Unsupervised approach for misinformation detection in Russia-Ukraine war news

Nina Khairova [1,2,*,†], Andrea Galassi[3,†], Fabrizio Lo Scudo[4,†], Bogdan Ivasiuk[3,†] and Ivan Redozub[2,†]

[1] *Umeå University, 90187, Sweden*

[2] *National Technical University "Kharkiv Polytechnic Institute", 2, Kyrpychova str., 61002, Kharkiv, Ukraine*

[3] *University of Bologna, Viale Risorgimento 2, Bologna, 40136, Italy*

[4] *University of Calabria, via Bucci, Rende, 87036, Italy*

Abstract

The Russian-Ukrainian war has attracted considerable global attention; however, fake news often obstructs the formation of public opinion and disseminates false information. To address this issue, we have curated the RUWA dataset, comprising over 16,500 news articles covering the pivotal events of the Russian invasion of Ukraine. These articles were sourced from established outlets in the USA, EU, Asia, Ukraine, and Russia, spanning the period from February to September 2022. The paper explores the use of semantic similarity to compare different aspects of articles from various web sources that cover the same events of the war. This unsupervised machine learning approach becomes crucial when obtaining annotated datasets is practically impossible due to the lack of real fact-checking during the ongoing war. The research goal is to uncover the potential of employing semantic similarity measures as a viable approach for detecting misinformation in news articles.

Keywords

Misinformation issues, fake news detection, Russian-Ukraine war, dataset, semantic similarity

1. Introduction

Misinformation, fake news, and disinformation have been present throughout human history. However, it was only after intense discussions around fake news during events such as the 2016 U.S. presidential election, the U.K. Brexit referendum, and the global spread of the coronavirus that the issue gained heightened attention. One indirect piece of evidence is the increasing number of scientific articles addressing this problem, especially in the last three years. Currently, according to the Scopus database, more than 22,500 scientific papers are related to the concept of 'misinformation'.

CLW-2024: Computational Linguistics Workshop at 8th International Conference on Computational Linguistics and Intelligent Systems (CoLInS-2024), April 12–13, 2024, Lviv, Ukraine

* Corresponding author.

† These authors contributed equally.

khairova.nina@gmail.com (N. Khairova); a.galassi@unibo.it (A. Galassi);); fabrizio.loscudo@unical.it (F. L. Scudo); bogdan.ivasiuk@studio.unibo.it (B. Ivasiuk); ivan.red707@gmail.com (I. Redozub)

0000-0002-9826-0286 (N. Khairova); 0000-0001-9711-7042 (A. Galassi); 0000-0001-9092-8918 (F. Lo Scudo); 0009-0003-1658-0804 (B. Ivasiuk); 0009-0008-3312-4700 (I. Redozub)

There are obvious technical and cognitive foundations that can explain why misinformation has become a significant issue in contemporary digital society. The development of information technology has increased the reliance of many individuals on online sources for their information and news. A recent Eurostat study revealed that 72% of internet users in the European Union now obtain their news online. Similar trends can be observed among adult Internet consumers in the USA. The trend of widespread and rapid dissemination of misinformation (fake news) leads to the influence of information or misinformation campaigns on large groups of people simultaneously.

The psychological and cognitive foundations influencing society's susceptibility to misinformation stem from the natural difficulties humans face in distinguishing between real and fake news. Two major factors make people naturally vulnerable to fake news. The first factor named Naive realism [1] suggests that people tend to believe that their perception of reality is the only true one, while others who disagree with this are considered ignorant, irrational, or biased. Furthermore, according to the theory known as Confirmation bias [2], it is challenging to correct a misperception once it has formed. Psychological studies indicate that attempting to correct false information, such as fake news, by presenting true, factual information is not only unhelpful in reducing misperceptions but can sometimes even exacerbate them, particularly among ideological groups [3].

All these technical and cognitive factors contribute to the potential for large-scale misinformation campaigns conducted by large corporations or even by certain governments to influence public opinion. The most sensitive areas affected by these campaigns often include social division, public health, and economic impact [4].

However, the most significant threats may arise from the political consequences of misinformation campaigns. These campaigns can aim not only to alter election outcomes but also to influence the course of wars. One notable and early example of the political consequences of misinformation was highlighted by a spokesman for the German government in January 2017. He stated that they confronted a wide array of Russian propaganda tools used to conduct disinformation campaigns aimed at destabilizing the German government. He remarked, 'We are dealing with a phenomenon of a dimension that has never been seen before".

Furthermore, an indirect consequence of misinformation, in general, is that it is possible to disrupt the authenticity balance within the news ecosystem, thereby altering people's perceptions and responses to real news. These fosters doubt and confusion, making it more challenging for individuals to differentiate between truth and falsehood. This year the European Council has recognized misinformation, especially the one carried out by Russia, as a "a long-term challenge for European democracies and societies".

2. Background

Numerous definitions of misinformation exist; however, the crux of the matter can be succinctly encapsulated as follows: Misinformation is intentionally and verifiably false information published or posted to mislead readers [4, 5]. This definition comprises three pivotal concepts. Authenticity revolves around the verification of information as either real or false. An illustrative instance of misinformation may manifest in the form of unfounded claims or rumors disseminated through social media platforms regarding medicines or health remedies for treating or preventing the coronavirus. However, the veracity of such information can be substantiated or debunked through reliance on credible and reputable sources, such as the World Health Organization (WHO).

The second indicator of misinformation involves intentionality, signifying the deliberate dissemination of inaccurate information to achieve specific goals. The final critical aspect incorporated into the definition of misinformation pertains to the dissemination—spreading false or misleading information through various mediums, such as articles, social media posts, websites, or any platform where information is publicly shared, is a mandatory requirement for misinformation.

2.1. Approaches for Misinformation Detection

Currently, the majority of studies focused on detecting misinformation utilize Machine Learning or Deep Learning approaches [6]. These studies typically involve four main steps: data source selection, data collection, data cleaning, and the application of classification or clustering techniques. In the case of Machine Learning approaches an additional step for feature extraction is included. Figure 1 illustrates the general schema for misinformation detection approaches.

Most research focuses on specific types of data sources, often concentrating either on misinformation detection in social media posts [7] or on fake news in articles on news websites [8]. Additionally, a group of studies considers utilizing machine learning approaches for misinformation detection by processing existing datasets of fake news.

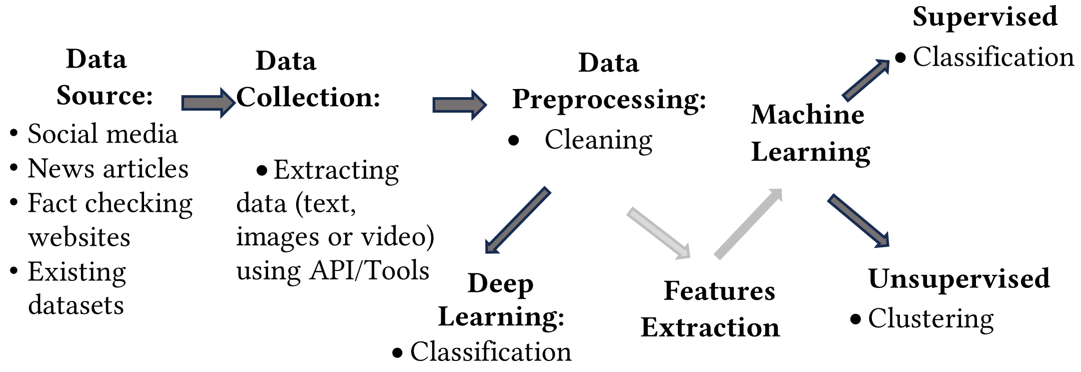

Figure 1: The general schema for misinformation detection approaches

The selection of a specific data source type has an impact on the features that can be utilized by machine learning models. For instance, features relevant to the propagation properties of information can be extracted specifically from the social media context. This features group includes users' profiles and various aspects of user demographics, such as registration age, number of followers, number of tweets that the user has authored, and the average number of followers, etc [9]. Network-based features, which are extracted to represent relationships among relevant users and posts, also pertain to the group of propagation properties.

Meanwhile, obtaining the propagation feature type from the news articles on the website is nearly impossible. For misinformation detection in these data sources, *style-based* or *knowledge-based* features are typically extracted [10]. Style-based methods aim to identify fake news by analyzing the manipulative elements present in the writing style of news content. The extraction of *style-based* features relies on the assumption that information created to intentionally deceive the public must sound 'more persuasive' compared to text without such intentions. These specific characteristics serve to reinforce deceptive statements or claims in news content, encompassing both psychological and linguistic aspects. Generally, for misinformation detection ML models apply the same linguistic-based features as for the general NLP tasks, such as text classification and clustering, or, for example, specific applications for author identification [11]. For instance, the average characteristics per sentence, subjective verbs (e.g. "feel", "believe"), report verbs (e.g. "announce"), positive/negative words, anxiety/angry/sadness words, and so on [12]. Some rhetorical techniques, such as repetitions, appeal to authority, exaggeration, or minimization, can be considered as psychological features [13].

The *knowledge-based* features for misinformation detection rely on factual knowledge. Traditionally, the unified standard definition of knowledge for their automated extraction is that knowledge consists of a set of (Subject, Predicate, Object) triples extracted from the given information, which well represents the provided information [14]. In the context of knowledge-based fake news detection, a commonly employed approach is fact-checking.

Fact-checking involves evaluating the authenticity of news by comparing statements extracted from the content to be verified with established factual knowledge. Utilizing *expert-based manual* fact-checking, often employed in the creation of fake news datasets, yields highly accurate results. However, this approach is expensive and becomes less efficient as the volume of news content to be checked increases. It is relatively uncommon to utilize a crowd-sourced approach for manual fact-checking. *Crowd-sourced fact-checking* involves a large number of ordinary individuals who contribute as fact-checkers. For instance, the publicly available large-scale fake news dataset, CREDBANK, was created in that way [15]. However, currently, this approach, similar to automatic fact-checking, faces challenges such as redundancy, invalidity, conflicts, and incompleteness, leading to relatively low accuracy and credibility [16, 17].

Many contemporary strategies for detecting misinformation center around extracting the mentioned features and integrating them into supervised classification models. These models are often based on Naïve Bayes, decision trees, logistic regression, k nearest neighbor (KNN), and support vector machines (SVM). The final selection of the classifier is typically based on comparing the performance of all utilized models [16, 17, 18].

All these supervised approaches require a pre-annotated fake news ground truth dataset or truth/false-annotated dataset to train a model. However, obtaining a reliable fake news dataset is a very time-consuming process as that often requires expert skilled annotators to conduct a meticulous examination of claims, along with assessing additional evidence, and reports from authoritative sources. This highlights the primary reason why, despite supervised classification methods potentially yielding more accurate models with a well-curated ground truth dataset for training, unsupervised models can be more practical due to the ease of obtaining unlabeled datasets.

Moreover, the exploration of embedding techniques, such as word embedding and deep neural networks, has attracted considerable attention in the extraction of textual features, showing potential for yielding positive outcomes in misinformation detection. Within the realm of Deep Learning (DL) models, news content is frequently subjected to word-level embedding as an initial step [19]. Subsequently, a proficiently trained neural network processes this embedding [20].

While DL models offer various advantages, they similar to supervised Machine Learning (ML) often perform better with large labeled datasets for training. However, acquiring and reliably annotating such datasets can be challenging and is not always addressed in misinformation detection. Moreover, DL models, especially when trained on limited or biased datasets, may be prone to overfitting. This means that the model may perform well on the training data but struggle to generalize to new data related to slightly offset topics that can be very considerable for fake news or misinformation-covered broad themes [7].

In recent years, the proliferation of visual content has become a significant tool for propagating fake news. Visual features extracted from images are crucial indicators in discerning fake news. At the same time, the rise of images and videos generated by neural networks, commonly known as 'deep fake videos', adds a new layer of complexity. Distinguishing between real and fake visual content becomes increasingly challenging. For these studies, new techniques for following the trace of revision and generation in a video are required [21].

2.2. Existing dataset

As mentioned in Section 2.1, statistical approaches to misinformation detection are generally constrained by the significant limitation of lacking labeled benchmark datasets.

Existing labeled datasets primarily focus on political news and are annotated through manual efforts [22] or by leveraging fact-checking websites like PolitiFact or GossipCop [23]. For instance, the Buzzfeed dataset [24] consists of 1,627 articles verified through manual fact-checking by professional journalists at BuzzFeed. These articles were sourced from nine prominent political publishers, three from the mainstream, hyperpartisan left-wing, and hyperpartisan right-wing categories. In total, the corpus includes 299 instances of fake news, with 97% originating from hyperpartisan publishers.

In certain instances, authentic news sources were selected from a designated group of reliable outlets, whereas fake news sources were drawn from known fake news lists, such as "Business Insider’s Zimdars Fake News list" [25]. Another annotation approach for the fake news dataset involved the AMT dataset [26], which comprises 480 articles annotated as either fake or true. In this dataset, fake news articles were intentionally crafted by journalists, while genuine news pieces were sourced from various domains.

The datasets focusing on fake news related to conflicts or wars exhibit a distinct nature. Take, for instance, the FA-KES dataset [27], which encompasses 804 news articles related to the Syrian war gathered from sources like Reuters, Etilaf, and others. To determine the veracity of the information, the creators employed a crowdsourcing platform, soliciting individuals for details on the number of casualties, and when, and where the events occurred. The obtained data was then compared with information from the Syrian Violation Documentation Center (VDC), which meticulously records all deaths during specific events.

Additionally, The CheckThat! initiative [5] [28], over its six iterations, has produced several datasets to address specific subtasks of the fact-checking problem, such as recognizing if a sentence should be checked [29] or if it contains a subjective perspective [30].

A summary of the most renowned annotated fake news datasets and their annotation methodologies is provided in Table 1.

Understandably, we were not able to detect false/true annotated datasets related to the Russian-Ukraine war, especially during the ongoing conflict. When considering the broader issue of fact-checking, additional challenges arise, including biases in sources and the 'fog of war' effect [28]. However, several researchers tackled the issue of dataset collection from social networks, mostly from Twitter, in the specific context of propaganda and fake news detection related to the Russian invasion of Ukraine.

Table 1

Comparison of existing fake news data sets

Dataset	Size	Domain	Annotation method	Labels set
Buzzfeed	1627 articles	US political news	fact-checked by journalists	Mostly true / mixture of true and false / mostly false/no factual content
LIAR	12.8K short statements	Economics, history, job-accomplish, foreign policy, etc.	manually labelled by PolitiFact fact-checkers	False / barely true / half-true / mostly true/true
FA-KES dataset	804 news articles	News events around the Syrian war	A semi-supervised fact-checking labelling approach based on VDC	Fake/ true
AMT	480 articles	The set of domains	The imitation of fake news	Fake/ true
Kaggle	208,00 texts	Mostly US political news	Based on unreliable sources	Fake/ true
FakeNews Net	>900 and < 17000 articles	The political campaign, gossip about the celebrities	Articles from fact-checking websites	Fake/ true
FakeCovid	5182	The set of domains	Manually annotated	True/ false/ partially false, etc.

For instance, the authors of [29] created a dataset containing 349,455 messages from Twitter with pro-Russian hashtags and a pro-Russian stance. These messages were posted by 132,131 different users, of which 250,853 messages (71.78%) were retweets. Additionally, the dataset includes 9,818,566 messages posted by 2,079,198 users, categorized as pro-Ukraine. The majority of these messages (80.93%) were written in English and posted the period between February 2022 and July 2022. The creation of this dataset enabled the authors to develop an approach for detecting bots on Twitter and suggested the presence of a large-scale Russian propaganda campaign on social media, especially at the beginning of the Russian invasion of Ukraine.

In the paper [30] provides a Twitter dataset of the 2022 Russo-Ukrainian war. The dataset contains over 1.6 million tweets shared during the first week of the crisis. Over the past year, a few studies with similar approaches and findings have been published.

3. Methodology

In our study, we directed our attention to the scrutiny of disinformation campaigns related to the ongoing Russia-Ukraine war. The articles were disseminated through established media outlets as integral components of information warfare and propaganda endeavors.

3.1. Data Collection

We curated the RUWA (Russian-Ukraine WAr) dataset [31], which compiles news articles covering key events related to the Russian-Ukraine war.

To ensure a balanced representation of public opinion and journalistic perspectives, we sourced texts from reputable global outlets spanning various regions. These include *BBC*, *Euronews*, and *The Guardian* (European region); *NBC News*, CNN, and *Bloomberg* (USA region); *Ukrinform* and *Censor.net* (Ukraine); and *Russia Today*, *News-front.info* (Russia), as well as Al *Jazeera* and *Reuters*.

To mitigate the risk of generating a topic detection model instead of a misinformation detection model, we proactively identified nine widely acknowledged events in the global press, such as *'The Beginning of the War*', *'Bucha Massacre*', and so on. Subsequently, articles about these specific events were obtained from the sites above.

The selection of articles for each event adhered to predefined criteria, including the publication time interval and keyword lists. The time interval typically spanned from the date of the specific event and extended three to four weeks thereafter. This approach aligns with the common pattern in media, where dedicated coverage of a particular event tends to last no more than two to three weeks.

The keyword list comprised approximately 100 keywords. Due to distinct narratives, terms, and concepts used in Ukraine, Russia, and the Western press to describe the same war events, we identified keywords for each news website separately. Subsequently, we aggregated all these keywords to effectively highlight articles across the sites. Primarily key words encompassed geographical names (e.g., "*Bucha*" or "*Olenivka*"), specific buildings (e.g., "*Kramatorsk train station*" or "*Mariupol theatre*"), organizational entities (e.g., "*Red Cross*", "*Nuclear Power Plant*"), personal names of politicians (e.g., "*Zelenskyi*"), and analogous proper nouns and phrases. Certain keywords possessed the capacity to unequivocally identify specific events; for example, the presence of the keyword "*Moscow ship*" within a defined temporal window accurately attributed an article to the event "*Sinking of the Moskva*."

However, a substantial subset of keywords pertained to general themes associated with the Russian-Ukraine war (e.g., "war crimes", "evacuation", "a special operation"). These terms did not serve as selective criteria for categorizing articles into distinct topics or events. In cases where these general keywords coexisted with specific event-related keywords within an article, we classified the article under the corresponding event. Conversely, articles lacking the conjunction of specific event-related keywords, despite being published during the stipulated timeframe and containing general keywords, remained unclassifiable about our predefined topics or events. Thus, a considerable number of articles, exceeding 14,000, belonged to the overarching theme of the "Russian-Ukraine war." However, these were systematically excluded from further consideration. Table 2 displays the distribution of articles from various websites in the RUWA dataset based on events and their descriptions, which correspond to the particular headlines of leading news agencies related to each event.

Table 2
Articles distribution due nine events in RUWA

Event	Description	Source of the event definition	Number of articles
Azovstal	Russia says Azovstal siege is over, in full control of Mariupol	*Al Jazeera*	1,816
Beginning	NATO officials say Russian attack of Ukraine has begun	*CBS News*	6,490
Bucha	Killing of civilians in Bucha and Kyiv condemned as ‘terrible war crime’	*The Guardian*	1,429
Nuclear Plant	Evacuations from Zaporizhzhia renew concerns for nuclear power plant safety	*CNN*	3,373
Prisoners	‘Absolute evil’: inside the Russian prison camp where dozens of Ukrainians burned to death	*The guardian*	578
Railway	Ukraine missile attack: Dozens killed at Kramatorsk railway station	*Al Jazeera*	1,466
Moskva Sinking	Russia is losing the battle for the Black Sea	Economist	175
Kremenchug Supermarket	Russian missile strike kills 16 in shopping mall, Ukraine says	Reuters	436
Mariupol Theatre	Russia bombs theater where hundreds sought shelter and ‘children’ was written on grounds	CNN	761
Total			16526

At present, the RUWA dataset comprises over 16,500 news articles documenting events of the Russian-Ukraine war from February 2022 to September 2022. Figure 2 illustrates the percentage distribution of articles across selected news websites.

Clearly, Ukrainian websites published the highest number of articles related to the war events that we distinguished. The Ukrainian website Ukrinform produced 6,750 articles, while Censor.net produced approximately 5,100 articles. In contrast, the Russian websites, Russia Today and News-front.info, together produced about 1,100 articles.

Notably, the Reuters agency devoted more attention to the events of the war than any other EU or USA news website, publishing around 2,000 articles covering the nine well-known events of the Russian invasion of Ukraine.

3.2. Data Analysis

As previously mentioned, acquiring information with one hundred percent certainty about events during an ongoing war is practically unattainable. Any narrative or description of an event inherently carries potential bias and reflects the author's perspective. Consequently, the creation of a true/false annotated dataset covering the Russia-Ukraine war poses significant challenges due to the inherent subjectivity and variability in how events are reported and interpreted.

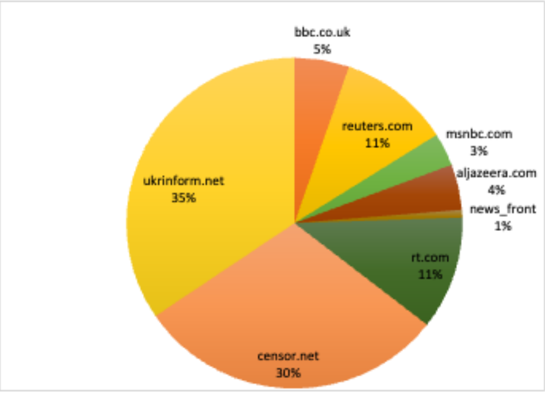

Figure 2: RUWA dataset composition by source

Our approach involved constructing the events-aligned RUWA dataset, followed by the application of unsupervised machine learning methods to address semantic similarity tasks. Fig.3 provides an overview of the architecture of our approach.

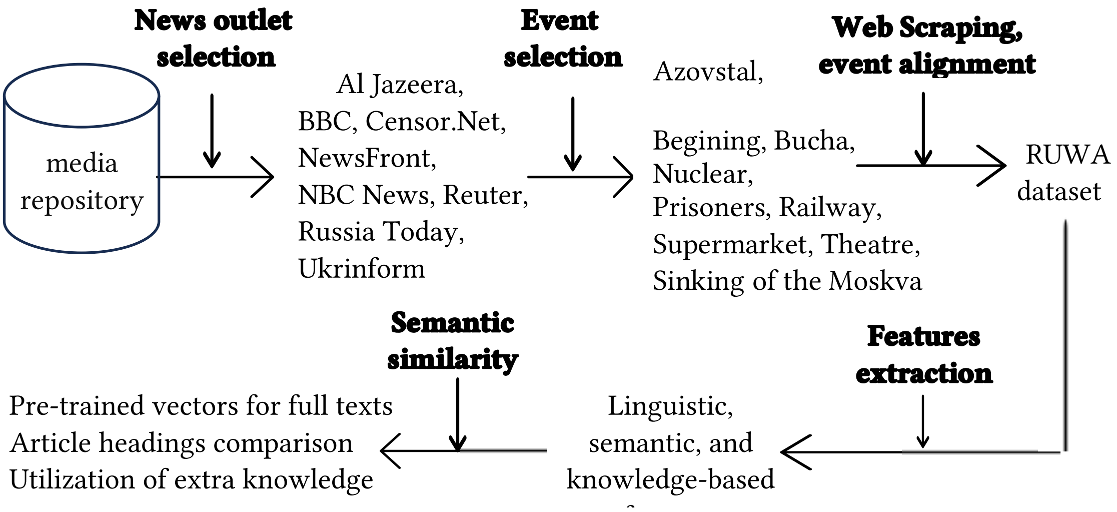
News outlet selection
media repository
Al Jazeera, BBC, Censor.Net, NewsFront, NBC News, Reuter, Russia Today, Ukrinform
Event selection
Azovstal, Begining, Bucha, Nuclear, Prisoners, Railway, Supermarket, Theatre, Sinking of the Moskva
Web Scraping, event alignment
RUWA dataset
Features extraction
Linguistic, semantic, and knowledge-based
Semantic similarity
Pre-trained vectors for full texts
Article headings comparison
Utilization of extra knowledge

Figure 3: Architectural Overview of the Approach.

In our study, where we focused on addressing misinformation detection through semantic similarity, we based our approach on several hypotheses. Primarily, we propose that news shared by media outlets in the two nations actively involved in the conflict is likely to display considerable differences. Information variations may be significant, even leading to conflicting accounts of events, such as the acknowledgment or denial of incidents like residential area bombings or civilian casualties. Consequently, we can expect that the semantic similarity coefficient between texts from Russian and Ukrainian outlets should be minimal.

Furthermore, we hypothesize that the semantic similarity coefficients among articles covering a specific event from various outlets, excluding one or two websites, are generally high. However, when comparing the semantic similarity of these one or two specific websites with all others, we can observe a significant divergence. This discrepancy suggests that these specific websites are likely to be untrustworthy.

That involves three types of experiments for detecting semantic similarity: (1) comparing the full texts of the articles, (2) analyzing article headings, and (3) comparing semantically meaningful sentences within the articles. For the first experiment, we aggregated all articles from the same source that covered one event as a single textual document and then pairwise evaluated the semantic similarity of all outlets' articles. In the second experiment, we calculated the similarity between two sources by analyzing sets of article titles that covered the same event. We achieved this by comparing each title from one source with the corresponding titles of comparable articles from the other source, all related to the particular event.

In the third step, we assessed the semantic similarity of semantic significant sentences from various sources. These sentences contained keywords associated with the event under consideration or verbs representing the actions linked to specific events. Compiling these lists for each event, we relied on the existing list of words associated with the Russian-Ukrainian war from [32] and added verbs extracted from the articles covering each specific event. For example, for the "Moskva sinking" event, the list of related verbs comprises over 120 verbs, while the "Mariupol Theatre" event includes about 60 verbs. This approach allowed us to focus on texts that exclusively provided information about a particular event, excluding phrases with similar meanings typically found in news articles from various web news sites like "witnesses report" or 'it doesn't appear evident' and so on.

For linguistic preprocessing, we employed stemming and stop-word removal. Additionally, we eliminated numerous specific symbols commonly found in web-wrapped texts.

To generate pre-trained vectors, we employed two types of language models (LM), based on Spacy and FastText. The 'en_core-web_lg' language model provided by SpaCy consists of 300-dimensional vectors and encompasses a vocabulary of 685,000 words. This model is trained on diverse datasets, including content from Wikipedia, OSCAR (Open Super-large Crawled Aggregated coRpus) totaling 1342 GB, and News-crawl data comprising 16.9 GB.

While the model was trained on various datasets, including web news, that must have resulted in minimal Out-of-Vocabulary (OOV) words in our articles, we observed that even after preprocessing, our dataset might still contain words lacking proper lexicon or improperly tokenized. To address this, we applied vectors generated by the fastText subworld-based pre-trained vectors from Facebook AI [33] in the subsequent step. In contrast to other LM, FastText LM excels in predicting subwords or character n-grams. This model is particularly adept at handling the challenges posed by scraped news outlets' texts, which may include disruptions from pictures, links, quotes, and other insertions by default. Consequently, the texts may potentially contain misspelled words, numbers, partial words, and single characters. For our study, we utilized the FastText "wiki-news-300d-1M-subword" LM, encompassing 2 million word vectors trained with subword information from the Common Crawl corpus, comprising 600 billion tokens.

4. Results and findings

We conducted a comprehensive evaluation of semantic similarity among pairs of outlets for nine events by analyzing full texts, article headings, and selected sentences from the articles. To avoid building a topic model instead of a misinformation detection model, each of the nine events was individually examined. Our experiments revealed that training vectors on the FastText language model produced better distinguishability results compared to using the 'en_core-web_lg' model provided by SpaCy. As an example, in Table 3, we present the semantic similarity scores for the full texts of the articles related to the *Azovstal topic*, utilizing the *en_subwords_wiki_lg* LM.

Table 3
Semantic similarity scores for the full texts of the articles related to the Azovstal topic, utilizing the en_subwords_wiki_lg LM

	bbs	reuters	nbcnews	aljazeera	censornet	news-front	rt	ukr-inform
bbs	-	99.8%	99.4%	99.1%	99.3%	98.6%	*97.0%*	99.5%
reuters	99.8%	-	99.6%	99.5%	99.4%	99.0%	*95.9%*	99.5%
nbcnews	99.4%	99.6%	-	99.8%	99.2%	99.4%	*95.8%*	99.4%
aljazeera	99.1%	99.5%	99.8%	-	98.9%	99.4%	*95.8%*	99.3%
censornet	99.3%	99.4%	99.2%	98.9%	-	99.2%	*96.1%*	99.8%
news-front	98.6%	99.0%	99.4%	99.4%	99.2%	-	*95.1%*	99.2%
rt	*97.0%*	*95.9%*	*95.8%*	*95.8%*	*96.1%*	*95.1%*	-	*97.0%*
ukrinform	99.5%	99.5%	99.4%	99.3%	99.8%	99.2%	*97.0%*	-

We observed that evaluating the semantic similarity of headlines encounters challenges, particularly when dealing with distributive semantic similarity scores. Even headlines from articles covering the same event and belonging to the same outlet yield relatively low similarity values. Several factors contribute to this outcome. Primarily, the efficacy of comparing article titles is significantly influenced by the number of articles published by each outlet for a specific event. The RUWA dataset, however, is not well-balanced across events. In certain cases, a website may have produced only a few articles related to a particular event, impacting the reliability of the semantic similarity assessment. Furthermore, each headline frequently not only neutrally conveys or describes an event but also mirrors the subjective perspectives and sentiments of certain authors.

As mentioned above, our third experimental direction focuses on processing the targeted, relevant, and topic-specific portions of texts, steering clear of broad or generalized content in articles. To achieve this goal, we specifically chose sentences containing keywords and verbs related to the considered event. This approach allowed us to generate more specific and directly relevant texts that are closely tied to the subject of the event Table 4 demonstrates an example of the calculation of semantic similarity scores for texts obtained by concatenating all sentences containing keywords related to the Mariupol Theatre topic from every outlet. Table 5 illustrates an example of the semantic similarity for texts obtained by concatenating all sentences containing particular verbs due to the Sinking of the warship Moskva topic.

5. Discussion

The conducted experiments collectively validate our hypotheses. Specifically, an analysis of news articles from outlets representing the two countries engaged in the war conflict, including Cersornet, Ukrinform, News-front, and RT, reveals significant disparities in most events. These differences are systematically reflected in the semantic similarity coefficients, underscoring the distinctiveness in the reporting styles and perspectives adopted by these outlets in the context of the ongoing conflict. We may infer with a certain degree of confidence that the semantic similarity coefficient's value correlates with the likelihood of conveying a certain degree of misinformation.

Table 4
Semantic similarity scores for sentences containing keywords related to the Mariupol Theatre

	bbs	reuters	nbcnews	al-jazeera	censor-net	news-front	rt	ukr-inform
bbs	-	99.8%	99.7%	99.4%	99.5%	97.5%	*99.0%*	*99.5%*
reuters	99.8%	-	99.6%	99.5 %	99.5%	97.7%	*99.1%*	99.6%
nbcnews	99.7%	99.6%	-	99.3%	99.6%	97.0 %	*98.7%*	99.5%
aljazeera	99.4%	99.5 %	99.3%	-	99.0%	97.5%	*99.7%*	99.3 %
censornet	99.5%	99.5%	99.6%	99.0%	-	98.0%	*98.9%*	99.7%
news-front	97.5%	97.7%	97.0 %	97.5%	98.0%	-	99.7%	98.2%
rt	*99.0%*	*99.1%*	*98.7%*	*99.7%*	*98.9%*	*97.8 %*	-	*99.2%*
ukrinform	*99.5%*	99.6%	99.5%	99.3 %	99.7%	98.2%	*99.2%*	-

Table 5
Semantic similarity scores for sentences containing particular verbs due to the Sinking of the warship Moskva topic

	theguardian	reuters	aljazeera	censor-net	edition.cnn	ukr-inform	rt
theguardian	-	16.8%	40.9%	18.6%	24.7%	*25.2%*	*17.6%*
reuters	16.8%	-	14.7 %	17.6%	10.1%	*7.0%*	19.4%
aljazeera	40.9%	14.7%	-	15.5%	24.6 %	*21.5%*	16.4%
censornet	18.6%	17.6 %	15.5%	-	7.2%	*9.3%*	8.1 %
edition.cnn	24.7%	10.1%	24.6.0%	7.2%	-	*42.8%*	9.7%
ukrinform	25.2%	7.0%	21.5%	9.3%	42.8%	-	11.5%
rt	*17.6%*	19.4%	16.4 %	8.1%	9.7%	*11.5%*	-

However, the experiments revealed a significant impact of both the number of articles covering an event and the size of the text used to formulate vectors on the semantic similarity score. Specifically, outcomes for events like *Kremenchuk Supermarket* and *Moskve Sinking*, which are covered by only a small amount of news (Tab. 2), often deviate from the general trends observed.

Additionally, we observed that the semantic similarity coefficients consistently fell within the narrow range of 91% to 99%. This tight clustering suggests a high degree of similarity among the articles, implying that they not only revolve around the same topic but also share a similar stylistic approach. All these articles must adopt a journalistic style in presenting news events.

The second group of experiments did not yield significant results. The comparison of articles' headings revealed a notable dependence of semantic similarity coefficients on the number of articles associated with a particular event. Furthermore, titles often incorporate authors' biased opinions and feelings, aligning with the genre-specific nature of the outlet's titles.

The incorporation of additional knowledge related to an event resulted in the optimal handling of web news. Almost all nine events in the final experimental group, which involved additional knowledge regarding actions specified by concrete verbs, provided a clear confirmation of our initial hypothesis. This indicates that the semantic similarity coefficient is notably lower between established outlets from countries engaged in the war on opposing sides.

In our assessment, this finding not only underscores the distinctiveness and divergence in the reporting styles and perspectives of news outlets representing countries with conflicting interests in the ongoing war but also suggests the potential dissemination of misinformation by one country regarding a specific event.

6. Conclusion

In our study, we introduced an innovative dataset focused on the Russian-Ukrainian war. This RUWA dataset involves above 16,500 web news articles from established world outlets, covering nine significant events of the Russian invasion of Ukraine that occurred from February to September 2022. In order to avoid topic modeling and focus on misinformation detection modeling, as well as to improve semantic coherence among articles from various news sources, we aligned the dataset articles based on events. The dataset offers a comprehensive view of diverse journalistic narratives surrounding the Russian-Ukrainian war, providing valuable support for future research.

Furthermore, our research contributes to illustrating how unsupervised machine learning approaches, such as semantic similarity scores, can offer insights into potential misinformation within news coverage of widely reported events across various outlets. We critically examined the pros and cons of multiple methods for assessing the semantic similarity of news articles discussing the same event across diverse reputable news outlets. Additionally, we showed that while relying solely on semantic similarity analysis may not be enough for effective misinformation detection, it offers valuable insights that can be synergistically combined with other techniques to enhance overall accuracy in detection.

Even though this exploration allows deepening our comprehension of the intricacies associated with pinpointing misinformation in the context of the Russia-Ukraine war, it has some limitations, namely:

1. The dataset is centered around nine specific events concerning the Russian-Ukrainian war occurring between February and September 2022. It comprises articles from a limited set of well-known news outlets. However, it is essential to note that this selection, while encompassing major events and reputable news sources, may not capture every relevant occurrence during the specified timeframe. Moreover, the chosen

outlets could introduce a bias, potentially overlooking alternative perspectives and regional nuances.

2. While semantic similarity is explored as a means of detecting misinformation, it has inherent limitations. It may not capture nuanced contextual differences, and the approach might be less effective in identifying subtle misinformation strategies.
3. Outcomes related to semantic similarity for misinformation detection may be specific to the characteristics of the chosen events and news sources. Generalizing the approach to different conflicts or regions requires caution.
4. Additionally, the absence of labeled data for training models limits the ability to assess the performance of the proposed approach against a ground truth.

Acknowledgements

This work was supported by the EU H2020 ICT48 project “Humane AI Net" under contract #952026 and partially supported by the European Commission Next Generation EU programme, PNRR-M4C2-Investimento 1.3, PE00000013-"FAIR" Spoke 8 as well.

References

[1] L. Ross and A. Ward, Naive realism in everyday life: Implications for social conflict and misunderstanding, Values and knowledge, pp. 103-135, 1997.

[2] R. S. Nickerson, Confirmation bias: A ubiquitous phenomenon in many guises, Review of general psychology, vol. 2, no. 2, pp. 175--220, 1998.

[3] B. Nyhan and J. Reifler, When corrections fail: The persistence of political misperceptions, Political Behavior, vol. 32, no. 2, pp. 303--330, 2010.

[4] K. Shu, S. Amy, S. Wang, J. Tang and L. Huan, Fake news detection on social media: A data mining perspective, ACM SIGKDD explorations newsletter, pp. 22-36, 2017.

[5] A. Galassi, F. Ruggeri, A. Barrón-Cedeño, F. Alam, T. Caselli, M. Kutlu, J. M. Struß, F. Antici, M. Hasanain, J. Köhler, K. Korre, F. Leistra, A. Muti, M. Siegel and Türkmen, Overview of the CLEF-2023 CheckThat! Lab: Task 2 on Subjectivity in News Articles, 24th Working Notes of the Conference and Labs of the Evaluation Forum, CLEF-WN, pp. 236--249, 2023.

[6] S. Rastogi and D. Bansal, A review on fake news detection 3T’s: typology, time of detection, taxonomies, International Journal of Information Security, vol. 22, no. 1, p. 177–212, 2023.

[7] M. R. Islam, S. Liu and G. Xu, Deep learning for misinformation detection on online social networks: a survey and new perspectives, Social Network Analysis and Mining, vol. 10, pp. 1-20, 2020.

[8] J. C. S. Reis, A. Correia, F. Murai, A. Veloso and F. Benevenuto, Supervised learning for fake news detection, IEEE Intelligent Systems, vol. 34, no. 2, pp. 76-81, 2019.

[9] A. Jarrahi and L. Safari, Evaluating the effectiveness of publishers’ features in fake news detection on social media, Multimedia Tools and Applications, vol. 82, no. 2, pp. 2913-2939, 2023.

[10] X. Zhou and R. Zafarani, A survey of fake news: Fundamental theories, detection methods, and opportunities, ACM Computing Surveys (CSUR), vol. 53(5), pp. 1-40, 2020.

[11] A. Choudhary and A. Anuja, Linguistic feature based learning model for fake news detection and classification, Expert Systems with Applications, vol. 169, p. 114171, 2021.

[12] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff and B. Stein, A stylometric inquiry into hyperpartisan and fake news, arXiv preprint arXiv:1702.05638, 2017.

[13] K.-H. Huang, K. McKeown, P. Nakov, Y. Choi and H. Ji, Faking fake news for real fake news detection: Propaganda-loaded training data generation, arXiv preprint arXiv:2203.05386, 2022.

[14] N. Maximilian, L. Rosasco and T. Poggio, Holographic embeddings of knowledge graphs, Proceedings of the AAAI conference on artificial intelligence, vol. 30, no. 1, 2018.

[15] T. Mitra and E. Gilbert, Credbank: A large-scale social media corpus with associated credibility annotations, Proceedings of the international AAAI conference on web and social media, vol. 9, no. 1, pp. 258-267, 2015.

[16] Z. Khanam, B. Alwasel, H. Sirafi and R. Mamoon, Fake news detection using machine learning approaches, IOP conference series: materials science and engineering, vol. 1099, no. 1, p. 012040, 2021.

[17] S. Kumar and B. Arora, A review of fake news detection using machine learning techniques, 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 1-8, 2021.

[18] B. Guo, Y. Ding, L. Yao, Y. Liang and Z. Yu, The future of misinformation detection: new perspectives and trends, arXiv preprint arXiv:1909.03654, 2019.

[19] E. Aïmeur, S. Amri and G. Brassard, Fake news, disinformation and misinformation in social media: a review, Social Network Analysis and Mining, vol. 13, no. 1, p. 30, 2023.

[20] V.-I. Ilie, C.-O. Truică, E.-S. Apostol and A. Paschke, Context-Aware Misinformation Detection: A Benchmark of Deep Learning Architectures Using Word Embeddings, IEEE Access, vol. 9, pp. 162122--162146, 2021.

[21] P. Yu, Z. Xia, J. Fei and Y. Lu, A survey on deepfake video detection, Iet Biometrics, vol. 10, no. 6, pp. 607--624, 2021.

[22] W. Y. Wang, "liar, liar pants on fire": A new benchmark dataset for fake news detection, arXiv preprint arXiv:1705.00648, 2017.

[23] K. Shu, D. Mahudeswaran, S. Wang, D. Lee and H. Liu, Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media, Big data, vol. 8, no. 3, pp. 171-188, 2020.

[24] A. Choudhary and A. Arora, Linguistic feature based learning model for fake news detection and classification, Expert Systems with Applications, vol. 169, p. 114171, 2021.

[25] M. Janicka, M. Pszona and A. Wawer, Cross-domain failures of fake news detection, Computación y Sistemas, vol. 23, no. 3, pp. Cross-domain failures of fake news detection, 2019.

[26] M. Potthast , J. Kiesel, K. Reinartz, J. Bevendorff and B. Stein, A stylometric inquiry into hyperpartisan and fake news, arXiv preprint arXiv:1702.05638, 2017.

[27] F. K. A. Salem, R. A. Feel, S. Elbassuoni, M. Jaber and M. Farah, Fa-kes: A fake news dataset around the syrian war, Proceedings of the international AAAI conference on web and social media, vol. 13, pp. 573--582, 2019.

[28] A. Barrón-Cedeño , F. Alam, T. Caselli, G. Da San Martino, T. Elsayed, A. Galassi, F. Haouari, F. Ruggeri, J. M. Struß, R. Nath Nandi, G. S. Cheema and D. Azizov, The clef-2023 checkthat! lab: Checkworthiness, subjectivity, political bias, factuality, and authority, European Conference on Information Retrieval, pp. 506--517, 2023.

[29] F. Alam, A. Barrón-Cedeño, G. S. Cheema, G. K. Shahi, S. Hakimov, M. Hasanain, C. Li, R. Míguez, H. Mubarak, W. Zaghouani and P. Nakov, Overview of the CLEF-2023 CheckThat! lab task 1 on check-worthiness in multimodal and multigenre content, Working Notes of CLEF, 2023.

[30] A. Barron-Cedeno, F. Alam, A. Galassi, G. D. S. Martino, P. Nakov, T. Elsayed, D. Azizov, T. Caselli, G. S. Cheema, F. Haouari, M. Hasanain, M. Kutlu, C. Li, F. Ruggeri and Struß, Overview of the CLEF–2023 CheckThat! Lab on Checkworthiness, Subjectivity, Political Bias, Factuality, and Authority of News Articles and Their Source, International Conference of the Cross-Language Evaluation Forum for European Languages, pp. 251-275, 2023.