Introduction

The phrase “Fake News” is one of the most phrases that have been circulated throughout social media platforms today. One dataset of Fake News from Kaggle contains 12,999 posts collected from 244 websites between 10/25/2016 to 11/25/2016. This illustrates how much of Fake News have been produced in a short time. Not just the volume of Fake News that matter, the magnitude of it does matter the most. For example, in 2013 The Associated Press Twitter account had been hacked and tweeted that President Barack Obama is injured in an explosion at The White House. In a matter of seconds, stock market lost $130 billion in stocks values. AP quickly addressed the issue and suspend its Twitter account. Although lost stocks value have been recovered, however the disrupted environment among traders and investors stay for awhile (Time,2013). Fake News today play major role in politics and business; it is an easy weapon that anyone can use to spark disruption and create a chaos. Over the past decade, many companies have been targeted with fake news.

After the election day on November 2016, PEPSI CEO says that some of her employees are worried after the President Trump victory. On November 12, many Fake News websites published fabricated PEPSI’s CEO statment titled “Pepsi CEO Indra Nooyi told Donald Trump supporters to "take their business elsewhere."(Snops.com). Then people went to social media and threatening to boycott PEPSI using to hashtags on Twitter, #boycottPepsi and #Pepsiboycott. The impact on PEPSI stock was huge following the published fake news about the CEO statement. PEPSI stock fell from $106 on November 11 to $102 on On November 15, which is a decrease of 3.75% which approximately equals to $4 billions, just because of Fake News. The following graph shows the impact of Fake News on PEPSI sentiment analysis and PEPSI stock price following the spread of Fake News about PEPSI CEO(Alva Group).

Figure 1 shows correlation between publishing the fabricated CEO statement and the big hit the stock had during that time period. In addition to the hit on PEPSI stock price, the company public image was damaged. From Figure1 the sentiment analysis PEPSI's daily US sentiment analysis fell down from 5.7 to 4(Alva Group,2017). Another example on Fake News disrupting business, on August 2, 2017 Starbucks ads were circulated in the internet and many social media platforms. These ads images advertising "Starbucks Dreamer Day," stating that Starbucks will give deep discount for undocumented immigrants in solidarity Dream Day Act. Starbucks attempted to fight back on all social media platforms by responding online to this matter. There were two hashtag on twitter calling to #boycottstarbucks over the #borderfreecoffee. Despite the fact that Starbucks issued a statement clarifying that these ads are false, some people who want to believe this add did boycott starbucks. Some people also mentioned Immigration and Customs Enforcement @ICEgov to go ahead and visit Starbucks locations on August 11 and 12 to arrest undocumented immigrants who came to have 40% coffee discount.

These two examples show how fake news can damage company's reputation and public image. The damage to a company does not just impact stock price, but goes beyond to damage company's public image. The damage could be creating a chaos at the company stores or simply encourage customers to rethink about buying the company products due to Fake News that unchecked. In the case of Starbucks for example,

The Code: Why We Made It, and What It Does

As awareness of fake news in the public sphere continues to increase, various resources have been developed to help identify sources of fake news in the hopes of mitigating it. One such resource, OpenSources, is a curated list of websites that are frequent sources of misleading and blatantly false news stories established by a professor at Merrimack College [1].

Twitter and Reddit, two of the most prominent social media sites in the Western world, each have at least 250 million users [2] [3]. Due to their high traffic, they represent convenient avenues for the propagation of news stories by any interested party[4][5], as well as providing vectors for these websites to propagate fake news. While the websites can be used for benign purposes or for relatively normal marketing, their immense reach means that fake news articles that go viral will spread quickly across the internet.

Due to the speed with which viral fake news can have a negative impact on both the public image and the financial bottom line of businesses, we posit that a proactive approach to fake news postings on these websites before they can hoodwink unsuspecting readers represents a useful asset to the standard crisis communications toolbox. Our approach is to immediately recognize fake news articles pertaining to any interested company and identify them as such to all readers. It is often the case that users begin responding to and discussing articles based purely on the headline, without evaluating their contents; with this being true, alerting readers that they may be being misled is paramount [6]

We currently possess two primary avenues for achieving this. The first is predicated on the aforementioned list of fake news websites curated by OpenSources. By making the default assumption that any story hosted by these websites has a strong likelihood of being misleading or fake, we can simply preempt any significant discussion on Reddit or Twitter by announcing to

readers that the source is, at best, suspect.

Our second approach will be to evaluate the text inside the story as the basis for classification of news articles independent of their source. This approach will train software to evaluate text using a large corpus of previously-classified news stories. The fake news articles in this corpus were compiled from websites flagged by OpenSources, while the real news articles were collected from media outlets with a long history of reliable reporting [7].

Automating Responses Using Pre-Flagged Websites

The first approach mentioned above, oriented around the OpenSources list of fake news websites, requires automated comparison of the website hosting stories highlighted in Twitter tweets or Reddit posts to the Opensources list. We utilized the APIs provided by both Twitter and Reddit to design bots that can generate pre-programmed responses. The Twitter bot identifies tweets pertaining to our client companies using hashtags, and then inspects these tweets to see if they contain links to other pages. If these other pages are hosted by websites on the OpenSources list, a response to that tweet is automatically generated by the Twitter bot.

Figure 1 shows an example of the Twitter bot, called Bot Defender, in action. The bot was programmed to evaluate any tweet with a hashtag containing the word "Pepsi" and evaluate any links that tweet might contain. Because this tweet did contain a link to an external source that was flagged, an automated response was generated.

Similarly, we have developed a bot, again called Bot Defender, that can automatically scroll through Reddit traffic. This version of Bot Defender evaluates both the parent link in the title of a post, as well as the user commentary within the thread. In the first case, an automated response is triggered if the link is to a flagged source. In the second case, if any user generates a reply with a link, that link is investigated and an automated response is triggered if the link is to a flagged source.

Figure 2 shows an example of the Reddit-based Bot Defender in action. Here, as discussed above, an automated response is generated due to the nature of the link in the title. The user comment containing a link to a flagged source receives an automated response, while the user comment without a link does not.

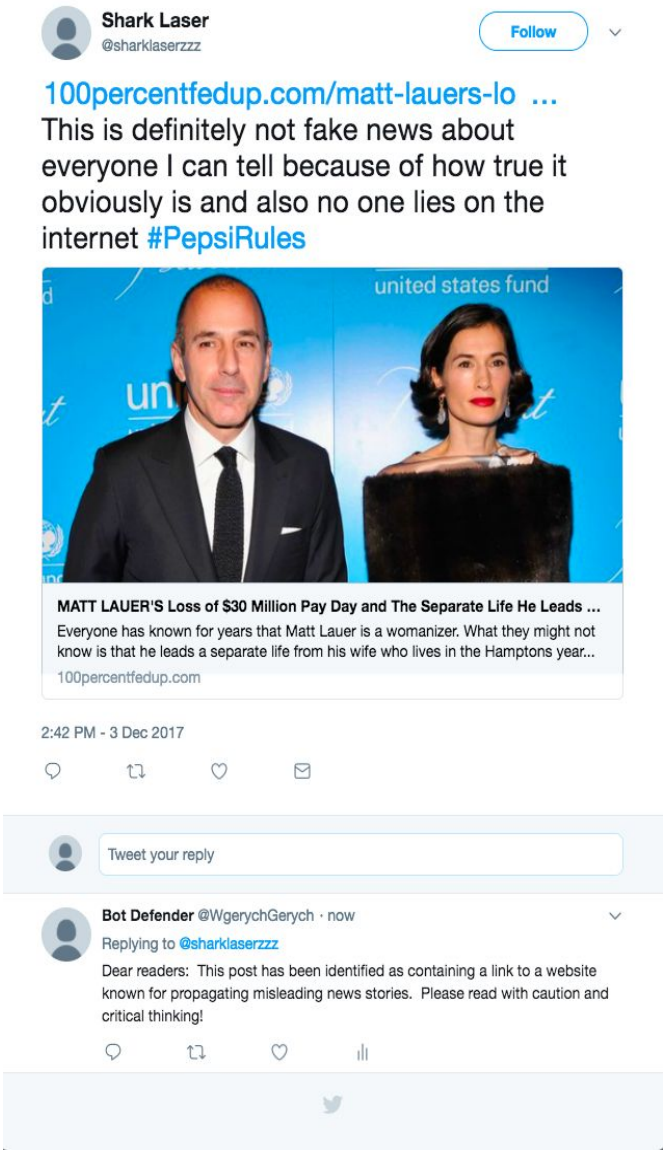

Figure 1: The automated Twitter bot at work

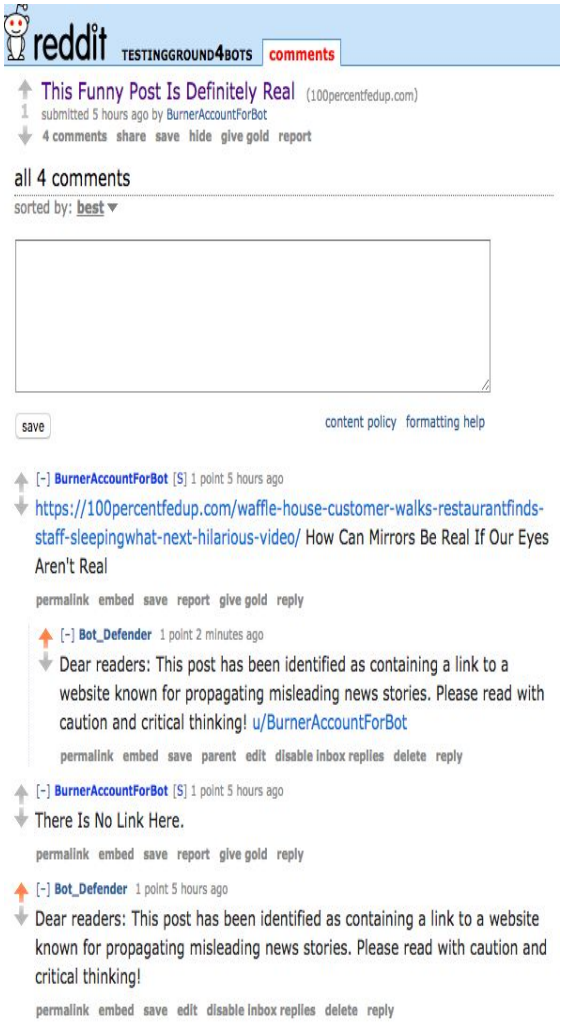

Figure 2: The automated Reddit bot at work

Automating Responses Using Independent Evaluation

While curated lists identifying sources of fake news are helpful, they are unlikely to identify every possible purveyor of fake news. Therefore, we require a method of evaluating individual stories based on their contents rather than their source. In the past few decades, artificial neural networks have achieved significant advances in natural language processing. Much of the code we use in this section is founded on work done by Gareth Dwyer and implemented with a neural networks API called Keras [8].

We will first implement what is called tokenization, which allows us to transform text into a set of numbers that can be more easily analyzed. Keras implements tokenization by transforming more commonly occurring terms into smaller integers; the most commonly occurring term is tokenized as 1. After doing so, we ensure that each document is treated by the tokenizer as having the same length. We then limit the neural network's analysis to the 20,000 most common words, as well as limiting our analysis to the 300 most highly weighted words from each document.

Keras allows easy development of neural network architecture, meaning that we can choose a system customized to our purposes. Because we are engaged in natural language processing, we rely on recurrent neural networks, which are commonly used in this context [9]. We use the long short-term memory method introduced by Hochreiter and Schmidhuber in 1997 to address some of the shortcomings in the methodology, as well as some other adjustments recommended by Dwyer [8][10]. After training the neural network with our corpus of fake and legitimate news stories, the network is capable of predicting the nature of new documents with approximately 88% accuracy.

To make use of this, we have programmed Bot_Defender to not only examine the source of news articles, but also to read the articles by itself and come to a conclusion about whether or not the article's contents were semantically similar to fake news articles. This can be performed on either Twitter and Reddit, and is demonstrated using our Reddit-based Bot Defender.

Figure 3: Reddit Bot classifying an article as fake news

Figure 4: Reddit Bot classifying an article as factual

For those interested, we give a brief summary of the functioning of the neural network. The neural network consists of 6 layers, as shown in Figure 5 [4].

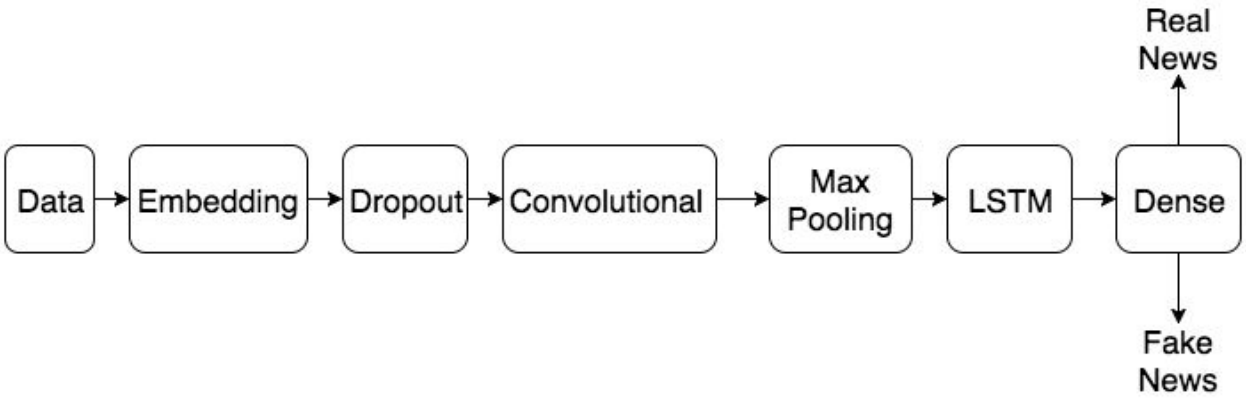

Figure 5: An outline of the neural network utilized

The embedding layer transforms each individual word into a vector of a given size (in our use case, a vector of 128 words). As the neural network is trained, the mapping function used to transform the words learns which words are similar, and adjusts its transformations to map similar words into similar vectors.

The dropout layer is similar in concept to the random forest technique applied in ensemble learning methods. It randomly disables neurons in subsequent layers in order to ensure that they are able to learn a comprehensive variety of lessons and nuances rather than re-learning the same lesson over and over. Because learning mechanisms may fall prey to local minima and fail to discover the true response (for example, always concluding that 'couple' refers to a pair of people and never learning that it may be used idiomatically), dropout layers are necessary to providing a more complete and generalisable approach for the neural network.

The convolutional layer works in tandem with the max pooling layer to train the classifier using pertinent n-grams instead of limiting the trainer to single worlds. The convolutional layer assigns a filter to subsets of the text to obtain what can be seen as a weighting, while the max pooling layer collects the most significantly weighted blocks from each convolution.

The recurrent neural network - long short-term memory layer (RNN - LSTM) layer is the heart of our learning method, as the other layers are simply dedicated to providing input to this layer. RNNs function by parsing sequential data and passing information gained from each datum on to the next iteration. Because our data is linguistic, where each word has an impact on the interpretation of the subsequent words in the phrase, it is well-suited to our purposes. The LSTM modification allows the function to retain information for particularly relevant content.

The dense layer is the final layer and translates the RNN-LSTM output into an actual classification. This classification is then integrated into the Reddit Bot code for application.

Business Model:

The business model for the application is monthly subscription based. Our customers can subscribe to one of the three different plans we that we offer. The following table shows the

different plans we offer and their features and cost.

	Subscribtion Level		
Featuer	Basic	Pro	Advance
Monthly Cost	$3,000	$5,000	$6,000
Keyword	2	5	10
Mention/month	3000	6000	10000
Extra Mention	$0.95	$0.75	$0.45
Reddit	X	X	X
Twitter	X	X	X
Automatic Respons	X	X	X
Custom Respons	NA	NA	X
Users	1	3	10
Dashboard	NA	X	X
Sentiment Analysis	NA	NA	X
Influencers Dashboard	NA	NA	X
Data Exports	NA	NA	X
Social Media Monitoring	NA	NA	X
Mainstream Media	NA	NA	X

The Keyword is a keyword or set of Keywords that Defender will use to monitor and crawl the web and social media for as specified for each plan. The keyword or the set of keywords will be provided by the customer. For example, P&G is a single keyword, but P&G, Gillette and Tide are a set of keywords for one customer. All three plans include Twitter and Reddit. Pro and Advance plans offer more features such as dashboard and more users from the customer's side. The dashboard deliver real time activity monitoring for the customer accounts on the subscribed plan. The user on the Pro and Advance plan allow our customers to take advantage of other services available with their subscription such as social media marketing and search engine optimization. The Advance Plan offers more features to our customers such as sentiment analysis on Facebook, Instagram, and Telegram. In addition to that, it includes mainstream media monitoring and analysis for our customer’s keyword. Advance plan also includes custom response in which the customer will work with assigned customer support to provide custom response to and Fake News related to the customer.

We priced our services based on the current market. We locked at our competitors in the market and evaluated their services. Although they are not providing the same core services that we are providing(Fake News Detection), they compete in providing sentiment analysis for their clients. Brandwatch, Indico, Lexalytics and Agorapulse are some of the big players in sentiment analysis area. They target small local business and corporates, and their prices started from $500 monthly to $5000 just as starting plan. If a customer need more services or add on

features, the customer need to pay extra. Defender is targeting medium size and corporate clients, therefore we charge premium prices for our services.

Our finance to start the service is as follow:

	Costs for 36 months
Expenses for 36 Months	
Rent for Space	$90,000
4 Employees	$864,000
Powerfull Computers	$30,000
Twitter Premium API	$54,000.00
Maintaining The Services	$50,000
Futer Resarch	$200,000
Dashboard	$50,000
Data Storage	$5,000
Social Media Monitoring	$15,000
Total	$1,358,000
Asking for	45%

[1]http://www.opensources.co/
[2]https://about.reddit.com/advertise/
[3]https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/
[4]https://www.washingtonpost.com/news/the-switch/wp/2017/06/29/twitter-is-looking-for-ways-to-let-users-flag-fake-news/?utm_term=.d1eaacbc334e
[5]https://www.forbes.com/sites/jaymcgregor/2016/12/14/how-we-bought-reddit-for-200/#4fe1cfc144a8
[6]https://www.washingtonpost.com/news/the-intersect/wp/2016/06/16/six-in-10-of-you-will-share-this-link-without-reading-it-according-to-a-new-and-depressing-study/?utm_term=.c8121e5654bb
[7]https://opendatascience.com/blog/how-to-build-a-fake-news-classification-model/
[8]http://www.developintelligence.com/blog/2017/06/practical-neural-networks-keras-classifying-yelp-reviews/

[9]https://pdfs.semanticscholar.org/93eb/7d6592290e0b0fa6ee1cde7cb08d2f4aceb6.pdf
[10]http://www.bioinf.jku.at/publications/older/2604.pdf

[11]http://www.alva-group.com/us/fake-news-affect-corporate-reputation/

[12]https://www.pulsarplatform.com/blog/2016/brand-dig-pepsi-new-balance-and-facebook-battle-fake-news/