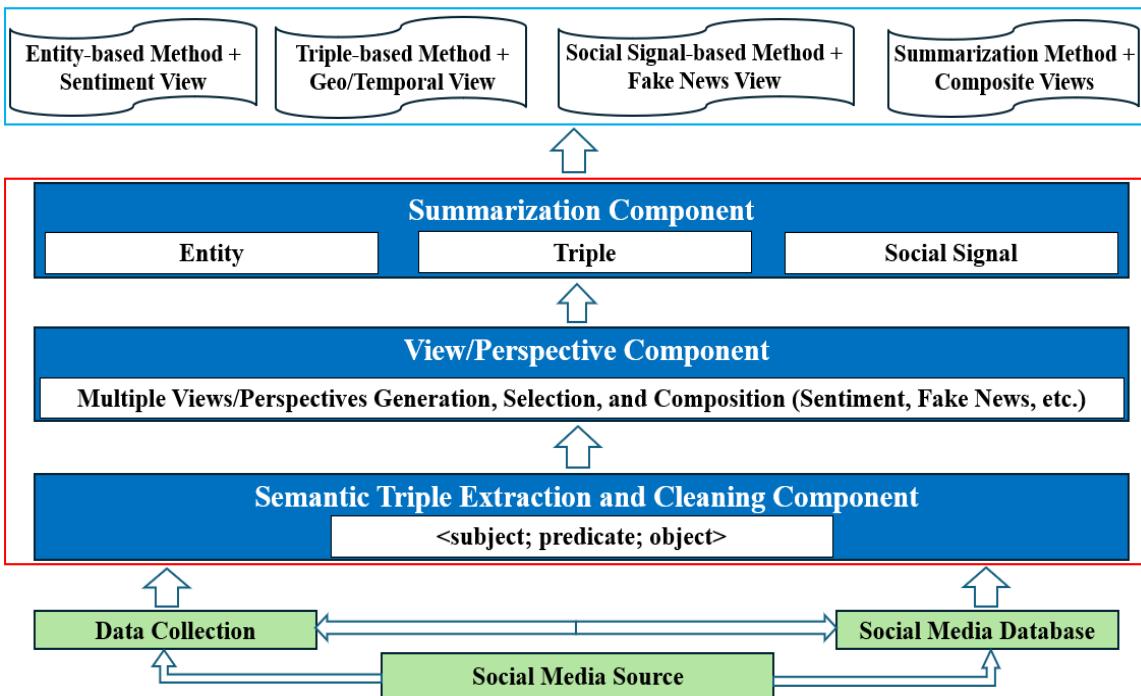
Gunaratna et al. [35] selected related features among entities while maintaining the diversity and saliency of features within entity sets. By selecting (i) inter-entity facts that are similar and (ii) intra-entity facts that are important and diverse, the approach summarizes facts about a collection of entities. Entity summarization has been categorized into extractive and non-extractive methods. Among extractive methods, the summarization is further divided into single-entity and multi-entity categories. For single-entity categories, Gunaratna et al. [36] proposed FACES to incorporate diversity in summarization. For multi-entity categories, FACES-E [37] uses focus term detection and aligns these focus terms with ontology classes and entities present in knowledge graphs [38]. FACES-E showed the usefulness of type-computed literals in creating comprehensive entity summaries. For the non-extractive categories, REMES was introduced [39] to maximize relatedness of facts between entity summaries and importance and diversity of facts within each entity summary. One of our methods performs entity-based summarization.

Amplayo et al. presented an abstractive opinion summarization model that generates input aspect-based summaries for a set of product or hotel reviews [83]. Their aspect-based summaries use transformer models, and for generation of summaries users need to input a specific aspect code or keywords. Li and Chaturvedi [82] presented the Rationale-based Opinion Summarization system (RATION) that generates summaries of user reviews by extracting the representative opinions as well as one or more corresponding rationales as supporting details for opinions. They use transformer based opinion extraction and Gibbs Sampling to sample a user-specified number of sentences as rationales by approximating this joint probability distribution.

Unlike most of the approaches in the literature that focus on a single aspect, our framework focuses on separate or combined perspectives expressed in the social media posts, and ranks the extracted information based on respective criteria to obtain fine-grained extractive summaries.

## 3. Multiple View Summarization Framework

Our MVSF addresses the diverse aspects of social media posts. The framework can summarize the same social media dataset on one particular topic into multiple summaries based on different user perspectives. In this paper, we use the terms “views” and “perspectives” synonymously. The general architecture of multiple view summarization in Figure 1 has following components:



**Figure 1.** Multiple View Summarization Framework (MVSF).

**Data Cleaning and Semantic Triple Extraction Component:** We start by collecting data on a particular topic from a microblogging site such as Twitter (now “X”). We applied standard text-preprocessing to clean the dataset, by removing non-ASCII characters, redundant spaces, and URLs. Social media users are often more interested in posts that describe events or activities [40]. Therefore, we apply *semantic triple extraction* to get a set of  $\langle \text{subject } (S), \text{ predicate } (P), \text{ object } (O) \rangle$  triples for each posting that captures events or activities in the data. The extracted triples represent structured content by identifying entities, relationships, actions, and events. The triples and their original posts serve as the base for generating multiple perspectives.

**View/Perspective Component:** This component analyzes multiple perspectives expressed in each textual post, such as *sentiment analysis*, which captures emotional trends and public perceptions [44]; and or *fake news detection modeling*, flagging misleading information to address the spread.

**Summarization Component:** This component generates summaries of input sets of social media posts filtered by one or more perspectives. We present three summarization methods: *Entity-based (EbS) Summarization* that highlights key individuals, organizations, or places mentioned in the dataset; *Triple-based (TbS) Summarization* that groups post related to specific events or statements as a unit; *Social Signal-based (SbS) Summarization* that utilizes social media interaction features, including counts of user followers and retweets to capture socially popular content.

**Output Component:** The output component can generate a variety of summaries, based on users’ desired summarization methods and perspectives. For example, users can choose to generate a sentiment summary that tracks the emotions across posts over time, or a fake news summary that summarizes the posts with misleading information. Moreover, users can choose to generate a composite summary by combining multiple perspectives.

Therefore, given a set of social media postings and one specific perspective (negative sentiments only), our framework can generate a summary with the requested perspective. With the same input but with a different perspective, it will generate another summary with that different perspective (e.g. fake news only). Therefore, it can generate multiple summaries of the same set of social media posts that convey different perspectives. One can choose to combine one or more perspectives from the View/Perspective Component to generate customized summaries that express multiple perspectives, e.g., a summary of negative and fake opinions.

This flexibility allows the same set of postings to be summarized in diverse summaries expressing different perspectives. Our MVSF demonstrates the power and adaptability to discover

diverse insights from large social media data, tailored to users' needs. The MVSF is able to accommodate different analytical scenarios, and deliver comprehensive summary reports.

#### 4. Semantic Triple Extraction and Cleaning

The *Semantic Triple Extraction and Cleaning Component* (in Figure 1) is a fundamental and important component in the MVSF. The semantic triple extraction transforms unstructured social media text data into structured knowledge triple representations. Social media content usually contains irregular sentence structures, hashtags, URLs, and abbreviations [81]. Thus, it is challenging to perform analysis directly on social media content. To address this issue, we extract meaningful insights by capturing relationships between key entities in the form of  $\langle \text{subject}; \text{predicate}; \text{object} \rangle$  (*SPO*) *triples*. The triples capture events and actions expressed in the text, and they serve as the foundation for generating diverse summaries in our framework. For instance, a post like "*The CDC announced new vaccine guidelines*" would yield a triple  $\langle \text{CDC}; \text{announced}; \text{new vaccine guidelines} \rangle$ , which showcase the key information in the sentence.

We used the Stanford Open Information Extraction (Open IE) model [41] to extract the triples. The model first breaks down sentences into shorter logical clauses. The model then identifies the  $\langle \text{subject}; \text{predicate}; \text{object} \rangle$  for each clause, and predicts whether an edge in the dependency graph should form a new independent clause [42]. For example, the sentence "The CDC and WHO announced that vaccines will be distributed by March" can be broken into two triples:  $\langle \text{CDC}; \text{announced}; \text{vaccines distribution} \rangle$  and  $\langle \text{WHO}; \text{announced}; \text{vaccines distribution} \rangle$ . Each extracted triple is then converted to lowercase to achieve a unified text representation, ensuring consistency across the dataset.

#### 5. Perspective Analytics for Summarization

The Perspective Detection Component in our framework is used to identify different perspectives expressed in each social media post. Perspective detection can use different methods. For instance, a negative perspective can be identified by sentiment analysis, a fake content view can be detected by a machine learning model to classify a post as fake or not. These perspectives are essential to provide unique insights, such as a summary with a particular sentiment, or a summary of a particular political bias. In this Section, we provide perspective analysis approaches we used. With the proliferation of AI models available with APIs, our framework assumes that the appropriate perspective detection model is available to be utilized for the summarization.

##### 5.1. Sentiment Perspective Analysis

We chose the Stanford Sentiment Analyzer [43] as our sentiment analysis tool based on the evaluation of [44]. An input phrase will be labeled either as "(Very) Negative," "Neutral," or "(Very) Positive." With sentiment labels, we can choose to focus only on the  $\langle S, P, O \rangle$  triples with a specific sentiment, e.g., triples expressing negative sentiments towards mandatory COVID-19 policies.

##### 5.2. Fake News or Real News Perspectives

Social media platforms have been flooded with misinformation, so-called fake news, which confuse citizens, cause conflicts, and drown out authentic information [45]. Therefore, summarizing large fake news item sets can produce a readable summary. This would enable concerned users to counteract the spread of misleading information. This can be achieved by training a Machine Learning model to classify the social media posts and collecting those labelled as fake news.

##### 5.3. Political Bias Perspectives

To systematically discern the political biases in a posted text, we can employ a machine learning model trained on data labeled either as 'left' or 'right' political viewpoints or biases. This binary classification approach, inspired by approaches in [79], allows detection of political orientations within content. By integrating this method into our framework, we enhance the granularity of our summarization, ensuring a comprehensive representation of the political landscape within the data.

In addition to these, one can consider contextual perspectives such as temporal or location-specific perspectives for summaries of the public reactions to restrictive government COVID-19 policies, e.g., right after their introduction, and again after six months. This could evaluate the acceptance or rejection of a policy, which would allow government agencies to fine-tune it. This view can be derived from time stamps of social media posts. A summary based on specific locations, or at different administrative levels, such as country, state/province, etc. can capture location-specific perspectives that can be compared with summaries from different locations to determine differences.

#### 5.4. Composition of Multiple Perspectives

The power of our MVSF lies in the ability to select and compose multiple views to generate customized summaries. For example, when a user is interested in a *fake news-only view with negative sentiments*, the framework's View/Perspective component can generate this combined view. The *composite perspectives* generate composite views using the following expression:

$$cv = \text{compose}(v_1, v_2, \dots)$$

where  $v_i$  is a single view or perspective. Such capability of multiple perspective composition demonstrates the framework's flexibility and adaptability. This allows users to generate summaries and insights from big data customized for their needs.

While views/perspectives are composable, the summarization methods are not. Each method operates independently, and focuses on different tasks. For example, the Entity-based method highlights key entities mentioned. The Triple-based method groups specific events. The Social Signal-based method utilizes social media features to emphasize socially popular content. These methods have different summarization goals and are applied individually. The difference between view composition and summarization method independence demonstrates our MVSF's adaptability. The framework allows users to generate summaries using a method that provides targeted insights and combining it with multiple perspectives, therefore presenting a wide range of summarization scenarios.

## 6. Microblogging Summarization Methods

### 6.1. Entity-Based Summarization (EbS)

In this method (EbS), the entities define the primary subjects and objects for obtaining summaries. Unlike in other research where the user provides the entities of interest (e.g., "Joe Biden"), we discover the most prominent entities mentioned in the social media posts by first finding the triple verbs that contain the same semantic events, and by identifying the salient entities in triple subjects and objects.

8. **Identify same events in triples using WordNet:** Important entities can appear in semantically similar events that are expressed by different predicates/actions. Thus, we need to identify predicates expressing similar or identical meanings with different verbs, such as "offer," "pass," and "transfer," or different forms of the same verb, e.g., "provide," and "provided," all of which express a similar meaning of "giving something to someone." Groups of semantically similar verbs can be subsumed by one *root verb*, using the appropriate synset from WordNet [46]. A synset is a set of one or more synonyms. WordNet organizes synset into generalization hierarchies. A verb in a hierarchy that is the most general is referred to as a root verb. For triples using verbs expressing similar/identical meanings with different words or tense forms, we replaced the verbs as follows. To disambiguate the meaning of a predicate and find the closest synsets, we compared two methods and used a human evaluator to determine the better one. The first method is the Lesk algorithm [47]. Given a verb and the triple where it occurs, Lesk returns a synset that represents its context meaning. However, Lesk often failed at finding the correct synset. Among a set of 80 randomly picked triples, only 37 verb synsets were correctly identified, according to the human evaluator. The second method was that we selected "v.01" (primary meaning returned by WordNet) as the verb synset. Although this is a simple approach, it produced a better result (78/80) according to the human review. Therefore, we used the "v.01" meaning for each verb to find the root synset. For each synset vx.v.01 of the verb vx, we followed

its hypernym (superclass) chain upward until reaching its root synset. We identified the frequent events by selecting triples whose root verbs occur more often than a threshold  $\theta$ . Given a triple set  $TS = \{ts_1, ts_2, \dots\}$  where  $ts_i = \langle s, p, o \rangle$ , find  $ts_j = \langle s, \text{root}(p), o \rangle$ . We select triples that have frequent root verbs  $\text{root}(p)$ , i.e.,  $\text{frequency}(\text{root}(p)) > \theta$ . We present the program to identify the synset hierarchy in Table 1, and an example synset hierarchy in Figure 2, where all verbs are summarized by “give.”

9. **Identifying Salient Entities:** The second task in this method is to identify important entities to focus on. Each triple includes two entities, a subject, and an object. As noted, this method is based on the entities (subject/object) that occur most often in the triple sets. To identify the frequencies of *meaningful* words, for each entity, we removed numbers, stop words, and words with fewer than three characters, and performed lemmatization. If a pre-processed entity became an empty string, we removed its triple from the triple set. If a pre-processed subject consisted of more than two words, we removed all words except for the last two words (which are likely to contain “most of the meaning”).

**Table 1.** Pseudo Code to identify the Root Verb Synset by Using WordNet.

```

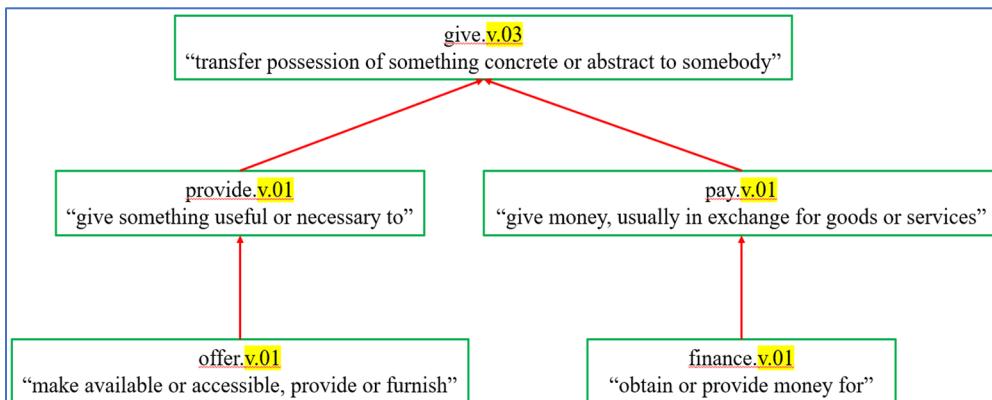
highestSuperclass = []

for v in tripleVerb:
    currentV = v + '.v.01' # we apply v.01 (most natural) synset to all triple verbs

    while (len(wordNet.synset(currentV).hypernyms()) == 1):
        currentV = wordNet.synset(currentV).hypernyms()[0] # replace current
        synset with its direct superclass synset

    highestSuperclass.append(currentV)

```



**Figure 2.** Example of Synset for All Verbs related to the root “give.”.

Each subject is labelled with its frequency in the triple set. If there is more than one word in a subject, we obtain the subject score based on the word with the higher frequency. For example, if there are 77 mentions of “covid” and 100 of “vaccine,” the subject score of “covid” is 77, of “vaccine” is 100, and of “covid vaccine” is also 100. The same method applied to objects for computing object scores. The preprocessing and score calculation of triple entities is shown in Table 2.

**Table 2.** Pseudo Code of identifying top Entities.

---

```

def preprocess (entity):
    entity = remove stop words and numbers from entity
    if (length of entity) < three characters:
        entity = empty string
    entity = lemmatization (entity)
    return entity

subjectWord = [ ] // a list that collects all words in triple subjects
objectWord = [ ] // a list that collects all words in triple objects

for each subject s, object o, triple t in triple set ST: // (s, o) is the (subject, object) of triple t in ST,
    s = preprocess (s)
    o = preprocess (o)
    if s == empty string or o == empty string:
        remove t from ST

    s = s.split(" ")[:-2] // if s have more than two words, remove all except for the last two words

    subjectWord.extend (s.split(" "))
    objectWord.extend (o.split(" "))

for each word x in subjectWord:
    label x with an integer score of its frequency in subjectWord

for each word y in objectWord:
    label y with an integer score of its frequency in objectWord

for each subject s, object o, triple t in triple set ST: // (s, o) is the (subject, object) of triple t in ST
    Subject Score of s = highest score of the word in s
    Object Score of o = highest score of the word in o

```

---

To identify the summary with the best accuracy, we experimented with assigning different weights ( $\alpha$ ) to the subjects and  $(1-\alpha)$  to the objects. We select the top-scoring  $\mathbf{m}_E$  triples and the corresponding original sentences. The parameter  $\mathbf{m}_E$  controls the length of the summary.

$$\mathbf{TripleScore} = \alpha * (\text{Subject Score}) + (1 - \alpha) * (\text{Object Score}) \quad (1)$$

where  $\alpha \in [0.0, 1.0]$

### 6.2. Triple Based Summarization (TbS)

In this method, we capture the important contextual meaning (the whole statement instead of entities) by using triple sentence representations. To create sentence representations, we use BERT [48] sentence embeddings. BERT reads the entire input sequence at one time, which allows BERT to learn the contextual information of each word based on its neighboring (left and right) words. We used an autoencoder [49] to learn a 32-dimensional vector representation to ensure that we achieve a lower dimensional representation [24]. Instead of focusing on specific entities, we produced a summary based on important statement information in triples. We use the distance measures to the centroid of the triple vectors to select the salient triples for summarization.

For each triple t, its vector representation v is:

$$v_t = [v_{t1}, v_{t2}, v_{t3}, \dots, v_{t32}] \quad (2)$$

The centroid vector  $v_c$  of all triple vectors is calculated:

$$v_c = \frac{1}{n} (\sum_{t=1}^n v_t) = [v_{c1}, v_{c2}, v_{c3}, \dots, v_{c32}] \quad (3)$$

where  $n$  is the triple count.

The Euclidean distance  $d$  from a triple  $t$  to the centroid  $c$  is:

$$d_{tc} = \sqrt{(\sum_{i=1}^{32} (v_{ti} - v_{ci})^2)} \quad (4)$$

We selected triples with the shortest distances to the centroid  $c$  as top summarization triples. The original sentences corresponding to the selected triples were recovered to form the summary.

### 6.3. Social Signal-Based Summarization (SbS)

In this method (SbS), we exploit the social signals (i.e., a tweet's retweet count, and its poster's followers count) to identify the saliences of tweets. A tweet is more important if it is (1) posted by a user with many followers, and (2) retweeted many times, according to [40], where the salience score of a tweet is the multiplication of retweet count, user follower count and readability. Our goal is to identify summaries with social prominence on social media. Thus, we modified the formula and defined the salience score of a tweet:

$$\text{Salience Score} = \begin{cases} \text{follower} + \text{retweet} * 707, & \text{follower} > 0 \\ 0, & \text{follower} = 0 \end{cases} \quad (5)$$

where 707 is the average number of followers of a Twitter user [30]. When a post is retweeted, there will be on average 707 people who will see it. We rank triples based on the scores of their original tweets, and select the  $msr$  top-scoring triples. The corresponding original sentences of the triples are selected to form the summary.

## 7. Evaluation of Microblogging Summarization Methods

To evaluate our summarization approaches, we compared them with extractive and abstractive summarization models from the literature. The extractive models are BertSum [13] and SBert [14], variants of BERT used in our method (TbS). Abstractive summarization models are Bart-large-CNN [5] and T5 [6], which according to [44], generate better summaries than TextRank [20] and GPT-2 [12].

In our pursuit of evaluation, we encountered a challenge: we could not locate a social media dataset having a gold standard summary. Nonetheless, we could perform an evaluation by leveraging an alternative dataset – the BBC news items set [51]. Originally designed as a benchmark for machine learning research, we repurposed this dataset to evaluate the performance of summaries in our specific task. The dataset comprises 2,225 documents, each sourced from the BBC news website and spanning five diverse topical areas, namely business, entertainment, sport, politics, and technology. This dataset serves as a trustworthy foundation for our assessment of summary generation by providing human extractive summaries for each document.

We selected 20 news documents from the business category. We summarized them using our Entity-based (EbS) and Triple-based summarization (TbS) approaches, with and without sentiment perspective (S). Because this dataset lacks social signals, we excluded the SbS method from our evaluation process. Each of our generated summaries adheres to a consistent length of approximately 300 words, aligning with both our chosen methods and the published results.

We employed Rouge scores [3] as our evaluation metrics. The scores enable a comparison between each generated summary and the corresponding gold standard summary, in 1-gram, 2-gram, and longest common subsequence (LCS) units. Our summaries based on Entity and Triple with or without a perspective outperformed those by the benchmarking models (Table 3). The Triple based Summaries with sentiment view, i.e., TbS + S performed best with the highest score for 1-grams and LCS, while the method TbS performed best for 2-grams. This result shows that our summaries are more consistent with the gold standard summaries than existing models.

**Table 3.** Rouge Scores of Summarization Models.

Rouge Score	1-gram	2-gram	LCS
Method (EbS), $\alpha = 0.7$	0.465	0.310	0.444

Method (TbS)	0.557	<b>0.427</b>	0.551
TbS with View (S)	<b>0.58</b>	0.416	<b>0.572</b>
EbS with View(S), $\alpha = 0.7$	0.459	0.299	0.433
BertSum	0.406	0.174	0.394
SBert	0.444	0.152	0.428
bart-large-cnn	0.402	0.150	0.394
T5	0.304	0.104	0.304

## 8. Application of MVSF to COVID-19 Vaccine Tweets

We show a number of examples of our summarization framework to a dataset of COVID-19 vaccine-related tweets as a case study.

### 8.1. Data

We used 18,047 tweets [52] about widely used COVID-19 vaccines worldwide, Pfizer/BioNTech, Sinopharm, Sinovac, Moderna, AstraZeneca (AZ), Covaxin, and Sputnik V. The tweets were posted between December 2020 and November 2021 (This dataset may not fully represent all tweets concerning COVID-19 vaccines). The dataset's columns include *tweet ID*, *tweet content text*, *date time* (when the tweet was posted), *user ID*, *retweets* (how many times the tweet has been retweeted by other users), *likes* (how many times other users clicked like on the tweet post), *hashtags*, and *user\_loc* (location of a user according to their profile).

### 8.2. Preprocessing

We removed non-ASCII codes, URLs, redundant spaces, and punctuations from the post contents. We assigned each post an integer *post index*. We then performed sentence tokenization [55] to split posts into sentences, because a tweet may contain multiple sentences. It is beneficial to work with post fragments (sentences) rather than entire posts [25]. We assigned to each sentence an integer *sentence index*, as this sentence tokenization is used for tracing back to the corresponding original sentences from the selected knowledge triples. Table 4 shows an example of a post with two sentences. We obtained 28,242 sentences from 18,047 posts.

**Table 4.** A post with two sentences in the dataset.

post index	post text	sentence index	sentence text
484	A shipment of Sputnik V vaccine arrived in Vietnam.	0	A shipment of Sputnik V vaccine arrived in Vietnam.
	The handover ceremony took place at Noi Bai airport.	1	The handover ceremony took place at Noi Bai airport.

### 8.3. Triple Extraction and Cleaning Process

Using triple extraction methods (OpenIE) [41] from the Stanford CoreNLP package [56], we extracted 101,432 triples, and linked them to their sentences and their original posts.

We further filtered out triples with auxiliary verbs that carry little meaning, e.g., “be” or “have.” This reduced the triples from 101,432 to 67,049. In a number of cases, there were overlapping triples from the same sentence, but of different lengths. For triples from the same post and sentence, and using the same verb, we eliminated all but the one triple retaining the most information (words). The final 16,270 triples were used as the input for each summarization method.

### 8.4. Experimental Results

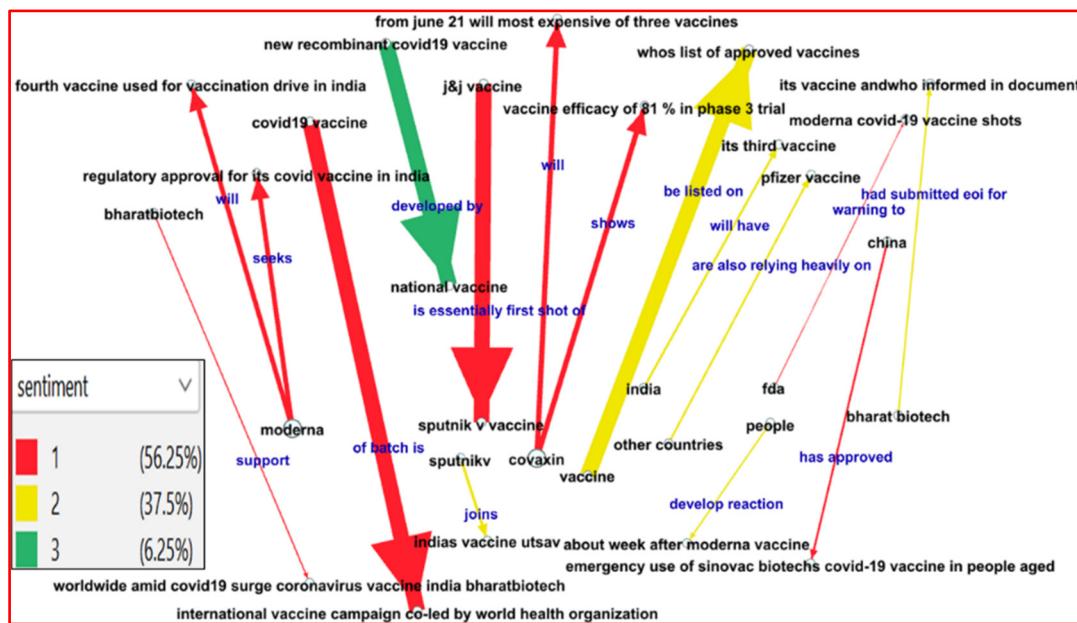
#### 8.4.1. Summaries Without Perspectives

We use all the 16,270 triples (from subsection 8.3) to generate the summary for each method (EbS, TbS, SbS), respectively.

Figure 3 shows the knowledge graph [38] of the summary triples using entity-based summarization (EbS). We used Gephi [69] as our visualization tool. The size of each node is proportional to its degree, with a size range (10, 20) implemented in Gephi. The edge color expresses the sentiment expressed by the triple, with red (=1) for “Negative,” yellow (=2) for “Neutral,” and green (=3) for “Positive.” The edge thickness represents the triple’s rank. The thicker an edge is, the higher its rank (*i.e.*, *higher (triple/salience) score, or shorter distance to the centroid*) is in the summary set. For the summary generation, we selected the top-scoring  $m_e$  triples from the triple set, with their corresponding original sentences, such that the total number of words in the summary was approximately 300. The details of the summary triples are in Table 5. The corresponding original sentences of the selected summary triples are in Table 6 in temporal order of the original tweets.

**Table 5.** Summary Triples of Method (EbS), their Triple Scores, Sentiments, and Vaccines, in Rank Order.

triples	sentiment	tripleScore	vaccine
j&j vaccine is essentially first shot of sputnik v vaccine	1	1646	sputnik
new recombinant covid19 vaccine developed by national vaccine	3	1646	sinopharm
covid19 vaccine of batch is international vaccine campaign co-led by world health organization	1	1646	sinopharm
vaccine be listed on whos list of approved vaccines	2	1646	covaxin
covaxin shows vaccine efficacy of 81 % in phase 3 trial	1	1153.2	covaxin
covaxin will from june 21 will most expensive of three vaccines	1	1153.2	covaxin
moderna seeks regulatory approval for its covid vaccine in india	1	1129.4	moderna
moderna will fourth vaccine used for vaccination drive in india	1	1129.4	moderna
sputnikv joins indias vaccine utsav	2	1028.6	sputnik
india will have its third vaccine	2	1002.7	sputnik
china has approved emergency use of sinovac biotech's covid-19 vaccine in people aged	1	999.2	sinovac
other countries are also relying heavily on pfizer vaccine	2	994.3	pfizer
bharat biotech had submitted eoi for its vaccine and who informed in document	2	978.2	covaxin
people develop reaction about week after moderna vaccine	2	974	moderna
bharatbiotech support worldwide amid covid19 surge	1	962.8	covaxin
coronavirus vaccine india bharatbiotech			
fda warning to moderna covid-19 vaccine shots	1	939.7	moderna pfizer



**Figure 3.** Knowledge Graph Visualization of Summary Triples of View (EbS).

**Table 6.** Entity-based Summary (in Temporal Order).

Other countries in the Middle East, including Saudi Arabia, Qatar, Kuwait, and Oman, are also relying heavily on the Pfizer vaccine, developed by the US company PfizerVaccine. This is also another validation of Sputnik V pioneering technology as J&J vaccine is essentially a first shot of Sputnik V vaccine (Ad 26 - human adenoviral vector 26). Covaxin shows vaccine efficacy of 81% in phase 3 trial. About a week after the Moderna vaccine, some people develop a reaction. The new recombinant COVID19 vaccine, developed by the National Vaccine & Serum Institute, a R&D center of Sinopharm bioscience subsidiary the China National Biotec Group, got approval from the National Medical Products Administration on Apr. 9. SputnikV joins Indias vaccine utsav. Once DCGI approves SputnikV, India will have its third vaccine after Covishield and Covaxin. Bharat Biotech announces COVAXIN capacity expansion to support vaccination campaigns in India & worldwide amid COVID19 surge Coronavirus Vaccine India BharatBiotech. The first batch of COVID19 vaccine supplied by Chinas Sinopharm to COVAX, the international vaccine campaign co-led by the World Health Organization, was officially rolled off the production line on Tuesday. China has approved emergency use of Sinovac Biotechs COVID-19 vaccine in people aged between 3 and 17. From June 21 Covaxin will be the most expensive of the three vaccines which will be available in private hospitals. The vaccine is yet to be listed on the WHOs list of approved vaccines. FDA adds warning to Pfizer, Moderna Covid-19 vaccine shots to indicate the rare risk of heart inflammation after its use Pfizer Moderna COVID19 Vaccine US FDA. Moderna seeks regulatory approval for its Covid vaccine in India. Moderna will be the fourth vaccine to be used for the vaccination drive in India. Bharat Biotech had submitted EOI (Expression of Interest) on April 19 for its vaccine and WHO informed in a document that the assessment status for Covaxin is ongoing

Similarly, for methods (TbS) and (SbS), the goal was again to retain the  $m_t$  and  $m_s$  top-ranking triples that could be used to generate readable summaries of a length of approximately 300 words, respectively. The visualization of the method (TbS) and the method (SbS) and the selected triples are omitted, but we show the summary results in Table 7 and Table 8, respectively.

**Table 7.** Triple-based Summarization (in Temporal Order).

The Food and Drug Administration on Wednesday gave its approval for Sinovac use on the elderly after considering the recommendation of the experts and the current situation of high Covid-19 transmission and limited available vaccines. The Moderna Covid19 jab is now available at 11 of 38

vaccination centres in Singapore, while the rest are offering the PfizerBioNTech product. The CoronaVac vaccine developed by the Chinese biopharmaceutical company Sinovac Biotech has effectively reduced the risk of COVID19 symptoms in medical workers by 94%, showed a study by the Indonesian Health Ministry. On the picture Deepak Sapra, Global Head of Custom Pharma Services at drreddys Laboratories is getting a shot of Sputnik V in Hyderabad. unless you had J&J) Moderna Pfizer As of today 154199664 Americans are fully vaccinated, according to the CDC! The first validation samples taken from the produced batch will be shipped to the Gamaleya Center for quality control. The World Health Organization's pandemic programme plans to ship 100 million doses of the Sinovac and Sinopharm COVID19 shots by the end of next month, mostly to Africa and Asia, in its first delivery of Chinese vaccines, a WHO document shows. Your queries answered Deadline for booster dose for Sinopharm announced, if you received vaccine over six months PfizerBiontech COVID19. USA doesn't recognise Indian vaccine COVAXIN PM Modi has gone to US and possibly he took doses of COVAXIN Whether all Indians are allowed to visit US with COVAXIN doses? days after PM Modi's vaccine diplomatic push, Covaxin gets WHO nod; propaganda by anti-govt voices falls flat. Sinopharm approved for travel to UK From Nov. 22, the Sinopharm vaccine will be added to the UK's list of approved vaccines for inbound travel, benefiting more fully vaccinated people travelling from to Sinopharm is the leading vaccine administered in Sri Lanka.

**Table 8.** Social Signal-based Summary (in Temporal Order).

India gets third coronavirus vaccine as Russia's Sputnik V is cleared for emergency use CovidVaccine. Dr Reddys administers first dose of the Sputnik V vaccine in Hyderabad. The second consignment of Sputnik V arrives in Hyderabad, Telangana. There are plans to introduce single-dose vaccine soon in India-Sputnik Lite. JustIn. A PIL has been moved in Delhi High Court challenging Centres notification which has accorded permission to conduct the Phase II/III clinical trial of Covaxin in the age group 2 to 18 years to its manufacturer Bharat Biotech. Bharat Biotech Amid Travel Fears NDTVs Shonakshi Chakravarty reports Read more. Malta firm wants to supply 60 million doses of Sputnik V to Haryana, state government writes to Centre. Assam Covaxin Shortage Grows Into Crisis, Some Miss Second Dose Deadline. Made in India Covaxin is the third costliest vaccine globally. The appointment slots can be booked via CoWIN portal, according to the hospital administration. Delhi's Madhukar Rainbow Childrens Hospital will start administering Russia's COVID19 vaccine Sputnik V, tentatively by June 20. children are undergoing trials for the vaccine across the country. After Covaxin, Zydus Cadila is the second indigenously produced vaccine for children currently under trial in India. This will also benefit people travelling abroad for education, jobs or business. I request for your kind intervention so that an early approval is received for Covaxin from WHO. Moderna approved for emergency use, 4th vaccine okayed by India COVID19 Vaccine. Accept Covishield, Covaxin Or Face Mandatory Quarantine, India Tells EU. COVAXIN effective against DeltaPlus variant of COVID19, says Indian Council of Medical Research study. DCGI gives nod to study mixing of Covishield and Covaxin. Subject Expert Committee recommends Covaxin for kids aged between 2 and 18 NDTVs Meher Pandey reports. Covaxin Cleared By UK, Relief For Indian Students And Tourists.

As shown in the summary results, three different methods yielded completely different triple sets, and therefore produced different summaries. Each corresponding original sentence in our summaries is unique, meaning there is no common sentence among our methods. As desired, this shows that the three methods generate different results, focusing either on entities, or triple statements, or social features.

#### 8.4.2. Summarization with Sentiment Perspectives

In this subsection, we generate summaries based on negative and positive sentiments (view S) as follows. The different view compositions show the power of our framework, which provides summaries with different perspectives from a single post set. We performed sentiment analysis on each of the data items using the Stanford Sentiment Analyzer [43]. The Stanford Sentiment Analyzer

uses a fine-grained analysis based on both words and labeled phrasal parse trees to train a Recursive Neural Tensor Network (RNTN) model. The RNTN model computed the sentiments based on (1) the sentiment values of each word, and (2) the sentiment of the parse-tree structure composed from the sentiment values of words and sub-phrases. Each post will be labelled either as negative, very negative, neutral, positive, or very positive.

We present (1) S view + Social Signal-based method (S+SbS) (Table 9), (2) S view + Triple-based method (S+TbS) (Table 10), and (3) S view + Entity-based method (S+EbS) (Table 11). We compare summaries with negative sentiments and positive sentiments. Among the 16,270 triples from subsection 8.3, there are 926 positive or very positive triples, and 5,423 negative or very negative triples, which we combined into just two classes, positive and negative.

**Table 9.** Composite View Summaries of Positive (left) and Negative (right) Views with SbS.

<p>The 80-year-old three-time World Cup champion called it an unforgettable day and urged discipline to preserve lives. The COVID-19 vaccine developed by Chinas Sinopharm has been approved for emergency use in the Maldives, the Maldives Food and Drug Administration announced at a press conference Monday afternoon. Azerbaijan on Thursday received a batch of Sinovacs COVID-19 vaccines that it directly purchased from China. Serbian President Aleksandar Vucic received a dose of Chinese Sinopharm COVID-19 vaccine on Tuesday, encouraging more people to join the immunization, according to local media. Delhi high court refuses to stay Covaxin trial among children (RichaBanka reports) I request for your kind intervention so that an early approval is received for Covaxin from WHO The U.S. drug regulator on Friday added a warning to the literature that accompanies Moderna and Pfizer-BioNTech COVID-19 vaccine shots, indicating the rare risk of heart inflammation after its use. Accept Covishield, Covaxin Or Face Mandatory Quarantine, India Tells EU The Union health ministry cited a large-scale, real-life study conducted by the ICMR and said that two doses of Covidvaccines, irrespective of Covishield and Covaxin, were successful to extend 95% protection from death. Covaxin receives certificate of Good Manufacturing Practice from Hungarian authorities. ICMR study Watch for details We want to ensure equitable access of the vaccine to every Indian citizen, and the expansion of Covaxin production facilities by Bharat Biotech will take us closer to this goal. Dr Sumit Ray, Holy Family Hospital, on delay in Covaxin approval by the World</p>	<p>Covaxin 81% Effective, Works Against UK Variant, Claims Bharat Biotech Read more: India to get its third COVID19 vaccine; Subject Expert Committee recommends Russias SputnikV vaccine for Emergency Use Authorisation NDTVs Sukirti Dwivedi with the latest updates India gets third coronavirus vaccine as Russias SputnikV is cleared for emergency use CovidVaccine Dr Reddys administers first dose of the SputnikV vaccine in Hyderabad There are plans to introduce single-dose vaccine soon in India-Sputnik Lite. JustIn. A PIL has been moved in Delhi High Court challenging Centres notification which has accorded permission to conduct the Phase II/III clinical trial of Covaxin in the age group 2 to 18 years to its manufacturer Bharat Biotech Malta firm wants to supply 60 million doses of SputnikV to Haryana, state government writes to Centre. After Malta Firm Offers 6 Crore Sputnik Jabs, Haryana Seeks Centres Help NDTVs Mohammad Ghazali reports SputnikV Assam Covaxin Shortage Grows Into Crisis, Some Miss Second Dose Deadline VaccinationDrive CovidVaccine Assam Covaxin Shortage Grows Into Crisis, Some Miss Second Dose Deadline Made in India Covaxin is the third costliest vaccine globally Delhis Madhukar Rainbow Childrens Hospital will start administering Russias COVID19 vaccine SputnikV, tentatively by June 20. The appointment slots can be booked via CoWIN portal, according to the hospital administration. After Covaxin, Zydus Cadilla is the second indigenously produced vaccine for children currently under trial in India. Moderna approved for emergency use, 4th vaccine okayed by India COVID19Vaccine COVAXIN effective against DeltaPlus variant of COVID19, says Indian Council of Medical Research study Union Minister of Health and Family Welfare, Mansukh Mandaviya launches the first commercial batch of Bharat Biotechs Covaxin</p>
--	--

<p>Health Organisation. Bharat Biotechs Covaxin, the vaccine against Covid19, has received the recommendation of a SEC for use in children between the ages 2 to 18 PM NarendraModi jis visionary decision to back our scientists &amp; researchers is now a perfect Diwali Gift from to the World. This video fits the last almost 2 years into 2 minutes.</p>	<p>manufactured in Gujarat's Ankleshwar. Subject Expert Committee recommends Covaxin for kids aged between 2 and 18 NDTVs Meher Pandey reports The Lancet peer-review confirms the efficacy analysis of Bharat Biotechs Covaxin. As per phase-three clinical trials data, Covaxin demonstrates 77.8% efficacy against symptomatic COVID19.</p>
---	--

**Table 10.** Composite View Summaries of Positive (left) and Negative (right) Views with TbS.

<p>Only the PfizerBioNtech vaccine has received EUL approval, so far. Sinovac vaccine works on UK, South African variants - Brazil institute SLnews SriLanka Sinovac US President Joe Biden announces that US will share US-authorized vaccines doses of Pfizer, Moderna and Johnson &amp; Johnson, as they become available, with the rest of the world. India can't afford to fail because India will lead the fightback by mass production of vaccines for the developing world. Haryana government has received an expression of interest from an international pharmaceutical company headquartered in Malta to provide up to 60 million doses of SputnikV vaccine. NSTworld Israel will receive in return the doses that Pfizer is to send to the Palestinian Authority. The state has reportedly received a fresh consignment of 693210 doses of Covaxin and Covishield vaccines from the central govt. Top sources tell CNNnews18 Authorisation of Covaxin internationally will happen by end of Aug WHO rep Meeting Health Minister today COVID-19 Vaccination Drive to Resume In Mumbai From Tomorrow As BMC Receives Fresh Stock of Vaccines COVID19 COVID19 A new study by the Centers for Disease Control and Prevention revealed that Moderna's COVID-19 vaccine is somewhat more effective than those offered by Pfizer and Johnson &amp; Johnson's vaccines. Following the arrival of more Sinopharm vaccines to Sri Lanka, doses will be released to the North to begin administering on the 20-29 age group. Subject Expert Committee has given a recommendation to Official sources to ANI NSTnation. Individuals who have been vaccinated with Sinovac are allowed to perform the umrah in Saudi Arabia on the condition that they get a third dose of the vaccine. The last-minute addition comes less than one week before the US launches its new travel system, granting entry to travellers who have received a vaccine that has been approved by the FDA or WHO US</p>	<p>But with the general pool eligible for the vaccine expanding 2.5 times to 345 million from April 1, Covaxin will need to step up to service demand MintPlainFacts rashmi kundu. As you can see, the UK could conceivably restrict AstraZeneca vaccine to older people and lean on Moderna, Novavax and \$JNJ jabs for the younger crowd. My COVID vaccine volunteering experience Vaccinated GetVaccinated COVID Moderna Pfizer Maryland Central Govt recently gave BharatBiotech permission to test Covaxin on 2-18 yrs age group, marking it an important milestone in vaccine development. Indonesia aims to be regional vaccine making hub. According to researchers data, both Pfizer and Moderna vaccines remain highly effective at preventing severe illness and death, even amid surging Delta Variant cases and a booster is not required. The U.S. Food and Drug Administration on Thursday authorized a booster dose of COVID-19 vaccines from Pfizer Inc and Moderna Inc for people with compromised immune systems. US FDA COVID19 COVID19 Vaccine CDCgov now recommends COVID19 Booster Shots for all eligible Americans. Egypt to roll out a vaccination campaign with locally produced additional 5 million doses Chinese Sinovac vaccines in Sept, Consultant to Minister of Health and Population for Research Noha Assem stated in a TV interview. Your queries answered. Deadline for booster dose for Sinopharm announced, if you received vaccine over six months PfizerBiontech COVID19 BharatBiotech releases a statement after the Subject Expert Committee on COVID19 recommends the emergency use of Covaxin for children 2-18 years of age. If the US FDA signs off on Moderna's booster, the U.S. Centers for Disease Control and Prevention will make specific</p>
--	--

Covaxin Travel	recommendations on who should get the shots. Indias COVID19 vaccine Covaxin will be added to the UK governments approved list of vaccines for international travellers from November 22. BharatBiotechs Covaxin now recognized by HongKong COVID19
----------------	--

**Table 11.** Composite View Summaries of Positive (left) and Negative (right) Views with EbS.

Peru launched the first stage of its national vaccination campaign against the novel coronavirus disease on Feb. 9, using vaccines developed by Chinese company Sinopharm to immunize healthcare workers. Argentina on Sunday approved the COVID19 vaccine developed by Chinese company Sinopharm for emergency use. PM narendramodi was administered a shot of Covaxin, the vaccine fully Researched & Developed as well as Made in India by BharatBiotech to mark phase 2 of Covid-19 inoculation campaign. CDSCO expert panel recommends moving Covaxin out of clinical trial mode Covaxin The COVID-19 vaccine developed by Chinas Sinopharm has been approved for emergency use in the Maldives, the Maldives Food and Drug Administration announced at a press conference Monday afternoon. Mauritius arms itself with COVAXIN, Indias indigenously developed vaccine, in its fight against COVID19 India stands strong with Mauritius in these tough times Consignment to arrive tomorrow A time-tested and enduring partnership IndiaMauritius The new recombinant COVID19 vaccine, developed by the National Vaccine &Serum Institute, a R&D center of Sinopharms bioscience subsidiary the China National Biotec Group, got approval from the National Medical Products Administration on Apr. 9. Covaxin, developed completely in India, can effectively neutralise multiple variants of SARS-CoV-2. Covaxin BharatBiotech developed Covaxin effective on B.1.617 and B.1.1.7, emerging variants first identified in India and UK respectively - Study by Journal Clinical Infectious Diseases PIBKochi COVIDNewsByMIB PIB India KirenRijiju BSF India CISFHQrs CRPF sector GMSRailway A study published in NEJM found that a 3rd dose of the Moderna or Pfizer vaccine significantly improved its effectiveness in organ transplant recipients who take immunosuppressant drugs. Moderna created the COVID19 vaccine using	Cambodian Prime Minister Hun Sen emphasized that Chinas vaccines are very safe and effective, and China will become the safest and most stable supplier of COVID-19 vaccines. Sinopharms COVID19 vaccines have just arrived in this first EU country approving the Chinese vaccine. Bharat Biotech confirms deal with Brazil to supply 20 million doses of COVAXIN vaccine CoronavirusVaccine BharatBiotech Covaxin Brazil India on Friday began using SputnikV in its battle against COVID19 with the first dose of the vaccine from Russia administered in Hyderabad. BharatBiotech announces the quick ramp-up of additional manufacturing capacities for COVAXIN at Chiron Behring Vaccines, Ankleshwar, Gujarat, a wholly-owned subsidiary of Bharat Biotech. The Kings sister had just approved on Thursday the Sinopharm covid19 vaccine be imported into Thailand as alternative vaccines to help the nation cope with the pandemic. Centre recently gave BharatBiotech permission to test Covaxin on 2-18 yrs age group, marking it an important milestone in vaccine development. The first batch of COVID19 vaccine supplied by Chinas Sinopharm to COVAX, the international vaccine campaign co-led by the World Health Organization, was officially rolled off the production line on Tuesday. From June 21 Covaxin will be the most expensive of the three vaccines which will be available in private hospitals. The Drugs Controller General of India allowed Indian pharmaceutical Cipla to import the Moderna mRNA COVID19 vaccine on Tuesday making it the fourth vaccine that will be available to Indians. Moderna approved for emergency use, 4th vaccine okayed by India COVID19Vaccine Global coronavirus death toll surpasses 468 million COVID19 vaccines Moderna CovidVictoria An expert committee recommended a booster dose of Modernas anti-Covid vaccine in the United States for certain at-risk groups, a month after making a similar
--	---

the sequence data released on the Internet. New study shows which vaccine offers the best protection against COVID19. As of now, Covaxin seems to be the most effective & long lasting vaccine against the Delta variant!	decision for the Pfizer shot. He further added that 96 countries have recognized both Covishield and Covaxin vaccines.
---	--

In the (S+SbS) summary (Table 9), the positive-sentiment summary focuses on encouraging vaccine developments, such as approval and distribution progress, and highlights Sinovac and Sinopharm in countries like China and regions supportive of these vaccines. The negative-sentiment summary addresses topics about vaccine discussions, primarily focusing on Sputnik V and Covaxin. The negative-sentiment summary expresses concerns about vaccine supply issues, regulatory actions, and logistical challenges. For example, the summary discusses high-profile reports and decisions around emergency authorizations, delays, and shortages, reflecting a current public concern. This difference between sentiment-based summaries reveals that negative views emphasize obstacles and critical opinions, while positive views celebrate advancements and achievements in vaccine deployment.

For S+TbS (Table 10), the summaries illustrate how different sentiments shape the narrative around vaccine distribution and public policy. The negative-sentiment summary focuses on policies and actions by governments and institutions about vaccine regulations, meanwhile highlighting delays, supply chain issues, and public dissatisfaction.

On the other hand, the positive-sentiment summary emphasizes successful implementation and the progress of vaccine campaigns. This difference captures the essence of sentiment-specific discussions by focusing on the relational dynamics between policies and outcomes.

In the (S+EbS) summaries (Table 11), the negative-sentiment summary focuses on discussions about the supply, import, and regulatory approval of vaccines, particularly during the early stages of vaccine availability in Asian countries like China, India, and Thailand. The summary reflects a sensitivity to logistical and regulatory challenges, which aligns with prevalent concerns at that time.

In contrast, the positive-sentiment summary brings attention to the successful development and effectiveness of vaccines, highlighting endorsements by various entities and governments. The positive perspective often includes mentions of vaccine approvals and endorsements, which underscore trust in vaccine efficiency and safety.

Different sentiments resulted in different summaries. In a polarized environment, some users might prefer to focus on positive tweets, e.g., the governing party, while the opposition most likely would stress the ideas expressed by negative tweets.

## 9. Summary with Perspective of Fake and Real News

The perspective analysis component can use the fake news detection models to identify fake vs. real news, as shown in [45]. In this section, we present a summary comparison generated from a dataset including 4,080 fake news items and 4,480 real news items [80]. The dataset distinguishing between fake and real news about COVID-19 was curated through both automated and manual verification methods.

For real news, data collection was conducted using the Twitter API to collect tweets specifically from verified accounts, such as the World Health Organization (WHO), the Centers for Disease Control and Prevention (CDC), and other relevant government and medical organizations. Each data item was reviewed by human annotators to confirm its factual accuracy and relevance to COVID-19. The process included verifying content regarding updates on vaccine progress, government responses, and statistical data on the pandemic to ensure the integrity and reliability of the data items.

For fake news, the collection involved a manual verification of social media posts, news articles, and public statements that had been previously flagged or debunked. Fact-checking websites such as Politifact, Snopes, and Boomlive served to confirm the falsehoods. Content from various platforms, including Facebook and Twitter, was examined against original documents or credible sources to confirm their status as fake. The dataset has two columns, text and label (Table 12).

**Table 12.** Examples of fake news and real news.

Text	Label
Politically Correct Woman (Almost) Uses Pandemic as Excuse Not to Reuse Plastic Bag #coronavirus #nashville	Fake
Covid Act Now found "on average each person in Illinois with COVID-19 is infecting 1.11 other people. Data shows that the infection growth rate has declined over time this factors in the stay-at-home order and other restrictions put in place."	Real

We leverage our MVSF to explore how different summarization perspectives capture the nuances to differentiate deceptive content from factual reporting. Therefore, this section aims to enhance the MVSF's support for informed decision-making. We present the experimental result of summaries generated by method (EbS) and method (TbS), and perform a comparative analysis of fake news and real news.

### 9.1. Entity-Based (EbS) Summarization of Fake News and Real News

In this subsection, we generate Entity-based summaries of fake news and real news (Table 13). The real news summaries focus on systemic and governmental responses to public health situations. The prominent entities (*state, case, death*), supported by the consistent triple score of 311.5 (Table 14), show a high level of organization and relevance to public policy and safety. All except for one sentiment are neutral, which reflects the factual and informative nature of the content.

**Table 13.** Summary Sentences of Real News (left) and Fake News (right) using method (EbS).

Indias Total Recoveries continue to rise cross 32.5 lakh today 5 States contribute 60% of total cases 62% of active cases and 70% of total fatality reported in India. Sometimes when a state reports a large number of deaths it is because they caught up on a reporting backlog of deaths that occurred long in the past. Few states reported race and ethnicity data at the beginning of April. State decides testing rates of COVID-19 for pvt labs. Several states are seeing outbreaks of in meat and poultry processing facilities. To protect the lives of healthcare workers every state needs stay-at-home orders NOW. Together with the States of Uttar Pradesh and Tamil Nadu these 5 states contribute nearly 60% of the total active cases. Some states also provide data about the date of death. These States are seeing a sudden surge in the number of cases and some of them are also reporting high mortality. States contribute 60% of total cases 62% of cases and 70% of total repo. Seven states saw the number of people hospitalized rise by 100 or more today. States reported relatively low numbers of tests (713 k) and cases (60 k). Six states saw a rise of over 100 (FL CA TX AZ GA TN) in their number of currently hospitalized COVID-19 patients. But looking at other metrics today the state reported	A photo shows people infected with coronavirus lying on the sidewalk in China. Video shows coronavirus patients and doctors. Video of a doctor fainted on the floor after getting infected with coronavirus. A long message attributed to Bill Gates, the Microsoft billionaire, encouraging people to reflect positively on their lives during the coronavirus outbreak has been shared in multiple countries. Chinese government is burning down people infected with Coronavirus. People shouting Allahu Akbar in Europe after the coronavirus outbreak. A video shows a man spitting inside food packets during the coronavirus crisis. Video shows Canadian PMs wife talking about the effects of Coronavirus. People in Ukraine will be forcibly vaccinated against the new coronavirus. Video shows people behaving abnormally in China due to coronavirus. A video shows a new hospital for coronavirus patients in China. People in Ahmedabad tested positive, 11 in Kanpur and 8 in Lucknow after being exposed to vegetable infected with coronavirus. Video shows Bodies of dead novel coronavirus patients in Russia. Video showing dead bodies of coronavirus patients in Osmania Hospitals mortuary. Video shows coronavirus infected notes scattered on Indore streets being sanitized. People who have never died before are now dying from coronavirus. A video showing a police officer briefing about cases being registered against the WhatsApp group
---	---

record hospitalizations and its 2nd-highest number of deaths. Reporting gaps in 20+ states leave the public in dark about the true scope of the pandemic.	admins is shared in the context of coronavirus lockdown. People died in Hyderabad due to the 2019 coronavirus. People infected with coronavirus die in the street while doctors travel through the infection zone.
---	--

**Table 14.** Summary Triples of Real News using method (EbS).

triple	sentiment	triplescore
5 states contribute 60 % of total cases	2	311.5
state reports large number of deaths	2	311.5
few states reported race data at beginning of april	2	311.5
state decides testing rates of covid-19 for pvt labs	2	311.5
several states are seeing outbreaks of in meat	2	311.5
state protect lives of healthcare workers	2	311.5
5 states together contribute nearly 60 % of total active cases	1	311.5
states also provide data about date of death	2	311.5
states are seeing sudden surge in number of cases	2	311.5
states contribute 60 % of total cases	2	311.5
seven states saw number of people	2	311.5
states reported relatively low numbers of tests	2	311.5
six states saw rise of over 100 in their number of currently hospitalized covid-19 patients	2	311.5
state reported its 2nd-highest number of deaths	2	311.5
states leave public in dark about true scope of pandemic	2	311.5

On the other hand, fake news summaries are characterized based on the depiction of vivid scenarios such as people infected, or dramatic events related to health crises. The prominent entities (*people, video, coronavirus*) include individual actions and localized events, with triple scores ranging slightly and generally lower than real news. The sentiments expressed are slightly more negative than for real news, which suggests a focus on shocking the audience and emotional content (Table 15). These summaries appear to capture attention through specific and impactful imagery and scenarios.

**Table 15.** Summary Triples of Fake News using method (EbS).

triple	sentiment	triplescore
	t	e
people infected with coronavirus	2	175.3
video shows coronavirus patients	2	177.4
video fainted getting with coronavirus	2	177.4
people reflect positively on their lives during coronavirus outbreak	2	175.3
people infected with coronavirus	2	175.3
people shouting allahu akbar after coronavirus outbreak	1	175.3
video shows man spitting inside food packets during coronavirus crisis	1	177.4
video shows canadian pms wife talking about effects of coronavirus	2	177.4
people will forcibly vaccinated against new coronavirus	2	175.3
people behaving abnormally due to coronavirus	2	175.3
video shows new hospital for coronavirus patients in china	2	177.4
people tested exposed to vegetable infected with coronavirus	2	175.3
video shows bodies of dead novel coronavirus patients in russia	1	177.4

video showing dead bodies of coronavirus patients in osmania hospitals mortuary	2	177.4
video shows coronavirus infected notes scattered on indore streets	1	177.4
people are now dying from coronavirus	2	175.3
video is shared in context of coronavirus lockdown	2	177.4
people died due to 2019 coronavirus	2	175.3
people infected with coronavirus	2	175.3

The emphasized entities are apparently different between real news and fake news. Real news summaries focus on systemic and governmental responses, such as *state actions, case updates, and death tolls*. This shows a consistent approach in how these entities are scored, which is supported by the consistent triple score, 311.5. This consistency indicates a standardized method of evaluating the importance of each entity, reinforcing the objective and factual nature of real news reporting.

On the other hand, fake news summaries frequently spotlight individual, localized, or even imaginary events, including *specific people, dramatic videos, or aspects related to the coronavirus*. The triple scores in fake news vary between 175.3 and 177.4, reflecting a broader and less consistent approach to how information is valued. This variability suggests that the evaluation of entities in fake news does not adhere to a consistent standard. This inconsistency is likely due to the more sensational and diverse topics covered, aiming to engage the audience on an emotional level.

Understanding these differences in entity focus and scoring consistency is important. It helps to discern the reliability and intent of news content. This also helps emphasize the need for critical evaluation when consuming news, especially in distinguishing between credible and misleading information. Our analytical results underscore the importance of recognizing the veracity in news summaries.

#### 9.2. Triple-Based (TbS) Summarization of Fake News and Real News

Table 16 shows triple-based summaries of fake news and real news. The real news summary illustrates a focus on responses to current public health challenges. The summary discusses *strategies, initiatives, and evaluations surrounding public health crises*. The results reflect a systematic approach to managing and communicating about public health events. By choosing triples with shortest distances from the centroid, we ensure that the topics discussed are central to the ongoing discourse and aligned with the core issues. Therefore, the experimental result suggests a focused and relevant dissemination of information.

**Table 16.** Summary Sentences of Real News (left) and Fake News (right) using method (TbS).

Principal Secretary to Prime Minister directed all concerned for an evidence based preparedness of all aspects of with active participation of Districts and States for effectiveness. Testing a drug will determine if an emergency use authorization comes by late fall. Our estimates suggest that once an effective vaccine has been distributed and international travel and trade is fully restored the economic gains will far outweigh the \$38 billion investment required for the ACT Accelerator- Dr Tedros. India records more than 82000 Recoveries for two days in a row Total Recoveries cross 40 lakhs Recovered Cases exceed Active Cases by more than 30 lakhs. CDCMMWR finds steps that help slow the spread of may also reduce if widely practiced. We continue to call on all countries to use every tool at their disposal to suppress transmission &	During Lock down period such a fantastic natural scenery on sea beach near Chandrabhaga, Puri to Konark marine drive road. BBC replaces Nichola Sturgeons pandemic briefings with Jamie Oliver making curried haggis. Spain corrected their number of deaths by COVID19 from more than 26,000 to 2,000. Nigel Farage to teach kids climate change denial for balance in BBC Lockdown Learning Scheme. Canadas COVID alert app warns that the virus is calling from inside the house. Well, I think climate change still counts as the worst federal response to a national emergency in our nations history. Are you struggling to work out the difference between real and fake news during the crisis? We fact checked Night 3 of Trumps actions to prevent COVID19 Biden on school choice and
---	--

save lives until & after we have a vaccine- Dr Tedros. The recovered cases (4674987) exceed active cases (966382) by more than 37 lakh. In Cross River our Rapid Response Team (R) is supporting the state to enhance sample collection and ramp-up testing for The also worked with the Cross River State Response Team to assess a 100-bed isolation facility in Adiabo Tinapa. County health depts ( Maricopa AZ and St. Louis MO) provide both states most reliable source of LTC data. High level teams will assist State/UT in strengthening public health measures for surveillance drharshvardhan Prakash Javdekar PIB India. Across 50 states and DC we've tracked 16502 total 1953 positive 13419 negative and 1130 pending.	defunding the police. Military COVID infected 118984. A majority of COVID19 deaths in the United States happen in a medical facility but people die at home too. Mr. Mandetta singles out the presidents COVID19 denialism as the biggest problem in tackling the pandemic. Takes us back to our childhood when the first line of defence against a common cold sore throat was an iodine tincture, or Betadine gargle. Trump ensures Americas stimulus checks will bounce by writing his name on them. There are sterilization agents in the COVID vax which can cause sterility not only in the patient but also in the sexual partners of people who have taken the shot. Rock Legend Beats Bug. Hindu gods quarantined due to COVID19.
--	--

On the other hand, the fake news summary is characterized by the distortion of facts, including the misleading presentation and incorrect information under the guise of news. These narratives often include sensational or controversial topics to attract attention or provoke emotional responses. The content involves outrageous claims or conspiracy theories that diverge from genuine news reporting.

The difference in method (TbS) between real and fake news lies in the alignment and integrity of the presented information. Real news aligns with current public discourse, reflecting ongoing societal or governmental responses. The real news summary aims to provide data and updates to promote knowledge and safety. In contrast, fake news deviates from such discussions, instead introducing irrelevant topics to distort perceptions or manipulate readers' emotions. Fake news lacks integrity and focuses on generating engagement through controversial or sensational content at the expense of relevance and truthfulness.

## 10. Prototype Application

In this section, we present our web prototype (Figure 4) <http://ai4sg.njit.edu/ai4sg/Summarize>. This app allows users to easily summarize social media discussions and other input text. Users can input a segment of text such as a social media thread or article, and customize summarization preferences based on key elements of the MVSF framework including method (Entity-based or Triple-based), sentiment perspective (negative or positive), or political perspective (left or right).

We also offer an ongoing dataset that captures COVID-19 government health policies [44], organized on a policy and month basis. Additionally, the web site contains a built-in sample set of 100 social media posts of fake news [45] about the COVID-19 pandemic. Users can utilize these datasets to create summaries according to their preferred summarization techniques and viewpoints. With an input token limit of 4,096 for each month's policy tweet compilation, the application employs a random selection mechanism to ensure it stays within this token boundary.

**Summarize Social Media Data in Different Perspectives using ChatGPT**

[About and Contributions of this App](#)

Summarize Twitter Posts of a government policy about COVID-19 pandemic

Select a COVID-19 Policy  Select a Year  Select a Month

Enter text

Your selected input data will be shown here. You can also paste your own text to perform summarization. This textbox can be cursor-dragged to adjust its size.

**Choose perspectives on different dimensions (Optional)**

Summarization Focus on  Entity-based  Event-based

Sentiment Perspective in  Negative  Positive

Political Bias Perspective  Left  Right

**Summarize**

**Summarize Sample Data** (100 social media posts of misleading information about COVID-19 Pandemic)

**Figure 4.** Screenshot of Web Prototype, Upper Part (a), and Lower Part (b) of the Webpage.

Figure 5 shows the summary of tweets about the Business Closing policy in Dec. 2021 (Table 17), using the Entity-based method, combed with the perspectives of negative sentiment and left political bias, while Figure 6 shows the summary of same tweet set using the Triple-based method, combined with the perspectives of positive sentiment and right political bias.

**Table 17.** Summary of Business Closing Policy Tweets in 2021-12.

I managed a restaurant in BH during COVID, not only did we never close, but business was booming due to delivery services being the main option of that soft shutdown. It never felt safe the entire time

Senate passes stopgap funding bill, avoiding shutdown, despite TED CRUZ VOTING TO SHUT DOWN GOVERNMENT. @SenTedCruz

Its crazy how in the beginning of the pandemic we were closing down and disinfecting everything because of the chance of Covid now when everyone basically has it its business as usual.

we need to shutdown the nation and pay people and business owners to stay home and let covid-19 run its course. It's absolutely clear you can't make Americans take the jab. Capitalism only benefit the few. #COVID19

This is pretty important information. We are closing in on looking at COVID in a very different way, and the messaging should support that. Get vaccinated. Get boosted. Time to trust science and live your lives.

So we learned nothing from last year or the social justice work that was done? Business as usual @BroadwayLeague not considering industry shutdown amid COVID cancellations |

The restaurant i work at just got shutdown bc we dont get enough business due to covid and ion really know how im gonna pay for the hotel i live in this week, or even this month. really not tryna live n my car again but ill do what i got to. soon as im almost up they push me down

### Policy, Year, Month:

Business Closing in 2021-12

### Summary:

The Senate has passed a stopgap funding bill, despite Ted Cruz voting for a government shutdown, while a restaurant worker is facing financial hardship due to their workplace's closure due to the pandemic. The Broadway League is being criticized for not considering the effects of COVID cancellations. There is a call to shut down the nation and pay people and business owners to stay home to allow COVID-19 to run its course, while an appreciation of science and vaccination is being promoted. An example of a restaurant in Beverly Hills that stayed open despite the pandemic and was successful due to delivery services was also shared. It was highlighted how the earlier precautions taken due to the pandemic are now being ignored as business continues as usual.

### Selected Options:

- Method: Entity
- Sentiment: Negative
- Political Bias: Left

**Figure 5.** Summary Text of Tweets about COVID-19 Business Closing policy in Dec 2021, using Entity method, negative sentiment, and left political bias perspectives.

### Policy, Year, Month:

Business Closing in 2021-12

### Summary:

The right-leaning political bias in the given text is that despite Senator Ted Cruz's attempt to shut down the government, the Senate passed a stopgap bill to avoid it. Businesses in BH managed to stay open despite the pandemic and benefit from delivery services. People are now looking at COVID in a different way and the messaging should be in support of getting vaccinated. However, some businesses were forced to close due to lack of customers.

### Selected Options:

- Method: Event
- Sentiment: Positive
- Political Bias: Right

**Figure 6.** Summary Text of Tweets about COVID-19 Business Closing policy in Dec 2021, using Event method, positive sentiment, and right political bias perspectives.

In our experiments we were not able to coax ChatGPT into using extractive summarization. Thus, our algorithms differ from the web application built with ChatGPT.

The two summaries differ in emphasis and tone based on the requested perspectives. The first (Figure 5), leaning left, emphasizes entities and negative sentiments: it accentuates Ted Cruz's controversial vote, underscores pandemic-induced hardships, and critiques some pandemic responses. The second summary (Figure 6), in contrast, is leaning right, and highlights positive events: it applauds the Senate's actions and emphasizes business resilience during the pandemic. Through selective extraction and strategic phrasing, the same input is transformed to resonate with different biases and foci.

Leveraging the power of ChatGPT [11], the application can rapidly produce a customized summary highlighting the most relevant information tailored to the user's preferences. This provides an accessible and user-friendly tool for text summarization that can supplement the computationally demanding algorithms within our framework. The integration of ChatGPT enables high-quality summarization capabilities within the overall MVSF architecture, while opening up these techniques to a broader general audience.

## 11. Discussion and Conclusions

We presented the Multiple View Summarization Framework (MVSF) that enables generating summaries of diverse perspectives with diverse methods from the same dataset. Entity-based (EbS) summarization focuses on frequent entities, Triple-based (TbS) summarization emphasizes the

events by ranking prominent triples, while Social Signal-based (SbS) summarization captures social prominence of tweets. Additionally, combining methods and views enables users to narrow down summaries to specific topics or sentiments. For instance, a user may focus on a summary of entities that are related to negative sentiments or possibly derived with a fake news analysis [45]. A notable challenge for our evaluation was the lack of a dataset of social media posts with gold standard summaries. Despite this, MVSF's competitive performance against state-of-the-art models on a news dataset showcases its efficacy.

Our MVSF offers several key theoretical implications for the field of summarization. By combining multiple extractive summarization perspectives within a unified framework, the MVSF enables targeted information extraction from social media content. This provides a more nuanced understanding of the underlying data. Additionally, MVSF's ability to cater to specific user interests and viewpoints introduces a novel approach to personalized summarization. It contributes to the advancement of research in extractive summarization by demonstrating the effectiveness of a multi-perspective approach. The practical implications of the MVSF are to offer content curation and decision-making support in the era of vast social media datasets. By offering personalized and fine-grained summaries, MVSF empowers users to efficiently extract vital information aligned with their specific interests and requirements. This capability might be particularly valuable for professionals, journalists, politicians, and social media analysts, who deal with large volumes of daily information.

The MVSF sets itself apart from existing work in several key ways. While traditional transformer-based models [57] such as BertSum, SBert, T5, and Bart-Large-CNN are limited in handling long input sequences due to the performance of their attention mechanism [58], the MVSF is more flexible when accepting input text of greater length. The comparative evaluations of state-of-the-art models demonstrate the MVSF's better performance, with an average gain of 14% in Rouge scores.

One major challenge we encountered is the ongoing instability of Twitter (now X), our primary source of social media posts. With policies changing frequently and new restrictions being imposed, such as limitations on the number of tweets users can access, Twitter's dynamics have become unpredictable. Additionally, Twitter is now charging fees. Despite these challenges, we have tailored our methods to effectively handle Twitter's unique characteristics, including its short, disconnected, and often ungrammatical posts. Moreover, we believe that MVSF holds potential for applications beyond Twitter, including other social networks such as Reddit. As digital landscapes evolve, we anticipate adapting our framework to suit varying contexts. Looking ahead, MVSF's adaptability and user-centric approach render it a valuable tool for comprehending the diverse perspectives of online discussions. For example, our MVSF has incorporated a comparative analysis of fake and real news. This offers a unique method for discerning and contrasting the reliability of information on social media. It facilitates developing a deeper understanding of misleading information dynamics and enhancing media literacy among users. By addressing user preferences and offering comprehensive insights into social media content, our framework paves the way for a more informed and insightful exploration of social media data by lay audiences and decision makers alike.

In conclusion, the MVSF represents an advance in the field of extractive summarization, with its combination of multiple perspectives and extractive methods within a unified framework. The practical implications of the MVSF are that it offers insights for decision-making support and analysis, while its theoretical implications contribute to the understanding of multiple information extraction perspectives and methods within a unified framework.

We are currently integrating the disparate elements of our summarization methodology into a unified real-time summarization web application. The design of this web application will enable users who are not affiliated with this research group to *choose a set of keywords*, a time range, a geographical range, and a collection period for tweets from X. Common-sense limitations will be imposed, and password registration will be required. Users can then return after the end of the collection period and select the exact combinations of methods and views that can help them comprehend "what Twitter/X users say about their topic of interest." This web application will be based on our own algorithms, as opposed to ChatGPT.

The methods that we described are extractive. In the future, we would like to extend our work to a hybrid (abstractive and extractive) summarization method. Having to construct new sentences as opposed to reusing existing sentences is an additional challenge.

The introduction of ChatGPT has allowed us to build a rapid web frontend within our MVSF framework that is not inhibited by the hardware limitations of the system on which we are running our own summarization algorithms. It remains to be determined in a comparative study between ChatGPT and our implementation, whether our purpose-built algorithms perform better than ChatGPT. While tailored methods often have an edge over generalized implementations, the question is still open whether this is the case for MVSF text summarization. Furthermore, we plan to experiment with having ChatGPT generate triples from social media posts and then use our methods to summarize them. This exploration will further allow us to compare the effectiveness of our customized algorithms with ChatGPT's handling of similar tasks.

Lastly, we are carefully monitoring the rapid advancement of ChatGPT, Google Bard, and their kin. These generative AI tools have enabled many new approaches to all AI tasks, while reducing long-established AI methods to technological anachronisms. We intend to make full use of the power of these new tools in our research.

**Acknowledgments:** Michael Renda has substantially contributed to the implementation of the MVSF framework and the web application.

## References

1. Li, C., Chun, S. A., Geller, J. (2023). Multiple View Summarization Framework for Social Media. Proceedings of the 36<sup>th</sup> International Florida Artificial Intelligence Research Society Conference, FLAIRS. <https://doi.org/10.32473/flairs.36.133169>
2. Sintek, M., Decker, S. (2002). TRIPLE—A Query, Inference, and Transformation Language for the Semantic Web. In: Horrocks, I., Hendler, J. (eds) The Semantic Web — ISWC 2002. ISWC 2002. Lecture Notes in Computer Science, vol 2342. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/3-540-48005-6\\_28](https://doi.org/10.1007/3-540-48005-6_28)
3. Lin, C. -Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. In Text Summarization Branches Out, Association for Computational Linguistics, Barcelona, Spain, 74–81. Retrieved from <https://aclanthology.org/W04-1013>
4. Widyassari, A. P., Rustad, S., Shidik, G. F., Noersasongko, E., Syukur, A., Affandy, A., Setiadi, D. R. I. M. (2022). Review of automatic text summarization techniques & methods. Journal of King Saud University - Computer and Information Sciences, 34(4), 1029–1046. <https://doi.org/10.1016/j.jksuci.2020.05.006>
5. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2020). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 7871–7880. <https://doi.org/10.18653/v1/2020.acl-main.703>
6. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P. J. (2019). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. arXiv. <https://doi.org/10.48550/ARXIV.1910.10683>
7. Zhang, S., Niu, J., & Wei, C. (2021). Fine-grained Factual Consistency Assessment for Abstractive Summarization Models. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 107–116. <https://doi.org/10.18653/v1/2021.emnlp-main.9>
8. Ott, M., Edunov, S., Baevski, A., Fan, A., Gross, S., Ng, N., Grangier, D., Auli, M. (2019). fairseq: A Fast, Extensible Toolkit for Sequence Modeling. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations), 48–53. <https://doi.org/10.18653/v1/N19-4009>
9. Zhang, J., Zhao, Y., Saleh, M., Liu, P. J. (2019). PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization. arXiv. <https://doi.org/10.48550/ARXIV.1912.08777>
10. Chi, Z., Dong, L., Wei, F., Wang, W., Mao, X. L., Huang, H. (2019). Cross-Lingual Natural Language Generation via Pre-Training. arXiv. <https://doi.org/10.48550/ARXIV.1909.10481>

11. Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*, 74(5), 570–581. <https://doi.org/10.1002/asi.24750>
12. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. <https://api.semanticscholar.org/CorpusID:160025533>
13. Liu, Y. (2019). Fine-tune BERT for Extractive Summarization. arXiv. <https://doi.org/10.48550/ARXIV.1903.10318>
14. Reimers, N., Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. arXiv. <https://doi.org/10.48550/ARXIV.1908.10084>
15. Joshi, A., Fidalgo, E., Alegre, E., Alaiz-Rodriguez, R. (2022). RankSum - An unsupervised extractive text summarization based on rank fusion. *Expert Systems with Applications*, 200, 116846. <https://doi.org/10.1016/j.eswa.2022.116846>
16. Jia, R., Cao, Y., Tang, H., Fang, F., Cao, C., Wang, S. (2020). Neural Extractive Summarization with Hierarchical Attentive Heterogeneous Graph Network. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 3622–3631. <https://doi.org/10.18653/v1/2020.emnlp-main.295>
17. Kwon, J., Kobayashi, N., Kamigaito, H., Okumura, M. (2021). Considering Nested Tree Structure in Sentence Extractive Summarization with Pre-trained Transformer. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Online, 4039–4044. <https://doi.org/10.18653/v1/2021.emnlp-main.330>
18. Roush, A., Balaji, A. (2020). DebateSum: A large-scale argument mining and summarization dataset. arXiv. <https://doi.org/10.48550/ARXIV.2011.07251>
19. Gu, N., Ash, E., Hahnloser, R. H. R. (2021). MemSum: Extractive Summarization of Long Documents Using Multi-Step Episodic Markov Decision Processes. arXiv. <https://doi.org/10.48550/ARXIV.2107.08929>
20. Barrios, F., López, F., Argerich, L., Wachenchauzer, R. (2016) Variations of the Similarity Function of TextRank for Automated Summarization. arXiv 2016. <https://doi.org/10.48550/ARXIV.1602.03606>
21. Blei, D. M., Ng, A. Y., Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3, 993–1022. <http://dx.doi.org/10.1162/jmlr.2003.3.4-5.993>
22. Pennington, J., Socher, R., Manning, C. (2014). GloVe: Global Vectors for Word Representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1532–1543. <https://doi.org/10.3115/v1/D14-1162>
23. Hermann, K. M., Kočiský, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M., Blunsom, P. (2015). Teaching machines to read and comprehend. In Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'15). MIT Press, Cambridge, MA, USA, 1693–1701.
24. Khatri, C., Goel, E., Hedayatnia, B., Metanillou, A., Venkatesh, A., Gabriel, R., Mandal, A. (2018) "Contextual Topic Modeling For Dialog Systems," 2018 IEEE Spoken Language Technology Workshop (SLT), Athens, Greece, 2018, pp. 892-899, <https://doi.org/10.1109/SLT.2018.8639552>
25. Rudra, K., Ghosh, D., Ganguly, N., Goyal, P., Ghosh, S. (2015). Extracting Situational Information from Microblogs during Disaster Events: A Classification-Summarization Approach. Proceedings of the 24<sup>th</sup> ACM International on Conference on Information and Knowledge Management, 583–592. <https://doi.org/10.1145/2806416.2806485>
26. Yadav, A.K., Singh, A., Dhiman, M. (2022). Extractive text summarization using deep learning approach. *Int. j. inf. tecnol.* 14, 2407–2415 (2022). <https://doi.org/10.1007/s41870-022-00863-7>
27. Modhe, S. V., Panse, N. S., Rathi, M. U., Chavan, S. (2021). Extractive Based Approach For Microblogs Summarization. International Research Journal of Modernization in Engineering Technology and Science Volume:03/Issue:03/March-2021.
28. Geng, F., Liu, Q., Zhang, P. (2020). A time-aware query-focused summarization of an evolving microblogging stream via sentence extraction, *Digital Communications and Networks*, Volume 6, Issue 3, 2020, Pages 389-397, ISSN 2352-8648, <https://doi.org/10.1016/j.dcan.2020.04.014>

29. Saini, N., Saha, S., Bhattacharyya, P. (2019). "Multiobjective-based approach for microblog summarization," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 6, pp. 1219–1231, 2019. <https://doi.org/10.1109/TCSS.2019.2945172>

30. Dutta, S., Chandra, V., Mehra, K., Ghatak, S., Das, A.K., Ghosh, S. (2019). Summarizing Microblogs During Emergency Events: A Comparison of Extractive Summarization Algorithms. In: Abraham, A., Dutta, P., Mandal, J., Bhattacharya, A., Dutta, S. (eds) *Emerging Technologies in Data Mining and Information Security*. *Advances in Intelligent Systems and Computing*, vol 813. Springer, Singapore. [https://doi.org/10.1007/978-981-13-1498-8\\_76](https://doi.org/10.1007/978-981-13-1498-8_76) .

31. Olariu, A. (2014). Efficient Online Summarization of Microblogging Streams. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, volume 2: Short Papers, Association for Computational Linguistics, Gothenburg, Sweden, 236–240. <https://doi.org/10.3115/v1/E14-4046>

32. Sharifi, B., Hutton, M. A., Kalita, J. K. (2010). Experiments in Microblog Summarization. 49-56. <https://doi.org/10.1109/SocialCom.2010.17>

33. Inouye, D. (2010). Multiple post microblog summarization. *REU Research Final Report*. 2010 Jul;1:34–40.

34. Jin, X., Han, J. (2011). K-Means Clustering. In: Sammut C., Webb G.I. (eds) *Encyclopedia of Machine Learning*. Springer, Boston, MA. [https://doi.org/10.1007/978-0-387-30164-8\\_425](https://doi.org/10.1007/978-0-387-30164-8_425)

35. Gunaratna, K., Yazdavar, A. H., Thirunarayan, K., Sheth, A., Cheng, G. (2017) Relatedness-based multi-entity summarization. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence Main track*. Pages 1060-1066. <https://doi.org/10.24963/ijcai.2017/147>.

36. Gunaratna, K., Thirunarayan, K., Sheth, A. (2015). FACES: Diversity-Aware Entity Summarization Using Incremental Hierarchical Conceptual Clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*. 29, 1 (Feb. 2015). <https://doi.org/10.1609/aaai.v29i1.9180>

37. Gunaratna, K., Thirunarayan, K., Sheth, A., Cheng, G. (2016). Gleaning Types for Literals in RDF Triples with Application to Entity Summarization. In: Sack, H., Blomqvist, E., d'Aquin, M., Ghidini, C., Ponzetto, S., Lange, C. (eds) *The Semantic Web. Latest Advances and New Domains. ESWC 2016*. vol 9678. Springer, Cham. [https://doi.org/10.1007/978-3-319-34129-3\\_6](https://doi.org/10.1007/978-3-319-34129-3_6)

38. Li, C., Chun, S. A., Geller, J. (2021). Knowledge graph analysis of Russian trolls. In C. Quix, S. Hammoudi, & W. van der Aalst (Eds.), *Proceedings of the 10th International Conference on Data Science, Technology and Applications, DATA 2021* (pp. 335-342). (*Proceedings of the 10th International Conference on Data Science, Technology and Applications, DATA 2021*). SciTePress. <https://doi.org/10.5220/0010605403350342>

39. Gunaratna, K., Yazdavar, A. H., Thirunarayan, K., Sheth, A. and Cheng, G. (2017). Relatedness-based Multi-Entity Summarization. *IJCAI : proceedings of the conference*, 2017, 1060–1066. <https://doi.org/10.24963/ijcai.2017/147>

40. Liu, X., Li, Y., Wei, F., & Zhou, M. (2012). Graph-Based Multi-Tweet Summarization using Social Signals. *Proceedings of COLING 2012*, 1699–1714. <https://aclanthology.org/C12-1104>

41. Angeli, G., Premkumar, M. J. J., Manning, C. D. (2015). Leveraging Linguistic Structure For Open Domain Information Extraction. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 344–354. <https://doi.org/10.3115/v1/P15-1034>

42. Niklaus, C., Cetto, M., Freitas, A., Handschuh, S. (2018). A Survey on Open Information Extraction. *Proceedings of the 27th International Conference on Computational Linguistics*, 3866–3878. <https://aclanthology.org/C18-1326>

43. Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C.D., Ng, A., Potts, C. (2013). Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, WA, 18–21 October 2013. Available online: <https://aclanthology.org/D13-1170>

44. Li, C., Renda, M., Yusuf, F., Geller, J., Chun, S.A. (2022). Public Health Policy Monitoring through Public Perceptions: A Case of COVID-19 Tweet Analysis. *Journal of Multidisciplinary Digital Publishing Institute (MDPI)*. *Information* 2022, 13, 543. <https://doi.org/10.3390/info13110543>

45. Li, C. Y., Chun, S. A., Geller, J. (2022). Stemming the Tide of Fake News about the COVID-19 Pandemic. Proceedings of the International Florida Artificial Intelligence Research Society Conference, FLAIRS, 35. <https://doi.org/10.32473/flairs.v35i.130716>

46. Fellbaum, C. (2005). WordNet and wordnets. In: Brown, Keith et al. (eds.), Encyclopedia of Language and Linguistics, Second Edition, Oxford: Elsevier, 665-670.

47. Lesk, M. (1986). Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone. Proceedings of the 5th Annual International Conference on Systems Documentation, 24–26. <https://doi.org/10.1145/318723.318728>

48. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers) (pp. 4171–4186). Association for Computational Linguistics. <https://doi.org/10.18653/v1/n19-1423>

49. Bank, D., Koenigstein, N., Giryes, R. (2020). Autoencoders. arXiv. <https://doi.org/10.48550/ARXIV.2003.05991>

50. Petrov, C. (2023). 50+ Stunning Twitter Statistics You Need to Know in 2023. <https://techjury.net/blog/twitter-statistics/> Retrieved March 1<sup>st</sup>, 2023.

51. Greene, D., Cunningham, P. (2006). Practical Solutions to the Problem of Diagonal Dominance in Kernel Document Clustering. Proceedings of the 23rd International Conference on Machine Learning, 377–384. <https://doi.org/10.1145/1143844.1143892>

52. Preda, G. (2021). COVID-19 All Vaccines Tweets. Kaggle. <https://www.kaggle.com/datasets/gpreda/all-covid19-vaccines-tweets> Retrieved Oct 1<sup>st</sup>, 2022.

53. Google Translate API for Python. (2023). Retrieved from <https://py-googletrans.readthedocs.io/en/latest/> on March 13<sup>th</sup>, 2023.

54. GeoPy's documentation. (2023). Retrieved from <https://geopy.readthedocs.io/en/stable/> on March 13<sup>th</sup>, 2023.

55. Qi, P., Zhang, Y., Zhang, Y., Bolton, J., Manning, C. D. (2020). Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 101–108. <https://doi.org/10.18653/v1/2020.acl-demos.14>

56. Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 55–60. <https://doi.org/10.3115/v1/P14-5010>

57. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I. (2017). Attention Is All You Need. arXiv. <https://doi.org/10.48550/ARXIV.1706.03762>

58. Rohde, T., Wu, X., Liu, Y. (2021). Hierarchical Learning for Generation with Long Source Sequences. arXiv. <https://doi.org/10.48550/ARXIV.2104.07545>

59. Arthur, D., Vassilvitskii, S. (2007). K-means++: the advantages of careful seeding. In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms (SODA '07). Society for Industrial and Applied Mathematics, USA, 1027–1035.

60. Zhao, Y., Karypis, George. (2001). Criterion Functions for Document Clustering: Experiments and Analysis. Retrieved from the University of Minnesota Digital Conservancy, <https://hdl.handle.net/11299/215490>

61. Hahn, U., Mani, I. (2000) The challenges of automatic summarization," in Computer, vol. 33, no. 11, pp. 29-36, Nov. 2000, <https://doi.org/10.1109/2.881692>

62. Sharifi, B., Hutton, M. A., Kalita, J. (2010). Automatic Summarization of Twitter Topics. National Workshop on Design and Analysis of Algorithm, Tezpur, India, 2010. <https://api.semanticscholar.org/CorpusID:17725346>

63. Shirky, C. (2011). The Political Power of Social Media: Technology, the Public Sphere, and Political Change. Foreign Affairs, 90(1), 28–41. <http://www.jstor.org/stable/25800379>

64. Saveski, M., Gillani, N., Yuan, A., Vijayaraghavan, P., & Roy, D. (2022). Perspective-Taking to Reduce Affective Polarization on Social Media. Proceedings of the International AAAI Conference on Web and Social Media, 16(1), 885-895. <https://doi.org/10.1609/icwsm.v16i1.19343>

65. Bakshy, E., Messing, S., Adamic, L. (2015). Exposure to Ideologically Diverse News and Opinion on Facebook. *Science*, 348(6239), 1130–1132. <https://doi.org/10.1126/science.aaa1160>

66. Sunstein, C. R. (2017). *#Republic: Divided Democracy in the Age of Social Media*. Princeton: Princeton University Press. [https://doi.org/10.2307/j.ctv8xnhtd\\_](https://doi.org/10.2307/j.ctv8xnhtd_)

67. Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., & Michele Starnini. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), e2023301118. <https://doi.org/10.1073/pnas.2023301118>

68. Barberá, P. (2020). Social Media, Echo Chambers, and Political Polarization. In N. Persily & J. Tucker (Eds.), *Social Media and Democracy: The State of the Field, Prospects for Reform* (SSRC Anxieties of Democracy, pp. 34-55). Cambridge: Cambridge University Press.

69. Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>

70. Kitchens, Brent., Johnson, Steven L., Gray, Peter. (2020). "Understanding Echo Chambers and Filter Bubbles: The Impact of Social Media on Diversification and Partisan Shifts in News Consumption," *MIS Quarterly*, (44: 4) pp. 1619-1649. <https://doi.org/10.25300/MISQ/2020/16371>

71. Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why Differences Make a Difference: A Field Study of Diversity, Conflict, and Performance in Workgroups. *Administrative Science Quarterly*, 44(4), 741–763. <https://doi.org/10.2307/2667054>

72. Mutz, D. C., Martin, P. S. 2001. "Facilitating Communication across Lines of Political Difference: The Role of Mass Media," *American Political Science Review*, 95(1), 97-114. <https://doi.org/10.1017/S0003055401000223>

73. Chan, C. K., Zhao, M. M., & Lee, P. S. (2023). Determinants of escape from echo chambers: The predictive power of political orientation, social media use, and demographics. *Global Media and China*, 8(2), 155–173. <https://doi.org/10.1177/20594364221140820>

74. Rule, J. B. (2010). Cass R. Sunstein, Going to Extremes: How Like Minds Unite and Divide. *Social Service Review*, 84(2), 309–312. <https://doi.org/10.1086/655101>

75. Mourão, R., Thorson, E., Chen, W., Tham, S. (2018). Media repertoires and news trust during the early Trump administration. *Journalism Studies*, 19(13), 1945–1956. <https://doi.org/10.1080/1461670x.2018.1500492>

76. Vodolazova, T., Lloret, E. (2019). Towards Adaptive Text Summarization: How Does Compression Rate Affect Summary Readability of L2 Texts? *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, 1265–1274. [https://doi.org/10.26615/978-954-452-056-4\\_145](https://doi.org/10.26615/978-954-452-056-4_145)

77. Goyal, T., Durrett, G. (2021). Annotating and Modeling Fine-grained Factuality in Summarization. *North American Chapter of the Association for Computational Linguistics*. <https://doi.org/10.48448/v6sp-h865>

78. Li, S., Xu, J. (2023). MRC-Sum: An MRC framework for extractive summarization of academic articles in natural sciences and medicine. *Information Processing & Management*, 60(5), 103467. <https://doi.org/10.1016/j.ipm.2023.103467>

79. Chun, S., Holowczak, R., Dharan, K., Wang, R., Basu, S., Geller, J. 2019. Detecting Political Bias Trolls in Twitter Data. In *Proceedings of the 15th International Conference on Web Information Systems and Technologies (WEBIST 2019)*. SCITEPRESS - Science and Technology Publications, Lda, Setubal, PRT, 334–342. <https://doi.org/10.5220/0008350303340342>

80. Patwa, P., Sharma, S., Pykl, S., Guptha, V., Kumari, G., Akhtar, M. S., Ekbal, A., Das, A., & Chakraborty, T. (2021). Fighting an Infodemic: COVID-19 Fake News Dataset. In *Combating Online Hostile Posts in Regional Languages during Emergency Situation* (pp. 21–29). Springer International Publishing. [https://doi.org/10.1007/978-3-030-73696-5\\_3](https://doi.org/10.1007/978-3-030-73696-5_3)

81. Chowdhary, M., & Chahal, P. (2023, July). A Systematic Review of Challenges in Information Retrieval from Online Social Networking Platforms. In *International Conference on Artificial-Business Analytics*,

*Quantum and Machine Learning* (pp. 719-734). Singapore: Springer Nature Singapore. [https://doi.org/10.1007/978-981-97-2508-3\\_53](https://doi.org/10.1007/978-981-97-2508-3_53)

82. Li, H., & Chaturvedi, S. (2024). Rationale-based Opinion Summarization. *arXiv preprint arXiv:2404.00217*.
83. Amplayo, R. K., Angelidis, S., & Lapata, M. (2021). Aspect-controllable opinion summarization. *arXiv preprint arXiv:2109.03171*.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.