A Model for Evaluating Fake News

Dr. Char Sample

ICF Inc. LLC
Columbia, MD
char.sample@icf.com

Dr. Connie Justice
Purdue School of Engineering
& Technology, IUPUI
Indianapolis, IN
cjustice@iupui.edu

Dr. Emily Darraj
Cyber Security Department
Capitol Technology University
Laurel, MD
edarraj@captechu.edu

Abstract— **Fake News is slowly being recognized as a security problem that involves multiple academic disciplines; therefore, solving the problem of fake news will rely on a cross-discipline approach where behavioural science, linguistics, computer science, mathematics, statistics and cyber security work in concert to rapidly measure and evaluate the level of truth in any article. The proposed model relies on computational linguistics to identify characteristics between true news and fake news so that true news content can be quantitatively characterized. Additionally, the pattern spread of true news differs from fake news since fake news relies, in part on bots and trolls to saturate the news space. Finally, provenance will be addressed, not in the traditional that examines the various sources, but in terms of the historical evaluations of author and publication computational linguistics and pattern spread**

Keywords—fake news; computational linguistics; pattern spread; provenance; trust

I. INTRODUCTION

The term "fake news" was officially ushered into the lexicon when Oxford Dictionary added the term in 2017 [1]. While the term is frequently used and definitions vary, the problem of deceptive data is serious and exposes a profound and underlying flaw in information and network security models. This flaw is trust in entities, without verification of the content that they exchange.

"Trust but verify" [2] is an old proverb that until recently resulted in trust at the expense of verification. Trust in journalists historically resulted from reputation of the journalist as well as the news organization (publisher). However, publisher reputations of news organizations can vary widely, and the line between news and entertainment continues to blur [3]. The journalistic integrity of news organizations, while an interesting discussion, is not the focus of this effort; however, defining, measuring and characterizing fact-based news is.

Our historical method of placing trust in reporters or news organizations is under attack [4]. When a reporter can be discredited for $50,000 [4] and a news story can be staged for $200,000 [4] the facts within their context must be preserved and protected. Protection begins with understanding of the value asset that is to be protected. In the case of news the assets include the story (data and metadata), as well as the reporter and the publisher.

Insurance companies rely on statisticians to determine the value of items that they insure [5] allowing for reasonable prediction of repairs and replacements. Data in general could benefit from a similar model, and news data specifically needs an immediate solution that is accurate and efficient.

In some cases trust was assumed without any evidence of trustworthiness (Facebook, Twitter) resulting in large groups receiving news from social media sites [6]. In other cases, trust is granted based on reputation such as news sites [7]. In all cases, the changing role of the news media due to the Internet results in a rush to deliver news first.

Any solution to fake news (FN) must consider the full scope of information or the "totality of information" [8]. The customization of the fake narratives along with the targeted delivery demands that an effective solution must fuse non-technical disciplines with traditional technical responses. The attacks may originate from any source, although the Russian based attacks are the quite sophisticated [9] and have gathered quite a bit of attention.

There are aspects of the Russian approach that warrant inclusion into the framework, even if the implementation becomes uniquely western. The Russian term "protivoborstvo" describes the intentionally created rhetorical game that is foundational to FN, and this rhetorical games can partially be addressed using computational linguistics and machine learning, illustrating one example of interdisciplinary fusion.

FN or deceptive information campaigns can be thought of as an opening shot in future information conflict that supports hybrid warfare [10,11].This framework may provide guidance for countering deceptive information campaigns. We have no cyber equivalent for trust but verify (yet). The purpose of this paper is to introduce a framework for evaluating FN. This framework may provide the cyber equivalent of "trust but verify" [2] for FN. In addition to countering Russian FN efforts this framework provides the foundation for examining the quality of data and may assist analysts in evaluating other news stories or events.

II. BACKGROUND

Falsehoods and deception in political discourse are a long standing problem in an industry where words matter. Deception and propaganda have a long history, the Trojan horse serves as an example of one of the earliest deceptions [12]. The Internet

makes possible the ability to deliver the deceptive message to a larger audience, and social media data made possible the customization of deceptive data [13]. Data science techniques performed by Cambridge Analytica [14] made possible the rapid customization of messaging. Chatbots exacerbated the problem through the use of Artificial Intelligence (AI) software that could dynamically adjust to, and manipulate user responses [15]

There are many ways that facts can be distorted resulting in altered perceptions, but there are a limited number of ways that facts can remain faithful to their original creation. This provides an entry point into the solution. Thus, attempting to model deceptive data is similar to attempting to model malware or any other host of cyber security problems. Deception is unbounded therefore, attempting to model or predict deceptive data is difficult and subject to continual change. Facts, however, are constrained, allowing for more accurate modeling. While deceptive data may have several common features, all of these features should be examined in the context of the factual data that the deception is designed to conquer.

In order for deceptive data to be effective the data should elicit an emotional response [16], otherwise, the data would be quickly forgotten. The response does not always need to be strong; thereby, avoiding suspicion of hyperbole. Trust must be gained about deceptive data, and while trusted users can shorten the time required a little initial skepticism is normal. Russian deceptions build the FN foundation by offering an alternative view or narrative that is designed to sound reasonable [17-19].

A. What is propaganda

Propaganda is information or ideas that are spread by a group, including government, with the goal of influencing a targeted group or person's opinions, through omission of facts or by secretly emphasizing only one narrative of the facts [20]. Often times deliberately used to control, influence or change cognition of the targeted group; propaganda entwines fundamental elements of psychology and technology in service of the goal. Psychological aspects include campaigns of winning the minds, means and measures of message distribution, requiring a behavioral science understanding of message creation and application [21]. The past is understood through information that was recorded and left behind by the scribes of the particular time. Chronicles and annals written provide contextual understanding of the past; however, these writings contain the biases of the scholars, historians, clergy, rulers, and ordinary citizens in local communities [22].

In order for propaganda to be effective, the source or purveyor needs to fully understand the values of their target audience thus rendering the target's intellect ineffective. The most accomplished propagandist discerns and plays on their targets values, morals, needs, or fears [16]. This goal can be achieved subtly or overtly based on the values of the target.

Taylor [23] suggested the earliest form of propagandist imagery was found in the Neolithic age. The use of war propaganda may be found in Neolithic cave paintings where imagery carved in the wall commemorated their battles. The carvings in the cave walls were available for their clan and other tribes to view the victorious battles and potentially intimidate their enemies through the cave wall drawings [23].

From the Neolithic age, historians then look at ancient Greece where we see some of the first examples of propaganda. Speech was utilized for conveying persuasive messages. Ingram [16] provided the example of Confucius's writings, the *Analects*, in Ancient Greece, where these writings were used for persuasion. The targeted men who read these writings were supposed to live a more meaningful existence. From Ancient Greece, to Alexander the Great to the Pharaohs of Egypt, propaganda was a weapon of choice to change the cognition of their targets. Egyptian pharaohs propaganda messages were prestige, nobility, and imperial legitimacy exemplified with grandiose architecture [24].

The Roman Empire largely influenced civilization reaching into Italy, Mediterranean, Britain, North Africa, Portugal, and the Persian Gulf. Dating back to 48 BC, Gaius Julius Caesar (Julius Caesar), father of Caesar Augustus first emperor of Rome, used political manipulation to win popular support of the people [25]. Julius Caesar wrote war memoirs chronicling the achievements of the civil war between Gaul and Pompey (Gnaeus Pompeius Magnus) along with the spoils of victory. Caesar sent runners to deliver war memoirs to be read before a crowd in a public area firsthand as the battle progressed [25]. This action showed the recognition of both message craft and delivery speed since the common people, or plebeians were not literate, and Caesar knew he needed their support. This early form of propaganda resulted in Julius Caesar being heralded as a hero [25].

Patriotism was tied to the military and only Roman citizens could be members of the military force, thus the fighting force was respected and feared by all and Julius Caesar had a strong reputation of looking after the Roman legions. Caesar's focus on public opinion and strong concentration on providing for his military forces was a major reason the memoirs were successful and the strong public opinion pave the way for Caesar Augustus (Augustus) [25, 26]. Augustus used writings as means to deliver his public information, which manipulated the events to tell stories from Augustus' perspective. Statues, monuments, coins, were also used to spread Augustus' image of a strong military leader, a statesman, and a peace keeper [25].

In addition to the spoken word propagandist messaging was also done through imagery and this form of messaging remains popular today. In the early stone ages, depictions of war were carved on cave walls, and then later drawn on paper or in scrolls. As propaganda matured messaging was clothing incorporated through imagery on the clothes. This included stunning regalia and insignia laden outfits [21].

The 18th century propagandists successfully used political cartoons and caricatures to directly communicate with their intended audiences [16]. The caricatures and prints were biased in nature and often times made fun or poked an individual the propagandists had targeted. The convergence of humor and politics in this new approach and was well received.

Propaganda continued to be a means of influence in times of conflict. During the Civil War cartoons became a popular propaganda media. Animated movies and political or military cartoons became an attractive means for distributing propaganda [27]. The propagandist-cartoonist can be classified in two categories, cathartic and ad justice [28].

A cathartic cartoon was successful when the message convinced people that they had nothing to fear from the enemy. An ad justice cartoon was designed to spur action and could be considered successful when the message inspired voluntary enlistment in the Union forces for example. One famous cartoon was Thomas Nast's "Compromise with the South", September 1864. The propagandist-cartoonist used the symbols of his trade to guide those toward the certain predisposed objective, which condemned the idea of compromise through the message of previous lives sacrificed for the cause [28]

According to Hinklemann [28] Hitler's *Mein Kampf* was considered an advanced work on the use of propaganda as a way to collect large numbers of supporters. The book masterfully at pulled the audience into accepting only the author's views as true and shifted blame for previous failures away from himself to the other Germans leaders; thus creating a victim mentality [27, 28]. *Mein Kampf* appealed directly to emotions formed from values and biases rather than logic. The blending of anti-Semitism and nationalism provided a message as a way for Germany to survive, through anti-Semitism as a form of nationalism or love of country.

Hinklemann observed that Hitler believed that good propaganda targeted emotion and not intelligence or the facts. Hitler played on the hatred and despair felt by the lower-class Germans. By preying on the poor economic status and fear of providing food and clothing for one's family [28] Hitler elevated emotions with reason. Combining Hitler's emotional elevation with Western societies need to assign blame [29] and cultural [30] regarding uncertainty avoidance or fear of the unfamiliar, the Jewish population, along with other non-Aryan groups were assigned the blame for the economic problems of Germany.

The United States used radio and movies to disseminate propaganda during World War II. Both Japan and the US held competitions to create patriotic theme songs. Each soldier serving in the US military was issued a songbook with songs such as Anchor Aweigh and Marine Corps Hymn. The songbook was part of standard issue [31]. Japan, in an interesting twist, used American songs as propaganda to make American G.I.'s homesick and weaken the American forces. Furthermore, the Japanese government forbade the playing of American music at home [31].

The Japanese were portrayed negatively in the media by focusing on physical characteristics such as crossed eyes and buckteeth. The Japanese were referred to as "Japs" and "back-stabbing monkeys" and "sneaky yellow rats" [31] in an attempt to de-humanize with the goal of behavior shaping and inciting desired actions [16].

Music with propaganda lyrics serves as a way to bring meanings to the intended audience. The music will instill a general feeling or emotion and with the proper message will serve as a mechanism to transmit propaganda. Music in general makes messaging easier to retain due in part to patterns and repetition [33].

Similarly, movies are a natural medium for propaganda. A moviegoer becomes a type of "hypnotized person" vulnerable to the suggestions presented by a film at a particular time [31].

Movies with their high quality visual and audio displays serve to reinforce the message [32,33].

The message creation aspects of propaganda continue to evolve, growing more sophisticated and polished as technology improves and knowledge grows. As topics, phrases and various aspects of crafting a message change the important thing to remember is that the goal remains the same. Of equal importance a message has no value until delivered, received and interpreted.

The technologic aspects that entwine with the behavioral aspects warrant discussion, loosely grouped as message delivery. Advancements in communication, military strategy and technology, and fluctuating partisan elite rapport and populace all contribute to the changing landscape of message delivery [16]. Ingram [16] stated scholars and scientists in modern times studied, determined and understood propaganda campaigns and techniques equating said campaigns to daily societal issues.

"Falsehoods fly, and the truth comes limping after it" [34]. Beginning with the earlier runners used by Julius Caesar delivering false messages as battles raged and continuing through to written press and more recently images that can travel at line speed falsehoods continue to fly. Meanwhile, the actual facts researched in context of the event occurrence requires time before it can even be recorded and presented.

The invention of the printing press and print engravings, propagandists were able to mass print their messages. After 1880, messages were further impacted with the inclusion of photographs [21]. The photographs could be staged or real, and the black and white images, and then eventually full-color, made a real impact on the targets cognitive perception. Eventually motion pictures were created, first in black and white, and then later in color, and this captured society's attention [21].

During the literary age, propaganda was produced through pamphlets, newspaper articles, advertisements, fliers, billboards, and any other means where an individual's cognitive perception could be altered or changed. Later, satirical caricatures and cartoons were used for the target audience. The propaganda campaigns utilized a new visual element and it proved to be quite successful [16].

In additional to literary campaigns, and structural campaigns, propaganda messages have also been waged through radio, satellite, and broadband communications. During Vietnam, the radio airwaves were laden with propaganda both for the United States and Vietnam [35]. Radio stations included: The Voice of America, BBC, Radio Free Europe and Radio Liberty; moreover, these stations transmitted propaganda both attributed and unattributed messages to their targeted audience [16].

Previously, 20th century propagandists utilized multiple means of delivery for their messages, using advertising and other techniques, to convey the intended message to the target [21]. All of these methods were asynchronously delivered. In the 21st century technological advances in communications, computers, networks, smart phones, Internet of Things, provided a broader landscape and saturation in the

propagandists' arsenal, and social media's role in the Arab spring resulted in a new trusted news source for users. The rise of social media has made this new landscape both more user-friendly, and, perhaps most importantly, trustworthy.

More recently Internet usage introduced a 21st century feature: the ability to rapidly disseminate deceptive data both asynchronously and synchronously. The initial dissemination relied heavily on bots and tolls to establish a starting point, once the starting point was established the dissemination reached the targets directly through the trusted channels of social media and social media trust relationships. Furthermore, the timing of the release of deceptive data took advantage of the inability to rapidly discern truth, allowing the falsehoods to fly. This strategic timing release of deceptive information is also known as weaponized information [36]. This is timed mass release of weaponized information gives the source more control over the spread than in previous times. This synchronous component relies on a mixture of "true believers" (aka "useful idiots") acting as trolls, paid trolls and artificial intelligence controlled chatbots.

The new landscape continues to grow and the amount of information available in this new environment is so rich, that a new discipline, data science has emerged. New technology, utilizing data science techniques allowed for more accurate target identification and continuous bombardment of the special crafted messages from trusted or quasi-trusted sources. The volume of these messages that reinforce values can effectively alter the target's perception. When the targeted user seeks to verify the content of a message a large number of similar messages are returned, and the target now knows that other people share the same values and beliefs.

B. Countering Propaganda

Research into computational linguistics (CL) showed that news could be accurately separated into truth, falsehoods and satire through analysis of linguistic features [37]. The credibility toolkit [38] provides the ability to assess news articles along the axes of reliability and objectivity as well as potential social media communities that might be interested in the content of the article, along with visualization tools to assist in interpretation. Thus, CL may offer a means of performing preliminary tagging of a news article for rapid evaluation of that article's veracity. In addition to CL, a reputation analysis and pattern spread may also offer valid insights that assist in the evaluation of a new story's veracity.

The initial response to FN relied on fact checking through source such as Snopes [39], PolitiFact [40], or other fact checking sites. This method worked well for years but is time intensive and easily overwhelmed with the volume of FN stories that are generated throughout the course of a campaign. Fig. 1 provides an example of a hashtag associated with a fact-based narrative, and fig. 2 illustrates the hashtag associated with the fake narrative overlaid on top of the fact-based narrative [41]. Thus fig. 2 shows the fact-based narrative being easily overwhelmed by the fake narrative.
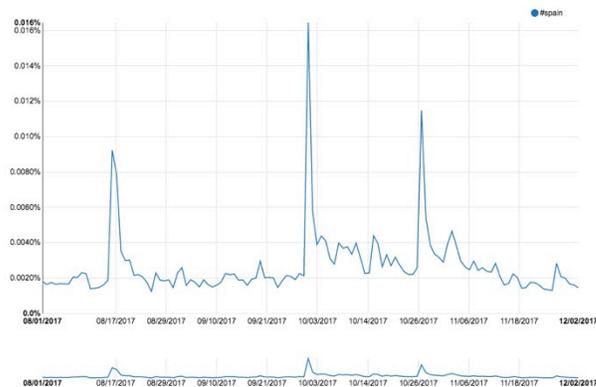

Figure 1: Hashtag associated with fact-based narrative

Fig. 1 shows the activity on #spain from 01 Aug 2017 to 01 Dec 2017, with the actual vote taking place in 01 Oct 2017 the highest of the three peaks [25].
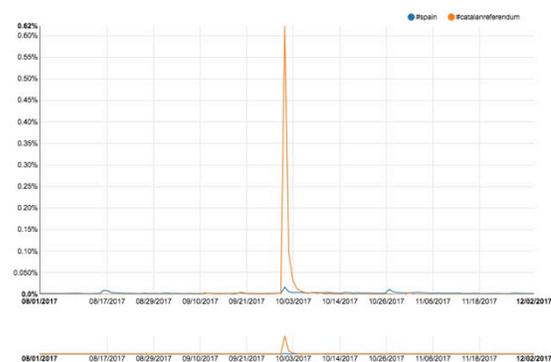

Figure 2: Fake hashtag overlaid on fact-based hashtag

The steep, peaked line in fig. 2 shows the rapid and intense injection of the #catalanreferendum hashtag that associated with the fake narrative in the same time window as #spain. Notice not only the high volume but also very short time line for the fake narrative [41]. By introducing the face narrative so close to the election (much like Clinton's e-mails [42] and Macron's e-mails [43]) the target has little to no time to respond; thus, the information at this point is weaponized, and active.

The volume associated with fake narratives is problematic because if the reader attempts to lookup the story, a large number of the same narratives will be returned, thus validating the fake narrative to the reader. Fact-checking sites take time to research and post their findings, and when the source of the fake story is a friend, or other close relationship biases are at play. Readers who are unable to determine the veracity of a news story often times lack the time and resources to look up the story in question, instead they rely on mental shortcuts such as biases [44] and source reputation to determine trustworthiness [45].

The reliance on reputation has been exploited in two manners. The first is through the use of popular social media applications, where through the targeting of groups multiple trusted entries become access points to individual members, into the larger society. Consider that social media's original mission of bringing together like-minded people to share information in the spirit of friendship and fellowship [46, 47];

however, these groups also provided a channel to distribute FN, since these sites also promoted news stories. These same sites, such as Twitter, became trusted news sources due to their role in the Arab Spring [48].

While social media sites have recently come under fire for the distribution of fake, these same sites were commended for their role in the Arab spring. The second manner in which reputation has been exploited is in the discrediting of reporters. A recent report on FN revealed that a journalist can be discredited for $50K and a news event, such as a protest, can be staged for $200K [4]. When these events are considered collectively the use of reputation analysis becomes problematic. Furthermore, reputation analysis is vulnerable to the flux problem that plagued DNS servers.

The issue of source trusting is complex and long standing as is also the history of verifying trust. The handshake was one of the earliest examples of verifying trust [45]. The verification of trust in the virtual environment is more difficult and is a problem that can has been referred to as data fidelity [49, 50]. The proposed solution will be discussed in detail in Section III of this paper.

Old models may serve as inspiration in the design of the newer models for FN evaluation. Trust in entities without verifying what those entities pass has been exploited. Trust in news sources continues to be manipulated. Untruths not only spread fast, but automated bots can persuade doubtful readers through interactive dialogue. The purveyors of propaganda have carefully profiled their targets values and beliefs before crafting their messages. The old adage about bringing a knife to a gun fight can now be replaced with bringing a gun to a bot fight.

III. PROPOSED MODEL

The model for evaluating FN relies on three areas, computational linguistics (CL), pattern spread (PS) and source provenance (SP). Each area will be discussed in greater detail in each subsection. CL findings can feed the PS and SP, while PS also can feed SP.

$$CL \rightarrow PS \; ; \; CL \; \& \; PS \subseteq SP$$

A. FN creation content analysis using computational linguistics (CL)

There are several unique characteristics associated with FN such as the size of the story in relation to the headline [4], the use of descriptive words and other features[16]. As noted earlier, attempting to model the numerous characteristics of the ever changing deceptive data is not an efficient method, but modeling the facts or ground truth data (GTD) is.

This study requires the researchers to enhance the existing RPI software with the intent of expanding the rating of the software to place event stories (AP, Reuters, Bloomberg) as GTD or **μ** in the distribution. As stories become embellished, the markers increase resulting in the growing value of deviations. We will use a number of other unreliable sources to assess which markers are especially useful to track. Conversely, when stories omit key pieces of information the markers will decrease or unrelated events will be used to fill the space, these unrelated events are often

times distractors commonly referred to as "whataboutism" [4]. Finally, traits such as repeating the same point three times [4] will be considered a deviation measure. Fig 3 depicts the proposed scoring scale.

Figure 3: Scoring scale for CL output

Context removed (-)		GT		Embellishment (+)
-10	5	0(**μ**)	5	10
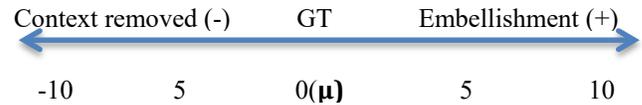
The proposed scoring scale uses the AP/Reuters/Bloomberg results as the basis for scoring (0) or μ. The deviations from this in either direction reflect the deviations from the original event. A deviation score of -10 would indicate the story was likely taken so far out of context, that the story is no longer recognized. Similarly a story with a +10 deviation score would indicate the story was so radically embellished that the original event may no longer be the central theme of the story.

The resultant score is an average of the criteria scores obtained from parts of speech, tone, ratios etc. The overall score becomes a label used for both storage and signal identification. The storage label becomes relevant for comparisons used in source provenance, and can be useful when a story originates without the GT sources. The second use for the overall CL score, signal identification is explained in the pattern spread subsection.

B. FN in motion the pattern spread (PS)

Earlier background discussion highlighted both temporal and volume aspects of FN distribution. From Julius Caesar's early runners to print media, to radio, television and more recently the Internet the speed of distribution has minimally kept pace with news and in some cases outpaced news [51]. Before the use of the Internet mass propaganda delivery tended to be asynchronous in nature but with the fusing of AI and data science chatbots could be deployed at a large scale allowing for interactive dissemination at scale. All events leave traces and digital events are no different. While some stories follow a meme-like pattern [52] the FN PS differs and will be further discussed in future work.

The PS of FN offers an opportunity to revisit and re-examine aspects of signal processing. The noise level on the Internet is very high making the signal more difficult to identify. This noisy environment provides an opportunity to identify and label the news stories in the environment using the techniques described in CL, in the environment. The labeling should result in the picture of the environment resembling a pre-painted paint by numbers picture where the various labeled items would ideally form clusters.

Fig.1 and fig. 2 showed the difference in spread of a fact-based hashtag and a FN associated hashtag. The indiscriminate use of bots and trolls resulted in the obvious signal, but China has shown that a different signal can be as effective [4]. Still in both Russian and Chinese cases automated software was used, and software is most efficient at patterned repetition.

Understanding the PS of a fact-based narrative is the first signal that requires identification. Once removed from the noise

field the remaining clusters can be identified and each deviation from the fact-based narrative can then be characterized, numerically. The numeric characterization can be used when performing analytics in SP processing.

By performing the operations in the specified order of CL preceding PS, an additional benefit may possibly be had in identifying the interactive behavior patterns. Ideally, twenty-one general PS signals should be initially detected. Naturally, machine learning algorithms will need to be evaluated to determine which one offers the greatest accuracy.

Detection of bots in general and chatbots specifically will require additional analysis and likely AI in determining the chatbot signals in those chatbots that are designed to increase amplification in efforts to persuade [53]. In those cases, once identified the dialogue can be analyzed through some of the CL techniques addressed previously. This particular aspect of FN is a separate but related effort.

C. Source provenance (SP)

Data provenance has a long history of research that precedes the introduction of FN. Data provenance details data origination and the process in which the data arrived into the system[54]. According to this definition data provenance can be likened to quality assurance processes surrounding software development where, once again, the job of reliability or in this case the veracity component of examining the content remains ignored. This area can also be revisited and the solution can be more robust than digital ledgers [55, 56] where entities can collude to lie.

By maintaining archive data on fact-based as well as FN the values that associate with these stories additional information can be extracted on authors and publishers. Publishers and authors reputations can be manipulated in attempted to discredit [4], or perhaps increase credibility. Thus reputation analysis becomes problematic. However, by maintaining the first two elements of CL and PS and associating those values over time with both authors and publications new patterns will also emerge.

In the simplest cases reporters who are artificially discredited will be easily detected when examining the body of an author's or publisher's work. Additionally, temporal analysis can detect trends in the same body of works indicating a trend toward FN or fact-based narratives. If authors move to different publishers or publishers change names in an attempt to hide bad reputations the characteristics of their previous work remains allowing for potential matches of emerging entities to existing bodies of work found in the archive.

This final area of FN determination relies heavily on the first two areas being developed. Both areas are in early stage development, so quite possibly other features will become relevant in the SP area. In spite of the lack of details the basic concept can be drawn, recognizing that changes will be incorporated.

IV. DISCUSSION

The proposed model, while not perfect, offers a robust approach that can be easily modified or presented in an easily understood manner. Most credible news stories will likely fall within two standard deviations of the fact-based event reported. By focusing on the wording of fact-based narratives and characterizing these narratives a certain robustness is built in for Byzantine behaviors, which may arise as the propagandists attempt to tailor messages to match the rules of fact-based narratives. This may be less a problem as the text will have to be less emotionally appealing possibly resulting in less efficacy.

A more significant problem with this model will likely be related to linguistic traits and slang expressions across languages, cultures, sub-cultures and tribes. The potential to inaccurately score an article is present. This work will benefit from involvement with a larger community including linguists and social scientists..

A more interesting and potentially more challenging scenario revolves around the improvement of automated behaviors to more accurately reflect human behaviors, this may affect the PS component of the model. Historically speaking, the behavior of presenting propaganda first may also offer insight into detection. As FN pattern signals become better understood, temporal analysis will also provide additional new insights. Finally, as the archive grows and more data analytics are performed on the archived data the ability to distinguish the fact-based narratives from the FN narrative will likely grow in sophistication.

V. EXAMPLE CASE

Before processing can begin, rules must be examined and tested. The first processing component relies on translation of propaganda rules into computational linguistic rules. An example of one of rules of propaganda is that message must appear interesting to the target and use an attention-getting distribution medium [57]. Thus, attention-grabbing headlines complete with pictures displayed on websites and social media sites would be an example.

In English, verbs are action words and adverbs are descriptor words, these words are used to convey what happened and provide details capable of eliciting a response. Thus, the article length and the rate of adverbs may provide a possible marker as a metric deviation from fact. Of course, these alone are not sufficient, but serve as an example for illustrative purposes. Also, considered, but not measured in this particular example, is the role of context in describing event news.

One assumption suggested the news wire (AP News, Reuters or Bloomberg) would report the fact-based narrative; and the model suggests that the fact-based narrative should serve as μ. The example chosen occurred in 2016 when candidate Clinton collapsed at the 9/11 ceremonies in New York City. The source sites were selected based on a Google search for "Clinton collapse 2016". The news sites were NBC News, The Washington Post, Fox News and the New York Post. An additional opinion piece was selected with the purpose of providing observational data on this type of publication.

```
IF (WordCount >= APWordCount) then
```

$$diff = \left(1 - \left(\left(WordCount - \frac{APWordCount}{APWordCount}\right)\right)\right)$$

else

$$diff = \left(\left(WordCount - \frac{APWordCount}{APWordCount}\right)\right)$$

Figure 4: % Difference equation

As expected the AP News story word count was in the middle of the group ranked third of six in order of low to high word count. The word count for the smallest story was 179 words and the largest story was 1331 words. Because only one story was selected for this effort, there are no average values for news stories and no standard deviations (σ). The results of the word count are shown in Table 1. The corresponding bar chart is depicted in fig. 5. Percent based differences were calculated using the logic displayed in fig. 4.

Table 1: News article word count

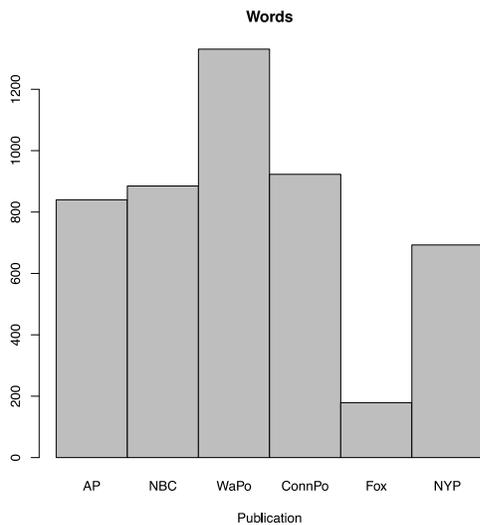| Source	Word Count	% Change
AP News	840	0
Connor Post	923	+10%
Fox News	179	-79%
NBC	885	+5%
New York Post	693	-17%
Washington Post	1331	+58.5%

Figure 5: News article word count

The most dramatic differences can be seen with Fox News and the Washington Post. The negative differences that associate with Fox, and New York Post suggest the potential for missing context, where part of the narrative may be missing. Missing context results in the reader having to mentally complete the story relying on existing cognitive biases. The further the measured distance from the μ, in the negative direction, the greater the potential for the reader to rely on cognitive biases. The New York Post, and to a lesser extent Connor Post show a significant deviation.

The Washington Post showed a large deviation in the opposite direction in terms of word count suggesting that minimally the publication embellishes but in the absence of σ measures the ability to determine normal variance is not yet available. Additionally, rules that separate context from propaganda will require translation into CL terms and software. Presently, much of the CL software will require modifications due to the cleaning of terms that are typically used in propaganda but are problematic for CL (e.g. them).

In both cases of strong deviation off the AP story there are no measures of variance or σ since this is an exemplar while the research continues into determining the optimal list of weighing factors. Word count does determine the positive or negative assignations and initial weight of the deviation. The weights will be modified over time as algorithms are tuned and the archive grows.

Another measure is the usage of adverbs. Because the stories are of varying lengths the measure uses percentage values obtained by dividing the total number of adverbs in an article by the total number of words for that article. Fig. 6 shows the equation used to calculate the percentage change from the AP adverb article percentages. Table 2 shows the resultant numeric differences and fig. 7 depicts the bar chart representation of the adverbs.

$$change = \left(1 - \left(\frac{\%Adverbs}{\%APAdverbs}\right)\right)$$

Figure 6: Equation to determine adverb distribution rate derivation from AP News

Table 2: Adverb rates per article

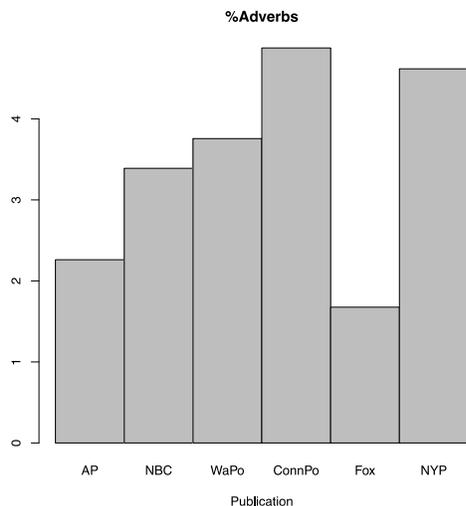| Source	# Adverbs	% Adverbs	% Change from AP
AP News	19	2.3	0
Connor Post	45	4.9	+113%
Fox News	3	1.7	-26%
NBC	30	3.4	48%
New York Post	32	4.6	100%
Washington Post	50	3.8	65%

%Adverbs

Figure 7: Adverb rates per article

The adverb usage helps to add context to the words findings. If verbs are considered action words, then adverbs are the action descriptors that are capable to adding urgency, confusion or other emotions designed to manipulate emotional responses. Descriptors make dramatic, the nouns and verbs that they define.

Working off the assumption that AP News provides the baseline. A quick examination of the values shows that the Washington Post has the most adverbs but when examined in the context of the number of words the deviance from AP both Connor Post (opinion article) and New York Post show a greater deviance from AP News. This positive deviance suggests embellishment, whereas the deviation by the Washington Post and NBC suggests a level of bias in the article. The Fox news deviation may possibly be too small to measure for this particular example; however, when combined with the small word count the possibility that a partial fact is being reported must be considered.

While not counted in this example other items of interest included the punctuation deviations and paragraph sentence counts. Connor Post and Fox both had the punctuation character "?" in their stories. This use of the "?" was of interest for two reasons, first because by asking a question the article provides an entry point to draw in a reader with the hope of engaging the targeted reader into the process. The second, most news stories use a period; thus, the use of "?" or "!" are used for special stories that are designed to elicit a response that is most likely emotional.

One other observation that may result in a marker for opinion pieces was the number of sentences in a paragraph. In the opinion piece (Connor Post) the vast majority of paragraphs contained three or more sentences. This finding was in contrast to all of the other news articles that typically contained single sentence paragraphs or two sentence paragraphs.

As the remaining criteria emerge and the distance from fact-based narratives (AP, Reuters, Bloomberg) the differences can be averaged creating the overall tag value for the article. This tagged value is used to assist in defining characteristics that will be used to define training data characteristics for use in machine learning algorithms. Table 3 contains the overall deviation values for the 6 sources. The values for this table were simple averages using equal weighting of inputs (words and verbs). The overall % values were divided by 10 to create the measure off µ used to fit the overall article scoring scale.

Table 3: Overall deviation measures

Source	Overall %Value	µ Distance
AP News	0	0
Connor Post	61.5%	6.15
Fox News	52.5%	-5.25
NBC	26.5	2.65
New York Post	58.5	-5.85
Washington Post	61.5	6.15

While some of the values suggest a high deviance they should be considered in the context of a single story lacking a corpus of data for base lining and comparisons. As mentioned earlier separating context from propaganda terms will improve the fidelity of the model, as will tuning of the algorithm weighting. One key finding; as suspected, AP News as the center point appears to be a valid assumption since in both word count and adverb rates there were entries above and below the AP News values.

In addition to fact checking trending may also be useful in tuning. The values seen in column three of table 3 represents the final value that gets tagged to the news article. This assigned number can be used to tag or identify the article for observation in the larger stream of articles that represent the second processing phase of PS analysis. The PS analysis will likely be highly dependent upon AI/ML techniques for both tagging and classifying.

The remaining description is of the archive where results from CL and PS are stored and made available for additional analytics. The actual archive has not yet been built, but there are certain pieces of information that are of interest to this area of research. Of note the archive is not designed to compete with existing archives, rather the archive is designed to augment existing archives. The archive primarily provides historic data of the meta-fields and should supplement other news archives. The archive is designed to encourage additional studies by other researchers. The actual design of the archive proposed here is preliminary in nature, and records will most likely be stored as CSV records. Table 4 provides a brief description for each of the fields.

Table 4: News Archive Fields

Field	Description
Identifier	Unique record identifier

Author(s)	Vector contains names of article author(s).
Author(s)-score	Average total CL score for author's other work
Publisher	Publisher name
Publisher-score	Average publisher CL score
Links	Vector with link information to the news story, and other archives.
Metadata profile	Vector containing the values
Topic	Story topic, and related information
Overall article score	Deviation score for the article from μ.
Comparative Scores	List of other CL scores for each of the components used in overall article score
Related stories	List of related stories.
Event date	Date of the news event
Publication Date	Date of publication

Putting together all three components of this model the three components the ability to evaluate any news story will ultimately be supported by all three components. The entire process is designed for both efficiency and the ability to use any single component with high assurance. Fig. 8 depicts the overall processing flow.
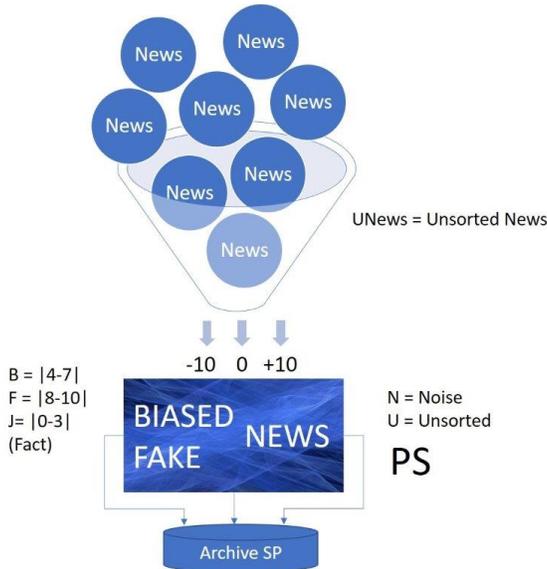

Figure 8: Oview of news processing

VI. Conclusion

The high efficacy along with the lost cost makes propaganda a good weapon in warfare.The ability to manipulate trust through various media relies on a flawed trust model that relies on object oriented constructs [29], resulting in a loss of context. The model presented provides contextual evaluation of news stories offers a rapid and less subjective way to evaluate any news article and provide an objective measure of distance between the GT (or fact-baesd) event and the narrative being presented.

Through the use of agreed upon event reporting metrics this model provides a starting point to evaluate FN in an objective manner. The ability of CL to identify FN has been shown to work [37] on a large scale in a similar model. The PS mechanism shows promise but has not been executed on a large scale to date [41]. The archives are being populated at several higher learning institutions and these institutions are expanding beyond English language based stories. The archive created for this project will agument existing archives by providing metadata characterizations and other relevant information that can add to other data mining efforts.

The ability to perform temporal analysis on the archives that are being built offers great promise because the findings can be combined with cultural and linguistic models that may ultimately identify vulnerable traits and ways in which populations can be quickly inoculated based on the identified traits.While Cambridge Analytica used data science techniques on personal data to identify potential targets, data science combined with cultural frameworks can be used for benevolent purposes.

We conclude with the observation that propaganda has been a long standing problem with FN on the Internet elevating the effectiveness of this tool. Standalone point solutions run the risk of repeating the mistakes of the signature-based model that prevailed during the early days of Internet security where those solutions created a false sense of security; therefore, the ultimate solution will likely take time, and require contextual evaluation of events. We suggest the model presented here will meet these new and comprehensive requirements.

VII. Acknowledgements

References

[1] Oxford dictionaries, website, https://www.oxforddictionaries.com/press/news/2016/12/11/WOTY-16. 2017.

[2] Collins dictionary, website, https://www.collinsdictionary.com/woty. 2017.

[3] D. L. Altheide. *"Creating fear: News and the construction of crisis"*. Routledge, 2017.

[4] L. Gu, V.Kropotov, & F. Yarochkin. "The Fake News Machine: How Propagandists Abuse the Internet and Manipulate the Public", Trend Micro, 2017.

[5] Cambridge dictionary, website, https://dictionary.cambridge.org/us/dictionary/english/actuary .

[6] J., Gottfried, and E. Shearer. "News use across social media platforms 2016", *Pew Research Center Journalism & Media.* Available: http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/, (2016 May 26).

[7] Oxford University website, https://medium.com/oxford-university/where-do-people-get-their-news-8e850a0dea03, 2018.

[8] M. Ristolainen. "Should 'RUNet2020' be taken seriously? Contradictory views about Cybersecurity between Russia and the West", *ECCWS conference proceedings*, Dublin, Ireland, 2017.

[9] A. Entous, E. Dwoskin, and C. Timberg. "Obama tried to give Zuckerberg a wake-up call over fake news on Facebook." *Washington Post* 2017.

[10] N. Verrall and D. Mason. "The Taming of the Shrewd", *The RUSI Journal*, DOI: 10.1080/03071847.2018.1445169. 2018.

[11] M. Connel and S. Vogler. "Russia's Approach to Cyber Warfare", *Center for Naval Analysis*, http://www.dtic.mil/dtic/tr/fulltext/u2/1032208.pdf, 2017.

[12] R. J. Hexter. "*What Was The Trojan Horse Made Of?*: Interpreting Virgil's Aeneid.": 109-131, 1990.

[13] A. Badawy, E. Ferrara, and K. Lerman. "Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign." *arXiv preprint arXiv:1802.04291* 2018. https://arxiv.org/pdf/1802.04291.pdf.

[14] H. Grassegger, and M. Krogerus. "The data that turned the world upside down." *Vice Motherboard* 28,2017.

[15] S. Rosenblatt. "Excerbating our fake news problem: Chatbots" https://www.the-parallax.com/2018/03/26/fake-news-chatbots/, (March 26, 2018).

[16] H. Ingram. "A brief history of propaganda during conflict: Lessons for counter-terrorism strategic communications, 2016". Retrieved from https://www.icct.nl/wp-content/uploads/2016/06/ICCT-Haroro-Ingram-Brief-History-Propaganda-June-2016-2.pdf.

[17] BBC News https://www.bbc.com/news/world-europe-29478415.

[18] P. Pomerantsev and M. Weiss. "The menace of unreality: How the Kremlin weaponizes information, culture and money". *The Interpreter*, *22*, 2014.

[19] H. Volodymyr. The "Hybrid Warefare" Ontology. *Фахове видання з економічних, філософських, політичних наук Затверджено постановами Президії* ВАК України від 26 січня 2011 р. № 1. 2016.

[20] C. Sample, J. McAlaney, J.Z. Bakdash and H. Thackray. "A Cultural Exploration of Social Media Manipulators", *Proceedings of the 17th European Conference on Cyber Warfare and Security,* Oslo, Norway, pp. 342 – 341, 2018.

[21] G. Jowett, and V. O'Donnell. Propaganda and Persuasion, 5th Edition. Chapter 2 – "*Propaganda Through the Ages*." Thousand Oaks, CA: Sage Publications, Inc. 2011.

[22] R. Marlin. "*Propaganda and the ethics of persuasion*", in Editor (Ed)., *Book Propaganda and the Ethics of Persuasion,* 2013.

[23] P. Taylor. *Munitions of the Mind: A history of propaganda from the ancient world to the present day*. Manchester, NY: Manchester University Press, 2003.

[24] P.M. Taylor. *War and the Media: Propaganda and persuasion in the Gulf War,* 1992 https://books.google.com/books?hl=en&lr=&id=V9tRAQAAIAAJ&oi=fnd&pg=PR7&dq=taylor+1990+propaganda+persuasion&ots=AG3yYpVQTP&sig=9ru8pRygFicbOYd2J_RGKbrzERs#v=onepage&q=taylor%201990%20propaganda%20persuasion&f=false.

[25] P.A. Hayward. *Factoids, Dishonesty and Propaganda in the Middle Ages,* 2018.

[26] A. Pollok, and D.U. California. *Roman Propaganda in the Age of Augustus,* Dominican University of California, 2017.

[27] J. Lively. *"Propaganda Techniques of Civil War Cartoonists"*, *The Public Opinion Quarterly*, Vol. 6 (1), pp. 99-106, 1942.

[28] D. Hinklemann, *Struggle for Power: By Any Means,* Create Space Indpendent Publishing Platform, 2016.

[29] . Nisbett. *The Geography of Thought: How Asians and Westerners Think Differently*. Simon and Schuster, 2010.

[30] G. Hofstede, G.J. Hofstede, and M. Minkov. *Cultures and organizations,* New York, NY: McGraw-Hill Publishing, 2010

[31] W.A. Sheppard. "An exotic enemy: Anti-Japanese musical propaganda in World War I Hollywood, *Journal of the American Musical Society,* Vol. 51 (12) pp. 303-357, 2001.

[32] M. J. Marks. "Teaching the Holocaust as a Cautionary Tale." *The Social Studies* 108.4 : 129-135., 2017.

[33] A. Yuhaniz, F. Zainon, Z. Ghazali, N. Man, F. M. Alipiah, and M. Y. M. Yunus."The Influence Of Music On Memorization Performance Of Mathematics Students." *Proceedings of the ICECRS* 1.2 2018.

[34] J. Swift. The art of political lying", *The Examiner 14, 1710.*

[35] J. B. Whitton. *Cold War Propaganda,* 2018.

[36] E. Darraj, C. Sample and J. Cowley. "Information Operations: The use of weaponized information in the 2016 US presidential election", *Proceedings of fthe 16th European Conference on Cyber Warfare and Security*, Dublin, Ireland, pp. 113-119, 2017.

[37] B. Horne and S. Adali. "This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News", *Association for the Advancement of Artificial Intelligence.* 2017.

[38] Website: http://nelatoolkit.com

[39] Snopes website, https://www.snopes.com

[40] Website: https://www.politifact.com

[41] Website: https://osome.iuni.iu.edu/tools/trends/#

[42] R. Faris, H. Roberts, B. Etling, N. Bourassa, E. Zuckerman, and Y. Benkler. "Partisanship, propaganda, and disinformation: Online media and the 2016 US presidential election." (2017). https://dash.harvard.edu/bitstream/handle/1/33759251/2017-08_electionReport_0.pdf

[43] M. Burgess. "The Emmanuel Macron email hack warns us fake news is an ever-evolvong beast", *Wired*, https://www.wired.co.uk/article/france-election-macron-email-hack

[44] M. Tanis, and T. Postmes. "A social identity approach to trust: Interpersonal perception, group membership and trusting behaviour", European Journal of Social Psychology, Vol. 35, no. 3, pp. 413-424, 2005.

[45] J. H.Cho, K. Chan, and S. Adali. "A survey on trust modeling", *ACM Computing Surveys,* vol. 48, No. 2, Article 28, October, 2015. DOI: http://dx.doi.org/10.1145/2815595.

[46] Facebook website: https://www.facebook.com/pg/facebookcareers/photos/?tab=album&album_id=1655178611435493

[47] J. Fox. "Why Twitter's mission statement matters", *Harvard Business Review,* 2014. https://hbr.org/2014/11/why-twitters-mission-statement-matters

[48] O.Parent and A Zouache. "Role of peer effects in social protest. Evidence from the Arab spring.". http://erf.org.eg/wp-content/uploads/2018/02/Olivier-Abdallah.pdf. 2017.

[49] C. Sample, T. Watson, S. Hutchinson, B. Hallaq, J. Cowley and C. Maple. "Data fidelity: Security's soft underbelly", *Proceedings of the 11th International Confereince on Research Challenges in Information Systems,* Brighton, UK, pp. 315 – 321, 2017.

[50] M. DeLucia, S. Hutchinson and C. Sample. "Data fidelity in the post-truth era: Network Data", *Proceedings of the 13th International Conference on Cyber Warfare and Security,* Washington, DC, pp. 149-158, 2018,

[51] S. Visoughi, D. Roy and S. Aral. "The spread of true and false news online:, *Science,* Vol. 359 no. 6380, pp. 1146 – 1151, 2018.

[52] L. Shifman. "Memes in a digital world: Reconciling with a conceptual troublemaker", *Journal of Computer-Mediated Communication*, Vol. 18, pp. 362-377, 2013. Available at: https://academic.oup.com/jcmc/article/18/3/362/4067545.

[53] S. Rosenblatt. "Exacerbating our fake news problem:Chatbots, March 26, 2018. Available at: https://www.the-parallax.com/2018/03/26/fake-news-chatbots/.

[54] P. Buneman, S. Khanna, and W. Tan. "Why and Where: A Characterization of Data Provenance", In *International conference on database theory pp.316-330.* Springer, Berlin, Heidelberg.

[55] S. H. Ammous."Blockchain technology: what is it good for?" Available at: http://capitalism.columbia.edu/files/ccs/workingpage/2016/ammous_blockchain_technology_.pdf.

[56] M. Staples. "Blockchain is useful for a lot more than bitcoin", *The Conversation*, 2016. Available at: http://theconversation.com/blockchain-is-useful-for-a-lot-more-than-just-bitcoin-58921

[57] PsyWarrior Website: http://www.psywarrior.com/Goebbels.html