Article

Not peer-reviewed version

Developing Chatbots for Cyber Security: Assessing Threats through Sentiment Analysis on Social Media

Amit Arora * , Anshu Arora , John McIntyre

Posted Date: 3 August 2023

doi: 10.20944/preprints202308.0329.v1

Keywords: chatbot; cyber security; artificial intelligence; threats; vulnerability; data manipulation; social media; sentiment analysis

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 August 2023 doi:10.20944/preprints202308.0329.v1

Article

Developing Chatbots for Cyber Security: Assessing Threats through Sentiment Analysis on Social Media

Amit Arora [1,*], Anshu Arora [1] and John McIntyre [2]

[1] University of the District of Columbia; amit.arora@udc.edu
[2] Georgia Institute of Technology; john.mcintyre@scheller.gatech.edu
* Correspondence: amit.arora@udc.edu

Abstract: In recent years, groups of cyber criminals/hackers have carried out cyber-attacks using various tactics with the goal of destabilizing web services in a specific context for which they are motivated. Predicting these attacks is a critical task that assists in determining what actions should be taken to mitigate the effects of such attacks and to prevent them in the future. Although there are programs to detect security concerns on the internet, there is currently no system that can anticipate or foretell whether the attacks will be successful. This research aims to develop sustainable strategies to reduce threats, vulnerability, and data manipulation of chatbots, consequently improving cyber security. To achieve this goal, we develop a conversational chatbot, an application that uses artificial intelligence (AI) to communicate, and deploy it on social media sites (e.g., Twitter) for cyber security purposes. Chatbots have the capacity to consume large amounts of information and give an appropriate response in an efficient and timely manner, thus rendering them useful in predicting threats emanating from social media. The research utilizes sentiment analysis strategy by employing chatbots on Twitter (and analyzing Twitter data) for predicting future threats and cyber-attacks. The strategy is based on a daily collection of tweets from two types of users: those who use the platform to voice their opinions on important and relevant subjects, and those who use it to share information on cyber security attacks. The research provides tools and strategies for developing chatbots that can be used for assessing cyber threats on social media through sentiment analysis leading to a global sustainable development of businesses. Future research may utilize and improvise on the tools and strategies suggested in our research to strengthen the knowledge domain of chatbots, cyber security, and social media.

Keywords: chatbot; cyber security; artificial intelligence; threats; vulnerability; data manipulation; social media

---

1. Introduction

Cybersecurity is a multidisciplinary field and can have far-reaching economic, environmental, and social consequences [1–3]. Cybersecurity statistics indicate that there are 2,200 cyber-attacks per day, with a cyber-attack happening every 39 seconds on average. In the US, a single data breach costs an average of $9.44M, and cybercrime is predicted to cost $8 trillion in 2023 [4]. As governments and businesses become more reliant on new communication technologies and social media, the threat of cyber-attacks on such organizations has increased tremendously. To counter such threats, governments and businesses have increased their investments in cybersecurity [5]. Advances in Natural Language Processing (NLP) and Machine Learning (ML) techniques have led to chatbots (also known as conversational agents) becoming capable of extracting meaningful information regarding cybersecurity threats [6] on social media. Rapid deployment of artificial intelligence (AI) coupled with digitalization of a globalized economy has produced a vast amount of textual data through social media. Chatbot applications along with technology-enabled solutions lead to a sustainable development of global businesses and economies. Governments, businesses, and political parties depend on the sentiments and opinions expressed on social media sites to gauge the mood of public in real time [7]. This is also a vital source of information related to security threats to a nation and business organizations. Consequently, it becomes imperative for intelligence and security communities to delve deeper into cybersecurity to protect national security and economic interests.

 doi:10.20944/preprints202308.0329.v1

Social networks on the internet have enabled people to interact with each other in real-time. Microblogging platforms, such as Twitter, have emerged as the most popular communication tool since it allows a wide variety of expressions, such as interactive short texts, pictures, emojis, etc., with relative ease [8,9]. Such platforms act as a social public square where users express their feelings, sentiments, ideas, and opinions on wide ranging topics. Research has shown that analyzing these feelings and sentiments expressed on social networks and platforms is an effective way to forecast a variety of events such as market trends, election results, brand image, etc. [8,10]. Sentiment analysis can be performed quickly on a large amount of textual data available on social platforms. However, there is dearth of research to assess sentiments in detecting probable cybersecurity threats.

Previous research has shown that the best practice to combat threats in cyber security is to develop strategies that are complementary to each specific threat [11]. In this research we develop strategies to reduce threats, vulnerabilities, and data manipulation of chatbots, consequently improving cyber security. Specifically, we develop a chatbot on Bot Libre, an open-source platform, and deploy it on Twitter. The research also focuses on sentiment analysis of tweets generated by Twitter users conversing with our developed chatbot.

The objectives of this research are to 1) survey the existing state-of-the-art multilingual Chatbot tools, 2) develop and test this Chatbot Testbed on Twitter, 3) conduct sentiment analysis of textual data generated through tweets, and 4) create documentation and materials so that this toolbox can be used by a variety of users with sustainable development goals. The Chatbot Testbed was created by integrating existing open-source and commercial tools to effectively create a solution that is usable for understanding problems in influence, information operations, and insider threat.

2. Background and Related Work

Chatbot is an application that uses artificial intelligence (AI) to communicate. Artificial intelligence is the automation of intelligent behavior which allows machines to simulate anthropomorphic conversations. Chatbots has been programmed to use artificial intelligence and concepts such as Natural Language Processing (NLP), Artificial Intelligence Markup Language (AIML), Pattern Matching, Chat Script, and Natural Language Understanding (NLU) to communicate with users, analyze the conversation and use the extracted data for marketing, personal content, to target specific groups, etc. The knowledge domain, the service provided, the goals, the input processing, and response generation method, the human-aid, and the build method are some of the categories under which chatbots can be classified.

The knowledge domain classification considers the knowledge a chatbot can access as well as the amount of data it is trained on. Closed domain chatbots are focused on a certain knowledge subject and may fail to answer other questions, but open domain chatbots can talk about various topics and respond effectively [12]. Conversely, the sentimental proximity of the chatbot to the user, the quantity of intimate connection, and chatbot performance are factors in the classification of chatbots based on the service provided. Interpersonal chatbots are in the communication area and offer services such as restaurant reservations, flight reservations, and FAQs. They gather information and pass it on to the user, but they are not the user's companions. They are permitted to have a personality, be nice, and recall information about the user, however, they are not required or expected to do so [12]. [12] (pp. 373-383) states that "Intrapersonal chatbots exist within the personal domain of the user, such as chat apps like Messenger, Slack, and WhatsApp. They are companions to the user and understand the user like a human does. Inter-agent chatbots become omnipresent while all chatbots will require some inter-chatbot communication possibilities. The need for protocols for inter-chatbot communication has already emerged. Alexa-Cortana integration is an example of inter-agent communication".

Informative chatbots, such as FAQ chatbots, are designed to offer the user information that has been stored in advance or is available from a fixed source. The manner of processing inputs and creating responses is taken into consideration when classifying based on input processing and response generation. The relevant replies are generated using one of three models: rule-based, retrieval-based, and generative. Another classification for chatbots is based on how much human-aid is included in its components. Human computation is used in at least one element of a human-aid chatbot. To address the gaps produced by the constraints of completely automated chatbots, crowd workers, freelancers, or full-time employees can incorporate their intelligence in the chatbot logic. [12] (pp. 373-383) examines the main classification of chatbots as per the development platform permissions, where the authors defined 'development platforms' as "…open-source, such as RASA, or can be of proprietary code such as development platforms typically offered by large companies such as Google or IBM."

Two of the main categories that chatbots may fall into as it relates to their anthropomorphic characteristics are Error-free and Clarification chatbot. Anthropomorphism is "the attribution of human characteristics or traits to nonhuman agents" [13] (p. 865). Anthropomorphic chatbots are perceived to be more palatable to consumers since consumers perceive the chatbots to be humanlike, rather than how firms design chatbots as humanlike [14]. An Error-free chatbot can be defined as a hypothetically flawless chatbot while a clarification chatbot has difficulties inferring meaning and therefore asks for clarification from the user. Clarification chatbots are seen as more anthropomorphic since clarification by the chatbot is seen as giving care and attention to the needs of the customer. According to [15], "The error-free chatbot offers no indication that it is anything but human. It correctly interprets all human utterances and responds with relevant and precise humanlike utterances of its own." On the first parse, the clarification chatbot does not have the intelligence to accurately interpret all human utterances. The chatbot, on the other hand, is clever enough to identify the root of the misunderstanding, referred to as a difficulty source, and request an explanation. Since seeking clarification is a normal element of interpersonal communication, clarification chatbots' anthropomorphic characteristics increase with their ability to recognize a problem source and display intersubjective effort. There is no current commercial application of the error-free chatbot, however, clarification chatbots are currently being used by companies such as Amazon, Walmart, T-Mobile, Bank of America, and Apple, as first contact customer service representatives.

Threats and vulnerability are key factors (and dangers) affecting the cyber security of chatbots. Cyber threats can be characterized as methods in which a computer system can be hacked. Spoofing, tampering, repudiation, information disclosure, denial of service, privilege elevation, and other threats are examples of chatbot threats. Conversely, vulnerabilities are ways in which a system can be harmed that aren't appropriately mitigated. When a system is not effectively maintained, has bad coding, lacks protection, or is subject to human mistake, it becomes vulnerable and accessible to assaults. Self-destructive messages can be used in conjunction with other security measures such as end-to-end encryption, secure protocol, user identity authentication, and authorization to reduce vulnerabilities. Another method to ensure the security of chatbots is the use of User Behavioral Analytics (UBA).

A vulnerability is defined as a weakness in a system's security protocols, internal controls, or implementation that could be exploited or activated by a threat source. Secure Development Lifecycle refers to the process of incorporating security components into the Software Development Lifecycle (SDLC). SDLC, on the other hand, is a thorough plan that outlines how companies construct applications from conception through decommission. According to [11], implementing Security Development Lifecycle (SDL) related activities into the development lifecycle is one of the most effective ways to mitigate vulnerabilities. Planning and needs, testing the code and outcomes, architecture and design, test planning, and coding are phases commonly followed by all models for Secure Development Lifecycle. This reduces the vulnerabilities and openness to attacks. User Behavioral Analytics (UBA) is a method of analyzing user activity patterns through the use of software applications. It allows them to use advanced algorithms and statistical analysis to spot any unusual behavior that could be a security risk. The use of this analytical software will allow for easy identification of other bots being used to infiltrate a secure system through hacking. Hence this reduces the risk of a cyber-attack.

As previously mentioned, cyber threats can be characterized as methods in which a computer system can be hacked. Spoofing, tampering, repudiation, information disclosure, denial of service, privilege elevation, are examples of threats. To reduce the impacts of these threats, specific approaches need to be taken for each particular threat. Spoofing is done to gain information and use it for the impersonation of something or someone else. To abate this, correct authentication such as a strong password is required to secure sensitive data. Tampering is a threat where the hacker aims to maliciously modify data. Here, the mitigation strategy is to use digital signatures, audit trails, Network Time Protocol, and log timestamps. Denial of Service is another category of threats in which the attacker intends to deny access to valid users. In this instance, the best strategies to reduce this threat are filtering and throttling [11].

As of December 2022, Twitter had 368 million monthly active users worldwide (statista.com/statistics/303681/twitter-users-worldwide), providing a chance to gather a large amount of data in near-real time. In this research, we focus on the development and deployment of a chatbot on a social media platform in order collect a large sample of textual data in the form of tweets and perform sentiment analysis using algorithmic techniques to forecast certain threats and vulnerabilities related to cybersecurity.

3. Methods and Data Analysis

This section focuses on the two main aspects of the research: a) development and deployment of a conversational chatbot on a social media site; and b) conducting sentiment analysis on the vast amount of textual data from a social media site.

3.1. Development and Deployment of Chatbot

Initially, the project team focused on building a chatbot on SAP open-source platform. However, it is hard to use SAP Conversational AI chatbot outside SAP S/4 Hana cloud. After considering other open-source platforms like Botpress, our conversational chatbot was developed on Bot Libre, an open source end-to-end chatbot-building platform. It can be used to build, train, connect, and monitor the chatbot on a social media site. Bot Libre chatbot uses both text and images and is categorized as Communication Channels chatbot [11,12]. This platform allows the chatbot to be deployed on various social media sites like Twitter, WhatsApp, Facebook, Discord, Kik, etc. The language modeling, which is a part of personalizing how the bot communicates with specific users allows the bot to interact with users in multiple languages, can be tailored to include English, French, Russian Spanish, Italian, Japanese, among other languages.

Currently, our chatbot can converse in English language only on Twitter platform. The automation feature allows the bot to tweet over an extensive period. For example, in the month of March, the chatbot was programmed to tweet “Happy Women’s History Month” every 24 hours. It utilizes the ‘conversational feature’ by initiating and maintaining conversations with other users of Twitter. Its ‘Informational Effect’ and ‘Data Effect’ are highlighted by its ability to collect data from conversations it has with other users as well as extract information from the platform based on key terms searched. For example, the chatbot can search the key terms “Putin”, “Nuclear Weapon” and “Russia” and extract all tweets associated with these key terms. The goal of the chatbot is to communicate and extract information/intelligence from users on Twitter which can be used by intelligence and security communities. Any keyword that can be deemed as a threat (e.g., hate speech, defense related, etc.) can be searched on Twitter platform using the chatbot. The information is collected using the Application Programming Interface (API) keys. This monitoring of information on a social media platform will aid in cyber security within the United States. The analytics feature of Bot Libre platform can provide useful information about chat conversations conducted by the chatbot during a specific day, week, month, or any specified period. Figure 1 illustrates the analytics feature of Bot Libre platform. Data that can be analyzed includes conversations, messages, conversation length, response time, connects, chats, errors, etc.

 doi:10.20944/preprints202308.0329.v1

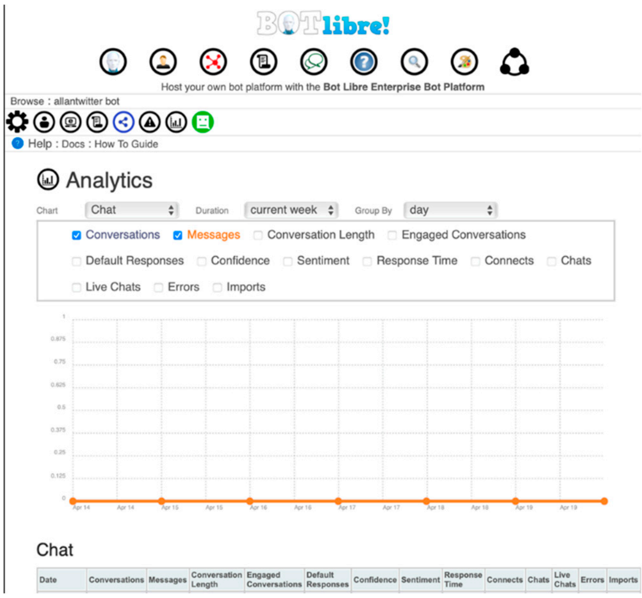

Figure 1. Snapshot of analytics feature of chatbot developed on Bot Libre platform.

Next, the project team focused on conducting sentiment analysis on the vast amount of textual data collected from Twitter.

3.2. Sentiment Analysis

Previous research has shown that written text on social media sites is impacted by the emotions, intentions, and thoughts of the user [16,17]. Thus, written text is a useful source of information about the user. This section describes the process of data collection, cleaning, and analysis in detail.

3.1.1. Data Collection

First, we discuss the collection of data. The Twitter API academic research access was used to collect global, real-time, and historical textual data in the form of tweets. The collected tweets were processed in JSON and added to corpus *C* [8,18], identified as:

$$\mathcal{C} = c_i \in \{tweet_{id_i}, tweet_{text_i}, tweet_{date_i}, tweet_{language_i}\}_{i=1}^{n} \quad \forall\, t_{i,4} = en \tag{1}$$

c_i represents the i^{th} tweet in the corpus. Each tweet is identified by four objects: id, text, date, and language. *C* is stored locally in MySQL, a relational database. A primary key, $tweet_{id}$, is allocated to each tweet in *C*, which is utilized to identify unprocessed tweets.

The code for Twitter API credentials and extraction of tweets is given in the appendix.

3.1.2. Data Pre-Processing

Next step is data pre-processing which consists of speech tagging and noise removal. Since the tweets posted by users on Twitter platform are informal in nature, the raw textual data in the form of tweets tends to have grammatical errors, may be unstructured, and generally considered noisy. This could potentially make it difficult to interpret the data correctly. Hence, the data should be pre-processed before analyzing and determining the user's sentiment. Speech tagging is a process of assigning a specific tag to each word in the corpus of tweets. These tags divide the textual data, e.g., tweets, into verbs, adverbs, nouns, adjectives, etc. which can be used as potential markers to determine the polarity (or sentiment) of tweets. The polarity (or sentiment) can be positive, neutral, or negative. Exclamation marks, question marks, and emoticons are also considered for determining

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 August 2023 doi:10.20944/preprints202308.0329.v1

polarity. The tags utilized in this study are an optimized version of *Penn Treebank's* compendium [19]. Below is an example of speech tagging of a sample of tweet [8].

$$\underbrace{I}_{O}\ \underbrace{hate}_{V}\ \underbrace{\#ISIS}_{\wedge}\ \underbrace{and}_{\&}\ \underbrace{everything}_{N}\ \underbrace{that}_{P}\ \underbrace{they}_{O}\ \underbrace{stand}_{V}\ \underbrace{for}_{P}\ \underbrace{.}_{,}$$

Sample	Tag	Attribute
I	O	pronoun (personal / no possessive)
hate	V	copulative verb
#ISIS	^	proper name
and	&	coordinating conjunction
everything	N	common noun
that	P	preposition
they	O	particle verb
stand	V	copulative verb
for	P	preposition
,	,	punctuation

Figure 2. Example of speech tagging of a tweet.

Next, we focus on removal of noise from data. Sometimes tweets may include text and other markers that seem to be unrelated to the expressed sentiment, which need to be removed before data analysis. Noise in twitter data may be in the form of URLs, replies to other users, retweets, and common stop words which do not add value to the meaning of a tweet [8,20]. To eliminate these occurrences, a noise removal procedure is used. The code used for this purpose is given in Appendix A.

3.1.3. Sentiment Extraction

The process of determining the emotional tone (positive, negative, or neutral) in a body of text is commonly referred to as sentiment extraction. This is done by locating and identifying candidate markers written by most users in their tweets. To achieve this, the Apriori algorithm is utilized. This is an algorithm for mining frequent item sets and learning association rules [8,21]. If a collection of textual data contains a minimum of 1% support as the frequency of occurrence, it is classified as featuring frequent markers [8,9].The goal of association rules is to uncover latent words that can be used as frequent markers.

$$\Psi^{f} = \{\psi \colon \psi \in \Psi \wedge minSupport(\psi) \tag{2}$$

According to the methods presented in [8,22], candidate markers that the Apriori algorithm cannot identify are removed.

The National Initiative for Cybersecurity Careers and Studies (NICCS) lexicon dictionary [23] δ for information security is used as a reference to identify tweets related to cybersecurity. This dictionary contains key cybersecurity terms which provide a comprehensive understanding of definitions/terminology pertaining to cybersecurity [8,24].

The following is computed [8,25] to acquire the samples that contain words in δ:

$$\mathcal{H}_i = \{\Psi_{i,1}, \Psi_{j,2}\} \text{ if } \delta_i(\Psi_{j,2}) > 1 \tag{3}$$

Or the following:

$$\mathcal{B}_i = \{\Psi_{i,1}, \Psi_{j,2}\} \tag{4}$$

where $\mathcal{H}$ is the collection of samples that contain at least one word from δ and $\mathcal{B}$ is the collection of the remaining samples, i.e., tweets that do not contain specific content about security issues:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 August 2023 doi:10.20944/preprints202308.0329.v1

$$\mathcal{H} = \{\psi : \psi \in \Psi \wedge \delta(\psi) > 1\} \quad (5)$$

$$\mathcal{B} = \{\psi : \psi \in \Psi \wedge \delta(\psi) < 1\} \quad (6)$$

3.1.4. Sentiment Orientation and Analysis

Sentiment orientation stage is carried out by analyzing the frequent markers in Ψ_f in $\mathcal{H}_i$ and $\mathcal{B}_i$ where i is the i'th sample. The polarity is determined by scores previously defined in the SentiWordnet compendium [8,26]. SentiWordnet is a lexical resource compendium for opinion mining. Each word set consists of three sentiments: positive, negative, and neutral [8,27]. These sentiments are based on relationships and associations between words such as antonyms, synonyms, and hyponyms. These are used to develop certain rules to identify the polarity (or sentiment) of the text in consideration [8,28].

The findOrientation algorithm is used to identify tweets having a negative orientation, with $\mathcal{H}$ and $\mathcal{B}$ being the only compendiums containing negative markers.

The algorithm is given in Appendix A.

The $\mathcal{H}^1$ and $\mathcal{B}^1$ compendiums obtained using the findOrientation algorithm are linked to their respective primary key $tweet_{idi}$ of corpus *C*, which is utilized to establish the original tweet's creation date, $tweet_{datei}$.

The tweets are sorted by date and then combined to get a daily total score for $\mathcal{H}^1$ and $\mathcal{B}^1$ as shown below [8,29]:

$$\sum_{p=1}^{n} \mathcal{H}', \sum_{p=1}^{n} \mathcal{B}' \quad (7)$$
$$n = number\ of\ tweets\ per\ day$$

Sentiment analysis was performed on a sample of Twitter text. Google Colaboratory was used as our platform for machine learning specific code in the Python language. The consumer key, consumer secret, access token, access token secret and bearer token were downloaded from the Twitter project account with academic access and stored in a .csv file. These are necessary to give permission to retrieve the tweets needed for our analysis. The Tweepy Python library was imported for reducing the amount of code that it takes to perform certain actions, such as authentication, to allow access to the Tweets from the internal Twitter database.

We used TextBlob (https://textblob.readthedocs.io/en/dev/), a Python library, for processing and analyzing the Twitter data. It provides an open-source API for speech tagging, sentiment analysis, etc. A text message or a tweet is a collection of words. Each word has its own intensity and semantic orientation that defines its sentiment. Overall sentiment of a text is calculated by taking the average of the sentiments of all words in the text. Sentiment is a function of polarity and subjectivity. Polarity of a text is a floating value between [-1,1] where -1 represents a highly negative sentiment and 1 represents a highly positive sentiment. 0 value indicates a neutral sentiment. Subjectivity of a text is a floating value between [0,1] where 0 represents least subjective and highly factual text, while 1 represents most subjective and least factual text. Subjectivity is quantified as a measure of the amount of personal opinion vs factual information in a text. TextBlob supports this complex analysis on text data and returns both polarity and subjectivity of a text.

The code used to classify subjectivity and polarity, and to visualize the words of a tweet in the form of a word cloud is given in Appendix A. Word cloud is based on the frequency of words in a text (for e.g., a collection of tweets). Also included in Appendix A is the code of a scatter plot of subjectivity vs polarity of tweets.

Figure 3 illustrates an example of a word cloud created from the most prominent words from the Twitter text data.

 doi:10.20944/preprints202308.0329.v1

Figure 3. Word cloud created from prominent words in tweets.

Figures 4 and 5 illustrate an example of a scatter plot created with subjectivity and polarity of a sample of tweets, and a bar graph representing the count of neutral, positive, and negative tweets. The area of interest from cybersecurity point of view is the upper left quadrant in Figure 4 which represents specific tweets high on subjectivity and having a negative polarity. These may be referred to as outliers in the data which deserve a deeper analysis.

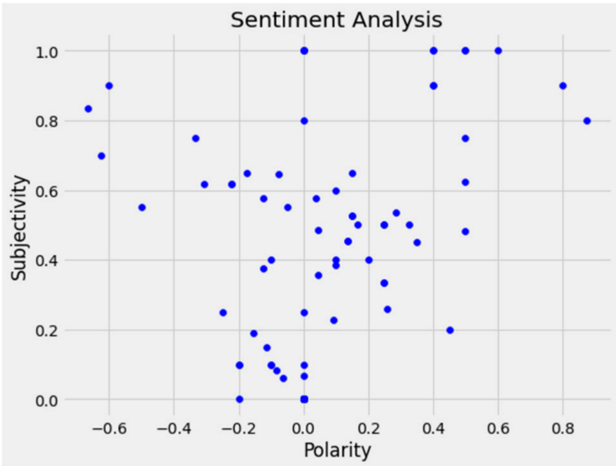

Figure 4. Subjectivity vs Polarity of a sample of tweets.

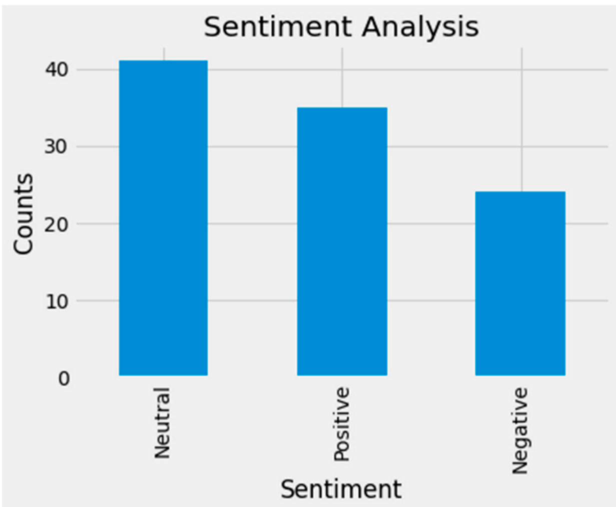

Figure 5. Classification of sample of tweets.

4. General Discussion

Organizations today are incorporating better AI and AI-based chatbots for new avenues of sustainability, innovation, cost savings, business and revenue growth, and overall sustainable development of businesses. Sustainable businesses have a positive impact on economy, community, society, or environment. AI-based chatbots are tools for sustainable development of businesses since they increase efficiency, automate processes, and provide sustainable solutions for environmental, economic, and social issues, three pillars of sustainability [30]. These environmental, economic, and social factors should be tested by designers and developers for creating the engaging uses of chatbots that can be utilized by global businesses sustainably. Developers should create reliable, sustainable and practical chatbots by providing interactive functions needed to develop humanized and natural conversations with these AI-based agents. These conversational agents should be based on natural language processing and machine learning capabilities for creating a sustainable global impact on businesses resulting in the best user experience and sustainable business development [31].

This research develops chatbots that can be used for assessing cyber threats on social media through sentiment analysis leading to a global sustainable development of businesses. Our Bot Libre chatbot, developed as an open-source platform, is deployed on Twitter, and helps in sentiment analysis of tweets generated by Twitter users conversing with our developed chatbot, thus assessing issues (risks and challenges) of cybersecurity and national security. Social media chatbots are helpful for global businesses for sending mass messages and updates. These conversational agents help understand consumer preferences (both online and offline) better along with enhancing customer interaction rates through social media chatbots. For example, Facebook messenger is powered by a computer program over AI, and this FB messenger chatbot track social media analytics, website traffic, and thus boost consumer confidence and consumer loyalty with brands and businesses. Another example is the homeowners and renters' insurance provider, Lemonade using a customer onboarding chatbot, Maya that can onboard customers in 90 seconds, as compared to online traditional insurers who take 10 minutes. In addition to Maya, Lemonade's claims chatbot, Jim settles insurance claims within seconds, while traditional insurers may take anywhere between 48 hours and 12 months to settle home insurance claims. Similarly, Marriott International's chatbot, ChatBotlr is available through Facebook Messenger and Slack, and it helps Marriott Rewards members book their travel (plan for upcoming trips with suggestions linked from Marriot's digital magazine Marriott Traveler) to more than 4,700 hotels.

Technological advancements may result in positive or negative impacts on sustainability. However, we argue that AI and AI-related chatbots help in making a positive impact. Through our research on social media AI-based chatbots and cybersecurity, we strive to benefit consumers through enhanced privacy and convenience along with greater security, which positively impacts the sustainable development of businesses worldwide. Incorporating sustainable business practices into the development process of AI and AI-related technologies (e.g., chatbots on social media) will help in ensuring the alignment of technologies with sustainable development principles. Current and future social media chatbots and cybersecurity researchers may wish to focus on the following research questions.

- What AI related technological advancements are better suited to promote environmental, economic, and social sustainability for global organizations?
- To what extent do privacy and security risks affect sustainability? How can social media chatbots prevent these risks and challenges?
- To what extent do technical factors (e.g., infrastructure, design) affect environmental, economic, and social sustainability for global organizations?
- Which sustainable business models can be developed and evaluated for AI and AI-related technology adoption, including circular and sharing economy models?
- What are the drivers and barriers to promoting sustainable AI-based systems, and how can AI-technology adoption support sustainable business practices globally?

 doi:10.20944/preprints202308.0329.v1

5. Conclusion and Future Directions

Social media has made it possible for people around the world to communicate with each other freely and has reduced time and space constraints. At the same time, it has proved to be a useful tool to detect threats, both national and organizational, and subvert them in a timely manner. Future work entails automating the process of retrieving tweets from Twitter space and automating the sentiment analysis process. Expanding the work to other social media sites, such as Reddit, etc. will help expand the scope of the project. In a global world, threats can emanate from any part of the world and in any language. Future work needs to be done in terms of language modeling in languages other than English with specific focus on Russian, Chinese and Arabic. The chatbot developed on Bot Libre platform needs to be refined in order to converse more naturally on social media. It needs to be more accurate in starting chat conversations with potential threatening individuals and organizations in order to extract more information from these potential malicious sources. We expect future researchers to come up with innovative ideas and methods to fill the gaps in the current knowledge domain.

Funding: This research was supported by the Office of the Undersecretary of Defense for Research and Engineering (OUSD (R&E)), United States Department of Defense (DoD)] under the following agreement: HQ003421F0013, University of Maryland, College Park, "Pilot Projects for the ARLIS Intelligence & Security University Research Enterprise (INSURE) Academic Consortia."

Appendix A

The code for Twitter API credentials and extraction of tweets is below.

```
# Twitter API credentials
consumerKey = log['key'][0]
consumerSecret = log['key'][1]
accessToken = log['key'][2]
accessTokenSecret = log['key'][3]
bearer_token = log['key'][4]

accessToken

# Create the authentication object
authenticate = tweepy.OAuthHandler(consumerKey, consumerSecret)
# Set the access token and access token secret
authenticate.set_access_token(accessToken, accessTokenSecret)
# Create the API object while passing in the auth information
api = tweepy.API(authenticate, wait_on_rate_limit = True)

# Extract 2000 tweets
posts = [status for status in tweepy.Cursor(api.search, q='russia', tweet_mode='extended',
lang='uk', retweeted= False, truncated=False).items(20)]
```

The code for noise removal is below.

```
#Create a function to clean the tweets
def cleanTxt(text):
text = re.sub(r'@[A-Za-z0-9]+', '', text) # Removed @mentions
text = re.sub(r'#', '', text) #Removing the "#' symbol
text = re.sub(r'RT[\s]+', '', text) # Removing RT
text = re.sub(r'https?:\/\/\S+', '', text) # Remove the hyper link
text = re.sub(r':[\s]+', '', text) # Removing:
text = text.lstrip()
```

 doi:10.20944/preprints202308.0329.v1


```
return text
#Cleaning the text
df['Tweets']= df['Tweets'].apply(cleanTxt)
#Show the cleaned text
df
```

The algorithm for finding tweets with negative orientation is below.

Procedure **findOrientation**
$(frequent\_markers, compendium\_of\_samples)$
begin
for each marker a_i in $frequent\_markers$
for each sample b_j in $compendium\_of\_samples$
if($a_i = b_j$){
if($\mathbf{searchForNegtive}(a_i)! = \text{false}$){
$compendium\_of\_negative_{makers} = b_j$}
endfor
endfor
end

Procedure **searchForNegative**
$(marker, sentiwordnet)$
begin
for each synset s_i in $sentiwordnet$
if($marker = s_i$){
if ($s_{negative} > s_{positive}$)
return $marker$}
else {
return false}
endfor
end

The codes to classify subjectivity and polarity of tweets, form a word cloud, and scatterplot of subjectivity vs polarity is below.

```
# Create a function to get the subjectivity
def getSubjectivity(text):
return TextBlob(text).sentiment.subjectivity

# Create a function to get the polarity
def getPolarity(text):
return TextBlob(text).sentiment.polarity


#Create two new columns
df['Subjectivity'] = df['Tweets'].apply(getSubjectivity)
df['Polarity'] = df['Tweets'].apply(getPolarity)

#Show the new dataframe with the new columns
df

# Plot The Word Cloud
allwords = ' '.join([twts for twts in df['Tweets']])
wordCloud = WordCloud(width = 1000, height=600, random_state = 21, max_font_size = 119).generate(allwords)
plt.imshow(wordCloud, interpolation = "bilinear")
plt.axis('off')
plt.show()

#Create a function to compute the negative, neutral and positive analysis
def getAnalysis(score):
if score < 0:
```

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 August 2023 doi:10.20944/preprints202308.0329.v1


```
return 'Negative'
elif score == 0:
return 'Neutral'
else:
return 'Positive'

df['Analysis'] = df[ 'Polarity' ].apply(getAnalysis)

#Show the dataframe
df

# Print all of the positive tweets
j=1
sortedDF= df.sort_values(by=['Polarity'])
for i in range(0, sortedDF.shape [0]):
if (sortedDF['Analysis'][i] == 'Positive'):
print(str(j) + ') '+sortedDF['Tweets'][i])
print()
j = j+1

# Print all of the negative tweets
j=1
sortedDF= df.sort_values(by=['Polarity'], ascending='False')
for i in range(0, sortedDF.shape[0]):
if (sortedDF['Analysis'][i] == 'Negative'):
print(str(j) + ') '+sortedDF['Tweets'][i])
print()
j = j+1

# Plot the polarity and subjectivity
plt.figure(figsize=(8,6))
for i in range(0, df. shape[0]):
plt.scatter(df['Polarity'][i], df['Subjectivity'][i], color='Blue')
plt.title('Sentiment Analysis')
plt.xlabel('Polarity')
plt.ylabel('Subjectivity')
plt.show()

# Get the percentage of positive tweets
ptweets = df[df.Analysis== 'Positive']
ptweets = ptweets['Tweets']

round( (ptweets.shape[0] / df.shape[0]) *100 , 1)

# Get the percentage of negative tweets
ntweets = df[df.Analysis== 'Negative']
ntweets = ntweets['Tweets']

round( (ntweets.shape[0] / df.shape[0]) *100 , 1)

#Show the value counts
```

 doi:10.20944/preprints202308.0329.v1


```
df['Analysis'].value_counts()

#plot and visualize the counts
plt.title('Sentiment Analysis')
plt.xlabel('Sentiment')
plt.ylabel ('Counts')
df['Analysis'].value_counts().plot(kind='bar')
plt.show()
```

The code for developing Chatbot on Bot Libre platform is below.

```
<script type='text/javascript' src='https://www.botlibre.com/scripts/sdk.js'></script>
<script type='text/javascript' src='https://www.botlibre.com/scripts/game-sdk.js'></script>
<script type='text/javascript'>
SDK.applicationId = "6191571217345391239";
SDK.backlinkURL = "http://www.botlibre.com/login?affiliate=allanmuir1";
var sdk = new SDKConnection();
var user = new UserConfig();
user.user = "allanmuir1";
user.token = "1393605116044980714";
sdk.connect(user, function() {
var web = new WebChatbotListener();
web.connection = sdk;
web.instance = "41557310";
web.instanceName = "allantwitter bot";
web.prefix = "botplatform";
web.caption = "Chat Now";
web.boxLocation = "bottom-right";
web.color = "#009900";
web.background = "#fff";
web.css = "https://www.botlibre.com/css/chatlog.css";
web.gameSDKcss = "https://www.botlibre.com/css/game-sdk.css";
web.buttoncss = "https://www.botlibre.com/css/blue_round_button.css";
web.version = 8.5;
web.bubble = true;
web.backlink = true;
web.showMenubar = true;
web.showBoxmax = true;
web.showSendImage = true;
web.showChooseLanguage = true;
web.avatar = true;
web.chatLog = true;
web.popupURL =
"https://www.botlibre.com/chat?&id=41557310&embedded=true&chatLog=true&facebookLogin=false&application=6191571217345391239&user=allanmuir1&token=1393605116044980714&bubble=true&menubar=true&chooseLanguage=true&sendImage=true&background=%23fff&prompt=You+say&send=Send&css=https://www.botlibre.com/css/chatlog.css";
web.createBox();
});
</script>
```

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 August 2023 doi:10.20944/preprints202308.0329.v1

References

1. Cutter, S., T. J. Wilbank, (ed.): Geographical Dimensions of Terrorism, Taylor & Francis, Inc. (2003)
2. Kennedy, L. W., Lunn, C. M.: Developing a Foundation for Policy Relevant Terrorism Research in Criminology (2003)
3. Reid, E., Qin, J., Chung, W., Xu, J., Zhou, Y., Schumaker, R., ... & Chen, H. (2004, June). Terrorism knowledge discovery project: A knowledge discovery approach to addressing the threats of terrorism. In International Conference on Intelligence and Security Informatics (pp. 125-145). Springer, Berlin, Heidelberg.
4. James, N. (2023). 160 Cybersecurity Statistics 2023. https://www.getastra.com/blog/security-audit/cyber-security-statistics/ Retrieved June 11, 2023.
5. Gartner Research, "Forecast Analysis: Information Security Worldwide 2Q18 Update", March 2020, [online] Available: https://www.gartner.com/en/documents/3889055.
6. Franco, M. F., Rodrigues, B., Scheid, E. J., Jacobs, A., Killer, C., Granville, L. Z., & Stiller, B. (2020, November). SecBot: a business-driven conversational agent for cybersecurity planning and management. In 2020 16th International Conference on Network and Service Management (CNSM) (pp. 1-7). IEEE.
7. Thapa, B. (2022). Sentiment Analysis of Cybersecurity Content on Twitter and Reddit. arXiv preprint arXiv:2204.12267.
8. Hernández-García, Á., & Conde-González, M. A. (2016). Bridging the gap between LMS and social network learning analytics in online learning. Journal of Information Technology Research (JITR), 9(4), 1-15.
9. Chatzakou, D., Koutsonikola, V., Vakali, A., & Kafetsios, K. (2013, September). Micro-blogging content analysis via emotionally-driven clustering. In 2013 humaine association conference on affective computing and intelligent interaction (pp. 375-380). IEEE.
10. Achrekar, H., Gandhe, A., Lazarus, R., Yu, S. H., & Liu, B. (2011, April). Predicting flu trends using twitter data. In 2011 IEEE conference on computer communications workshops (INFOCOM WKSHPS) (pp. 702-707). IEEE.
11. Pinard, P. (2019, December 3). 4 chatbot security measures you absolutely need to consider - dzone security. dzone.com. Retrieved October 29, 2021, from https://dzone.com/articles/4-chatbots-security-measures-you-absolutely-need-t
12. Adamopoulou, E., & Moussiades, L. (2020, June). An overview of chatbot technology. In IFIP International Conference on Artificial Intelligence Applications and Innovations (pp. 373-383). Springer, Cham.
13. Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. Psychological review, 114(4), 864.
14. Blut, M., Wang, C., Wünderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: a meta-analysis of physical robots, chatbots, and other AI. Journal of the Academy of Marketing Science, 49(4), 632-658.
15. Sheehan, Ben, Hyun Seung Jin, and Udo Gottlieb. "Customer service chatbots: Anthropomorphism and adoption." Journal of Business Research 115 (2020): 14-24.
16. Feine, J., Morana, S., & Gnewuch, U. (2019). Measuring service encounter satisfaction with customer service chatbots using sentiment analysis.
17. Küster, D., Kappas, A.: Measuring Emotions Online: Expression and Physiology. In: Holyst, J.A. (ed.) Cyberemotions: Collective Emotions in Cyberspace, pp. 71–93. Springer International Publishing, Cham (2017)
18. Yang, Y., Li, H., & Deng, G. (2011, July). A case study: behavior study of chinese users on the internet and mobile internet. In International Conference on Internationalization, Design and Global Development (pp. 585-593). Springer, Berlin, Heidelberg.
19. Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank.
20. Choy, M. (2012). Effective listings of function stop words for Twitter. arXiv preprint arXiv:1205.6396.
21. Agrawal, R., & Srikant, R. (1994, September). Fast algorithms for mining association rules. In Proc. 20th int. conf. very large data bases, VLDB (Vol. 1215, pp. 487-499).
22. Hu, Ming, Guannan Zheng, and Hongmei Wang. "Improvement and research on Aprioriall algorithm of sequential patterns mining." 2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering. Vol. 2. IEEE, 2013.
23. U.S. Department of Homeland Security (DHS). (2020). National Initiative for Cybersecurity Careers and Studies. Glossary. Retrieved from https://niccs.us-cert.gov/glossary
24. Framework, N. C. W. (2017). National Initiative for Cybersecurity Careers and Studies.
25. Achrekar, H., Gandhe, A., Lazarus, R., Yu, S. H., & Liu, B. (2011, April). Predicting flu trends using twitter data. In 2011 IEEE conference on computer communications workshops (INFOCOM WKSHPS) (pp. 702-707). IEEE.
26. Feng, S., Wang, D., Yu, G., Yang, C., & Yang, N. (2009). Sentiment clustering: A novel method to explore in the blogosphere. In Advances in data and web management (pp. 332-344). Springer, Berlin, Heidelberg.

 doi:10.20944/preprints202308.0329.v1

27. Andreevskaia, A., & Bergler, S. (2006, April). Mining wordnet for a fuzzy sentiment: Sentiment tag extraction from wordnet glosses. In 11th conference of the European chapter of the Association for Computational Linguistics.
28. Fei, G., Liu, B., Hsu, M., Castellanos, M., & Ghosh, R. (2012, December). A dictionary-based approach to identifying aspects implied by adjectives for opinion mining. In Proceedings of COLING 2012: Posters (pp. 309-318).
29. Owoputi, O., O'Connor, B., Dyer, C., Gimpel, K., Schneider, N., & Smith, N. A. (2013, June). Improved part-of-speech tagging for online conversational text with word clusters. In Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: human language technologies (pp. 380-390).
30. Ahamer, G. (2016). GISS and GISP facilitate higher education and cooperative learning design. In Geospatial research: Concepts, methodologies, tools, and applications (pp. 810–833). IGI Global. https://doi.org/10.4018/978-1-4666-9845-1.ch037.
31. Al-Sharafi, M. A., Al-Emran, M., Iranmanesh, M., Al-Qaysi, N., Iahad, N. A., & Arpaci, I. (2022). Understanding the impact of knowledge management factors on the sustainable use of AI-based chatbots for educational purposes using a hybrid SEM-ANN approach. Interactive Learning Environments, 1-20.