information

Article

Social Media Analytics on Russia–Ukraine Cyber War with Natural Language Processing: Perspectives and Challenges

Fahim Sufi

School of Public Health and Preventive Medicine, Monash University, 553 St. Kilda Rd., Melbourne, VIC 3004, Australia; fahim.sufi@monash.edu

Abstract: Utilizing social media data is imperative in comprehending critical insights on the Russia–Ukraine cyber conflict due to their unparalleled capacity to provide real-time information dissemination, thereby enabling the timely tracking and analysis of cyber incidents. The vast array of user-generated content on these platforms, ranging from eyewitness accounts to multimedia evidence, serves as invaluable resources for corroborating and contextualizing cyber attacks, facilitating the attribution of malicious actors. Furthermore, social media data afford unique access to public sentiment, the propagation of propaganda, and emerging narratives, offering profound insights into the effectiveness of information operations and shaping counter-messaging strategies. However, there have been hardly any studies reported on the Russia–Ukraine cyber war harnessing social media analytics. This paper presents a comprehensive analysis of the crucial role of social-media-based cyber intelligence in understanding Russia's cyber threats during the ongoing Russo–Ukrainian conflict. This paper introduces an innovative multidimensional cyber intelligence framework and utilizes Twitter data to generate cyber intelligence reports. By leveraging advanced monitoring tools and NLP algorithms, like language detection, translation, sentiment analysis, term frequency–inverse document frequency (TF-IDF), latent Dirichlet allocation (LDA), Porter stemming, n-grams, and others, this study automatically generated cyber intelligence for Russia and Ukraine. Using 37,386 tweets originating from 30,706 users in 54 languages from 13 October 2022 to 6 April 2023, this paper reported the first detailed multilingual analysis on the Russia–Ukraine cyber crisis in four cyber dimensions (geopolitical and socioeconomic; targeted victim; psychological and societal; and national priority and concerns). It also highlights challenges faced in harnessing reliable social-media-based cyber intelligence.

Keywords: cyber analytics; analyzing cyber threat; cyber war; social media analytics; Russian cyber incident; Ukrainian cyber incident

Citation: Sufi, F. Social Media Analytics on Russia–Ukraine Cyber War with Natural Language Processing: Perspectives and Challenges. *Information* **2023**, *14*, 485. https://doi.org/10.3390/info14090485

Academic Editor: Vincenzo Moscato

Received: 8 August 2023
Revised: 29 August 2023
Accepted: 30 August 2023
Published: 31 August 2023

1. Introduction

The Russian–Ukraine cyber war has witnessed an escalation in the use of cyberspace as a battleground, where information warfare and cyber attacks are employed to further geopolitical objectives. The academic studies and news reports in [1–16] collectively suggest that cyber warfare has played a significant role in the Russia–Ukraine conflict. These studies describe various incidents of cyber attacks, espionage, and propaganda by both sides, targeting each other's government, military, and civilian infrastructure. The attacks include election interference, power grid disruption, destructive malware, surveillance, website defacement, and email leaks. The studies also provide insights into the strategies, impacts, and implications of cyber operations, highlighting the growing sophistication of cyber capabilities and the challenges of attribution and deterrence. Overall, the studies highlight the importance of cybersecurity in modern conflicts and the need for international norms and cooperation to prevent and mitigate cyber threats. The following are a few notable events as per [1–16]:

- Russia has engaged in cyber warfare as a part of its military strategy in the ongoing conflict with Ukraine.
- The cyber attacks conducted by Russia have targeted a wide range of Ukrainian organizations, including government agencies, media outlets, and critical infrastructure.
- The cyber attacks have caused significant damage to Ukrainian organizations, including disruptions to IT networks, power outages, and data theft.
- Russian cyber operations have sought to influence political events in Ukraine, including attempts to manipulate election results and conduct surveillance on political figures.
- Ukrainian hackers have retaliated against Russian cyber attacks by targeting Russian organizations, including media outlets and government agencies.
- Hacktivist groups have also targeted Belarusian infrastructure as a means of disrupting Russian troop movements towards Ukraine.
- Despite efforts by researchers to decipher Russian cyber strategies, the exact goals and motivations behind these attacks remain unclear.

In the contemporary context of the Russo–Ukrainian cyber war, social media platforms have emerged as crucial sources of valuable intelligence, owing to their expansive reach, real-time data, and user-generated content. This paper aims to comprehensively analyze the indispensability of social-media-based cyber intelligence in understanding and countering Russia's cyber threats during the ongoing conflict with Ukraine. Recent studies [17–25] have highlighted that social media platforms enable rapid information dissemination, providing cybersecurity professionals with real-time tracking and analysis capabilities for cyber incidents, facilitating timely response and mitigation strategies. Additionally, these platforms serve as hubs for user-generated content, such as eyewitness accounts, multimedia evidence, and open-source intelligence, which play a vital role in corroborating and contextualizing cyber incidents, aiding in the identification of attackers. By monitoring social media, analysts can also gauge public sentiment, track the spread of propaganda, and identify emerging narratives, thus offering insights into information operations and counter-messaging strategies. Furthermore, social media monitoring tools and algorithms (such as sentiment analysis, entity recognition, word frequency calculation, and topic analysis, as depicted in [17,19,22]) empower analysts to detect and analyze cyber threats in real-time, enabling proactive defense measures and the attribution of cyber attacks by identifying patterns, tracking malware propagation, and uncovering digital footprints left by threat actors.

In conclusion, social-media-based cyber intelligence constitutes an essential component of modern cybersecurity efforts amidst the Russo–Ukrainian cyber war. Leveraging the capabilities of these platforms to disseminate real-time information, verify incidents, assess public sentiment, and detect cyber threats empowers nations to enhance their defensive postures and mitigate the risks posed by Russia's cyber activities in the context of the ongoing conflict with Ukraine. The following are the core contributions of this paper:

- This paper reported the first critical analysis of Twitter-based critical cyber analytics on the Russia–Ukraine cyber war using data obtained through the Twitter API.
- Using natural language processing (NLP) algorithms, like language detection, translation, sentiment analysis, latent Dirichlet allocation (LDA), term frequency–inverse document frequency (TF-IDF), Porter stemming, n-grams, and others, on live tweets, this study reported an innovative approach in social-media-based cyber intelligence.
- Using a comprehensive literature survey, this paper generated a four-dimensional cyber intelligence framework composed of a geopolitical and socioeconomic perspective; targeted victim perspective; psychological and societal perspective; and national priority and concerns perspective
- This paper used 37,386 tweets that originated from 30,706 users in 54 different languages from 13 October 2022 to 6 April 2023, for automatically generating a cyber intelligence report in four cyber dimensions.

- Finally, this paper reported 12 different challenges of using NLP algorithms on social media for harnessing reliable social-media-based cyber intelligence

In the next section, background and contextual information on the Russia–Ukraine cyber war, multidimensional analysis of cyber threats, and NLP-based tweet analysis are provided. Then, in Section 3 (i.e., Materials and Methods), the detailed steps, flow chart, and algorithms are provided for NLP-based tweet analysis for harnessing Russia–Ukraine cyber intelligence. Finally, results, discussions, and concluding remarks are detailed in Sections 4–6, consecutively.

2. Background Context and Literature

This section provides a succinct background on the Russia–Ukraine cyber conflict followed by multidimensional analysis of cyber threats with NLP.

2.1. Context of Russia–Ukraine Cyber War

Before the conflict in 2022, Russia used offensive cyberspace only twice, according to [1]. The blocking of the Georgian internet access was the closest thing to a significant interruption, with the majority of them being low-level cyber vandalism. Russia attempted to disrupt services and introduce dangerous malware into Ukrainian networks during the Russia–Ukraine war in 2022. This included phishing, denial of service attacks, and the use of software flaws [1]. One company discovered eight distinct families of damaging software that Russia deployed in these attacks, according to a different piece from the Center for Strategic and International Studies [2]. In its most recent confrontations, including its invasions of Georgia in 2008 and Crimea in 2014, Russia reportedly used cyber attacks, according to [3]. Since then, Russia's cyber activities have used Ukraine as a "training ground" [3].

According to a Carnegie Endowment for International Peace report, the biggest ongoing cyber risk to Ukraine as the war rages on is likely Russian intelligence gathering [4]. If Russian hackers are successful in gathering highly valuable intelligence that Moscow then effectively uses, they might theoretically still have a greater influence [4]. Figure 1 provides a brief background on the ongoing cyber war between Russia and Ukraine starting as early as 2014 as it becomes evident from both academic and nonacademic perspectives [1–5]. In the last 10 years, Russia's cyber attacks have caused a greater level of threats compared to Ukrainian attacks [5].

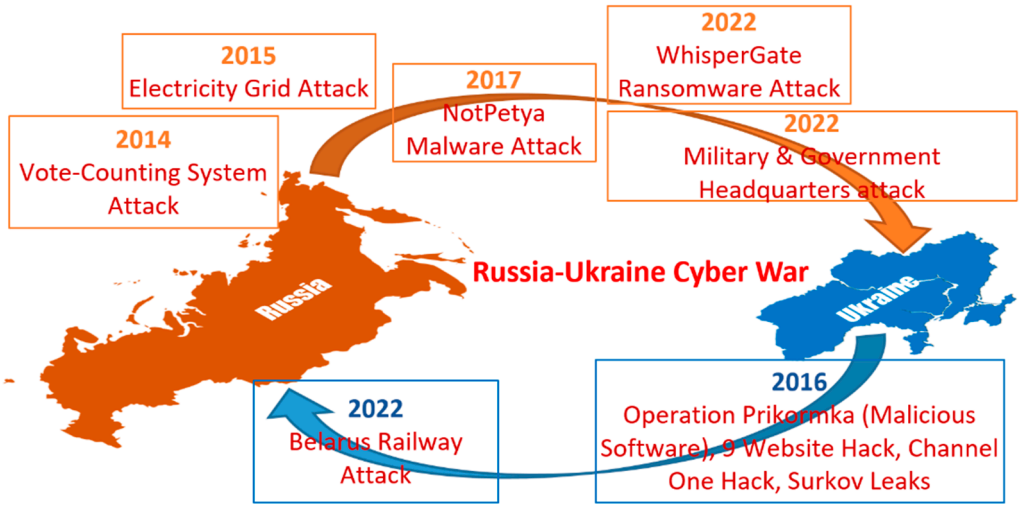

Figure 1. Contextual background of Russia–Ukraine cyber war.

On the day of the general elections (in 2014), Russian cyber attackers broke into the Ukrainian vote-counting system, destroying electronic records and forcing the authorities to physically tally the ballots [6]. A cyber strike during a mission attributed to a group

associated with Russian military intelligence knocked down electricity for several hours in western Ukraine and parts of Kiev in 2015 using a malware called "Industroyer2" [1,7]. It was the first blackout brought on by a cyber attack in history. In 2017, the NotPetya attack took place, which was carried out by the same gang connected to Russian military intelligence [8]. Before spreading globally, the malware package managed to infect roughly 10% of all Ukrainian computer systems. According to a US estimate, one of the most damaging cyber attacks in history resulted in damages for businesses totaling close to USD 10 billion. On 15 January 2022, Microsoft revealed harmful software, disguised as ransomware, known as WhisperGate, which was aimed at several Ukrainian government and nonprofit organizations as well as IT institutions [9]. A cyber attack on Global Affairs Canada (GAC) on 19 January 2022, occurred after Canadian authorities pledged their support for Ukraine [10]. Early in February 2022, Microsoft disclosed that the Actinium group, which is thought to be connected to the Russian secret services, had been targeting Ukrainian military headquarters and government networks [11]. This targeting, which started in October 2021, attempts to eavesdrop on people and collect intelligence.

A series of attacks at much smaller scales were also reported from Ukraine towards Russia during 2016 [12–15]. Among these attacks, Operation Prikormka involved the release of malicious software that displayed a list of fishing bait prices. The exact degree of harm this malicious program caused remains unknown [12]. Another operation involved nine successful hacks of the websites of the separatist movement "Donetsk People's Republic", as well as sites and networks of Russian private military enterprises and Russian sites for anti-Ukrainian propaganda [13]. The "Channel One" attack saw the Ukrainian cyber alliance of hackers FalconsFlame, Trinity, and Rukh8 hack the server of the Russian ChannelOne [14]. Moreover, the Surkov Leaks, which exposed attempts to annex Crimea and incite separatist violence in Donbas in October 2016, involved the leak of 2337 emails and hundreds of attachments [15]. In order to slow down the transit of Russian soldiers through the Republic of Belarus and to the frontiers of Ukraine, a recent cyber attack by a Ukrainian organization on 24 January 2022 damaged the functioning of the railway system in Belarus [16].

2.2. Multidimensional Analysis of Cyber Threats

Recent research has underscored the significance of conducting a comprehensive critical analysis of cyber warfare from multiple vantage points. Such an analysis entails exploring the cyber domain through the lenses of geopolitical and socioeconomic perspectives, with a particular emphasis on ascertaining the actors behind these attacks, their motivations, and the underlying reasons precipitating such cyber assaults [26]. Additionally, to gain deeper insights into the repercussions faced by the targeted victims, the dimension of the "Targeted Victim" assumes paramount importance [27,28]. An in-depth understanding of the broader impact on society and the variegated ways in which different social segments perceive cyber warfare necessitates the incorporation of the "Psychological & Societal" dimension, as previously exemplified in [29].

To quantitatively gauge societal perceptions of cyber warfare, researchers have availed themselves of sentiment analysis techniques, primarily involving the examination of targeted social media messages [17,22]. Of equal significance is the dimension of "National Priority and Concerns", which serves as a pivotal pillar for the comprehensive analysis of cyber warfare [29,30]. In essence, these four interrelated dimensions constitute indispensable components for the thorough examination of the Russia–Ukraine cyber conflict, as visually depicted in Figure 2. A detailed exposition of the progressive development of the four-dimensional cyber intelligence model, meticulously tailored for this study, can be observed in Table 1. This model has been formulated through a meticulous assimilation and synthesis of existing research works [26–32].

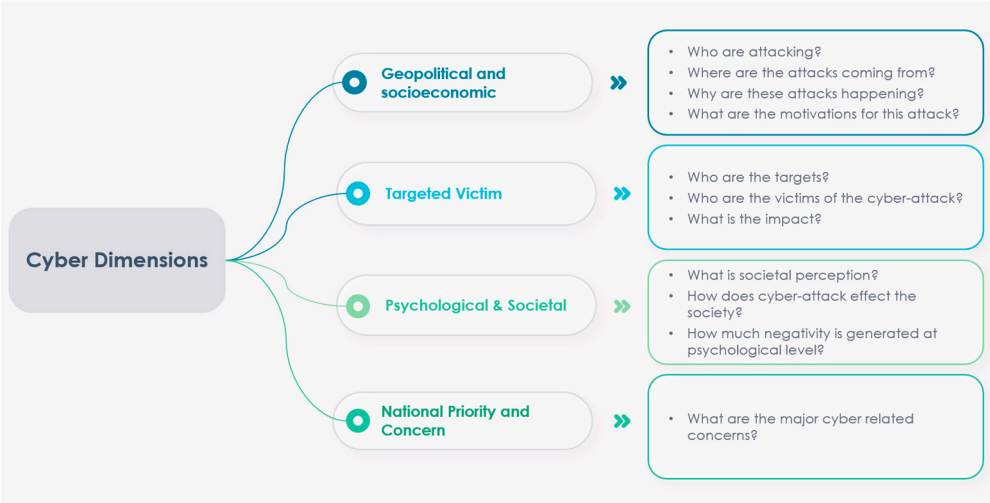

Figure 2. Multidimensional analysis of Russia–Ukraine cyber war.

Table 1. The four dimensions of cyber intelligence used in this study.

Cyber Dimensions	Existing Literature
1. Geopolitical and socioeconomic	[26]
2. Targeted victim	[27,28]
3. Psychological and societal	[29]
4. National priority and concerns	[29,30]

2.3. NLP-Based Analysis of Tweets

Analysis of cyber-related social media posts from Twitter started almost 10 years ago [33]. However, these studies did not utilize the power of AI-based techniques for automated critical analysis of social media posts. The methodology described in [18] uses sentiment analysis of tweets to forecast cyber attacks. They discussed how hacktivists responded to the candidates' comments and actions in major events (i.e., presidential elections of the US in 2016) as they examined their methods and provided examples.

The term frequency–inverse document frequency (TF-IDF) is used in [23] to extract features from a dataset of 2000 tweets that are evenly split between bully and nonbully tweets. The paper compares the performance of five different classifiers based on metrics, including precision, recall, F1-score, accuracy, specificity, MCC, fall out, miss rate, and mean square error. The classifiers are a support vector classifier (SVC), logistic regression, multinomial naive Bayes, a random forest classifier, and a stochastic gradient descent classifier [23]. Finally, the study in [23] comes to the conclusion that logistic regression, which achieved 91% precision, 96% recall, a 93% F1-score, and 93% accuracy for detecting bullying tweets, is the best classifier among the five. The paper also makes some recommendations for future research, including reporting and automatically deleting tweets that abuse people as well as taking harsh action against them. The research article in [24] gathers information about darknet traffic from a variety of sources and uses TF-IDF to identify features in the packet payloads. In order to categorize the packets into various cyber attacks, including worms, denial of service (DOS) attacks, backdoor attacks, distributed denial of service (DDoS) assaults, spam, and malicious contents, the paper uses a Light Gradient Boosting Machine (LGBM) [24]. Finally, the study assesses the LGBM's performance and compares it to other algorithms, concluding that the LGBM performs better than the other algorithms based on testing results [24]. In [20], two new machine-learning-based classifiers were utilized to collect a large-scale Twitter dataset and analyze the themes inside it using LDA. Wi-Fi, smartphones, laptops, smart home technology, financial security, help-seeking, and the roles of many stakeholders were among the themes covered. Additionally, sentiment analysis was performed for the study, and all themes had generally negative sentiments.

When combined with the benchmarked cyber keywords, the next level of sensitive keywords that could lead to vulnerability were identified in a study using LDA [21]. In order to extract sensitive terms from Twitter data, the study developed a framework combining cyber keywords with LDA. The research study in [25] analyzes the spread of COVID-19 misinformation via social media platforms like Twitter. The study gathers and analyzes 227,067 tweets from 2020 that utilize these hashtags and investigates their sentiment, emotion, topic, and user attributes. The majority of the tweets, according to the study, are negative, afraid, angry, and mistrustful. One in five users who used these hashtags had their Twitter accounts suspended by January 2021 [25]. The paper concludes that the tweets show a denial of the COVID-19 epidemic and the dissemination of false or misleading information that jeopardizes public health initiatives. Even though papers in [21,23–25] sporadically used various NLP-based techniques, this study provides the most comprehensive use of NLP-based techniques in a systematic manner. Hence this study leads to gradual improvements upon the research work presented in [21,23–25,27–29].

3. Materials and Methods

The current study employs an advanced social media analytics methodology to investigate Twitter data focusing on cyber-related topics. The data collection involves the retrieval of tweets containing the keywords "cyber" or "hack" from the Twitter platform. Upon obtaining the tweet corpus, a multistage analysis pipeline is deployed to gain valuable insights. The initial step involves the detection of the language used in each tweet. This task is accomplished by leveraging the state-of-the-art Microsoft Cognitive Services' Text Analytics API [34], which employs sophisticated natural language processing techniques to accurately ascertain the language of the tweet. Tweets are then categorized into two subsets: those composed in the English language and those composed in non-English languages. For the set of English tweets, a sentiment analysis procedure is applied, aiming to gauge the emotional tone and polarity of the tweets. This sentiment analysis, based on advanced machine learning models, provides a fine-grained sentiment score for each English tweet, enabling the determination of positive, negative, or neutral sentiment conveyed in the textual content. To handle the non-English tweets, a robust translation mechanism is invoked utilizing the Microsoft Cognitive Services' Text Analytics API [34], which facilitates the seamless translation of non-English tweets into English. The translated tweets are then subjected to the same sentiment analysis procedure as the English tweets, thereby unifying the sentiment analysis process across the entire tweet corpus. Following the sentiment analysis stage, tweets are further categorized based on mentions of country names within their textual content. This process aids in the formation of country-specific tweet groups, where each group comprises tweets related to a particular country or region. Subsequently, term frequency analysis is conducted on each country-specific tweet group, a widely employed statistical technique to identify the relative importance of terms within the textual data. This analysis provides insights into the frequency of occurrence of particular terms and aids in understanding the prominent topics or themes prevalent within each country group. Finally, latent Dirichlet allocation (LDA) topic modeling is deployed on the term frequency data of each country group. LDA is a sophisticated probabilistic model that unveils latent topics within textual data by identifying word co-occurrence patterns. By applying LDA, the study discerns underlying thematic patterns in the tweets, offering a comprehensive and granular perspective on the cyber-related discussions and conversations within distinct geographic contexts. It should be mentioned that generic preprocessing steps like transforming all the posts into lower case texts, removing stop words, removing hypertext markup tags, and tokenizing were performed similarly to recent studies in [17,22,35].

Overall, this highly refined and systematic approach enables a nuanced analysis of cyber-related discussions on Twitter, shedding light on sentiment dynamics across languages and regions while uncovering key topical themes. The findings gleaned from this methodological framework hold significant implications for understanding the global discourse surrounding cyber issues and contribute to the broader domain of social media

analytics research. This process is mathematically represented in the following 7 steps with the contextual notion definitions of Table 2:

1. Obtain tweets with the keywords "cyber" or "hack":

$$T = getTweetsWithKeywords(\text{“}cyber\text{”}, \text{“}hack\text{”}) \tag{1}$$

2. Categorize tweets into English and non-English tweets:

$$T^{english} = \{t \in T \mid L(t) == \text{“}English\text{”}\} \tag{2}$$

$$T^{non\_english} = \{t \in T \mid L(t) \neq \text{“}English\text{”}\} \tag{3}$$

3. Translate non-English tweets to English:

$$T^{translated} = \{translateToEnglish(t, L(t)) \mid t \in T^{non\_english}\} \tag{4}$$

4. Perform sentiment analysis on English and translated tweets:

$$S(t) = performSentimentAnalysis(t) \text{ for } t \in T^{english} \cup T^{translated} \tag{5}$$

5. Group tweets by country names:

$$G = groupTweetsByCountry(T) \tag{6}$$

6. Perform term frequency calculation on each country group:

$$TF(g) = calculateTermFrequency(g) \text{ for } g \in G \tag{7}$$

7. Perform LDA topic analysis on each country group's term frequency data:

$$LDA(g) = performLDATopicAnalysis(TF(g)) \text{ for } g \in G \tag{8}$$

Table 2. Description of mathematical notations.

Notation	Definition
T	Set of all tweets containing the keywords "cyber" or "hack"
$T^{english}$	Set of English tweets in T
$T^{non\_english}$	Set of non-English tweets in T
$T^{translated}$	Set of translated English tweets obtained from $T^{non\_english}$
$S(t)$	Sentiment score of tweet t
$L(t)$	Language of tweet t
$C(t)$	Country name mentioned in tweet t
G	Set of country groups, where each group contains tweets that belong to a specific country based on the country name mentioned in the tweet

Finally, we will have the analyzed LDA topics for each country group in LDA(g). This is also demonstrated with the flow chart in Figure 3. The pseudocode provided in Algorithm 1 provides a more detailed outline of the process. It should be noted that other NLP algorithms, like porter stemming, n-grams, etc., could also be used to perform social-media-based critical cyber analytics, as shown in Figure 4, Table 3, and Algorithm 2.

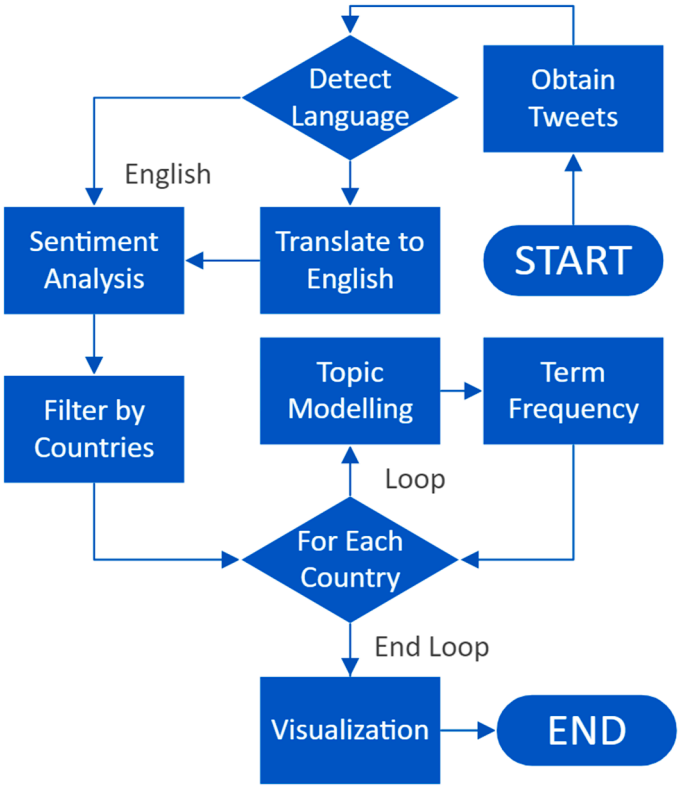

Figure 3. Flow chart of the proposed AI-driven cyber analytics solution with NLP.

Algorithm 1: Pseudocode of NLP-Based Cyber Intelligence Solution Using Twitter Feed

```
FUNCTION socialMediaAnalytics(keyword)
    tweets = getTweetsByKeyword(keyword) // Obtain tweets with the given keyword
    // Initialize data structures to store categorized tweet
    englishTweets = []
    FOR EACH tweet IN tweets
        language = detectTweetLanguage(tweet) // Detect the language of the tweet
        IF language == "English"
            sentiment = performSentimentAnalysis(tweet) // Perform sentiment analysis on English tweets
            englishTweets.append({tweet: sentiment})
        ELSE IF language != "English"
            translatedTweet = translateToEnglish(tweet, language) // Translate non-English tweets to English
            sentiment = performSentimentAnalysis(translatedTweet) // Perform sentiment analysis on translated tweets
            nonEnglishTweets.append({tweet: sentiment})
    // Group tweets by country names
    countryGroups = groupTweetsByCountry(tweets)
    // Initialize data structures to store analyzed country-grouped tweets
    analyzedCountryGroups = []
    FOR EACH countryGroup IN countryGroups
        termFrequency = calculateTermFrequency(countryGroup) // Calculate term frequency for the tweets
        ldaTopicAnalysis = performLDATopicAnalysis(termFrequency) // Perform LDA topic analysis on the term frequency data
        analyzedCountryGroups.append({countryGroup: ldaTopicAnalysis})
    RETURN analyzedCountryGroups // Return the final analyzed data for each country group
END FUNCTION
```

Figure 4. Use of NLP algorithms in analyzing Russia–Ukraine cyber war.

Table 3. Steps within the new NLP-based solution for analyzing Russian and Ukrainian Cyber Related Tweets (✔ denotes APIs used and X denotes APIs not used).

Name of Steps	Use of External API	Name of the Algorithm	References
Analyzing Sentiment	✔	Microsoft Text Analytics [34]	[18,20,36–39]
English Translation	✔	Microsoft Text Analytics [34]	[36–39]
Modeling Topics	X	LDA	[20,21]
Analyzing Term Frequency	X	TF-IDF	[18,20,21,23,24]
Analyzing Term Frequency	X	Porter Stemming	[18]
Analyzing Term Frequency	X	N-Gram	[18,20,21]

Algorithm 2: Using NLP on Tweets Concerning Cyber Issues for Russia and Ukraine

For Each x_i *Tweet in N, Multilingual Tweets*
  If Detect_Language(x_i)<> *'English Language'*
    y_i = *Perform_English_Translattion*(x_i)
  Else
    $y_i = x_i$
  End If
End Loop
For Each y_i *Tweet in N, English Tweets*
  s_i = *Analyse_Sentiment*(y_i)
  If y_i *Contains 'Russia' Or 'Ukraine'*
    $\{c_r, y_i, t_{c_r}\} = y_i$
  End If
End Loop
For Each c_r *in C, Country Names (i.e., for Russia & Ukraine)*
  $\{\{w_j, f_{w_j}\}, \ldots\}$ = *Perform_TF-IDF*(*Tokenization*(y_i))
  $\{\{w_k, f_{w_k}\}, \ldots\}$ = *Perform_PorterStemming*(*Tokenization*(y_i))
  $\{\{\mathrm{U}_1^l w_l, f_o\}, \ldots\}$ = *Perform_N-Gram*(*Tokenization* (y_i))
  $\{\{v_p, \{\{w_q, f_{w_q}\}\ldots\}\}, \ldots\}$ = *Perform_LDA*(*Tokenization* (y_i))
End Loop

4. Results

After implementing the proposed method using Algorithm 1, we evaluated the system from 13 October 2022 to 6 April 2023, with 37,386 tweets from 30,706 users. During this period, tweets in 54 different languages were acquired and analyzed. In total, 8199 http

requests were made for translating non-English Tweets. Table 4 provides information on Twitter data over a period of 7 months, from 2022 October to 2023 April. Within Table 4, the column "Month" provides the time frame of the data collected. Then, "Tweets" shows the number of tweets posted during that period. "Users" shows the number of unique users who posted the tweets. "Geo-Spatial Locations" depicts the number of unique locations mentioned in the tweets. "No. of Languages" shows the number of unique languages used in the tweets. Next, "Retweets" displays the number of times the tweets were retweeted. "Avg. Negative Sentiment", "Avg. Neutral Sentiment", and "Avg. Positive Sentiment" demonstrate the average confidence scores of negative, neutral, and positive sentiment analyses of the tweets, respectively (on a scale of 0 to 1). Finally, "English Translations" provides the number of tweets that were translated. Overall, Table 4 demonstrates that between October 2022 and April 2023 the quantity of tweets, distinct individuals, and distinct locations gradually rose. However, despite a minor increase over the previous month, the number of languages used remained stable. Each month saw a different variation in the overall number of retweets, with November 2022 seeing the greatest total. Over the course of the seven months, the average confidence scores for neutral, positive, and negative sentiment analyses remained stable, with negative sentiment having higher confidence levels than neutral and positive sentiment. It is also important to note that just a small portion of tweets were translated, with December 2022 seeing the largest percentage. It should be mentioned that the presented solution was deployed in desktop, mobile and tablet platforms (in iOS, Android, and Windows). Figure 5 shows the average daily sentiment scores within a deployed mobile app within an Apple iPad 9th Generation. Figure 6 represents the scores at a monthly average scale on a desktop computer. Unlike Figure 5, Figure 6 only shows the negative sentiment score. Cyber-related posts with Negative sentiments are considered to be alerts [17,38]. Finally, in Figure 7, the average negative sentiments for worldwide, Russia, and Ukraine during the entire monitoring period (i.e., 13 October 2022 to 6 April 2023) are summarized. As seen from Figure 7, Russian cyber issues were perceived to be more negative compared to worldwide or Ukrainian statistics.

Table 4. Processing of tweets for AI-based cyber threat intelligence.

Month	Tweets	Users	Geo-Spatial Locations	No. of Languages	Retweets	Avg. Negative Sentiment	Avg. Neutral Sentiment	Avg. Positive Sentiment	English Translations
Oct-22	3954	3556	1588	38	3,727,756	0.36	0.43	0.21	941
Nov-22	6470	5875	2358	38	9,981,856	0.34	0.43	0.23	1283
Dec-22	6512	5544	2225	42	7,565,946	0.35	0.42	0.23	1533
Jan-23	6685	5785	2364	40	7,802,301	0.36	0.40	0.24	1419
Feb-23	5976	5053	2114	43	4,276,479	0.37	0.42	0.21	1373
Mar-23	6634	5749	2357	41	4,799,540	0.36	0.43	0.21	1469
Apr-23	1155	1083	538	27	713,083	0.40	0.41	0.20	258
Total	37,386	30,706	10,178	54	38,866,961	0.36	0.42	0.22	8199

Figure 5. Average of tweet sentiments per day (8 November 2022 portrayed the maximum negative sentiment).

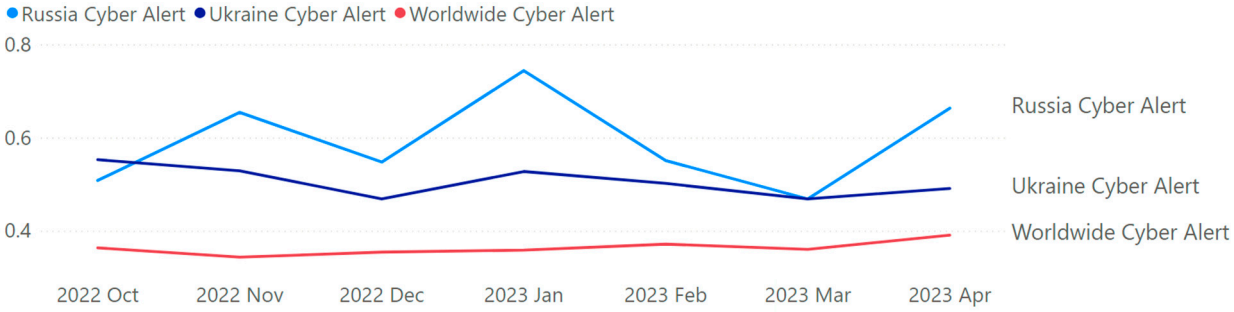

Figure 6. Monthly average negative sentiment of Russian and Ukrainian cyber-related tweets (in comparison to worldwide).

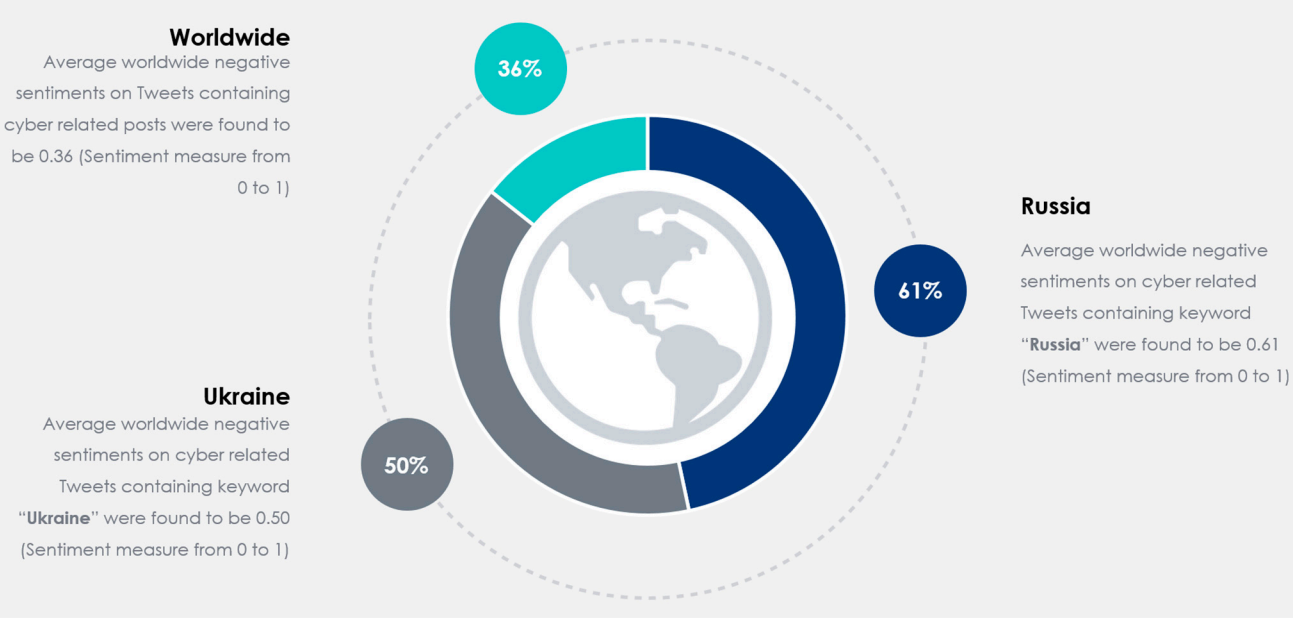

Figure 7. Average negative sentiment of Russian and Ukrainian cyber-related tweets (in comparison to worldwide).

5. Discussion

In general, the subject, vocabulary, and nation-specifics included in the posts varied significantly between Australia and China according to the topic analysis of cyber-related social media posts. With various articles focusing on Australian-specific incidents or worries, such as scams targeting Australians and the Australian health system, there is a heavy emphasis on cyber attacks and cybersecurity for Australia. Other subjects touch on more general cybersecurity issues, including data leaks and breaches. The Australian police and their efforts to prevent cybercrime are another noteworthy subject.

Tables 5 and 6 display the outcome of a topic analysis conducted on Russian cyber-related tweets. By critically examining the table, several meaningful insights emerge, shedding light on prevalent themes and discussions within this domain.

Within the topic analysis process, we configured the LDA algorithm to produce seven topics for both Russia and Ukraine. For each of these topics, LDA was configured to identify the five most-used keywords ranked by weight as show in Table 5.

Table 5. Number of topics with corresponding weights (Wgt) for topic analysis of Russian and Ukrainian cyber-related issues.

	No 1	Wgt	No 2	Wgt	No 3	Wgt	No 4	Wgt	No 5	Wgt	No 6	Wgt	No 7	Wgt
Russia	Russian	72	cyber	65	Russian	65	hack	70	Russia	60	Russia	97	Russian	60
	cyber	65	Russian	44	hack	25	Russia	36	invades	50	hacked	19	Putin	59
	attack	49	Ukraine	24	Shellenberger MD	14	Russian	30	cyber	42	cyber	18	using	58
	blame	27	McGonigal	22	hacking	14	Russians	27	attacks	33	helped	16	Trump	57
	threat	26	FBI	19	amp	13	DNC	16	Darth Putin KGB	26	new	16	story	57
Ukraine	state	3	The Study of War	2	says	3	role	3	country	5	Ukraine	117	leaks	2
	absolutely	2	FBI	2	GicAriana	2	OMC Ukraine	2	loser	3	cyber	76	cyber warfare	2
	threat	2	air	2	need	2	anonymous link	2	brigade	2	Russian	31	cyber attacks	2
	report	2	infrastructure	2	do not	2	council	2	hacker	2	Ukrainian	28	red	2
	cross	2	one	2	security	2	Ukraine–Russia War	2	awareness	2	hack	28	never	2

Table 6. Performance comparison of LDA-based topic analysis for Russian and Ukrainian cyber-related issues.

Performance Parameters	Measured Values for Russia	Measured Values for Ukraine
LogLikelihood	−57,933.967	−23,251.897
Perplexity	458.384	1016.203
Average tokens	1165.143	392
Average document_entropy	4.495	4.364
Average word-length	6.143	7.229
Average coherence	−13.754	−14.672
Average uniform_dist	2.677	2.009
Average corpus_dist	1.614	1.925
Average eff_num_words	98.33	179.378
Average token-doc-diff	0.001	0.007
Average rank_1_docs	0.772	0.174
Average allocation_count	0.85	0.16
Average exclusivity	0.597	0.461
AlphaSum	0.118	8.434
Beta	0.127	0.642
BetaSum	386.22	1039.923

5.1. Analysis of Russian Cyber-Related Tweets

5.1.1. Russian Topic 1

Topic 1, with words such as "Russian", "cyber", "attack", "blame", and "threat", provides valuable insights into the key concepts and concerns dominating the discourse surrounding Russian cyber activities.

The prominence of the term "Russian" with a weight of 72 suggests that discussions within the analyzed tweets frequently revolve around Russia's involvement in cyber-related incidents. This highlights the significance of Russia as a focal point in the context of cyber activities and signals the attention and interest given to the nation's actions within the cyber realm. Moreover, the high weight assigned to "cyber" (65) underscores the central focus on cyber-related topics in the analyzed tweets. It indicates a strong emphasis on various aspects of cybersecurity, encompassing discussions on cyber attacks, threats, and the broader landscape of digital security. The presence of the term "attack" with a weight of 49 highlights the prevalence of discussions related to cyber attacks. This suggests that the analyzed tweets frequently address specific incidents involving cyber assaults, potentially implicating Russian actors or attributing responsibility to them. The weight assigned to "attack" indicates the considerable attention given to these offensive actions and their potential ramifications. Furthermore, the term "blame" carries a weight of 27, indicating discussions surrounding attributions and assigning responsibility for cyber incidents. This suggests that the analyzed tweets frequently delve into debates and speculations regarding who should be held accountable for cyber attacks, with potential implications for Russia's involvement. The inclusion of the term "threat" with a weight of 26 suggests that the analyzed tweets often discuss the broader landscape of cyber threats. This encompasses considerations of potential risks, vulnerabilities, and the evolving nature of cyber threats posed by Russian actors or targeting Russian entities.

Overall, the findings from this topic analysis provide valuable insights into the prevailing themes and discussions surrounding Russian cyber-related tweets. The prominence of terms such as "Russian", "cyber", "attack", "blame", and "threat" underscores the focus on Russia's involvement in cyber activities, particularly in terms of cyber attacks, attributions, and the broader threat landscape.

5.1.2. Russian Topic 2

Topic 2, characterized by the terms "cyber", "Russian", "Ukraine", "McGonigal", and "FBI", contributes to our understanding of key concepts and relationships surrounding Russian cyber activities.

The weight assigned to "cyber" (65) highlights the central focus of the analyzed tweets on cyber-related topics. This suggests a substantial emphasis on various aspects of cybersecurity, including discussions on cyber attacks, defense strategies, and the broader landscape of digital security. The term "Russian" carries a weight of 44, indicating a significant presence in the analyzed tweets. It signifies the attention given to Russia's role and involvement in cyber activities, potentially implicating Russian actors in cyber incidents or discussing Russia's cybersecurity policies and initiatives. The inclusion of "Ukraine" with a weight of 24 suggests that discussions within the analyzed tweets often revolve around the cyber dynamics between Russia and Ukraine. This could involve topics such as cyber attacks targeting Ukrainian entities, the attribution of cyber incidents to Russian actors, or broader geopolitical implications of cyber activities in the region. Additionally, the term "FBI" carries a weight of 19, suggesting discussions related to the involvement or perspective of the Federal Bureau of Investigation (FBI) in the context of Russian cyber activities. This could include discussions on investigations, collaborations, or the attribution of cyber attacks to Russian actors.

Overall, the findings from this topic analysis provide valuable insights into the prevalent themes and relationships surrounding Russian cyber-related tweets. The prominence of terms such as "cyber", "Russian", "Ukraine", "McGonigal", and "FBI" underscores the focus on cybersecurity, Russia's involvement, and the specific dynamics between Russia, Ukraine, and external entities such as the FBI.

5.1.3. Russian Topic 3

Topic 3, characterized by the terms "Russian", "hack", "ShellenbergerMD", "hacking", and "amp", provides significant insights into the key concepts and relationships surrounding Russian cyber activities.

The weight assigned to "Russian" (65) indicates a dominant presence in the analyzed tweets, underscoring the focus on Russia's involvement in cyber-related incidents. This suggests a considerable emphasis on discussions pertaining to Russian actors, their tactics, motivations, and potential implications for cybersecurity. The term "hack" carries a weight of 25, indicating discussions centered around hacking activities within the analyzed tweets. This suggests that cyber breaches and unauthorized access to systems or data are common topics of interest, potentially involving Russian actors or targeting Russian entities. The presence of "ShellenbergerMD" and "hacking" with equal weights of 14 suggests the mention of these terms in the analyzed tweets, albeit to a lesser extent. Additionally, the term "amp" with a weight of 13 indicates a minor presence in the analyzed tweets. Further research is warranted to determine the specific context in which this term appears and its relationship to Russian cyber activities.

Overall, the findings from this topic analysis provide valuable insights into the prominent themes and relationships surrounding Russian cyber-related tweets. The weight assigned to terms such as "Russian" and "hack" underscores the focus on Russia's involvement in cyber activities, specifically hacking incidents.

5.1.4. Russian Topic 4

Topic 4, characterized by the terms "hack", "Russia", "Russian", "Russians", and "DNC", provides significant insights into key concepts and relationships surrounding Russian cyber activities.

The term "hack" carries a weight of 70, indicating a dominant presence in the analyzed tweets. This highlights a strong focus on discussions surrounding hacking activities, potentially involving Russian actors or targeting various entities. The prominence of "hack" underscores the significance of cyber breaches and unauthorized access as central

themes within the analyzed tweets. The presence of "Russia" with a weight of 36 suggests discussions that focus on the broader context of Russian involvement in cyber activities. This term indicates attention given to Russia as a nation-state, possibly exploring its policies, capabilities, and motivations within the cyber realm. Additionally, the terms "Russian" and "Russians" with weights of 30 and 27, respectively, further emphasize the attention given to Russian actors within the analyzed tweets. These terms may allude to discussions related to the tactics, motivations, and potential implications of cyber activities carried out by individuals or groups associated with Russia. The inclusion of "DNC" with a weight of 16 suggests discussions concerning the Democratic National Committee (DNC), which may be targeted by cyber attacks or implicated in broader discussions on Russian cyber activities. This term highlights a specific entity or context associated with cyber incidents and potential political ramifications.

Overall, the findings from this topic analysis provide valuable insights into the prominent themes and relationships surrounding Russian cyber-related tweets. The weight assigned to terms such as "hack", "Russia", "Russian", "Russians", and "DNC" underscores the focus on hacking activities, Russia's involvement, and potential political implications within the analyzed tweets.

5.1.5. Russian Topic 5

Topic 5, characterized by the terms "Russia", "Invades", "Cyber", "attacks", and "DarthPutinKGB", offers meaningful insights into key concepts and relationships surrounding Russian cyber activities.

The weight assigned to "Russia" at 60 indicates a dominant presence in the analyzed tweets. This suggests a significant focus on discussions centered around Russia's involvement in cyber-related incidents. It highlights the importance of understanding Russia's role and impact within the broader context of cybersecurity. The inclusion of "Invades" with a weight of 50 suggests discussions related to instances or allegations of Russia invading or intruding upon specific territories, possibly through cyber means. This term implies the potential crossing of boundaries or encroachments by Russian actors, contributing to the discourse surrounding cyber activities. The term "Cyber" carries a weight of 42, indicating discussions that focus on the broader domain of cybersecurity. This encompasses various aspects, such as cyber attacks, defense strategies, and the evolving landscape of digital security. The presence of "Cyber" underscores the importance of understanding and addressing the challenges posed by cyber threats in the context of Russian cyber activities. The term "attacks" with a weight of 33 highlights discussions concerning cyber attacks. This suggests a significant emphasis on analyzing and discussing incidents involving cyber assaults, potentially implicating Russian actors or targeting entities related to Russia. The weight assigned to "attacks" indicates the attention given to these offensive actions and their potential consequences. Additionally, the term "DarthPutinKGB" with a weight of 26 implies references to a specific persona or narrative associated with Russian cyber activities. Further investigation is necessary to determine the specific context and significance of this term within the broader discourse on Russian cyber-related incidents.

Overall, the findings from this topic analysis provide valuable insights into the prominent themes and relationships surrounding Russian cyber-related tweets. The weight assigned to terms such as "Russia", "Invades", "Cyber", "attacks", and "DarthPutinKGB" underscores the focus on Russia's involvement, cyber attacks, and the broader landscape of cybersecurity.

5.1.6. Russian Topic 6

Topic 6, characterized by the terms "Russia", "hacked", "cyber", "helped", and "new", offers significant insights into key concepts and relationships surrounding Russian cyber activities.

The weight assigned to "Russia" at 97 indicates a dominant presence in the analyzed tweets, suggesting a substantial focus on discussions related to Russia's involvement in

cyber-related incidents. This signifies the significance of understanding Russia's role and impact within the broader context of cybersecurity. The term "hacked" carries a weight of 19, suggesting discussions centered around cyber breaches or unauthorized access to systems or data. The presence of "hacked" in this topic suggests that hacking incidents, potentially involving Russian actors or targeting entities related to Russia, are a key focus within the analyzed tweets. Additionally, the term "cyber" with a weight of 18 highlights discussions on the broader domain of cybersecurity. This encompasses various aspects, such as cyber attacks, defense strategies, and the evolving landscape of digital security. The prominence of "cyber" emphasizes the importance of understanding and addressing the challenges posed by cyber threats in the context of Russian cyber activities. The terms "helped" and "new" with equal weights of 16 suggest their mention in the analyzed tweets, although to a lesser extent.

Overall, the findings from this topic analysis provide valuable insights into the prevalent themes and relationships surrounding Russian cyber-related tweets. The high weight assigned to "Russia" underscores the focus on Russia's involvement, while the presence of "hacked" and "cyber" indicates attention given to hacking incidents and the broader domain of cybersecurity.

5.1.7. Russian Topic 7

Topic 7, characterized by the terms "Russian", "Putin", "using", "Trump", and "story", provides significant insights into key concepts and relationships surrounding Russian cyber activities.

The term "Russian" carries a weight of 60, indicating a dominant presence in the analyzed tweets. This suggests a substantial focus on discussions concerning Russia's involvement in cyber-related incidents. The prominence of "Russian" underscores the importance of understanding Russia's role and impact within the broader context of cybersecurity. The term "Putin" carries a weight of 59, highlighting discussions specifically centered around Russian President Vladimir Putin. These discussions may explore his role, influence, or potential involvement in cyber activities. The presence of "Putin" signifies the attention given to the leadership and decision-making aspects associated with Russian cyber-related incidents. The term "using" carries a weight of 58, suggesting discussions related to the employment or utilization of various tactics, tools, or strategies within the context of cyber activities. This may encompass discussions on the methods used by Russian actors in carrying out cyber attacks or their utilization of cyber capabilities for specific purposes. Additionally, the terms "Trump" and "story" both have a weight of 57. This suggests discussions that involve former US President Donald Trump and narratives associated with Russian cyber-related incidents. These discussions may encompass topics such as alleged ties or collaborations between Trump and Russia or narratives surrounding Russian cyber activities that gained attention in the media or public discourse.

Overall, the findings from this topic analysis provide valuable insights into the prevalent themes and relationships surrounding Russian cyber-related tweets. The weights assigned to terms such as "Russian", "Putin", "using", "Trump", and "story" underscore the focus on Russia's involvement, leadership dynamics, tactics, and potential narratives within the analyzed tweets.

5.2. Analysis of Ukrainian Cyber-Related Tweets

5.2.1. Ukrainian Topic 1

Topic 1, characterized by the terms "State", "absolutely", "Threat", "report", and "Cross", offers valuable insights into key concepts and relationships surrounding Ukrainian cyber activities.

The term "State" carries a weight of 3, indicating its presence in the analyzed tweets. This suggests discussions related to the involvement or actions of state entities within the realm of Ukrainian cyber activities. It may encompass topics such as government initiatives, policies, or state-sponsored cyber operations. The term "absolutely" appears with a weight

of 2, suggesting its mention in the analyzed tweets, albeit to a lesser extent. The term "Threat" carries a weight of 2, indicating discussions related to cyber threats in the context of Ukraine. This suggests that the analyzed tweets may discuss potential risks, vulnerabilities, or challenges posed by cyber incidents to Ukrainian entities or infrastructure.

Additionally, the terms "report" and "Cross" both have a weight of 2. Overall, the findings from this topic analysis provide valuable insights into prevalent themes and relationships surrounding Ukrainian cyber-related tweets. The weights assigned to terms such as "State", "absolutely", "Threat", "report", and "Cross" signify discussions related to state involvement, potential risks, and the reporting of cyber incidents within the context of Ukraine.

5.2.2. Ukrainian Topic 2

Topic 2, characterized by the terms "TheStudyofWar", "FBI", "air", "infrastructure", and "one", provides meaningful insights into key concepts and relationships surrounding Ukrainian cyber activities.

The term "TheStudyofWar" carries a weight of 2, suggesting its presence in the analyzed tweets. The term "FBI" also appears with a weight of 2, indicating its mention in the analyzed tweets. This suggests discussions related to the involvement, perspective, or potential collaborations between the Federal Bureau of Investigation (FBI) and Ukrainian cyber-related incidents. The terms "air", "infrastructure", and "one" each have a weight of 2, suggesting their presence in the analyzed tweets, albeit to a lesser extent.

Overall, the findings from this topic analysis provide initial insights into the prevalent themes and relationships surrounding Ukrainian cyber-related tweets. The weights assigned to terms such as "TheStudyofWar", "FBI", "air", "infrastructure", and "one" suggest discussions related to various aspects of Ukrainian cyber activities, including potential collaborations with international entities and considerations of critical infrastructure and its vulnerabilities.

5.2.3. Ukrainian Topic 3

Topic 3, characterized by the terms "says", "GicAriana", "need", "don't", and "Security", offers valuable insights into key concepts and relationships surrounding Ukrainian cyber activities.

The term "says" carries a weight of 3, indicating its presence in the analyzed tweets. This suggests discussions related to statements, opinions, or reported information concerning Ukrainian cyber-related incidents. It signifies the focus on conveying and sharing information within the context of Ukrainian cyber activities. The term "GicAriana" appears with a weight of 2, suggesting its mention in the analyzed tweets. The terms "need" and "don't" each have a weight of 2, indicating their presence in the analyzed tweets, albeit to a lesser extent. These terms suggest discussions related to requirements, recommendations, or expressions of opinions concerning Ukrainian cyber activities. Additionally, the term "Security" carries a weight of 2, suggesting its presence in the analyzed tweets. This implies discussions concerning cybersecurity measures, practices, or concerns within the context of Ukrainian cyber-related incidents.

Overall, the findings from this topic analysis provide valuable insights into the prevalent themes and relationships surrounding Ukrainian cyber-related tweets. The weights assigned to terms such as "says", "GicAriana", "need", "don't", and "Security" suggest discussions related to information sharing, opinions, recommendations, and cybersecurity practices within the context of Ukrainian cyber activities.

5.2.4. Ukrainian Topic 4

Topic 4, characterized by the terms "role", "OMC_Ukraine", "Anonymous_Link", "Council", and "UkraineRussiaWar", provides meaningful insights into key concepts and relationships surrounding Ukrainian cyber activities.

The term "role" carries a weight of 3, indicating discussions related to the functions, responsibilities, or contributions of various actors within the context of Ukrainian cyber-related incidents. This term (i.e., "role") suggests a focus on understanding and analyzing the specific roles played by different entities in the cyber landscape. The terms "OMC_Ukraine" and "Anonymous_Link" both appear with a weight of 2, suggesting their presence in the analyzed tweets. The term "Council" carries a weight of 2, indicating its presence in the analyzed tweets. This suggests discussions concerning a council or a specific body related to cyber activities within the Ukrainian context. Additionally, the term "UkraineRussiaWar" appears with a weight of 2, implying discussions related to the ongoing conflict between Ukraine and Russia and its intersections with cyber activities. This term highlights the broader geopolitical dynamics influencing cyber-related incidents within the Ukrainian context.

Overall, the findings from this topic analysis provide valuable insights into the prevalent themes and relationships surrounding Ukrainian cyber-related tweets. The weights assigned to terms such as "role", "OMC_Ukraine", "Anonymous_Link", "Council", and "UkraineRussiaWar" suggest discussions related to understanding roles, specific entities, and the geopolitical context within Ukrainian cyber activities.

5.2.5. Ukrainian Topic 5

Topic 5, characterized by the terms "country", "loser", "brigade", "hacker", and "awareness", provides meaningful insights into key concepts and relationships surrounding Ukrainian cyber activities.

The term "country" carries a weight of 5, indicating discussions related to the nation-state context within the analyzed tweets. This suggests a focus on the role of countries, potentially including Ukraine and other nations, in cyber-related incidents. Discussions may involve national cybersecurity strategies, cyber defense capabilities, or the impact of cyber activities on national interests. The term "loser" appears with a weight of 3, suggesting discussions related to derogatory references or negative sentiments toward certain actors or entities within the context of Ukrainian cyber-related incidents. The terms "brigade" and "hacker" both carry a weight of 2, indicating their presence in the analyzed tweets. These terms suggest discussions related to organized groups or individuals engaged in cyber activities. Additionally, the term "awareness" carries a weight of 2, indicating discussions related to raising awareness of cyber threats, best practices, or education within the context of Ukrainian cyber-related incidents. This term (i.e., awareness) suggests a focus on improving knowledge and preparedness to mitigate cyber risks.

Overall, the findings from this topic analysis provide valuable insights into the prevalent themes and relationships surrounding Ukrainian cyber-related tweets. The weights assigned to terms such as "country", "loser", "brigade", "hacker", and "awareness" suggest discussions related to the nation-state context, negative sentiments, organized groups, and cybersecurity awareness within the analyzed tweets.

5.2.6. Ukrainian Topic 6

Topic 6, characterized by the terms "Ukraine", "cyber", "Russian", "Ukrainian", and "hack", offers significant insights into key concepts and relationships surrounding Ukrainian cyber activities.

The term "Ukraine" carries a weight of 117, indicating a dominant presence in the analyzed tweets. This suggests a strong focus on discussions specifically related to Ukraine within the context of cyber-related incidents. The prominence of "Ukraine" underscores the significance of understanding the country's role, challenges, and experiences in the cyber realm. The term "cyber" carries a weight of 76, indicating discussions centered around various aspects of cybersecurity. This includes discussions on cyber attacks, defense strategies, emerging threats, and the broader landscape of digital security. The high weight assigned to "cyber" signifies its importance and prominence within the analyzed tweets. The term "Russian" appears with a weight of 31, suggesting discussions related to

Russia's involvement in Ukrainian cyber-related incidents. This may involve attributing cyber attacks to Russian actors, analyzing their tactics and motivations, or exploring the geopolitical dynamics between Ukraine and Russia in the cyber domain. The term "Ukrainian" carries a weight of 28, indicating discussions specifically focused on Ukrainian actors, entities, or perspectives within the realm of cyber activities. This suggests an emphasis on understanding the Ukrainian context, cybersecurity initiatives, or responses to cyber threats. Additionally, the term "hack" carries a weight of 28, highlighting discussions related to cyber breaches, unauthorized access, or hacking incidents within the analyzed tweets. The weight assigned to "hack" indicates the attention given to these offensive actions and their potential impact on Ukrainian entities.

Overall, the findings from this topic analysis provide valuable insights into the prevalent themes and relationships surrounding Ukrainian cyber-related tweets. The high weights assigned to terms such as "Ukraine" and "cyber" underscore the focus on Ukraine's experiences and challenges in the cyber domain. The presence of "Russian", "Ukrainian", and "hack" further contributes to understanding the dynamics, actors, and incidents related to Ukrainian cyber activities.

5.2.7. Ukrainian Topic 7

Topic 7, characterized by the terms "Leaks", "cyberwarfare", "cyberattacks", "Red", and "never", provides significant insights into key concepts and relationships surrounding Ukrainian cyber activities.

The term "Leaks" carries a weight of 2, indicating its presence in the analyzed tweets. This suggests discussions related to the unauthorized disclosure of sensitive information within the context of Ukrainian cyber-related incidents. The terms "cyberwarfare" and "cyberattacks" both appear with a weight of 2, suggesting their mention in the analyzed tweets. These terms highlight discussions related to the use of cyber capabilities as a means of warfare and the occurrence of cyber attacks within the Ukrainian cyber landscape. These discussions may involve strategies, countermeasures, or the analysis of specific cyber incidents. The term "Red" also carries a weight of 2, suggesting its presence in the analyzed tweets. Additionally, the term "never" carries a weight of 2, indicating discussions related to the absence or prevention of certain events or outcomes within the context of Ukrainian cyber-related incidents.

Overall, the findings from this topic analysis provide insights into the prevalent themes and relationships surrounding Ukrainian cyber-related tweets. The weights assigned to terms such as "Leaks", "cyberwarfare", "cyberattacks", "Red", and "never" suggest discussions related to unauthorized disclosures, the use of cyber capabilities in warfare, cyber attacks, and the absence of certain outcomes within the analyzed tweets.

5.3. Overall Outcome of Topic Analysis

The analysis portrayed in Sections 5.1 and 5.2 contributes to the academic understanding of Russian and Ukrainian cyber activities by highlighting the key topics and concepts present in the analyzed tweets. It helps researchers, policymakers, and cybersecurity professionals gain a deeper insight into the discussions, relationships, and potential implications surrounding Russian cyber-related incidents, as summarized in Figure 8. Figure 8 shows the summarized outcome of Russian and Ukrainian cyber intelligence through the lenses of geopolitics, targets, societal impact, and national priorities. As seen from Figure 8, the dynamics of the Russian cyber war include the involvement of leaders (which is apparent from the mentions of world leaders like Putin, Trump, and others). Tweets related to the Russian cyber war also involve government entities and spy agencies, like the KGP, FBI, and others. Russian cyber activities often target external entities like the DNC as a form of offensive attack. Russian cyber activities are part of a governmental strategic goal of achieving worldwide information supremacy (i.e., business-as-usual activity). As highlighted previously in Figure 7, Russian cyber activity generates a relatively higher level of negative perception compared to Ukrainian or worldwide averages. During the

monitored period, the average negative sentiment on Russian cyber-related tweets was 0.61 (compared to the worldwide average of 0.36 and Ukrainian average of 0.50).

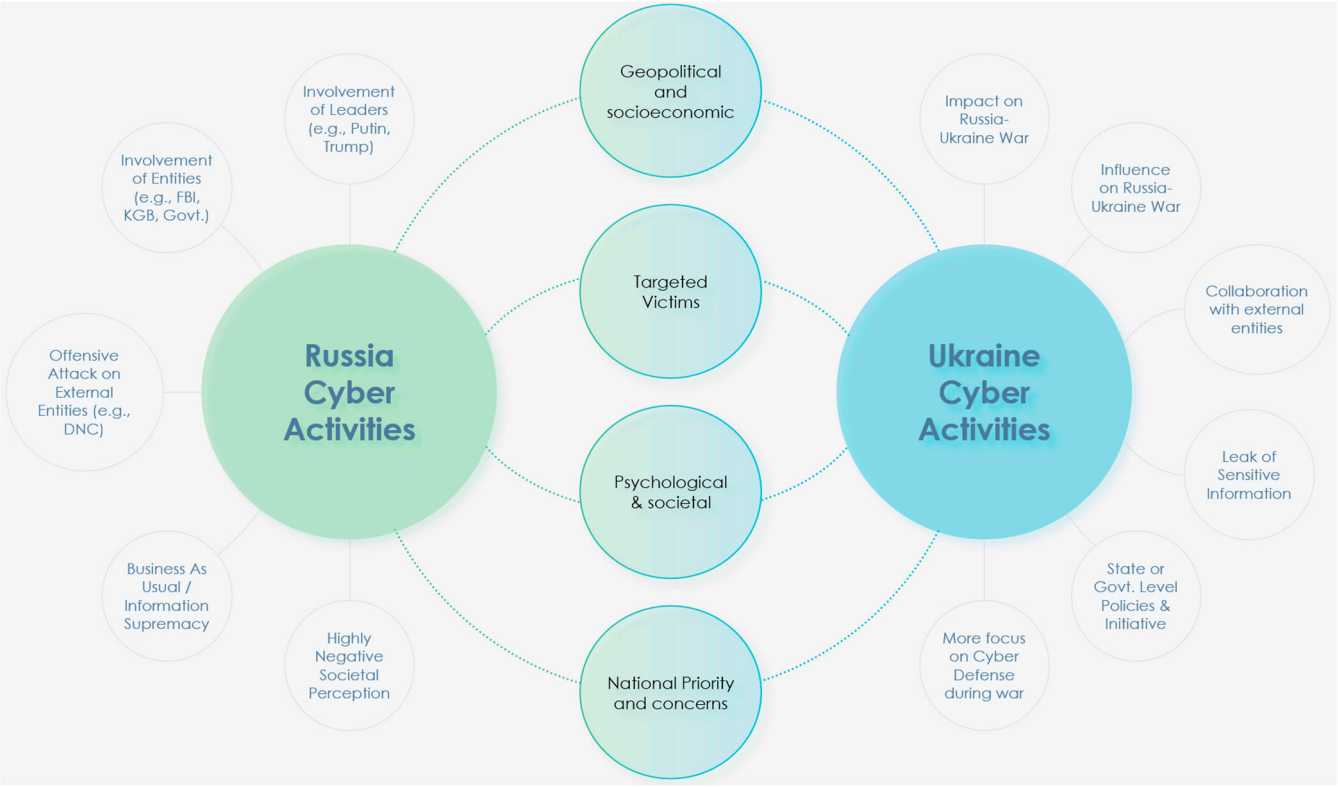

Figure 8. Outcome of NLP-based Russia–Ukraine cyber war Tweets.

In contrast, Ukrainian cyber activities are more dependent on the impact of the Russian–Ukrainian conflict. Ukrainian cyber activities mostly originate for influencing the Russian–Ukrainian conflict. Their focus is mostly on cyber defense against Russian attacks and protecting the leak of sensitive information. Ukrainian cyber activities also collaborate with international entities (e.g., US cyber authorities) for launching collaborative attacks directed towards Russia.

As seen from Figure 9, the presented system could be deployed in a mobile platform (e.g., Samsung S23 Ultra mobile), facilitating sending immediate insights to a mobile strategic decision maker. As decision makers need to make evidence-based strategic decisions being completely mobile, the presented system was deployed in almost all mobile platforms (i.e., iOS version 16 as shown in Figure 5 and Android version 14 as shown in Figure 9).

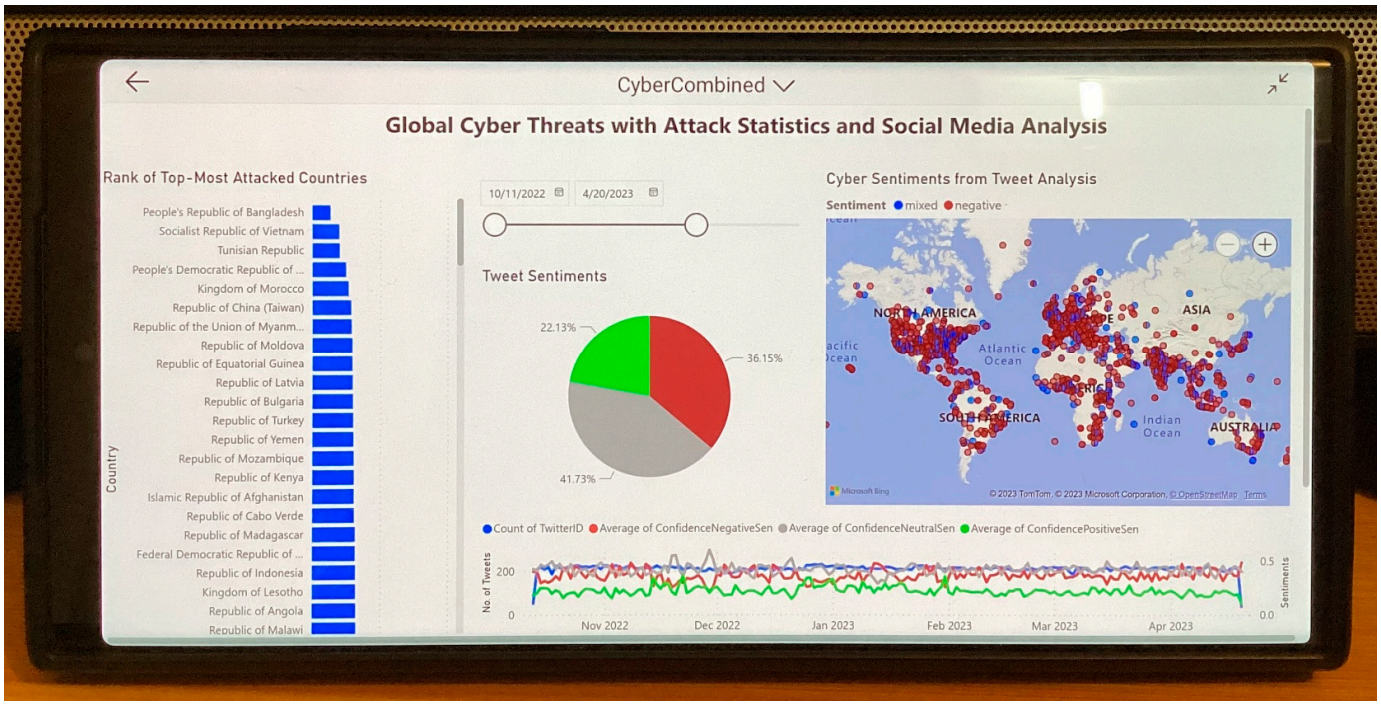

Figure 9. NLP-based cyber intelligence solution deployed in Samsung S23 Ultra Mobile Phone (running Android Version 14).

5.4. Challenges of Social-Media-Based Cyber Intelligence

In the previous section, an innovative approach on obtaining social-media-driven cyber intelligence was presented. Using this innovative method, a magnitude of NLP algorithms were applied to tweets for obtaining instant cyber intelligence on Russia and Ukraine. However, there are several drawbacks on using tweets as a single source of truth for cyber intelligence. These drawbacks fall into three main categories, qualitative, technical, and ethical, as portrayed in Figure 10. The qualitative category includes data quality issues (e.g., misinformation and false information coming out of fake accounts), inaccuracies (fake information, hoaxes, propaganda-oriented tweets), noises (e.g., tweets containing images, videos, emojis, and other nonsense characters). There are also technical challenges of changing formats or structures (i.e., complexity), the requirements of sophisticated tools and big data analytics methodologies (i.e., technical limitations), the massive quantity of data (i.e., data overload), or even the challenge of dealing with irrelevant topics and information (i.e., data relevance). Moreover, NLP-based algorithms, like sentiment analysis (as shown in [40–49]) and entity recognition (as shown in [44,45,50–52]), suffer from misclassifications or errors. Minimizing the errors with optimized solutions incurs additional computational burdens and complexities.

Apart from technical and qualitative concerns, there are also challenges pertaining to ethical considerations, as shown in Figure 10. These challenges include privacy laws about the anonymous and unauthorized access of private social media posts, cultural and religious biases, ignoring privacy laws, ignoring human right concerns, etc. Moreover, social media messages only provide information on how the general public perceives a cyber event. This does not provide the full picture on who conducted a cyber attack or why they conducted the cyber attack. To obtain a comprehensive picture on a cyber event, information from other sources (such as attack databases, anti-virus vendors, etc. [17,19,22]) must be incorporated. While Figure 10 summarizes these challenges, Table 7 shows the detailed references from where more information on these challenges could be obtained.

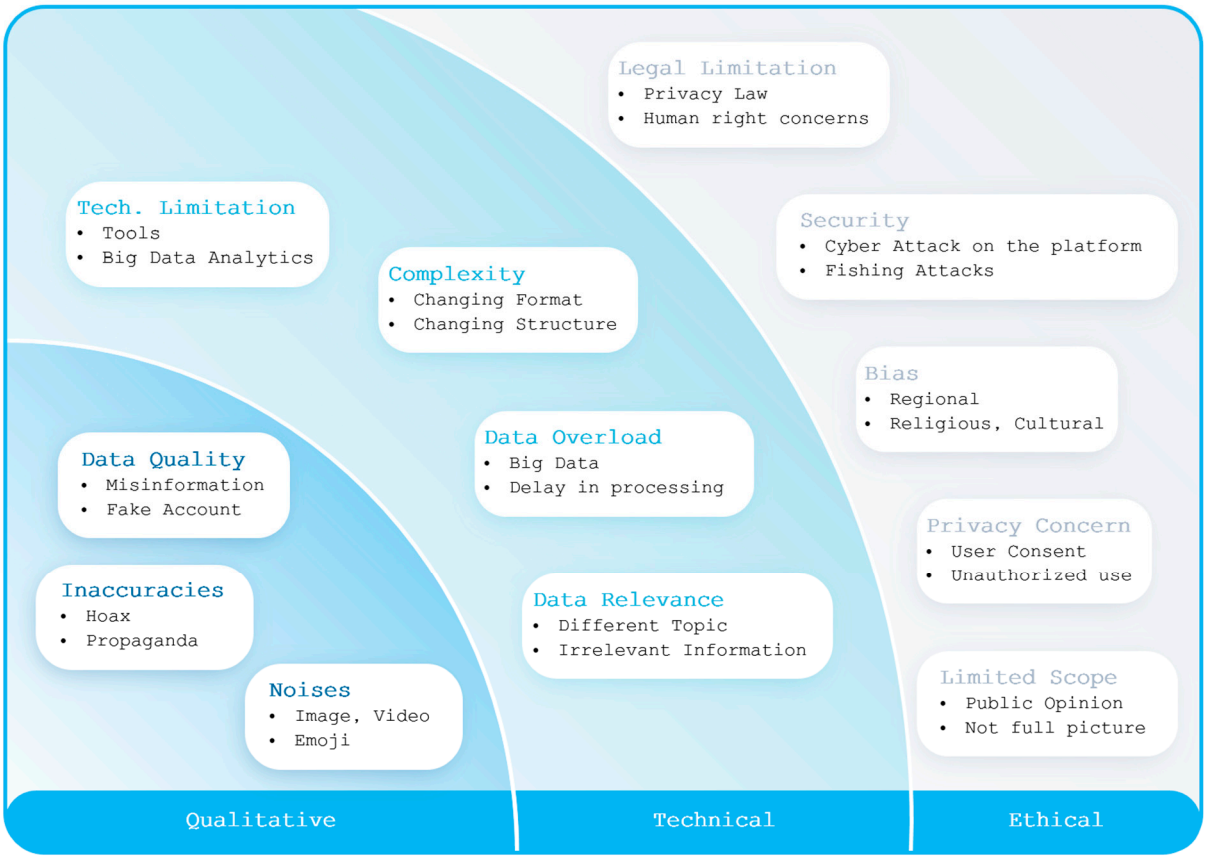

Figure 10. Challenges of NLP-based analysis of Russia–Ukraine cyber war using tweets.

Table 7. 12 identified challenges of social-media-based cyber intelligence supported by references (X means applicable).

References	Data Quality	Privacy Concern	Bias	Data Overload	Limited Scope	Technical Limitation	Complexity	Inaccuracies	Noise	Security	Legal Limitations	Data Relevance
[53]	X											
[54]			X									
[55]							X					
[56]		X										
[57]				X								
[58]												
[59]					X							
[60]			X				X	X	X			
[61]					X							
[62]				X								
[63]					X							
[64]										X		

Table 7. *Cont.*

References	Data Quality	Privacy Concern	Bias	Data Overload	Limited Scope	Technical Limitation	Complexity	Inaccuracies	Noise	Security	Legal Limitations	Data Relevance
[65]			X									
[66]						X						
[67]											X	
[68]				X								
[69]									X			
[70]					X							
[71]												X
[72]												
[73]			X									
[74]					X							
[75]							X					
[76]												
[77]		X										
[78]									X			
[50]	X											
[79]			X									
[80]												X
[81]		X						X		X	X	
[82]												
[83]										X		
[84]						X						
[85]								X				
[86]	X											
[87]				X								
[88]						X						
[51]										X		

6. Conclusions

This scholarly article introduces an innovative and groundbreaking concept revolving around the application of natural language processing (NLP) algorithms to extract valuable cyber intelligence from social media messages with a focus on cyber-related matters concerning different nations. The primary emphasis lies in investigating the cyber landscape of Russia and Ukraine by employing this NLP approach to delve into four distinct cyber dimensions, namely "Geopolitical and Socioeconomic", "Targeted Victims", "Psychological & Societal", and "National Priority and Concerns." It is essential to underscore that this comprehensive framework, encompassing four-dimensional cyber intelligence, has been developed through a meticulous LDA (latent Dirichlet allocation)-based topic analysis of the Russia–Ukraine cyber war. By analyzing a vast dataset comprising 37,386 tweets originating from 30,706 distinct users in 54 different languages, encompassing the period

from 13 October 2022 to 6 April 2023, this research provides an unprecedented and detailed multilingual exploration of the Russia–Ukraine cyber crisis. It is noteworthy that prior studies have not reported the existence of an autonomous cyber intelligence framework capable of harnessing the full potential of NLP algorithms to generate insights into the complexities of the Russia–Ukraine cyber conflict. However, the accuracy of social-media-based tweet analysis was detailed in recent research works [17,22]. In light of the inaccuracies and error rates that propagate through cyber intelligence solutions solely based on social media, the strategic users should validate information from other sources (like real-time country threat statistics, as depicted in [19]).

The devised system, ingeniously adaptable to multiple platforms, including Windows, Android, and iOS, has effectively demonstrated its capacity to furnish pervasive cyber intelligence. Such an accomplishment holds immense significance, as it opens new avenues for understanding and addressing cyber challenges across diverse domains. Furthermore, this article meticulously examines and elucidates the challenges inherent in implementing this NLP system to decipher social media messages, offering an exhaustive analysis by drawing upon the existing literature. In total, 12 distinct challenges have been discerned and carefully categorized into three main classes, thereby suggesting promising future directions in the realm of social-media-based cyber intelligence utilizing NLP methodologies. These 12 challenges briefly depict the limitations of the social-media-based cyber intelligence methodology that was portrayed in this study. The most crucial limitations of these studies are not identifying information coming out of fake Twitter users [89] and misinformation (or fake information) generated by organized entities as a part of information operations [90,91]. Another limitation of this study was its reliance on third party APIs and black box algorithms and that it was eventually notable to fine tune with hyper-tuning parameterizations for optimizations. Moreover, while analyzing social media posts from multiple social media platforms (e.g., Twitter, Facebook, Instagram, etc.), the information alignment becomes a critical challenge, and this study did not provide a solution to this critical limitation. In our future studies, we endeavor to address these limitations with innovative algorithms and methodologies.

In conclusion, this paper stands as a groundbreaking contribution to the realm of cyber intelligence by effectively proposing and implementing an NLP-based approach to dissecting cyber-related social media messages of various countries. By offering a comprehensive four-dimensional framework for cyber intelligence analysis pertaining to the the Russia–Ukraine cyber-war, this study opens up new possibilities for cross-platform deployment, thereby facilitating a deeper understanding of the intricacies surrounding such cyber conflicts.

Funding: This research received no external funding.

Data Availability Statement: Data will be provided upon request.

Acknowledgments: Special thanks to the COEUS Institute, MAINE, USA, where the author works as the Chief Technology Officer.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Willett, M. The Cyber Dimension of the Russia–Ukraine War. *Survival* **2022**, *64*, 7–26. [CrossRef]
2. Lewis, J.A. Cyber War and Ukraine. 16 June 2022. Available online: https://www.csis.org/analysis/cyber-war-and-ukraine (accessed on 2 May 2023).
3. Gibney, E. Where is Russia's cyberwar? Researchers decipher its strategy. *Nature* **2022**, *603*, 775–776. [CrossRef] [PubMed]
4. Bateman, J. Russia's Wartime Cyber Operations in Ukraine: Military Impacts, Influences, and Implications. 16 December 2022. Available online: https://carnegieendowment.org/2022/12/16/russia-s-wartime-cyber-operations-in-ukraine-military-impacts-influences-and-implications-pub-88657 (accessed on 3 May 2023).
5. Pearson, J.; Bing, C. The Cyber War between Ukraine and Russia: An Overview. 10 May 2022. Available online: https://www.reuters.com/world/europe/factbox-the-cyber-war-between-ukraine-russia-2022-05-10/ (accessed on 3 May 2023).

6. Rudenko, O. Authorities: Hackers Foiled in Bid to Rig Ukraine Presidential Election Results. 2014. Available online: https://www.kyivpost.com/post/7672 (accessed on 2 April 2023).
7. BBC News. Hackers Caused Power Cut in Western Ukraine—US. 2016. Available online: https://www.bbc.com/news/technology-35297464 (accessed on 2 April 2023).
8. Banerjea, A. NotPetya: How a Russian Malware Created the World's Worst Cyberattack Ever. 27 August 2018. Available online: https://www.business-standard.com/article/technology/notpetya-how-a-russian-malware-created-the-world-s-worst-cyberattack-ever-118082700261_1.html (accessed on 2 April 2023).
9. Microsoft Security. Destructive Malware Targeting Ukrainian Organizations. 15 January 2022. Available online: https://www.microsoft.com/en-us/security/blog/2022/01/15/destructive-malware-targeting-ukrainian-organizations/ (accessed on 2 April 2023).
10. Boutilier, A.; Stephenson, M. Global Affairs Canada Suffers 'Cyber Attack' Amid Russia-Ukraine Tensions: Sources. 24 January 2022. Available online: https://globalnews.ca/news/8533835/global-affairs-hit-with-significant-multi-day-disruption-to-it-networks-sources/ (accessed on 2 April 2023).
11. Microsoft Security. ACTINIUM Targets Ukrainian Organizations. 4 February 2022. Available online: https://www.microsoft.com/en-us/security/blog/2022/02/04/actinium-targets-ukrainian-organizations/ (accessed on 2 April 2023).
12. Kovacs, E. Ukraine Separatists, Politicians Targeted in Surveillance Operation. *Security Week*. 19 May 2016. Available online: https://www.securityweek.com/ukraine-separatists-politicians-targeted-surveillance-operation/ (accessed on 3 May 2023).
13. Shamanska, A. Hackers in Ukraine Deface Separatist Websites To Mark Victory Day. *Radio Free Europe*. 9 May 2016. Available online: https://www.rferl.org/a/hackers-ukraine-deface-separatist-websites-victory-day-opmay9/27724532.html (accessed on 3 May 2023).
14. Inform Napalm. Ukrainian Hackers Break into the Russian Channel One. 6 November 2016. Available online: https://informnapalm.org/en/ru-channel-one/ (accessed on 3 May 2023).
15. Walker, S. Kremlin Puppet Master's Leaked Emails Are Price of Return to Political Frontline. *The Guardian*. 26 October 2016. Available online: https://www.theguardian.com/world/2016/oct/26/kremlin-puppet-masters-leaked-emails-vladislav-surkov-east-ukraine (accessed on 3 May 2023).
16. Pietsch, B. Hacking Group Claims Control of Belarusian Railroads in Move to 'Disrupt' Russian Troops Heading near Ukraine. *Washington Post*. 25 January 2022. Available online: https://www.washingtonpost.com/world/2022/01/25/belarus-railway-hacktivist-russia-ukraine-cyberattack/ (accessed on 3 May 2023).
17. Sufi, F. A New Social Media-Driven Cyber Threat Intelligence. *Electronics* **2023**, *12*, 1242. [CrossRef]
18. Hernandez-Suarez, A.; Sanchez-Perez, G.; Toscano-Medina, K.; Martinez-Hernandez, V.; Perez-Meana, H.; Olivares-Mercado, J.; Sanchez, V. Social Sentiment Sensor in Twitter for Predicting Cyber-Attacks Using $\ell 1$ Regularization. *Sensors* **2018**, *18*, 1380. [CrossRef]
19. Sufi, F. Algorithms in Low-Code-No-Code for Research Applications: A Practical Review. *Algorithms* **2023**, *16*, 108. [CrossRef]
20. Pattnaik, N.; Li, S.; Nurse, J.R. Perspectives of non-expert users on cyber security and privacy: An analysis of online discussions on twitter. *Comput. Secur.* **2023**, *125*, 103008. [CrossRef]
21. Geetha, R.; Karthika, S. Sensitive Keyword Extraction Based on Cyber Keywords and LDA in Twitter to Avoid Regrets. In *Computational Intelligence in Data Science, ICCIDS 2020, IFIP Advances in Information and Communication Technology, Chennai, India, 20–22 February 2020*; Springer: Berlin/Heidelberg, Germany, 2020; Volume 578.
22. Sufi, F. A New AI-Based Semantic Cyber Intelligence Agent. *Futur. Internet* **2023**, *15*, 231. [CrossRef]
23. Shah, R.; Aparajit, S.; Chopdekar, R.; Patil, R. Machine Learning based Approach for Detection of Cyberbullying Tweets. *Int. J. Comput. Appl.* **2020**, *175*, 51–56. [CrossRef]
24. Rawat, R.; Mahor, V.; Chirgaiya, S.; Shaw, R.N.; Ghosh, A. Analysis of Darknet Traffic for Criminal Activities Detection Using TF-IDF and Light Gradient Boosted Machine Learning Algorithm. In *Innovations in Electrical and Electronic Engineering: Proceedings of ICEEE 2021*; Lecture Notes in Electrical Engineering book series; Springer: Singapore, 2021; Volume 756, pp. 671–681. [CrossRef]
25. Lanier, H.D.; Diaz, M.I.; Saleh, S.N.; Lehmann, C.U.; Medford, R.J. Analyzing COVID-19 disinformation on Twitter using the hashtags #scamdemic and #plandemic: Retrospective study. *PLoS ONE* **2022**, *17*, e0268409. [CrossRef]
26. Hagen, R.A. Unraveling the Complexity of Cyber Security Threats: A Multidimensional Approach. 15 April 2023. Available online: https://www.linkedin.com/pulse/unraveling-complexity-cyber-security-threats-approach-hagen/ (accessed on 25 April 2023).
27. Correia, V.J. An Explorative Study into the Importance of Defining and Classifying Cyber Terrorism in the United Kingdom. *SN Comput. Sci.* **2021**, *3*, 84. [CrossRef]
28. Li, Y.; Liu, Q. A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments. *Energy Rep.* **2021**, *7*, 8176–8186. [CrossRef]
29. Agrafiotis, I.; Nurse, J.R.C.; Goldsmith, M.; Creese, S.; Upton, D. A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate. *J. Cybersecur.* **2018**, *4*, tyy006. [CrossRef]
30. Bhaskar, R. Better Cybersecurity Awareness through Research. 2022. Available online: https://www.isaca.org/resources/isaca-journal/issues/2022/volume-3/better-cybersecurity-awareness-through-research (accessed on 1 April 2023).
31. Humayun, M.; Niazi, M.; Jhanjhi, N.; Alshayeb, M.; Mahmood, S. Cyber Security Threats and Vulnerabilities: A Systematic Mapping Study. *Arab. J. Sci. Eng.* **2020**, *45*, 3171–3189. [CrossRef]

32. Alkhalil, Z.; Hewage, C.; Nawaf, L.; Khan, I. Phishing Attacks: A Recent Comprehensive Study and a New Anatomy. *Front. Comput. Sci.* **2021**, *3*, 563060. [CrossRef]
33. Alim, S. Analysis of Tweets Related to Cyberbullying: Exploring Information Diffusion and Advice Available for Cyberbullying Victims. *Int. J. Cyber Behav. Psychol. Learn.* **2015**, *5*, 31–52. [CrossRef]
34. Microsoft Documentation. Text Analytics: A Collection of Features from AI Language that Extract, Classify, and Understand Text within Documents. 2023. Available online: https://azure.microsoft.com/en-us/products/ai-services/text-analytics (accessed on 6 August 2023).
35. Sufi, F. Novel Application of Open-Source Cyber Intelligence. *Electronics* **2023**, *12*, 3610. [CrossRef]
36. Sufi, F.K.; Khalil, I. Automated Disaster Monitoring from Social Media Posts Using AI-Based Location Intelligence and Sentiment Analysis. *IEEE Trans. Comput. Soc. Syst.* **2022**, *in press*. [CrossRef]
37. Sufi, F.K. AI-SocialDisaster: An AI-based software for identifying and analyzing natural disasters from social media. *Softw. Impacts* **2022**, *11*, 100319. [CrossRef]
38. Sufi, F.K.; Alsulami, M. Automated Multidimensional Analysis of Global Events with Entity Detection, Sentiment Analysis and Anomaly Detection. *IEEE Access* **2021**, *9*, 152449–152460. [CrossRef]
39. Sufi, F.K. AI-GlobalEvents: A Software for analyzing, identifying and explaining global events with Artificial Intelligence. *Softw. Impacts* **2022**, *11*, 100218. [CrossRef]
40. Pang, B.; Lee, L.; Vaithyanathan, S. Thumbs up?: Sentiment classification using machine learning techniques. *arXiv* **2002**, arXiv:0205070.
41. Turney, P.D. Thumbs up or thumbs down?: Semantic orientation applied. *arXiv* **2002**, arXiv:0212032.
42. Naseem, U.; Razzak, I.; Khushi, M.; Eklund, P.W.; Kim, J. COVIDSenti: A Large-Scale Benchmark Twitter. *IEEE Trans. Comput. Soc. Syst.* **2020**, *8*, 1003–1015. [CrossRef]
43. Li, L.; Zhang, Q.; Wang, X.; Zhang, J.; Wang, T.; Gao, T.-L.; Duan, W.; Tsoi, K.K.-F.; Wang, F.-Y. Characterizing the Propagation of Situational Information in Social Media During COVID-19 Epidemic: A Case Study on Weibo. *IEEE Trans. Comput. Soc. Syst.* **2020**, *7*, 556–562. [CrossRef]
44. Cameron, D.; Smith, G.A.; Daniulaityte, R.; Sheth, A.P.; Dave, D.; Chen, L.; Anand, G.; Carlson, R.; Watkins, K.Z.; Falck, R. PREDOSE: A semantic web platform for drug abuse epidemiology using social media. *J. Biomed. Inform.* **2013**, *46*, 985–997. [CrossRef] [PubMed]
45. Chen, X.; Faviez, C.; Schuck, S.; Lillo-Le-Louët, A.; Texier, N.; Dahamna, B.; Huot, C.; Foulquié, P.; Pereira, S.; Leroux, V.; et al. Mining Patients' Narratives in Social Media for Pharmacovigilance: Adverse Effects and Misuse of Methylphenidate. *Front. Pharmacol.* **2018**, *9*, 541. [CrossRef]
46. McNaughton, E.C.; Black, R.A.; Zulueta, M.G.; Budman, S.H.; Butler, S.F. Measuring online endorsement of prescription opioids abuse: An integrative methodology. *Pharmacoepidemiol. Drug Saf.* **2012**, *21*, 1081–1092. [CrossRef]
47. Al-Twairesh, N.; Al-Negheimish, H. Surface and Deep Features Ensemble for Sentiment Analysis of Arabic Tweets. *IEEE Access* **2019**, *7*, 84122–84131. [CrossRef]
48. Vashisht, G.; Sinha, Y.N. Sentimental study of CAA by location-based tweets. *Int. J. Inf. Technol.* **2021**, *13*, 1555–1567. [CrossRef]
49. Ebrahimi, M.; Yazdavar, A.H.; Sheth, A. Challenges of Sentiment Analysis for Dynamic Events. *IEEE Intell. Syst.* **2017**, *32*, 70–75. [CrossRef]
50. Evangelatos, P.; Iliou, C.; Mavropoulos, T.; Apostolou, K.; Tsikrika, T.; Vrochidis, S.; Kompatsiaris, I. Named Entity Recognition in Cyber Threat Intelligence Using Transformer-based Models. In Proceedings of the 2021 IEEE International Conference on Cyber Security and Resilience (CSR), Rhodes, Greece, 26–28 July 2021. [CrossRef]
51. Wu, H.; Li, X.; Gao, Y. An Effective Approach of Named Entity Recognition for Cyber Threat Intelligence. In Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 12–14 June 2020. [CrossRef]
52. Batbaatar, E.; Ryu, K.H. Ontology-Based Healthcare Named Entity Recognition from Twitter Messages Using a Recurrent Neural Network Approach. *Int. J. Environ. Res. Public Health* **2019**, *16*, 3628. [CrossRef]
53. Khandpur, R.P.; Ji, T.; Jan, S.; Wang, G.; Lu, C.-T.; Ramakrishnan, N. Crowdsourcing Cybersecurity: Cyber Attack Detection using Social Media. In Proceedings of the CIKM '17: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, Singapore, 6–10 November 2017.
54. Koloveas, P.; Chantzios, T.; Alevizopoulou, S.; Skiadopoulos, S.; Tryfonopoulos, C. inTIME: A Machine Learning-Based Framework for Gathering and Leveraging Web Data to Cyber-Threat Intelligence. *Electronics* **2021**, *10*, 818. [CrossRef]
55. Shin, H.-S.; Kwon, H.-Y.; Ryu, S.-J. A New Text Classification Model Based on Contrastive Word Embedding for Detecting Cybersecurity Intelligence in Twitter. *Electronics* **2020**, *9*, 1527. [CrossRef]
56. Zhao, J.; Yan, Q.; Li, J.; Shao, M.; He, Z.; Li, B. TIMiner: Automatically extracting and analyzing categorized cyber threat intelligence from social data. *Comput. Secur.* **2020**, *95*, 101867. [CrossRef]
57. Schellekens, J. Release the bots of war: Social media and Artificial Intelligence as international cyber attack. *Przegląd Eur.* **2021**, *4*, 163–179. [CrossRef]
58. Sun, N.; Zhang, J.; Gao, S.; Zhang, L.Y.; Camtepe, S.; Xiang, Y. Data Analytics of Crowdsourced Resources for Cybersecurity Intelligence. In *Network and System Security: 14th International Conference, NSS 2020, Melbourne, VIC, Australia, 25–27 November 2020, Proceedings 14*; Springer International Publishing: New York, NY, USA, 2020; Volume 12570, pp. 3–21.

59. Subroto, A.; Apriyana, A. Cyber risk prediction through social media big data analytics and statistical machine learning. *J. Big Data* **2019**, *6*, 50. [CrossRef]
60. Oosthoek, K.; Doerr, C. Cyber Threat Intelligence: A Product Without a Process? *Int. J. Intell. Counterintelligence* **2021**, *34*, 300–315. [CrossRef]
61. Van Hee, C.; Jacobs, G.; Emmery, C.; Desmet, B.; Lefever, E.; Verhoeven, B.; De Pauw, G.; Daelemans, W.; Hoste, V. Automatic detection of cyberbullying in social media text. *PLoS ONE* **2018**, *13*, e0203794. [CrossRef]
62. Paradise, A.; Shabtai, A.; Puzis, R.; Elyashar, A.; Elovici, Y.; Roshandel, M.; Peylo, C. Creation and Management of Social Network Honeypots for Detecting Targeted Cyber Attacks. *IEEE Trans. Comput. Soc. Syst.* **2017**, *4*, 65–79. [CrossRef]
63. Carley, K.M. Social cybersecurity: An emerging science. *Comput. Math. Organ. Theory* **2020**, *26*, 365–381. [CrossRef]
64. Yuvaraj, N.; Srihari, K.; Dhiman, G.; Somasundaram, K.; Sharma, A.; Rajeskannan, S.; Soni, M.; Gaba, G.S.; AlZain, M.A.; Masud, M. Nature-Inspired-Based Approach for Automated Cyberbullying Classification on Multimedia Social Networking. *Math. Probl. Eng.* **2021**, *2021*, 6644652. [CrossRef]
65. Shu, K.; Sliva, A.; Sampson, J.; Liu, H. Understanding Cyber Attack Behaviors with Sentiment Information on Social Media. In *Social, Cultural, and Behavioral Modeling: 11th International Conference, SBP-BRiMS 2018, Washington, DC, USA, 10–13 July 2018, Proceedings 11*; Springer International Publishing: New York, NY, USA, 2018; Volume 10899, pp. 377–388.
66. Sliva, A.; Shu, K.; Liu, H. Using Social Media to Understand Cyber Attack Behavior. In *Advances in Human Factors, Business Management and Society: Proceedings of the AHFE 2018 International Conference on Human Factors, Business Management and Society, Orlando, FL, USA, 21–25 July 2018*; Springer: New York, NY, USA, 2019; Volume 783, pp. 636–645. [CrossRef]
67. Du, Y.; Huang, C.; Liang, G.; Fu, Z.; Li, D.; Ding, Y. ExpSeeker: Extract public exploit code information from social media. *Appl. Intell.* **2022**, *53*, 15772–15786. [CrossRef]
68. Alves, F.; Bettini, A.; Ferreira, P.M.; Bessani, A. Processing tweets for cybersecurity threat awareness. *Inf. Syst.* **2020**, *95*, 101586. [CrossRef]
69. Mughaid, A.; Al-Zu'bi, S.; AL Arjan, A.; Al-Amrat, R.; Alajmi, R.; Abu Zitar, R.; Abualigah, L. An intelligent cybersecurity system for detecting fake news in social media websites. *Soft Comput.* **2022**, *26*, 5577–5591. [CrossRef] [PubMed]
70. Fang, Y.; Gao, J.; Liu, Z.; Huang, C. Detecting Cyber Threat Event from Twitter Using IDCNN and BiLSTM. *Appl. Sci.* **2020**, *10*, 5922. [CrossRef]
71. Tundis, A.; Ruppert, S.; Mühlhäuser, M. On the Automated Assessment of Open-Source Cyber Threat Intelligence Sources. In Proceedings of the Computational Science—ICCS 2020, Amsterdam, The Netherlands, 3–5 June 2020; Volume 12138. [CrossRef]
72. Sangwan, S.R.; Bhatia, M.P.S. Soft computing for abuse detection using cyber-physical and social big data in cognitive smart cities. *Expert Syst.* **2021**, *39*, e12766. [CrossRef]
73. Jacobs, G.; Van Hee, C.; Hoste, V. Automatic classification of participant roles in cyberbullying: Can we detect victims, bullies, and bystanders in social media text? *Nat. Lang. Eng.* **2022**, *28*, 141–166. [CrossRef]
74. Sánchez, J.R.; Campo-Archbold, A.; Rozo, A.Z.; Díaz-López, D.; Pastor-Galindo, J.; Mármol, F.G.; Díaz, J.A. Uncovering Cybercrimes in Social Media through Natural Language Processing. *Complexity* **2021**, *2021*, 7955637. [CrossRef]
75. Ho, S.M.; Li, W. "I know you are, but what am I?" Profiling cyberbullying based on charged language. *Comput. Math. Organ. Theory* **2022**, *28*, 293–320. [CrossRef]
76. Rezvan, M.; Shekarpour, S.; Alshargi, F.; Thirunarayan, K.; Shalin, V.L.; Sheth, A. Analyzing and learning the language for different types of harassment. *PLoS ONE* **2020**, *15*, e0227330. [CrossRef]
77. De Boer, M.H.T.; Bakker, B.J.; Boertjes, E.; Wilmer, M.; Raaijmakers, S.; van der Kleij, R. Text Mining in Cybersecurity: Exploring Threats and Opportunities. *Multimodal Technol. Interact.* **2019**, *3*, 62. [CrossRef]
78. Mendhurwar, S.; Mishra, R. Integration of social and IoT technologies: Architectural framework for digital transformation and cyber security challenges. *Enterp. Inf. Syst.* **2019**, *15*, 565–584. [CrossRef]
79. Basheer, R.; Alkhatib, B. Threats from the Dark: A Review over Dark Web Investigation Research for Cyber Threat Intelligence. *J. Comput. Netw. Commun.* **2021**, *2021*, 1302999. [CrossRef]
80. Mittal, S.; Das, P.K.; Mulwad, V.; Joshi, A.; Finin, T. CyberTwitter: Using Twitter to generate alerts for cybersecurity threats and vulnerabilities. In Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, CA, USA, 18–21 August 2016.
81. Thakur, K.; Hayajneh, T.; Tseng, J. Cyber Security in Social Media: Challenges and the Way Forward. *IT Prof.* **2019**, *21*, 41–49. [CrossRef]
82. Rodriguez, A.; Okamura, K. Social Media Data Mining for Proactive Cyber Defense. *J. Inf. Process.* **2020**, *28*, 230–238. [CrossRef]
83. Le, B.-D.; Wang, G.; Nasim, M.; Babar, M.A. Gathering Cyber Threat Intelligence from Twitter Using Novelty Classification. *arXiv* **2019**, arXiv:1907.01755.
84. Maisano, R.; Foresti, G.L. A Sentiment Analysis Anomaly Detection System for Cyber Intelligence. *Int. J. Neural Syst.* **2022**, *33*, 2350003. [CrossRef]
85. Lau, R.Y.K.; Xia, Y.; Ye, Y. A Probabilistic Generative Model for Mining Cybercriminal Networks from Online Social Media. *IEEE Comput. Intell. Mag.* **2014**, *9*, 31–43. [CrossRef]
86. Alevizopoulou, S.; Koloveas, P.; Tryfonopoulos, C.; Raftopoulou, P. Social Media Monitoring for IoT Cyber-Threats. In Proceedings of the 2021 IEEE International Conference on Cyber Security and Resilience (CSR), Rhodes, Greece, 26–28 July 2021.

87. Syed, R. Cybersecurity vulnerability management: A conceptual ontology and cyber intelligence alert system. *Inf. Manag.* **2020**, *57*, 103334. [CrossRef]
88. Lima, A.Q.; Keegan, B. Chapter 3—Challenges of using machine learning algorithms for cybersecurity: A study of threat-classification models applied to social media communication data. In *Cyber Influence and Cognitive Threats*; Academic Press: Cambridge, MA, USA, 2020; pp. 33–52.
89. Chen, B.; Chen, X. MAUIL: Multi-level Attribute Embedding for Semi-supervised User Identity Linkage. *Inf. Sci.* **2022**, *593*, 527–545. [CrossRef]
90. Zannettou, S.; Caulfield, T.; Bradlyn, B.; De Cristofaro, E.; Stringhini, G.; Blackburn, J. Characterizing the Use of Images in State-Sponsored Information Warfare Operations by Russian Trolls on Twitter. In Proceedings of the International AAAI Conference on Web and Social Media, Atlanta, GA, USA, 1–5 June 2020; Volume 14. [CrossRef]
91. Zannettou, S.; Caulfield, T.; De Cristofaro, E.; Sirivianos, M.; Stringhini, G.; Blackburn, J. Disinformation Warfare: Understanding State-Sponsored Trolls on Twitter and Their Influence on the Web. *arXiv* **2019**, arXiv:1801.09288. [CrossRef]