Article

Robust Benchmark for Propagandist Text Detection and Mining High-Quality Data

Pir Noman Ahmad [1,*], **Yuanchao Liu** [1], **Gauhar Ali** [2], **Mudasir Ahmad Wani** [2,*] **and Mohammed ElAffendi** [2]

1 School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China
2 EIAS Data Science and Blockchain Lab, College of Computer and Information Sciences,
   Prince Sultan University, Riyadh 11586, Saudi Arabia
* Correspondence: pirnoman@hit.edu.cn (P.N.A.); mwani@psu.edu.sa (M.A.W.)

Abstract: Social media, fake news, and different propaganda strategies have all contributed to an increase in misinformation online during the past ten years. As a result of the scarcity of high-quality data, the present datasets cannot be used to train a deep-learning model, making it impossible to establish an identification. We used a natural language processing approach to the issue in order to create a system that uses deep learning to automatically identify propaganda in news items. To assist the scholarly community in identifying propaganda in text news, this study suggested the propaganda texts (ProText) library. Truthfulness labels are assigned to ProText repositories after being manually and automatically verified with fact-checking methods. Additionally, this study proposed using a fine-tuned Robustly Optimized BERT Pre-training Approach (RoBERTa) and word embedding using multi-label multi-class text classification. Through experimentation and comparative research analysis, we address critical issues and collaborate to discover answers. We achieved an evaluation performance accuracy of 90%, 75%, 68%, and 65% on ProText, PTC, TSHP-17, and Qprop, respectively. The big-data method, particularly with deep-learning models, can assist us in filling out unsatisfactory big data in a novel text classification strategy. We urge collaboration to inspire researchers to acquire, exchange datasets, and develop a standard aimed at organizing, labeling, and fact-checking.

Keywords: misinformation; propaganda; fact-check; ProText; big data; social media

MSC: 68T50

1. Introduction

Social media, fake news, various propaganda techniques, and the inherent bias in the news are produced by people with social prejudices, and there has been an increase in misinformation on the internet over the preceding ten years [1]. Anyone with access to social media or the internet can build a website or blog and use these platforms as a source. Since social media platforms have advanced, anybody may now reach a large audience, as opposed to only major news companies in the past [2]. When a news organization's voice is heard by the public, it is a significant win because it raises the bar for freedom of speech and makes it possible for anybody to be heard. The news is inaccurate, spreads phony stories, and is propagandistic because of freedom of speech [3]. The reader needs to be well-versed in the subject area to spot propaganda. Recently, there has been increased interest in finding texts that are propaganda or extremely biased [4]. Propaganda detection occurs at various levels and phases; for instance, it starts at the document, phrase, and fragment/span levels [5,6]. It broadens the scope and highlights the uncertainty of social scientists' assumptions about conducting their interviews, categorizing their data, and creating a foundation for their study.

The term "propaganda" is frequently used interchangeably with "lies", "distortion", and "deception", as well as "distorted messages", whether purposefully or accidentally. Intentionally planned activities or opinions of people or groups are called propagandistic

actions or opinions [7]. Although social media has enabled the fastest spread in history [8], it is undeniable that it has also allowed the fastest distribution [9,10]. Concerns regarding these technologies' potential use of deep-learning models to disseminate erroneous, inaccurate, or misleading information surfaced as soon as they were developed. Although classifying propaganda tactics and examining news sources is used to distribute, propaganda is outside the purview of this study. However, the news source information is explained in Section 3.3 in detail. We aimed to collect a large quantity of the high-quality big data needed for the deep-learning model's training and testing. Researcher interest in propagandist text data from the news is rarely increased by significant occurrences and news reports (elections, pandemics, sports, and other information) [11], as data crawling is a difficult job [12] and requires more expertise and time [5]. Nevertheless, reliance on social media news without fact-checking risks introducing further biases, which are rarely acknowledged or discussed [13]. In the 2016 US presidential elections, concerns for supporting and opposite leading candidates may have spread the news widely and had a detrimental effect on the outcome [14]. Moreover, the Brexit vote caused unprecedented online misinformation and disinformation to propagate [15].

Additionally, the COVID-19 pandemic led to the publication of the first global infodemic news in 2020 [16]. Propaganda tactics are ineffective if the individuals are aware of the propaganda techniques employed in the news on social media [17]. These contain particular psychological and rhetorical techniques such as slogan, loaded_language, red herrings, etc. [18]. The effectiveness of propaganda attempts decreases as it is considerably harder to uncover and detect them [19]. However, a significant component of the problem is overlooked commonly as the mechanism by which misinformation is spread through propaganda techniques. The research is being conducted on the effects of propaganda news on many subjects of interest. The desire for financial gain by many online news providers has compounded concerns. Voting for your favored candidate in the election—whether you vote or not—is a motivator to defend your position, win the election, or accomplish your objectives. In this study, the Robust Optimized BERT Pre-training Approach (RoBERTa) tokenizer is used to truncate the propaganda datasets at the sentence level after autonomously gathering them from various news sources. The evaluation results are reported with given samples, as shown in Table 1.

Table 1. A collection of high-level propaganda samples, including head segmentation.

S_i^p	Sentence Sample
S_1^p	[B-5] Also the Left killed comedy [E-5].
S_2^p	[B-8] "I hope the American people can see through this sham [E-8]" Graham warned fellow GOPers about voting against the nomination.
S_3^p	President Donald J. Trump for the area, declaring him a [B-10] "bigot" [E-10]
S_4^p	[B-7] The plague is a lie, [E-7] Helene Raveloharisoa told the wire service.
S_5^{Np}	He continued by saying that an FBI would make him feel better at ease.

List of Propaganda Techniques		
0 Virtue_words	7 Exaggeration, minimization	15 Thought-terminating_Clichés
1 Beautiful_people	8 Flag_Waving	16 Whataboutism
2 Smears	9 Loaded_Language	17 Straw_Men
3 Cult_of_personality	10 Name_Calling, Labeling	18 Red_Herring
4 Repetition	11 Bandwagon	19 Obfuscation, intentional vagueness
5 Slogans	12 Reduction_ad-Hitlerum	20 Appeal_to-Authority
6 Doubt	13 Black and White Fallacy	21 Appeal_to-fear-prejudice
	14 Causal_Oversimplification	

This study posited that voters place greater faith in the news sources that back their preferred candidate or individual. Thus, voters sought to assess the exposure and engagement with news stories and discovered that they were widely disseminated and contributed

to validation bias. The problem is greater, as it encourages the idea that we should not believe what we hear on the radio, read online, or see on television. Furthermore, in the S_i^p illustrations, "p" represents propaganda, "np" represents non-propaganda, "s" signifies segmentation, and "i" shows an index. The sample sentences in $S_1^p - S_4^p$ represent the propaganda text, while S_5^p represents non-propaganda. The label is chosen from a list of 22 propaganda techniques with encoding values ranging from 0 to 21. Despite widespread news media use, social media still faces problems with inaccurate information and public-misleading propaganda [20]. Propaganda-containing news stories and texts are gathered, but evaluating and identifying them is a big data challenge. We tried to address the issue by collecting massive quantities of data, which spread news dangerously since it is so quick and simple to acquire. News articles tend to emphasize shocking news content, and because of our negative bias, we tend to focus more on unpleasant events. Similar issues are produced by the printing press, which offers a thorough historical examination.

This paper discusses the compilation process for ProText, a sizable but incomplete propagandist dataset, and the steps required to finish it. We concentrated on gathering information, making article complete texts available, and running experiments using ProText data. Additionally, we fact-checked addresses where propaganda reporting is inadequate and news information is deficient in the ProText. Therefore, collecting propagandist material (ProText) that is still insufficient and uneven to train deep-learning and machine-learning models is suggested and called for. The data are analyzed to define the classification methods to distinguish propagandist from non-propagandist news. Specifically, we describe the few stages of fact-checking structure (claim detection, claim verification, and evidence retrieval), consisting of verdict prediction and justification production. We make the following contribution:

- An annotated (labeled) repository with multiple labels, ProText, was proposed—every instance is based on the sentence level (sent-level) with a precise propaganda technique. These techniques are annotated (labeled) throughout the data automatically and manually.
- We comprehensively explain propaganda news identification by combining a deep-learning model with natural language processing (NLP) technologies. RoBERTa is fine-tuned and formulates specific techniques that classify multi-label propaganda from the news source into spans and techniques.
- The recognition of propaganda news articles is demonstrated using two general algorithms, automatic and manual fact-checking in each data sample.
- The researchers, social media companies, and news media companies must provide propagandist data as part of our deep-learning model. Therefore, we call for collecting propagandist text data to aid researchers in advancing their studies.

The rest of the paper is organized as follows: Section 1 briefly explains the introduction. Section 2 contains related work on approaching propaganda news, covering numerous methods from various viewpoints. Section 3 discusses our methodology combined with innovative classification processes. Section 4 provides the discussion and experiment analysis of the model. Section 5 addresses the conclusion and future perspectives.

2. Related Work

Propaganda news is multifaceted and problematic to comprehend as it is often confused with lies and distorted messages, intentionally or unintentionally [21]. Consequently, the central challenge is discovering propaganda news sources, dissemination, and when their conduct crosses the line into immorality and unethical behavior. Talking about morals and ethics in propaganda and communication is complicated [8]. However, the sheer tonnage is further complicated in the American context where everything from propaganda to outright lies and deception is protected as freedom of speech [22]. Significant technology corporations and social media platforms have proclaimed plans to employ content representatives due to public outcry and recognition that one platform inspires the spread of propaganda, false information, and fake news, as well as well-known insights that these

platforms are responsible for these things [23]. Abdullah et al. [24] developed a hybrid deep-learning model with the cutting-edge RoBERTa pre-trained language model to identify propaganda in 411 news items. Vlad et al. [25] employed neural network models with basic linguistic patterns to identify propaganda at the sentence level in a dataset of 350 news stories. Many NLP procedures are replaced by deep learning, which trains a large amount of data only in fields. Convolutional neural networks (CNNs), transformers, attention models, and in-text classification are tested with feature-oriented approaches [26]. However, the BERT-based model [27] to identify fakes and uncertain articles [28] uses ELMO embedding and the bidirectional long short-term memory (BiLSTM) layer. Moreover, back-translation, synonym substitution, and TF-IDF replacement were all incorporated into the ensemble BERT models [7]. Consequently, name entity recognition (NER) [29] was used in support vector machines (SVMs), deep neural networks (DNN), gradient-boosted trees (GB), and word and character embedding to analyze motivations based on history, healthcare information [30,31], religion, political ideology, and locality.

In addition, propaganda is an issue since many people receive their news from the internet. Propaganda text corpora (PTC) [11]; trusted satire, hoax, and propaganda (TSHP-17) [12]; and QProp [5] are some recent datasets on the topic. However, none focus on sentence-level propaganda, even though they are based on the document level [8,32]. Thus, TSHP was used to complete the standard four-class text classification assignment at the document level and examine news story trends. However, Proppy character n-grams are employed in online news as document-based propaganda, which was gathered and detected to be manually created [5]. Finally, according to Marín et al., approximately 1000 of the 2590 news pieces collected were considered to be the origin of the rumors [33]. Moreover, misinformation and propaganda on social media were detected in the framework, and it can be comprehensive in smart city contexts [34,35].

However, Cheng and Lin [36] composed the Reddit dataset, which was trained using fine-tuned contextualized embedding. NER, gated-wide learning systems, and long short-term memory (LSTM), and their claims for text classification assist as the foundation for the building of the recurrent neural network (RNN) architecture [37,38]. Since it is applied to downstream stages, sentence embedding in NLP is a topic of interest to researchers. Vorakitphan et al. [39] utilized PROTECT argumentation and the semantic structures of propaganda text as input to identify propaganda techniques. Therefore, Bafar et al. [40] collected news datasets from 30 trustworthy and 39 propagandistic news sources with statistics of 205,000 news articles. However, a sentence-level and span-level propaganda dataset was missing from computational propaganda and significant data concerns, so we had to tackle a few tasks [19,41]. This study seeks to fill this gap by gathering high-quality datasets initially annotated manually (remote supervision) rather than using labels from the news source. Additionally, we detected the span in the provided text and grouped each span into 22 propaganda tactics.

Li et al. [42] developed a pre-trained BERT model to separate the problem into span identification and technique classification. In the experiment, Chaudhari et al. [43] employed different supervised machine-learning methods that integrated a range of vectors and word embedding. A feature-engineering step for selecting the retrieval and extraction level is frequently incorporated into feature-based modeling. Furthermore, the pre-trained RoBERTa language model is used for features that require input tokens, and a smaller, perhaps separate model, such as LSTM or RNN, is used for features that require output tokens [44]. Therefore, the transformer employs the self-attention technique without an RNN structure, varying the significance of a subset of the incoming data [45]. However, RoBERTa improves on BERT's language masking approach by deleting the next-sentence pre-training objective and training with numerous learning rates and mini-batches [32]. RoBERTa is trained using extensive data gathered by big institutions and corporations (Washington University academics and Facebook). The RoBERTa-based model is developed with language masking (LM) and byte-level byte pair encoding (BPE). This model is used

to tokenize text documents at the sentence level. It makes it easier to recognize and classify each new phrase based on the type of claim it creates.

Furthermore, Media Bias/Fact Check (https://mediabiasfactcheck.com (accessed on 3 June 2023)) (MBFC) [46], Disinformation detector, Emergent (http://emergent.info (accessed on 3 June 2023)), Politifact (https://politifact.com (accessed on 3 June 2023)), Hoaxy (https://hoaxy.osome.iu.edu (accessed on 3 June 2023)), and Snopes (https://snopes.com (accessed on 3 June 2023)) are sources that verify facts and methods used in online misinformation [47,48]. Facebook collaborates with fact-checking groups to limit and appreciate the impact of propaganda news [49]. Nevertheless, it keeps track of the diffusion of information from independent fact-checking organizations and low-credibility sources. Moreover, fact-checking organizations identified several challenges. These challenges may arise due to, for example, depending on expert teams without a long-term plan or failing to measure the impact of one's work.

Moreover, researchers working alone and in groups have attempted various approaches to overcome this problem. To decrease or eliminate exposure to such material, they are improving people's awareness of potential news and modifying the structures [50,51]. Fact-checking is destructive since acquaintance with news articles or rumors fosters approval rather than disapproval [52]. Therefore, it is practicable on a big scale and, at the very least, saves moderators from filtering through unwanted stuff. This automatic fact evaluation focuses on the article's material, claims, and statements rather than information such as the source or rate of dissemination [53–55]. Furthermore, the ClaimRank technique identifies claims needing to be verified and refers to fact-checking websites that utilize manual or automated techniques to verify claims [56]. Thus, computational linguistics and artificial intelligence (AI) are big data and society systems that seek out news and tidbits of evidence pertinent to a claim to overcome the disinformation problem [57]. The algorithm can assist human verification specialists; however, it is unable to replace them [58]. Therefore, the effects of propaganda and fake news in specific situations are being documented.

This study aimed to develop an approach that allows fact-checkers to identify statements that may need a closer inspection by utilizing claim-type and RoBERTa-based classification. Thus, due to the scarcity of specific training datasets, many modern detectors are unsupervised or semi-supervised, which helps them to overwhelm the limitations of the supervised classifier. The next part, which uses datasets we found through a study of the past work, illustrates the challenges of creating such a dataset. However, it could seem that putting one together is straightforward. Therefore, this study's initial experiment lacks enough information to build a trustworthy system for detecting propaganda. Furthermore, the evaluation performance showed that training deep-learning models on real-world text news greatly improves the output compared to models trained on synthetic and open-domain information. However, misinformation, fake news, and propagandist stories are still spreading on social media and the online web, such as information related to COVID-19 [59,60]. While most fact-checking organizations use human validation of information, the ever-increasing amount of new information on the internet makes manual verification challenging, time-consuming, and costly [12]. Therefore, the deep-learning models trained on manually created data or claims are unlikely to be able to validate claims found on the web.

In addition, verifying claims in domains such as the social domain and healthcare, where expertise is needed, makes the task more challenging. Adapting fact-checking models trained on open-domain claims to health-related claims might not work well. Thus, to address the abovementioned issues, we recommend a novel dataset (ProText) collected from a trusted news source, compiled, and put through an automated or manual fact-checking model to ensure its validity. Compared to the existing efforts, we use naturally-occurring claims from the web and scientific articles for verification. This study benchmark posited that ProText provides a realistic and challenging dataset for future efforts on evidence-based fact-checking instances of propaganda-related news.

3. Materials and Methods

3.1. ProText Labeling and Matching Claims

The classifier reviews the text sentences anticipated to the novel label and is associated with earlier fact-checking statements. Merging the data poses proper questions, and social scientists must adopt professional social science standards while expanding their technical skills. The big data model predictions were frequently considered suitable aims on their own [61,62]. The sentences are extracted from the web server, and a deep learning system keeps track of the assertions made. Therefore, using manual and automatic fact-checking, our strategy used fact-checking to identify and match propaganda news claims [63]. As input documents, the architectural pipeline transforms the news from the source file. To evaluate the pipeline's relevance and reliability, we illustrate the manual and automated annotation process using specialized technology, as shown in Figure 1.
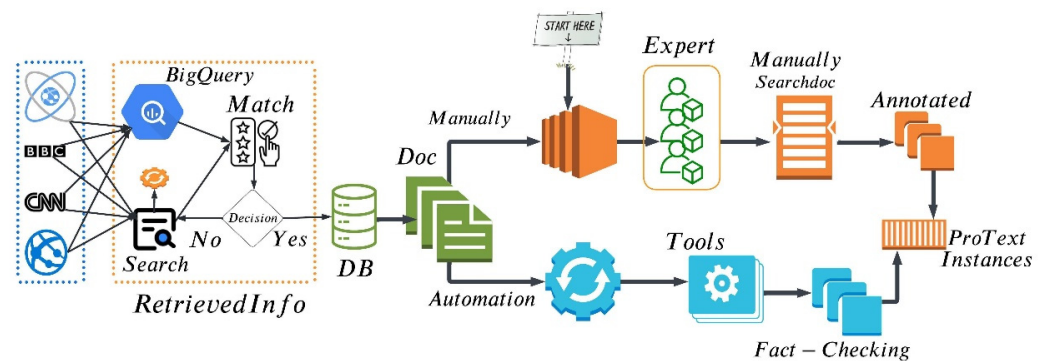

Figure 1. Joint pipeline of manual and automatic document collections from web server annotation claims/matches in architecture.

3.1.1. Manual Fact-Checking

Collecting data and their sources is a key component of fact extraction. The benefit of having human specialists supervise propaganda is that it ensures that assertions are thoroughly examined and draws attention to the facts. However, there are several possible drawbacks, such as the possibility of moderators prejudices spreading, the mental strain on those who carry out the checks, a lack of information, and low-level abilities [64]. A framework for trustworthiness indicators provides signs to help both automatic and manual systems determine the dependability of a piece of news content [65]. Thus, to facilitate human fact-checkers, we study what fact-checkers want and what research has been performed that can support propaganda in news articles. This is significant because manual fact-checking is time-consuming, going through numerous manual steps. The following typical sequence is described for fact-checking steps: (i) extracting declarations that are to be fact-checked, (ii) constructing relevant questions, (iii) gaining the pieces of evidence from relevant news sources, and (iv) making decisions using that evidence. Manual checking needs more labor, skill, and knowledge to teach the public about detecting and reducing the extent of misleading information, fake news, and propaganda. This experience is greatly needed to accomplish and train professionals who can spot, analyze, and stop the spread of propaganda news and misinformation using traditional methods.

Six professional annotators spent approximately 1200 h on this investigation, annotating and fact-checking the data [21]. These annotators are highly skillful in the NLP domain. This duration includes the average search of news mutating each propaganda instance. Finally, the average length of propaganda instances is 3–80 words (slogans have shorter instances). We encouraged the annotators to select relevant sentences that do not contain enough information to decide. In their manual technique, content moderators split the procedure into fact-extraction and fact-checking, as shown in Figure 2a.
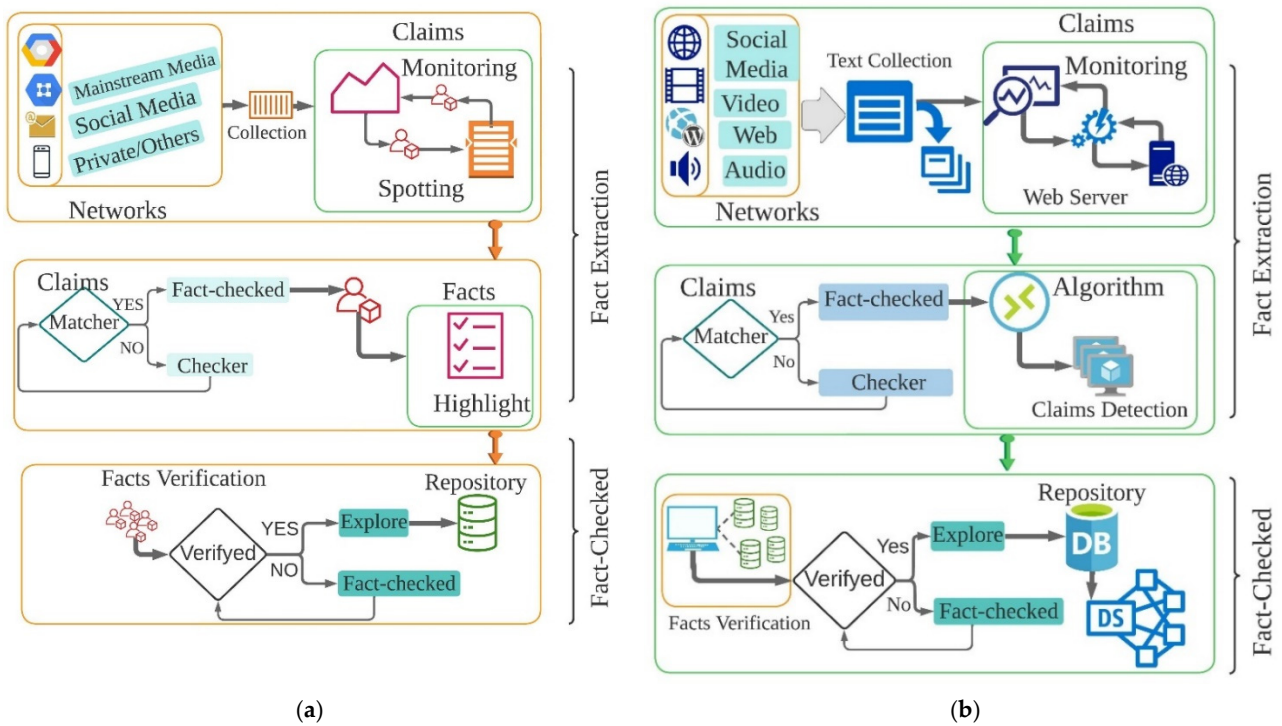
Figure 2. Prediction and label matching using (**a**) a manual fact-checking and extraction technique and (**b**) automated fact-checking and claims extraction for likelihood and similar architecture.

Experts pay close attention to each claim (claim verification) in the data taken from social media websites and strongly emphasize the fact-checking stored in the repository [66]. The research community can use this free resource to train deep-learning models. Any portion of the material supported by statements made on various topics, including quantifications, causes and effects, and forecasts, are considered claims.

3.1.2. Automatic Fact-Checking

The automatic fact-checking process incorporates machine-learning methods to develop a fact-checking model, resulting in the loss of textual information [67]. An automatic fact-checking system can use graphs, references, and context [68]. However, recent research reveals considerable training-set constraints while not explicitly specifying a context-aware model. To categorize claims in a multi-class classification task, the LSTM networks and pertinent text fragments from outside sources are employed [69]. According to the article, retrieving texts at the sentence and document levels is a complex procedure that has improved with more research. ClaimBuster is an alternative approach that keeps track of talks and finds claim items in repositories [70]. Contrarily, modern fact-checking research efforts on consistent data sources to gather proof assess the veracity of a claim and offer new tools for fact-checkers. In order to detect claims using deep learning techniques, the text gathered from observers is referred to as the claim-matching model, as shown in Figure 2b.

Another method of automatic verification involves analyzing the language used in the narrative itself, that is, looking for clues that point to exaggerated claims, excessively emotional language, or a style unusual in major news sources. The initial step is to import the 1200 news uniform resource locators (URL) from the URL repository into Tanbih (https://www.tanbih.org/propagandasubmit (accessed on 3 June 2023)) in sequence form. This model is significant because it is independent from the pre-processing stage.

In addition, after fetching the data, the next step is extracting text, which consists of a span or fragment of propaganda. Thus, the authors gathered propaganda at the sentence-level from the authentic web with related terms and labeled it with propaganda techniques to develop a collection of propaganda. These collections are stored in an excel file for

further calculation and comparison with sentiment analysis of propaganda techniques to create new domain-specific characteristics. The term matching with sentimental analysis is an excel file containing multiple features. The excel file is fed into the likeness score calculation and part of the algorithmic fact-checking method is fact-checked to generate fact-based features. Furthermore, all the basic features and excel are trained on six machine-learning classifier models, BiLSTM [71] and Distilroberta. DistilBERT [72] is the baseline, and XLNET [73], Albert [74], and RoBERTa [54] are proposed models, with performance metrics of confusion matrix, accuracy, precision, recall, and F1-score.

3.2. Propaganda Text Detection

We prearranged a text illustration and identified propaganda spans in the specified sample text. The propaganda technique is also necessary because text classification relies mainly on each sample text used in the linguistic characteristics of longer texts [7,27,29]. The sentence-level propaganda method, followed by information extraction, shows how RoBERTa analysis, design, and theme characteristics are integrated. However, the pruning feature is concentrated in BiLSTM model's last hidden layer before classification using RoBERTa to reach the SOTA results. As shown in Figure 3, the previous component gathers the likelihood propaganda label for each phrase to produce an M-count of prediction classifiers for each word token.
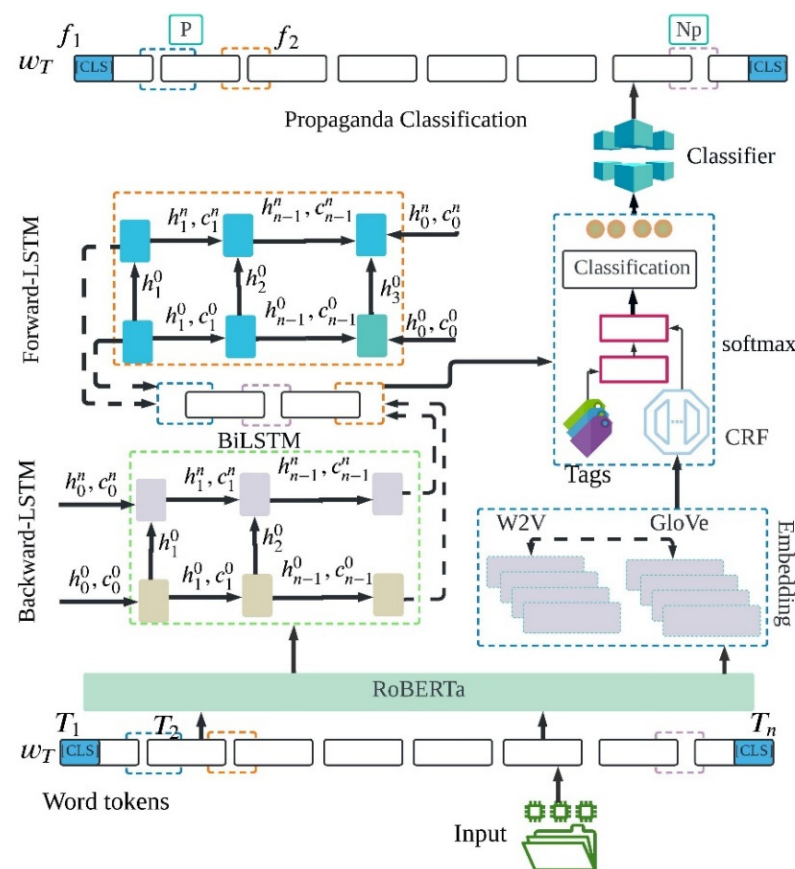

Figure 3. RoBERTa, embedding, BiLSTM, and propaganda text detection pipeline for the prediction of propaganda and non-propaganda.

Phenomena concerning the manipulation and dissemination of opinions, such as hate speech [75], fake news [62], chatbots, trolling, and social bots, are blended and become mixed up in the speech; fallacies and lexicon confusion are ubiquitous [76]. Furthermore, the list of hatred, such as #lockdown (the particular event occurred in lockdown 2020–2022), #buildthewall (US 2016 presidential election), and communal hashtags comprises those terms that attempt to spread hate and encourage violence among the people based on

their religion through logical fallacies (https://research.com/research/logical-fallacies-examples (accessed on 3 June 2023)), such as loaded language, slogans, etc. [46], for example, the term islamophobia (https://www.newsclick.in/Hashtags-Hate-Flood-Social-Media-Islamophobia-Grows (accessed on 3 June 2023)), coronavirus, or tweets that relate coronavirus with Islam. We include some types of hate speech under the misinformation category, as people are often targeted because of their affiliations or personal history. While the information can sometimes be based on reality (for example, targeting someone based on their religion using logical fallacies), it is being used intentionally to cause destruction. The continuous use of fake news to polarize public opinion, promote extremism, and spread hate speech has required activists to reassess the role of social media in activism and revise their communication strategies to tackle the challenges caused by propagandist news. Bersih (https://en.wikipedia.org/wiki/Bersih_2.0_rally (accessed on 3 June 2023)) activists are motivated to manage political conversations flowing within their communicative ecology with the help of social media distribution networks where they could develop counterclaims and critical narratives against predominant mainstream claims, fake news, misinformation, political propaganda, and hate speech [77].

Furthermore, we built several propaganda detection models utilizing DNN, SVM, and GB trees [30,31]. To be more precise, the likelihood models we created are based on semantic characteristics of news stories, and these pieces are graded according to the content. We periodically inserted a small number of hateful user votes and user profiles into the joint endorsement system to examine the dissemination of false and propaganda news [78,79]. Meanwhile, training time is also reduced by a small number. They adopted several guises or engaged users of social media and internet platforms (Google, Twitter, and Facebook).

3.3. Fact-Checking and Media Bias Topics

ProText utilizes both automatic and manual fact-checking techniques to validate the accuracy of the texts in the dataset tagging system that is now available. This dataset compiles the most widely debated and disseminated news subjects, including the 2016 US presidential election [14], COVID-19, Brexit [15], etc. The biggest issue is using previous data to train or test a fake news classifier that automatically links retrieved certain news from Snopes sites. Several web links are provided to lead the pages claim in perspective. This study finds that some phrases in an automatically compiled dataset are untrustworthy and are used as the only source of support for a claim. This section uses topic modeling to analyze the data from fact-checking websites (Hoaxy, MBFC, Snopes, and Politifact) to identify the news stories. Modeling subjects related to disinformation and fake news is crucial since skewed training datasets result in classifiers that cannot generalize to other topic distributions.

Moreover, the number of topics is altered to represent a distinct kind of news that may be visually investigated. As a result of their investigation, they discovered that earlier datasets were less likely to contain topic distributions related to sports, travel, tourism, economics, technology, and the environment. Therefore, they concluded that biased news websites are mainstream media's primary information source, independently determining its agenda. This study investigates the claim that pro-propaganda news sources and fact-checking websites may affect how the news media covers stories and the information they verify. These websites addressed most topics/themes and news sources, as shown in Figure 4.

This study evaluates the disinformation's veracity, which calls for a definition of propaganda that considers the integrity of its claims. Each text news sample was chosen randomly from a list of 60 reputable news sources that covered a range of subjects (sports, politics, education, etc.) and interests for each technique. In our data collection, we selected 20 news articles from each source. These challenges affect the topic distributions of the propaganda news datasets produced. These websites are freely available, i.e., Pakistan Times, ABC news, etc. However, a few news websites need a membership to access news

articles (i.e., New York Post), making it challenging to gather data. Furthermore, data from all news sources were utilized to annotate 22 propaganda techniques. It is critical to understand that each data source used a distinct approach.
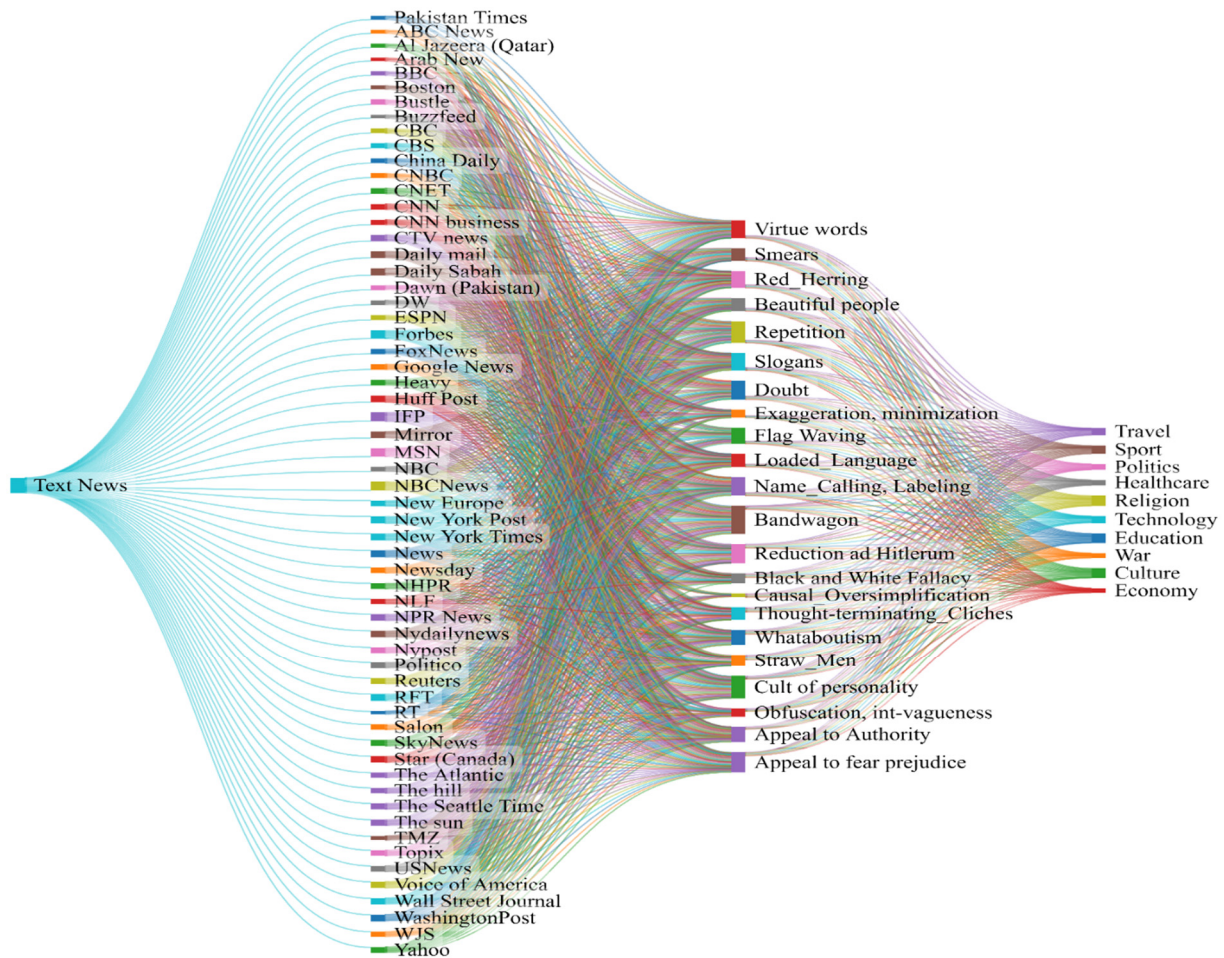

Figure 4. The combination of all relevant propaganda-related themes: (1) text news, (2) news sources, (3) propaganda techniques, (4) topics/themes of interest.

This study posited further data-gathering efforts, identifying data imbalances and gaps. Consequently, the ProText dataset increased in size and extended the range of themes. Although the three-level databases contain various news articles, sports, travel, tourism, the economy, technology, and the environment are underrepresented. As it frequently occurs in rumor news, it does not always indicate falsity. Therefore, it is still interesting. The models are trained on changes in propaganda reach and how themes influence the lexicon and storytelling methods employed in news reports. The dataset is utilized in significant text categorization research not covered in this article. When the train and test data originate from the same news sources, using data that is not evenly distributed results in high-accuracy classification. The high accuracy is deceptive since it requires a classifier to identify high-level traits that could be misinterpreted as fraud signals, even when the news articles do not mention fraud.

4. Results and Discussion

This section briefly overviews datasets, preprocessing, parameter configuration, and evaluation performance. The experiment is based on Nvidia GTX 1070Ti GPUs to train models, using the robust library as a PyTorch to represent neural networks. The data were preprocessed, and the coding environment was set up for deep-learning models.

4.1. Dataset Analysis

Propaganda aims to influence people's opinions and put forward a specific plan. It is difficult to determine how to monitor and where to gather propaganda data. Fact-checking starts by gathering information from dependable news sources and social media websites [33]. The information is gathered from social media and reputable internet news sources. The people's inherent barrier to critical thinking is decreased when misinformation is presented as news since it comes from various sources. Before we built a text system that identified propaganda with linguistic indications, the news stories were individually examined and categorized according to their level of authenticity. The statistics of all instances with technique-encoded data and the 22 mentioned propaganda tactics are provided in Figure 5.
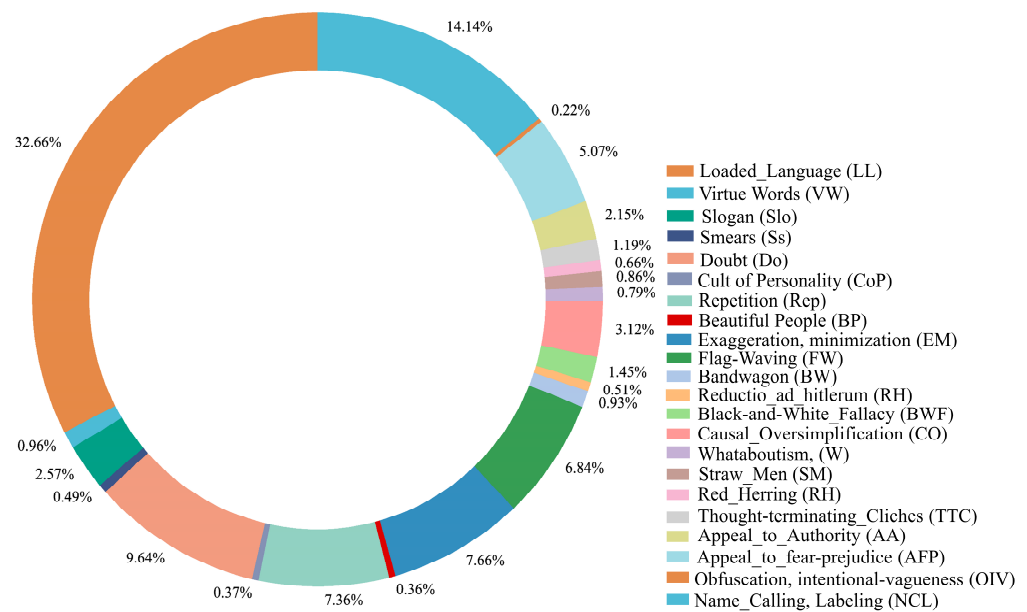

Figure 5. Propaganda techniques instances statistics with segment encoder.

The gathering of news is encoded into a segment, S_n^p, where "p" shows propaganda techniques, and "n" is an instance. The encoded propaganda instance samples a profusion of "Loded_Language" and "Name_Calling, Labelling"; however, the "Cult of Personality", "Smears", and "Beautiful people" labels are inadequate. The research on this topic recently revealed two key elements attempting to change people's perceptions of corpus annotation [41]. In them, individually targeted actions interact unexpectedly with structural trends through networked systems and sharing mechanisms [80]. ProText covers the big lie, voting fraud, sports-related problems, COVID-19 conspiracy theories, and social media [81]. Sentiment analysis in text news as propagandistic news in social media (positive or negative) is beyond the scope of this study [82]. Professional journalists may produce articles about propaganda, disinformation, and misinformation for various causes.

4.2. Statistics of Available Datasets

Annotated corpora are necessary; meanwhile, the current techniques for identifying propaganda in the text are supervised. The annotated corpus arguments of 1300 data pieces use Red-Herring, five-ad-hominem, and Appeal_to-Authority explanations [21]. TSHP-17 contains around 22,000 news articles, a balanced corpus with document-level annotation, and a compiled dataset of a reasonably large number of propagandist news stories [12]. Barron-Cede [5] proposed Qprop, an unbalanced dataset with statistics of 51,000 news articles at the document level, as the classifications made are more probable to model propaganda passages. QProp overcame the TSHP-17 limited sources number studied in each technique. Martino et al. [11] proposed a PTC corpus, which offered

536 news articles covering assets of the QProp and TSHP-17 datasets. The PTC corpus, annotated manually by experts in news articles with 14 techniques, including content and context-related propaganda techniques, i.e., exaggeration, minimization, fallacies reduction ad Hitlerum, and on the news websites, as shown in Table 2.

Table 2. Text data gathering of multi-label, multi-class datasets for each propaganda source.

Datasets	Sources	News Articles	Label	Propaganda Text	Level
PTC	49	536	14	7385	Doc
TSHP-17	11	22,580	4	5330	Doc
QProp	104	51,000	2	5737	Doc
ProText	60	1200	22	11,532	Span

SemEval (https://alt.qcri.org/semeval2020/ (accessed on 3 June 2023), https://github.com/strohne/Facepager (accessed on 3 June 2023)) and nlp4if were used to develop a standard for the collection of data approach related to propaganda news. Finally, ProText gathered accurate data on several topics, PTC, QProp, and TSHP, from trustworthy news sites and associated them with fake types to modify the material. ProText is completely well adjusted in span/fragment but imbalanced as news sources are chosen randomly from news websites regarding propaganda techniques. ProText randomly selected 1200 propagandist news articles as the document level for annotation is gathered manually, and 11,536 instances correlating to 5 to 10 topics/themes were retrieved. Due to the complexity of labeling the claim–evidence pairs and following previous efforts [83], we only evaluated the agreement between annotators on label assignment. We obtained a Cohen's Kappa of k = 0.68 [84], which indicates that the inter-annotator reliability is satisfactory, as the obtained k of 0.68 is above the commonly applied criteria of 0.64; it is also comparable to the 0.66 Cohen's Kappa reported in [83]. Using automated methods, the news articles are extracted and pitched using URLs from websites or libraries written in Python, such as those by highly qualified and trained experts.

4.3. Hyper-Parameters

Pre-trained sentence embedding or word embedding encode the model input into the embedding vector. The experiments are performed with RoBERTa [54] models, XLNET [73], Albert [74], DistilBERT [72], and BilSTM [71] using word2vec (w2v) [85], Sigmoid [86], GLU [87], and global vector (GloVe) [88] embedding. A sentence encoder builds a 768-dimensional hidden layer for individual phrase representation. These models are configured with two cases of hyperparameters (Case1 and Case2). TSHP-17, PTC, ProText, and Qprop all have the same hyperparameter tuning for Case2; however, Case1 has a different tuning. The best hyperparameter tuning range depends on the specific machine learning model and the dataset that is being used [89]. However, there are some general guidelines, as shown in Table 3.

Table 3. ProText, PTC, TSHP-17, and Qprop hyperparameter for two Cases.

Parameters	TSHP-17	Qprop	PTC	ProText	All Data
	Case1				**Case2**
Weight decay	0.1	0.1	0.1	0.1	0.1
n_clusters	0	2	2	2	0
Output Layer	GLU	Sigmoid	ReLU	PReLU	Softmax
Batch sizes	8	64	32	64	16
Epochs	25	18	20	12	15
kernel	1	2	1	1	2
Pre-Train model	24	24	12	12	12
Optimizer	GD	GD	Adam	RMSprop	AdamW
Embedding	TF/TF-IDF	TF-IDF	W-V	TF-IDF	GloVec
Hidden-Layers	768	768	768	768	768
Data-Train/test	0.7/0.3	0.9/0.1	0.6/0.4	0.85/0.15	0.8/0.2
Learning Rates	2×10^{-3}	2×10^{-3}	3×10^{-3}	3×10^{-3}	5×10^{-4}
Dropout	0.4	0.3	0.4	0.4	0.3

ProText, PTC, TSHP, and Qprop are used in different hyperparameter tunings for Case1, while it is the same in Case2 for all datasets. In Case1, other parameter configurations help to identify the error and misleading accuracy. The evaluation performance score of Case1 is compared to the performance of Case2.

4.4. Evaluation Performance Metrics

The proposed model is compared to numerous evaluation performance (confusion matrix, accuracy, precision, recall, and F1-score) criteria to calculate effectiveness. The F1-score is classified as a significant evaluation metric, whereas recall (R) and precision (P) are classified as minor, as shown in Equation (1).

$$\mathrm{F}_1 = 2 * \left(\frac{\mathrm{P} * \mathrm{R}}{\mathrm{P} + \mathrm{R}} \right) \tag{1}$$

where FP is false positive, FN is false negative, and TP is true positive, while R = recall $= \frac{\mathrm{TP}}{\mathrm{TP+FN}}$ and P = precision $= \frac{\mathrm{TP}}{\mathrm{TP+FP}}$ are metrics for classification, and accuracy = TP + TN/ TP + TN + FP + FN.

4.5. Results

We tested various deep-learning and pre-trained models to solve the propaganda detection issue. Consistent with the traditional wisdom, RoBERTa outperforms our model but needs additional tuning and training time. Feature selection must be executed carefully if this abundance is confirmed since simple models allow quick estimation of the state-of-the-art accuracy achieved on the ProText dataset. To detect propaganda, we tested our proposed model using ProText test data for Case1 and 2, as shown in Figure 6.

The results of all models are tested, including deep-learning models, in Cases1 and 2. In propaganda classification problems, the deep-learning model is trained on the training data, and efficiency is calculated by accuracy on the test data, which is the portion of appropriately predicted techniques. The objective of this experiment was to identify propaganda in text news and classify its techniques. The word error rate of the baseline model is very high, in this situation, even though at 0.2719, with a recall of 0.2697 and an F1 of 0.2708, and results of 0.2389, 0.2313, and 0.2242 on Case2, respectively, there is clearly only a small drop in the F1-score in Case1. For all the datasets, the observations were related. When clean training data were used, there was only a small drop in F1-score in Albert-BiLSTM, baselines, and BiLSTM at 20% in Case1; the drop became prominent in Case2.

We achieved text classification in our proposed model for test score in Case1 at the first level of hierarchy with a precision, recall, and F1 of 0.7197, 0.7039, and 0.7117, respectively, and 0.7577, 0.7343, and 0.7457 in Case2, respectively. As we have stated above, the results in these datasets are more informative, but it is best to perform experiments in the two settings described above to study such effects in benchmark datasets. These results recommend that a sentence's length and complexity effectively differentiate propagandistic sentences from non-propagandistic ones but not as effectively as LIWC, TF-IDF, and emotion do. Based on the above best parameters tuning (Case2), our classification systems finally obtained an F1-score of 0.7457, and the training procedures reserved approximately 24.3 using Nvidia GTX 1070Ti GPUs (training time depends on the number of iterations and batch size). Because the ProText in Case2 performed better in assessment, we used it for each propaganda technique, as shown in Figure 7.

ProText training and testing sets, in terms of F1 scores, achieved more, but several techniques on data did not achieve well. The loaded_language achieved F1-scores of 0.7698 and 0.6756 on training and test sets, respectively, on "Loaded_Language". In addition, Name_Calling performed at 0.6438 on the test set and 0.7130 on the training set. Similarly, flag-waving showed greater performance, with scores of 0.6832 and 0.5571 on training and test sets, respectively. Exaggeration and minimization obtained F1-scores of 0.6645 and 0.4349 on the given sets (training and test). While doing well on the training set—achieving scores

of 0.1307, 0.0825, 0.0433, and 0.1087, respectively—cult of personality, Beautiful people, repetition, and reduction ad Hitlerum were not executed well on the test set. The results of these extra data processing techniques were then described, and all the propaganda strategies were then further examined.
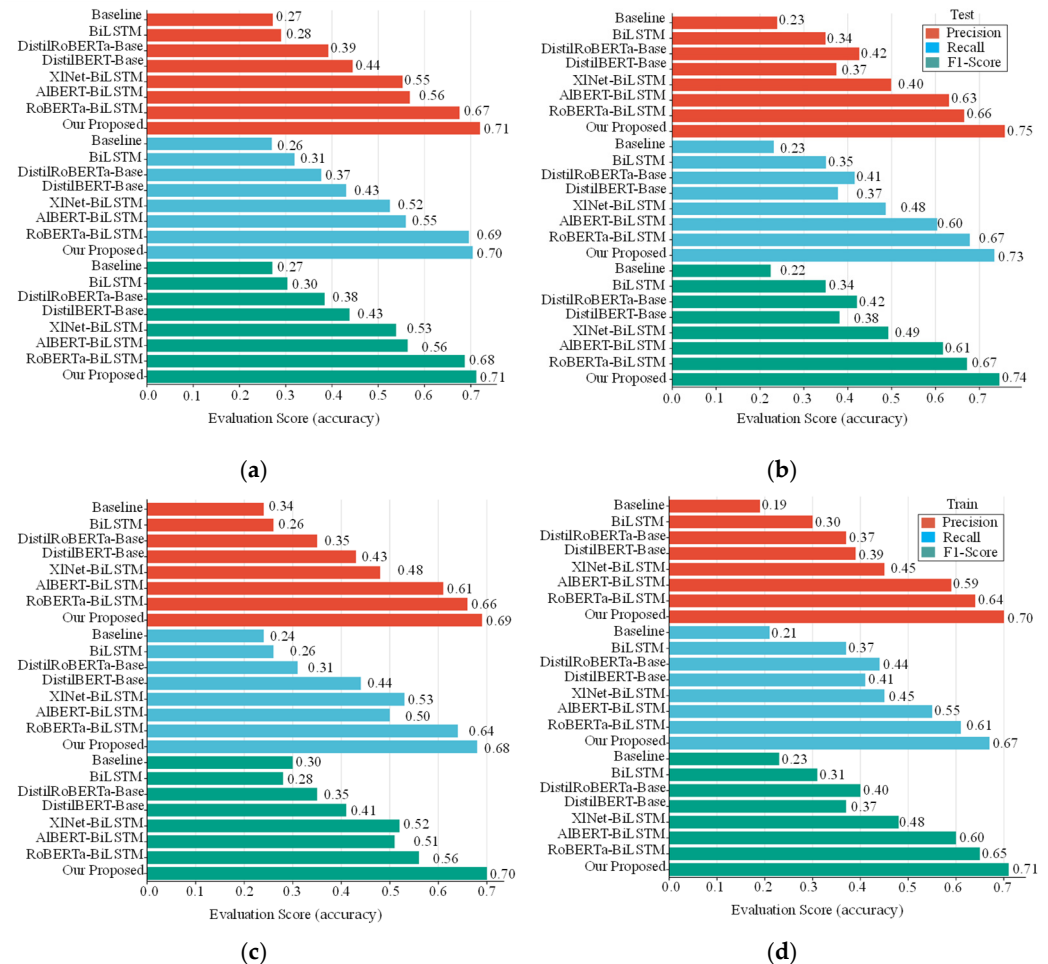

Figure 6. ProText train/test with Case1–2 hyperparameter tuning evaluation metric accuracy on proposed model. (**a**) Test score on Case1, (**b**), Test score on Case2, (**c**) Train score on Case1, and (**d**) Train score on Case1.

The optimum machine learning model, NLP strategy, and data processing approach were finally determined using Kruskal–Wallis and ANOVA tests. Given the disparity in the number of data points, we decided to use two Kruskal–Wallis trains and tests to compare the machine-learning models and to determine the maximum mean level (data processing and NLP technique) ensemble and RoBERTa, as shown in Figure 8.

According to an ANOVA ranking of the methods based on the mean in Figure 8a, the training technique has a higher rank than the testing approach. The findings of our Kruskal–Wallis train and test are shown in Figure 8b, and they indicate that there may be a significant difference between the techniques employed to test inside vs. across at the $p = 0.15$ level of significance with $p = 0.025$. The only way to compare honestly is to create a parallel corpus with clean dataset classification results (accuracy). After training, the models are evaluated on the original subset of the 100 instance sample and the training set. Figure 9 compares the evaluation performance accuracy of Baseline, BiLSTM, Distilroberta-base, Distlibert-base, Xlnet-BiLSTM, Albert-BiLSTM, RoBERTa-BiLSTM, and our proposed models to other data.
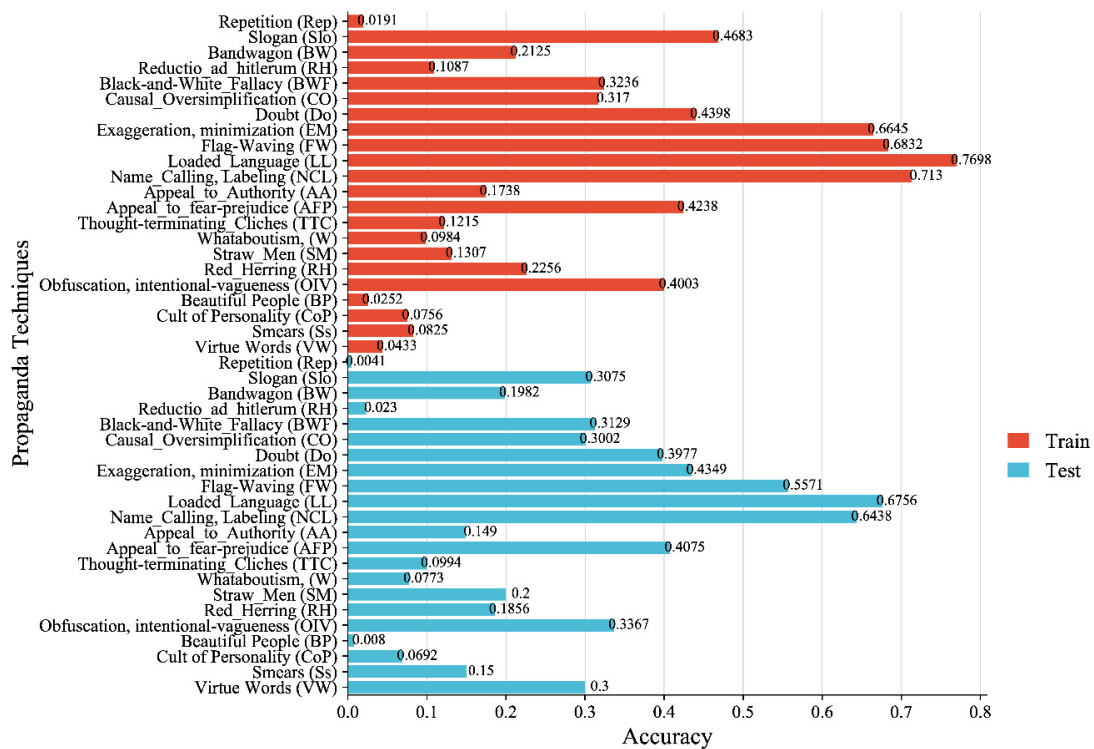
Figure 7. ProText with the Case2 hyperparameter setup for the evaluation performance of propaganda strategies.
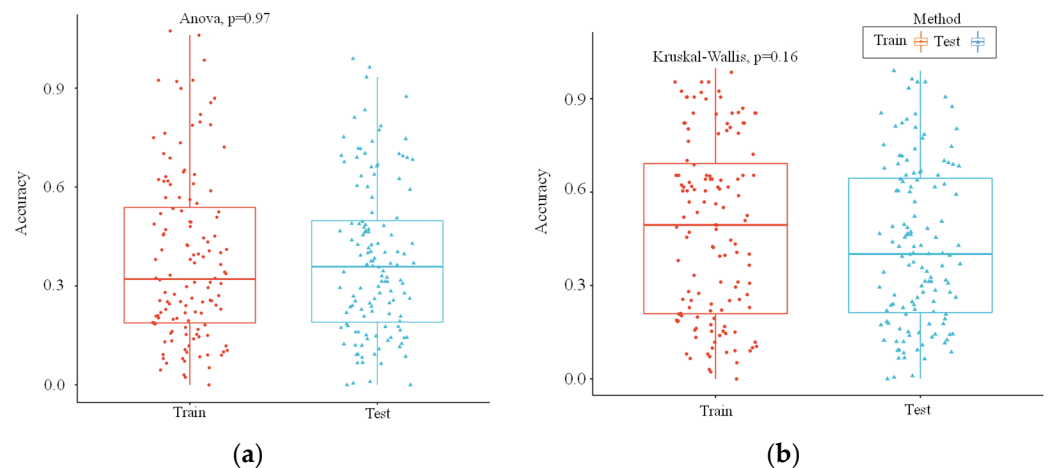

(**a**) (**b**)

Figure 8. The most successful machine learning model, comparison (**a**) ANOVA, and (**b**) Kruskal–Wallis trains and tests NLP methods.

As with the Case2 hyperparameter tuning, the classifiers detected all variants significantly better than transformers with Case1 parameter configuration. In addition, PTC, TSHP-17, and Qprop performed well against the baseline performance but less so with ProText. As a result, terms and tokens with unique values did not have a significant impact. Regarding accuracy, ProText, PTC, TSHP-17, and Qprop achieved 90%, 75%, 68%, and 65%, respectively. Our model was the best at detecting propaganda techniques with F1 = 0.74 in Case2 and 0.71 in Case1, based on the ProText test set. The average evaluation metric scores for Case1 and Case2 on the training and testing datasets were compared. As shown in Figure 10, the proposed model is compared to the propaganda technique detection evaluation of the preceding model.
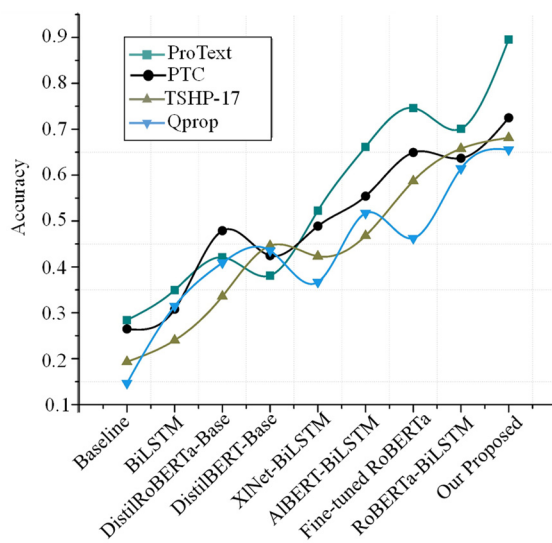
Figure 9. Comparison of various models' accuracy using the proposed model on Case2 in multiple datasets.
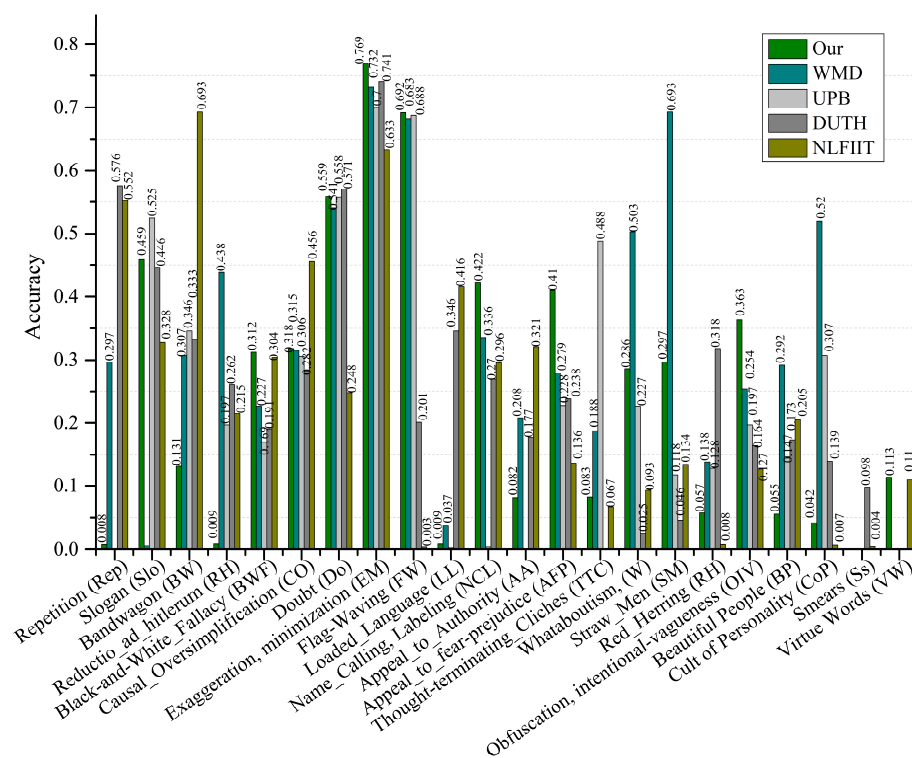

Figure 10. The evaluation of propaganda strategies in F1 compared to the prior model.

WMD [7] employed back-translation, synonym replacement, and TF-IDF replacement based on the TF-IDF score using joint BERT-based models together with SVMs, LSTM, random forest, GB, and embeddings (word/character). In DUTH [29], the convinced words were mapped into classes using NER with concentration on three entities (person, names, and gazetteers) comprising names, slogans, religions, ideologies, politics, and deviations of names of countries. The standard classifications were replaced by the category name in the input before passing the input to RoBERTa. UPB [27] used models based on BERT–they used masked language modeling to domain adapt with 9M bogus and suspicious news articles, whereas NLFIIT [28] used ELMO embeddings with BiLSTM. RoBERTa and the token segmentation with CRF were adjusted, and the outcome was that our model's

assessment performance F1 score increased—the confusion matrix of the propaganda techniques is shown in Figure 11.
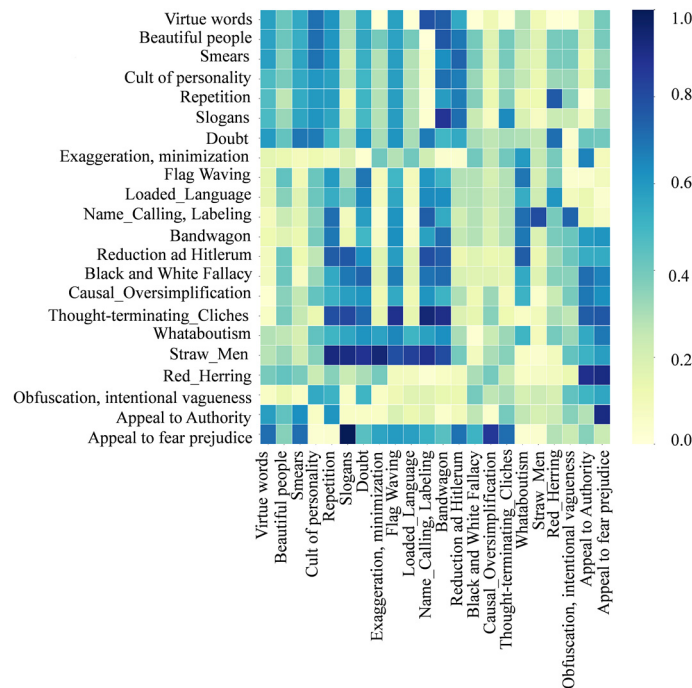

Figure 11. Comparison of the approach and the method for normalizing association strength using a confusion matrix.

The predicted severity levels of each technique are depicted in the *x*-axis, while the *y*-axis shows proper severity levels based on the test set samples. Blue represents downregulated, orange represents upregulated, and light dark represents no model changes. The analyses of differentially expressed models showed that 215 significant propaganda instances in differentially expressed models are classified in the RoBERTa group associated with the BiLSTM. Of these propaganda instances, 143 were significantly upregulated, and 72 were downregulated. A similar analysis is shown for baseline1, baseline2, and RoBERT-LSTM-CRF, as shown in Figure 12.
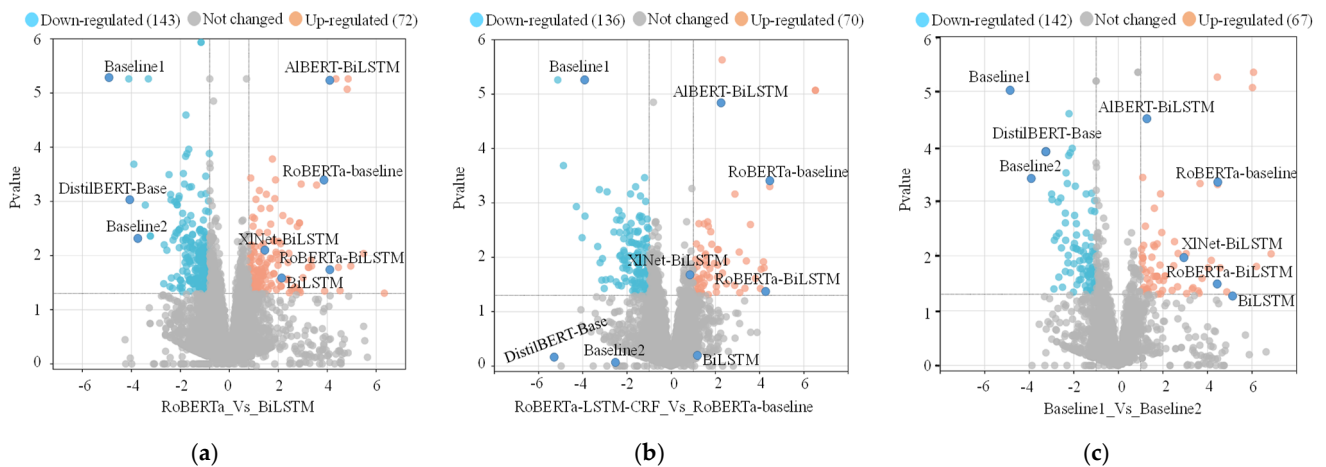

Figure 12. Compared to transformers, a volcano with a differentially expressed model analysis (**a**) RoBERTa_vs_BiLSTM, (**b**) RoBERTa−LSTM−CRF_vs_RoBERTa-baseline, and (**c**) baseline1_vs_baseline2, appears in propaganda contexts.

4.6. Discussion and Research Implications

4.6.1. Theoretical Contribution

This study's findings make theoretical and empirical contributions to the literature on propaganda detection in social media or news data. Theoretically, this study confirms that the theoretical implications of propagandist text detection and mining high-quality data are significant and can be understood in three ways.

First, the study contributes to the existing knowledge by validating the acute role of propaganda detection given the important influence of identification, and researchers can gain insights into how propaganda is used to manipulate public opinion. In the past, researchers focusing on the social media platform have used different, targeted measures to enhance text classification depending on content needs, such as propaganda [38,44]. Moreover, researchers have focused on the effects of online news textual content on user social projects [76]. However, when textual content is included, our results show that the amount indicating attention to the news increases. As a result, other elements of online reviews receive more attention, which evokes the emotions of the social media user.

Second, this study offers a deeper insight into the role of tracking the spread of propagandist text, and researchers can identify the websites, social media accounts, and other channels used to disseminate propaganda. This information can then be used to target these channels with counter-propaganda efforts. Most previous studies have taken news articles from social media or online websites collected as document-based datasets such as TSHP, Qprop, and PTC [5,11,12]. These datasets are used to train a deep-learning model to identify propaganda; however, it is time-consuming as the nature of the data is on the document level. Thus, this study posited novel datasets with a similar feature to the previous dataset and kept each instance at the sentence level, called ProText. The proposed comprehensive model advances the growing body of research on text classification and the literature on propaganda theory by supporting the importance of textual content in the news. The benefits, however, will be specific to information system experts and practitioners for further insight and improvements. Furthermore, this study calls for collaboration among researchers to acquire and exchange datasets and develop a standard for organizing, labeling, and fact-checking.

Third, propagandist text detection helps identify propaganda target audiences. This identification can help uncover hidden trends in spreading and using propaganda. It can also help to identify the source of the propaganda, which can be used to counter its spread. Finally, it can help to track changes in how propaganda is used over time. Propagandist text detection can help to track the effectiveness of propaganda campaigns. This information can then be used to inform countermeasures against the spread of propaganda. Additionally, it can be used to inform strategies to improve the effectiveness of future propaganda campaigns. The most popular propaganda campaigns are used in elections such as, "Yes we can" [12], advertisements such as "Just do it" [13], and healthcare, "Healthcare is wealth" [14]. These campaigns are designed to influence people's opinions and behavior and to shape public perception of issues and candidates. A campaign's success depends on its messaging's effectiveness and ability to reach the intended audience. This paper shows how to track the spread of propagandist text, thus; the benefits, however, are specific to organizers, and companies can see how effective propaganda campaigns are at achieving their goals. These theories provide new conceptual approaches for understanding novel data and extend the range of applications by validating their explanatory power in the current research context. In particular, this information can then be used to improve the design of future propaganda campaigns.

4.6.2. Practical Implication

This study provides important practical implications for information technology societies struggling to improve information system expert defensive performances. First, the feature set contains almost all the features used in the related literature, such as identifying misinformation [3,44], political bias, clickbait, and satire. There is a great deal

of early research in automatically identifying different writing styles and propaganda techniques employed by news sources [11]. Hence, the results of this study validated that a comprehensive dataset including many kinds of sources is instrumental in further re-fining these methods. Furthermore, we included Facebook engagement statistics for news articles (the number of shares and reactions). A variety of news credibility research strands benefit from this dataset. In particular, we argue that this dataset can not only test the generality of previous results in computational journalism, but also spark research in lesser-studied areas.

Second, a text analysis of propaganda can be used to improve machine-learning models for automatic detecting and classifying of propagandist text in social media. This detection can be conducted by adding new information to existing datasets and improving existing algorithms. In this way, machine-learning models can identify and classify propaganda more accurately and efficiently. This model would be beneficial for both researchers and practitioners in countering propaganda. This study's most important managerial implication includes that improved technology would be invaluable in protecting citizens from unintentional exposure to propaganda and for governments and organizations to better understand their adversaries' propaganda strategies. Therefore, this study's findings help to identify potential sources of propaganda, enabling quicker response times and countermeasures that are more effective.

Third, propagandist text detection and mining can provide a better understanding of the public's opinion and sentiment. It can also help identify and address potential risks to an organization or government. Thus, by detecting and analyzing propagandist text, users can make more informed decisions when dealing with the text data. This analysis will reduce the spread of misinformation and help create more honest and accurate conversations. Furthermore, this will foster an environment of trust and accountability. It will also help to ensure that users make decisions based on evidence and facts rather than on rumors and biased information. This study's findings also help create a more informed and responsible society. This study's findings will lead to more informed decision-making, positively affecting our lives and world. Overall, this study's findings show a more informed and conscious society.

5. Conclusions and Future Work

We examined many solutions to the issue of propaganda news and misinformation. We addressed the issue of manually and mechanically identifying propaganda texts and gathering new repositories to evaluate whether a specific news story is propaganda. Contemporary NLP and deep-learning algorithms need actual training data for categorizing propaganda materials. However, we do not think we can tell if a piece of writing is propaganda as computational linguists. Therefore, we advise using databases that feature articles that are fact-checked or verified and categorized by professionals. Unfortunately, we have discovered that such records are scarce since individual labeling takes time. However, the origins of such designations are fact-checking websites that offer services for the public benefit. The dataset's particular objects and labels were scraped, cleaned, and sorted by fact-checking sources. Additionally, topic analysis was carried out, and it was discovered that the datasets were skewed by themes, making text recognition challenging.

We offer ProText as an innovative dataset for propaganda that includes more than 11,000 examples of propaganda text with defined labels and spans. To train deep-learning and machine-learning models, however, our suggested collection of propagandist material is still inadequate and uneven. We welcome the researcher to combine and exchange datasets for collaborating on the deep-learning challenge in advance research because the additional effort will eventually be needed for text collections.

References

1. Ahmed, S.; Hinkelmann, K.; Corradini, F. Fact Checking: An Automatic End to End Fact Checking System. In *Combating Fake News with Computational Intelligence Techniques*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 345–366.
2. Hao, F.; Yang, Y.; Shang, J.; Park, D.-S. AFCMiner: Finding Absolute Fair Cliques From Attributed Social Networks for Responsible Computational Social Systems. In *IEEE Transactions on Computational Social Systems*; IEEE: Piscataway, NJ, USA, 2023.
3. Ahmad, P.N.; Khan, K. Propaganda Detection And Challenges Managing Smart Cities Information On Social Media. *EAI Endorsed Trans. Smart Cities* **2023**, *7*, e2. [CrossRef]
4. Khanday, A.M.U.D.; Wani, M.A.; Rabani, S.T.; Khan, Q.R. Hybrid Approach for Detecting Propagandistic Community and Core Node on Social Networks. *Sustainability* **2023**, *15*, 1249. [CrossRef]
5. Barrón-Cedeno, A.; Jaradat, I.; Da San Martino, G.; Nakov, P. Proppy: Organizing the News Based on Their Propagandistic Content. *Inf. Process. Manag.* **2019**, *56*, 1849–1864. [CrossRef]
6. Alhindi, T.; Pfeiffer, J.; Muresan, S. Fine-Tuned Neural Models for Propaganda Detection at the Sentence and Fragment Levels. *arXiv* **2019**, arXiv:1910.09702.
7. Daval-Frerot, G.; Weis, Y. WMD at SemEval-2020 Tasks 7 and 11: Assessing Humor and Propaganda Using Unsupervised Data Augmentation. In Proceedings of the Fourteenth Workshop on Semantic Evaluation, Barcelona, Spain, 12–13 December 2020; pp. 1865–1874.
8. Vosoughi, S.; Roy, D.; Aral, S. The Spread of True and False News Online. *Science* **2018**, *359*, 1146–1151. [CrossRef]
9. Pavleska, T.; Školkay, A.; Zankova, B.; Ribeiro, N.; Bechmann, A. Performance Analysis of Fact-Checking Organizations and Initiatives in Europe: A Critical Overview of Online Platforms Fighting Fake News. *Soc. Media Converg.* **2018**, *29*, 1–28.
10. Shao, C.; Ciampaglia, G.L.; Flammini, A.; Menczer, F. Hoaxy: A Platform for Tracking Online Misinformation. In Proceedings of the 25th International Conference Companion on World Wide Web, Geneva, Switzerland, 11–15 April 2016; pp. 745–750.
11. Da San Martino, G.; Yu, S.; Barrón-Cedeno, A.; Petrov, R.; Nakov, P. Fine-Grained Analysis of Propaganda in News Article. In Proceedings of the 2019 conference On Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 5636–5646.
12. Rashkin, H.; Choi, E.; Jang, J.Y.; Volkova, S.; Choi, Y. Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 9 September 2017; pp. 2931–2937.
13. Paudel, K.; Hinsley, A.; Veríssimo, D.; Milner-Gulland, E. Evaluating the Reliability of Media Reports for Gathering Information about Illegal Wildlife Trade Seizures. *PeerJ* **2022**, *10*, e13156. [CrossRef]
14. Chen, C.; Wu, K.; Srinivasan, V.; Zhang, X. Battling the Internet Water Army: Detection of Hidden Paid Posters. In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013), New York, NY, USA, 25–28 August 2013; pp. 116–120.
15. Cooper, G. *Populist Rhetoric and Media Misinformation in the 2016 UK Brexit Referendum*; Tumber, H., Waisbord, S., Eds.; Routledge: London, UK, 2021; pp. 397–410.
16. Amin, S.; Alharbi, A.; Uddin, M.I.; Alyami, H. Adapting Recurrent Neural Networks for Classifying Public Discourse on COVID-19 Symptoms in Twitter Content. *Soft Comput.* **2022**, *26*, 11077–11089. [CrossRef]
17. DiMaggio, A.R. Conspiracy Theories and the Manufacture of Dissent: QAnon, the 'Big Lie', COVID-19, and the Rise of Rightwing Propaganda. *Crit. Sociol.* **2022**, *48*, 1025–1048. [CrossRef]

18. Al-Khateeb, S.; Hussain, M.N.; Agarwal, N. Social Cyber Forensics Approach to Study Twitter's and Blogs' Influence on Propaganda Campaigns. In Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, Washington, DC, USA, 5–8 July 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 108–113.

19. Bolsover, G.; Howard, P. Computational Propaganda and Political Big Data: Moving toward a More Critical Research Agenda. *Big Data* **2017**, *5*, 273–276. [CrossRef]

20. Arocena, P.C.; Glavic, B.; Mecca, G.; Miller, R.J.; Papotti, P.; Santoro, D. Messing up with BART: Error Generation for Evaluating Data-Cleaning Algorithms. *Proc. VLDB Endow.* **2015**, *9*, 36–47. [CrossRef]

21. Da San Martino, G.; Cresci, S.; Barrón-Cedeño, A.; Yu, S.; Di Pietro, R.; Nakov, P. A Survey on Computational Propaganda Detection. In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, Yokohama, Japan, 7 January 2021; pp. 4826–4832.

22. Brunello, A.R. A Moral Compass and Modern Propaganda? Charting Ethical and Political Discourse. *Rev. Hist. Political Sci.* **2014**, *2*, 169–197.

23. Guo, N.; Wang, Y.; Jiang, H.; Xia, X.; Gu, Y. TALI: An Update-Distribution-Aware Learned Index for Social Media Data. *Mathematics* **2022**, *10*, 4507. [CrossRef]

24. Abdullah, M.; Altiti, O.; Obiedat, R. Detecting Propaganda Techniques in English News Articles Using Pre-Trained Transformers. In Proceedings of the 2022 13th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 21 June 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 301–308.

25. Vlad, G.-A.; Tanase, M.-A.; Onose, C.; Cercel, D.-C. Sentence-Level Propaganda Detection in News Articles with Transfer Learning and BERT-BiLSTM-Capsule Model. In Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda, Hong Kong, China, 4 November 2019; pp. 148–154.

26. Mosseri, A. News Feed Fyi: Addressing Hoaxes and Fake News. *Facebook Newsroom* **2016**, *15*, 12.

27. Paraschiv, A.; Cercel, D.-C.; Dascalu, M. Upb at Semeval-2020 Task 11: Propaganda Detection with Domain-Specific Trained Bert. *arXiv* **2020**, arXiv:2009.05289.

28. Martinkovic, M.; Pecar, S.; Šimko, M. NLFIIT at SemEval-2020 Task 11: Neural Network Architectures for Detection of Propaganda Techniques in News Articles. In Proceedings of the Fourteenth Workshop on Semantic Evaluation, Barcelona, Spain, 12–13 December 2020; pp. 1771–1778.

29. Bairaktaris, A.; Symeonidis, S.; Arampatzis, A. DUTH at SemEval-2020 Task 11: BERT with Entity Mapping for Propaganda Classification. *arXiv* **2020**, arXiv:2008.09894.

30. Zarour, M.; Alenezi, M.; Ansari, M.T.J.; Pandey, A.K.; Ahmad, M.; Agrawal, A.; Kumar, R.; Khan, R.A. Ensuring Data Integrity of Healthcare Information in the Era of Digital Health. *Healthc. Technol. Lett.* **2021**, *8*, 66–77. [CrossRef]

31. Jabeen, F.; Rehman, Z.U.; Shah, S.; Alharthy, R.D.; Jalil, S.; Khan, I.A.; Iqbal, J.; Abd El-Latif, A.A. Deep Learning-Based Prediction of Inhibitors Interaction with Butyrylcholinesterase for the Treatment of Alzheimer's Disease. *Comput. Electr. Eng.* **2023**, *105*, 108475. [CrossRef]

32. Victoria, V. How Fake News Spreads Online? *Int. J. Media Inf. Lit.* **2020**, *5*, 217–226.

33. García-Marín, D.; Elías, C.; Soengas-Pérez, X. Big Data and Disinformation: Algorithm Mapping for Fact Checking and Artificial Intelligence. In *Total Journalism*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 123–135.

34. Khattak, S.B.A.; Nasralla, M.M.; Marey, M.; Esmail, M.A.; Jia, M.; Umair, M.Y. WLAN Access Points Channel Assignment Strategy for Indoor Localization Systems in Smart Sustainable Cities. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Riyadh, Saudi Arabia, 19–22 February 2022; IOP Publishing: Bristol, UK, 2022; Volume 1026, p. 012043.

35. Khattak, S.B.A.; Jia, M.; Umair, M.Y.; Ahmed, A. Localization of a Mobile Node Using Fingerprinting in an Indoor Environment. In *Communications, Signal Processing, and Systems, Proceedings of the 2018 CSPS Volume II: Signal Processing 7th, Dalian, China, 14–16 July 2020*; Springer: Singapore, 2020; pp. 1080–1090.

36. Chang, R.-C.; Lin, C.-H. Detecting Propaganda on the Sentence Level during the COVID-19 Pandemic. *arXiv* **2021**, arXiv:2108.12269.

37. Chiu, J.P.; Nichols, E. Named Entity Recognition with Bidirectional LSTM-CNNs. *Trans. Assoc. Comput. Linguist.* **2016**, *4*, 357–370. [CrossRef]

38. Du, J.; Vong, C.-M.; Chen, C.P. Novel Efficient RNN and LSTM-like Architectures: Recurrent and Gated Broad Learning Systems and Their Applications for Text Classification. *IEEE Trans. Cybern.* **2020**, *51*, 1586–1597. [CrossRef]

39. Vorakitphan, V.; Cabrio, E.; Villata, S. PROTECT-A Pipeline for Propaganda Detection and Classification. In Proceedings of the Eighth Italian Conference on Computational Linguistics (CLIC-it 2021), Milan, Italy, 26–28 January 2022.

40. Barfar, A. A Linguistic/Game-Theoretic Approach to Detection/Explanation of Propaganda. *Expert Syst. Appl.* **2022**, *189*, 116069. [CrossRef]

41. Guo, Z.; Schlichtkrull, M.; Vlachos, A. A Survey on Automated Fact-Checking. *Trans. Assoc. Comput. Linguist.* **2022**, *10*, 178–206. [CrossRef]

42. Li, W.; Li, S.; Liu, C.; Lu, L.; Shi, Z.; Wen, S. Span Identification and Technique Classification of Propaganda in News Articles. *Complex Intell. Syst.* **2021**, *8*, 3603–3612. [CrossRef]

43. Chaudhari, D.; Pawar, A.V.; Barrón-Cedeño, A. H-Prop and H-Prop-News: Computational Propaganda Datasets in Hindi. *Data* **2022**, *7*, 29. [CrossRef]

44. Chadwick, A.; Stanyer, J. Deception as a Bridging Concept in the Study of Disinformation, Misinformation, and Misperceptions: Toward a Holistic Framework. *Commun. Theory* **2022**, *32*, 1–24. [CrossRef]

45. Lin, J.C.-W.; Shao, Y.; Djenouri, Y.; Yun, U. ASRNN: A Recurrent Neural Network with an Attention Model for Sequence Labeling. *Knowl.-Based Syst.* **2021**, *212*, 106548. [CrossRef]

46. Zareie, A.; Sakellariou, R. Minimizing the Spread of Misinformation in Online Social Networks: A Survey. *J. Netw. Comput. Appl.* **2021**, *186*, 103094. [CrossRef]

47. Ozturk, P.; Li, H.; Sakamoto, Y. Combating Rumor Spread on Social Media: The Effectiveness of Refutation and Warning. In Proceedings of the 2015 48th Hawaii International Conference on System Sciences, Kauai, HI, USA, 5–8 January 2015; pp. 2406–2414.

48. Wu, Y.; Agarwal, P.K.; Li, C.; Yang, J.; Yu, C. Toward Computational Fact-Checking. *Proc. VLDB Endow.* **2014**, *7*, 589–600. [CrossRef]

49. Jaradat, I.; Gencheva, P.; Barrón-Cedeño, A.; Màrquez, L.; Nakov, P. ClaimRank: Detecting Check-Worthy Claims in Arabic and English. *arXiv* **2018**, arXiv:1804.07587.

50. Margolin, D.B.; Hannak, A.; Weber, I. Political Fact-Checking on Twitter: When Do Corrections Have an Effect? *Political Commun.* **2018**, *35*, 196–219. [CrossRef]

51. Chen, Y.; Conroy, N.J.; Rubin, V.L. Misleading Online Content: Recognizing Clickbait as "False News". In Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection, Washington, DC, USA, 13 November 2015; pp. 15–19.

52. Zhang, A.X.; Ranganathan, A.; Metz, S.E.; Appling, S.; Sehat, C.M.; Gilmore, N.; Adams, N.B.; Vincent, E.; Lee, J.; Robbins, M. A Structured Response to Misinformation: Defining and Annotating Credibility Indicators in News Articles. In Proceedings of the Companion Proceedings of the The Web Conference 2018, Lyon, France, 23–27 April 2018; pp. 603–612.

53. Altiti, O.; Abdullah, M.; Obiedat, R. JUST at SemEval-2020 Task 11: Detecting Propaganda Techniques Using BERT Pre-Trained Model. In Proceedings of the Fourteenth Workshop on Semantic Evaluation, Barcelona, Spain, 12–13 December 2020; pp. 1749–1755.

54. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A Robustly Optimized Bert Pretraining Approach. *arXiv* **2019**, arXiv:1907.11692.

55. Liu, Y.; Wu, Y.-F. Early Detection of Fake News on Social Media through Propagation Path Classification with Recurrent and Convolutional Networks. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.

56. Mohtarami, M.; Baly, R.; Glass, J.; Nakov, P.; Màrquez, L.; Moschitti, A. Automatic Stance Detection Using End-to-End Memory Networks. *arXiv* **2018**, arXiv:1804.07581.

57. Mazza, M.; Cresci, S.; Avvenuti, M.; Quattrociocchi, W.; Tesconi, M. Rtbust: Exploiting Temporal Patterns for Botnet Detection on Twitter. In Proceedings of the 10th ACM Conference on Web Science, Boston, MA, USA, 30 June–3 July 2019; pp. 183–192.

58. Hu, Y.; Yang, B.; Duo, B.; Zhu, X. Exhaustive Exploitation of Local Seeding Algorithms for Community Detection in a Unified Manner. *Mathematics* **2022**, *10*, 2807. [CrossRef]

59. Pradeep, R.; Ma, X.; Nogueira, R.; Lin, J. Vera: Prediction Techniques for Reducing Harmful Misinformation in Consumer Health Search. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual, 11–15 July 2021; pp. 2066–2070.

60. Leng, Y.; Zhai, Y.; Sun, S.; Wu, Y.; Selzer, J.; Strover, S.; Zhang, H.; Chen, A.; Ding, Y. Misinformation during the COVID-19 Outbreak in China: Cultural, Social and Political Entanglements. *IEEE Trans. Big Data* **2021**, *7*, 69–80. [CrossRef]

61. Petrocchi, M.; Viviani, M. Report on the 2nd Workshop on Reducing Online Misinformation through Credible Information Retrieval (ROMCIR 2022) at ECIR 2022. In Proceedings of the ACM SIGIR Forum, Taipei, China, 23–27 July 2023; ACM: New York, NY, USA, 2023; Volume 56, pp. 1–9.

62. Djenouri, Y.; Belhadi, A.; Srivastava, G.; Lin, J.C.-W. Advanced Pattern-Mining System for Fake News Analysis. In *IEEE Transactions on Computational Social Systems*; IEEE: Piscataway, NJ, USA, 2023.

63. Koch, T.K.; Frischlich, L.; Lermer, E. Effects of Fact-Checking Warning Labels and Social Endorsement Cues on Climate Change Fake News Credibility and Engagement on Social Media. *J. Appl. Soc. Psychol.* **2023**, 1–3. [CrossRef]

64. Zhang, X.; Zhao, J.; LeCun, Y. Character-Level Convolutional Networks for Text Classification. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 649–657.

65. DiFonzo, N.; Robinson, N.M.; Suls, J.M.; Rini, C. Rumors about Cancer: Content, Sources, Coping, Transmission, and Belief. *J. Health Commun.* **2012**, *17*, 1099–1115. [CrossRef]

66. Li, X.; Wang, W.; Fang, J.; Jin, L.; Kang, H.; Liu, C. PEINet: Joint Prompt and Evidence Inference Network via Language Family Policy for Zero-Shot Multilingual Fact Checking. *Appl. Sci.* **2022**, *12*, 9688. [CrossRef]

67. Yaseen, M.U.; Nasralla, M.M.; Aslam, F.; Ali, S.S.; Khattak, S.B.A. A Novel Approach Based on Multi-Level Bottleneck Attention Modules Using Self-Guided Dropblock for Person Re-Identification. *IEEE Access* **2022**, *10*, 123160–123176. [CrossRef]

68. Saquete, E.; Tomás, D.; Moreda, P.; Martínez-Barco, P.; Palomar, M. Fighting Post-Truth Using Natural Language Processing: A Review and Open Challenges. *Expert Syst. Appl.* **2020**, *141*, 112943. [CrossRef]

69. Abdelnabi, S.; Hasan, R.; Fritz, M. Open-Domain, Content-Based, Multi-Modal Fact-Checking of Out-of-Context Images via Online Resources. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 14940–14949.

70. Kartal, Y.S.; Kutlu, M. Re-Think Before You Share: A Comprehensive Study on Prioritizing Check-Worthy Claims. *IEEE Trans. Comput. Soc. Syst.* **2022**, *10*, 362–375. [CrossRef]

71. Chang, G.; Gao, H.; Yao, Z.; Xiong, H. TextGuise: Adaptive Adversarial Example Attacks on Text Classification Model. *Neurocomputing* **2023**, *529*, 190–203. [CrossRef]

72. Sanh, V.; Debut, L.; Chaumond, J.; Wolf, T. DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter. *arXiv* **2019**, arXiv:1910.01108.

73. Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R.R.; Le, Q.V. Xlnet: Generalized Autoregressive Pretraining for Language Understanding. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 5753–5763.

74. Lan, Z.; Chen, M.; Goodman, S.; Gimpel, K.; Sharma, P.; Soricut, R. Albert: A Lite Bert for Self-Supervised Learning of Language Representations. *arXiv* **2019**, arXiv:1909.11942.

75. Kumar, G.; Singh, J.P.; Singh, A.K. Autoencoder-Based Feature Extraction for Identifying Hate Speech Spreaders in Social Media. In *IEEE Transactions on Computational Social Systems*; IEEE: Piscataway, NJ, USA, 2023.

76. Wani, M.A.; Agarwal, N.; Bours, P. Impact of Unreliable Content on Social Media Users during COVID-19 and Stance Detection System. *Electronics* **2020**, *10*, 5. [CrossRef]

77. Johns, A.; Cheong, N. Feeling the Chill: Bersih 2.0, State Censorship, and "Networked Affect" on Malaysian Social Media 2012–2018. *Soc. Media Soc.* **2019**, *5*, 2056305118821801. [CrossRef]

78. Chang, Y.; Keblis, M.F.; Li, R.; Iakovou, E.; White III, C.C. Misinformation and Disinformation in Modern Warfare. *Oper. Res.* **2022**, *3*, 1577–1597. [CrossRef]

79. Founta, A.; Djouvas, C.; Chatzakou, D.; Leontiadis, I.; Blackburn, J.; Stringhini, G.; Vakali, A.; Sirivianos, M.; Kourtellis, N. Large Scale Crowdsourcing and Characterization of Twitter Abusive Behavior. In Proceedings of the International AAAI Conference on Web and Social Media, Stanford, CA, USA, 25–28 June 2018; Volume 12.

80. Young, T.; Hazarika, D.; Poria, S.; Cambria, E. Recent Trends in Deep Learning Based Natural Language Processing. *IEEE Comput. Intell. Mag.* **2018**, *13*, 55–75. [CrossRef]

81. Yang, K.-C.; Varol, O.; Hui, P.-M.; Menczer, F. Scalable and Generalizable Social Bot Detection through Data Selection. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 1096–1103.

82. Echeverría, J.; De Cristofaro, E.; Kourtellis, N.; Leontiadis, I.; Stringhini, G.; Zhou, S. LOBO: Evaluation of Generalization Deficiencies in Twitter Bot Classifiers. In Proceedings of the 34th Annual Computer Security Applications Conference, San Juan, PR, USA, 3–7 December 2018; pp. 137–146.

83. Wadden, D.; Lin, S.; Lo, K.; Wang, L.L.; van Zuylen, M.; Cohan, A.; Hajishirzi, H. Fact or Fiction: Verifying Scientific Claims. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 16–20 November 2020; pp. 7534–7550.

84. Yilmaz, A.E.; Demirhan, H. Weighted Kappa Measures for Ordinal Multi-Class Classification Performance. *Appl. Soft Comput.* **2023**, *134*, 110020. [CrossRef]

85. Jang, M.; Kang, P. Sentence Transition Matrix: An Efficient Approach That Preserves Sentence Semantics. *Comput. Speech Lang.* **2022**, *71*, 101266. [CrossRef]

86. Solairaj, A.; Sugitha, G.; Kavitha, G. Enhanced Elman Spike Neural Network Based Sentiment Analysis of Online Product Recommendation. *Appl. Soft Comput.* **2023**, *132*, 109789.

87. Liu, X.; Chen, Q.; Liu, Y.; Siebert, J.; Hu, B.; Wu, X.; Tang, B. Decomposing Word Embedding with the Capsule Network. *Knowl. Based Syst.* **2021**, *212*, 106611. [CrossRef]

88. Sasaki, S.; Heinzerling, B.; Suzuki, J.; Inui, K. Examining the Effect of Whitening on Static and Contextualized Word Embeddings. *Inf. Process. Manag.* **2023**, *60*, 103272. [CrossRef]

89. Probst, P.; Wright, M.N.; Boulesteix, A.-L. Hyperparameters and Tuning Strategies for Random Forest. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2019**, *9*, e1301. [CrossRef]