

---

# ByWire Trust Indicator

Jetze Sikkema<sup>1</sup>

<sup>1</sup> ByWire, 35 Little Russell St, London, United Kingdom

---

June 19, 2021

In this article we describe various methods to determine the trustability of articles and how they are combined into the ByWire TrustIndicator.

## 1 Introduction

Fake news has been around as long as there has been news. Octavianus famously depicted his rival Marcus Antonius as a drunk and womanizer that would be a disaster for traditional roman values [4] (Apart from the fact that Brutus and co. sacrificed their lives to save the republic from a authoritarian leader by murdering Ceasar). While there are many excellent propaganda posters from WWI depicted the German soldiers as brutes while the Allied forces were very civilized and courageous (See figure 1). Finally the Nazi regime was excellent in producing fake news declaring the battle of Stalingrad won, while the victorious German troops were mopping up the remaining red army, completely ignoring the fact that sixth army had been cut off by 1.1 million red army soldiers [2]. On the opposite site the sniper, Vasily Zaytsev described in his memoirs of the battle of Stalingrad a sniper duel (which forms the core of the movie *Enemy at the Gates*), which most likely never took place at all [3], though without the duel he still remains the most successful marksman ever. So fake news has been around ever since, so what constitutes the current surge in interest? If one looks closely at the examples, it quickly becomes clear that they were issued by state actors that had control over the media and press. Nowadays the media landscape becomes much more fractured and either state actors can influence the media in other states (i.e. Russia buying Facebook ads in the US), or non-state actors can have a significant influence on the media (i.e. The Brexit campaign on Facebook). Also one does not need to control significant resources to start a new communication channel that can reach large parts of the



**Figure 1:** Various propaganda posters from WWI. They can either present an over negative image of the enemy or an over positive image of the own position (Imagine all the happiness you would feel, once you were at the WWI trenches at the front)

population (i.e. youtube influencers, websites as infowars). So in short, the growth of importance of fake news is defined by the loss of influence of state actors and media on their own state, which is replaced by the growth of influence of non-state actors who's identity is often easily hidden and who can have an outsized influence on public opinion for the budget they have. Independent of whether the fake news is created by traditional states, media or new non-state actors what are the ways in which it can be detected and fought? Fake news, especially in non-traditional media, needs to spread in order to effectively influence the public targeted. Research has show that people are much more likely to spread news in case it provokes anger or makes them feel good and confirms or strengthens existing biases and is easy to grasp. News that makes one feel sad, shamed, guilty or even jokes are much less spread. Especially news that sends a double message that is conflicting with ones own confirmation bias is much less likely to spread. This provides us several tools in which one can detect if a message was designed to be spread

- News that is unusual positive or unusually strong provoking anger are more likely to be designed to spread fake news.
- News that uses more than usual simplified language (words or sentence construction), or is shorter than usual is more likely to contain fake news.
- News containing many words statistically correlated with fake news are more likely to contain fake news
- News written by authors that publish often than usual fake news are more likely to publish fake news.

Of course all measures are determined by more than usual. One cannot expect the Guildford Evening Courier[1] to have the same style, word choice and article length as Foreign Policy. Finally statistically sampling for abnormal patterns can be very effective as shown by the use of Benfords law to detect accounting fraud. One can sample the words used and see if certain words are more likely to be used for faked news as for actual news.

One complicating factor is that fake news is often a sliding scale. Messages can be outright fabricated, but it much more common that part of the news is true but distorted or exaggerated or part of the facts are made up. For this reason we combine different detection techniques into a final score that tells how trustable we deem a news message.

What makes fake news a more relevant topic than before, however is that before the main actors producing fake news were the state and the major publishers. However it has become more easy to spread fake news, while at the same time due to falling budgets news publishers have less budgets to fund their own research to verify the stories.

- Actors can easily hide their identity and one actor can impersonate many identities to spread the same message, making it appear that many actors are involved in spreading the news.
- Actors can easily target large target populations through targeted ads on social media
- Actors can pay people with a large following to spread their news.
- Bots can easily make it appear that there is quite a buzz going on about a news item
- We can not go and check ourselves, so we need to trust someone to tell us whether it was true or not.

## 2 Definition of Fake News

Fake news is a general term that is hard to pinpoint as it contains several uses. In order to get a clear understanding we look at the different uses and define it in each instance what it mean. Starting from the in-

**Table 1:** *Different Uses of Fake News*

Name	
Group Identification	Believe to Belong Division Identification
Opponent Identification	
Mobilize Base	
Influence Opinion	
Distract & Deflect	
Create Unrest	

tended use we can more easily identify the techniques employed.

- **Group Identification.** Obvious false facts to identify as group. A prime example is the US where mask wearing has been used to identify political inclinations instead of recognizing the obvious truth that it is beneficial for all people to wear a mask. In order to belong one needs to make belief.
- **Opponent Identification.** Distorted facts to divide own public from the rest. An excellent example of this class is the statement that Black Live Matter protesters are looting scum. Which will of course drive a division between left and right.
- **Motivate Base.** Distorted facts to convince own public and or allies. Examples of this are the gulf of Tonkin incident also the “proof” of Iraq owning nuclear weapons.
- **Swing undecided.** This a relatively minor aim as mostly opponents are not easily convinced. However it can be used to swing the opinion of undecided voters.
- **Distract and Redirect.** Creating anger about an unrelated subject helps to deflect attention from the relevant facts.
- **Create Unrest.** This allows for a strongman to step in.

From this we can develop the following definition of Fake News. Fake news is an piece of information that.

- Is engineered to spread. This makes should make the news outstanding and evoke strong emotions (in particular anger and joy) [6].
- It needs to be understood immediately, which require simple language with respect to the target group and not allow for nuances or partial truths.
- It needs to create a clear distinction between opponents and proponents
- It needs to be factually incorrect (i.e. engineered).

## 3 Theory of Fake News Detection

Now that we have a definition of what fake news is we can identify the different points that distinguish it and

**Table 2:** Patterns to identify Fake News

Name	
Factual Checking	Manual
Style	Automatic
	Quantity
	Complexity
	Uncertainty
	Complexity
Propagation	Subjectivity
	Non-immediacy
	Diversity
	Specificity
	Sentiment
Source	Emotional Content
	Distribution Channels
	Spread on Social Media (Retweets)
Correlation	Author
	Publisher
	Words correlation
	Factual Similarities
	Source Similarities

Experts  
Crowd Checking  
Predicate Testing

**Table 3:** Style Elements to identify Fake News

Name	
Quantity	Character Count
	Word Count
	Noun Count
	Verb Count
	Number of noun phrases
Complexity	Sentence Count
	Paragraph Count
	Modifier Count (Adjectives/Adverbs/...)
	Avg. Number of clauses per sentence
	Avg. Number of words per sentence
	Avg. Number of characters per word
	Avg. Number of punctuation per sentence
	Uncertainty
	Percentage of modal verbs
	Percentage of certainty terms
Subjectivity	Percentage of generalizing terms
	Percentage of tentative terms
	Percentage of numbers and quantifiers
	Number of question marks
	Percentage of subjective verbs
Nonimmediacy	Percentage of report verbs
	Percentage of factive verbs
	Percentage of imperative commands
	Percentage of passive voice
	Percentage of rhetorical questions
Diversity	Self reference: 1st person pronouns
	Group reference: 1st per. plural pronouns
	Other reference: 2, 3 per. pronouns
	Number of quotations
	Number of exclamations marks
Specificity	Activation: dynamics of emotional state
	Lexical diversity (unique words)
	Content word diversity
	Reduncancy
	Informality Typographical error ratio
	Temporal Ratio
	Spatial Ratio
	Sensory Ratio
	Causation Terms
	Exclusive Terms
	Readability (Flesch-Kincaid/Gunning-Fog)

use these to detect when a news item is fake [7].

### 3.1 Factual Checking

Fact checking is of course the most obvious way to assess whether news is fake. It is the most important way at the moment to determine if news was faked. However it also has sever drawbacks

- Fact Checking is labour intensive as it involves a significant amount of manual checking. Even the machine learning algorithms will require a significant input of manually checked facts.
- The facts needs to come from somewhere, which means one would either need to trust experts, or reference works.
- As seen in the introduction fake news can be about the way facts are represented. So even though the content might be true they may be place into a context that misrepresents the facts.

Currently at bywire we are using fact checking only on an ad-hoc basis. The core algorithm comprises of indicators that can be easily employed on a statistical basis.

### 3.2 Style Checking

Style checking is an important indicator

### 3.3 Correlation

In order for fake news to be more effective the same message is repeated often in different forms. Also cer-

**Table 4:** Patterns to identify Fake News

	Name
Factual Checking	Manual
	Automatic
Style	Quantity
	Complexity
	Uncertainty
	Complexity
	Subjectivity
	Non-immediacy
	Diversity
	Specificity
Propagation	Sentiment
	Emotional Content
	Distribution Channels
	Spread/Retweets
Source	Author
	Publisher
Correlation	Words correlation
	Factual Similarities
	Social Media Pickup
	Source Similarities

**Table 5:** Patterns to identify Fake News

	Name
Factual Checking	Manual
	Automatic
Style	Quantity
	Complexity
	Uncertainty
	Complexity
	Subjectivity
	Non-immediacy
	Diversity
	Specificity
Propagation	Sentiment
	Emotional Content
	Distribution Channels
	Spread/Retweets
Source	Author
	Publisher
Correlation	Words correlation
	Factual Similarities
	Social Media Pickup
	Source Similarities

tain terms are more likely to provoke a strong response. At ByWire we use the following measures

- Dictionary with known correlation with fake news per word and word group. This dictionary is improved upon over time.
- Correlation with other articles in the bywire database.
- Correlation with other articles on the internet.
- Correlation with other articles by the same source (This is treated in the source section).

### 3.4 Propagation

An essential facet of fake news is that it is propagated to reach a significant audience. A perfectly engineered message is useless when it doesn't arrive at the target audience. At bywire we distinguished the following means to engineer propagation [6], [sivek2019].

- Messages provoking strong feelings of Anger and Trust
- Number of shares / retweets.
- Language used (this is taken care of by the style indicator).
- Number of search results when searching for the message text.

### 3.5 Source

One of the most important distinguishing factors in fake news detection is the source [5]. This is something

one knows intuitively, as one is likely to trust the Times more than Cosmopolitan. However this does not mean one is less likely to read Cosmopolitan. Since it depends also strongly on the topic (i.e. Cosmopolitan over the Times for celebrity gossip). At ByWire we implemented the following algorithm.

- For each source (newspaper and writer) calculate an overall score and an topic score based for all articles published.
- Calculate a score per topic.
- For each channel (twitter, site, facebook) calculate an overall score and an topic score based for all articles published.
- For authors combine the different papers for which they write in such a way that most discerning power is obtained.
- Combine each score in a way that adds most discerning power.

## 4 Methodology

The algorithm to detect fake news a bywire was constructed using standard data science techniques. First a training and test set were developed using manually labelling fake and non-fake messages. Based on this test set

- Optimize the coefficients of each individual measure.
- Determine of each individual measure whether it has predictive power to discern between fake news and factual news.

- Test on test data to see if the predictive power holds up.
- Calculate a score per category to determine which elements of a fake news message are pronounced.
- Combine the individual measures to improve the predictive power.
- Test on test data to see if the predictive power holds up.
- Test on out of sample data to validate that the constructed model is stable and not an artifact of the dataset (overfitting).

## 5 Results

### Bibliography

- [1] Douglas Adams. *The Hitchhikers Guide to the Galaxy*. 1979.
- [2] Jay W. Baird. “The Myth of Stalingrad”. In: *Journal of Contemporary History* 4.3 (1969), pp. 187–204. DOI: 10 . 1177 / 002200946900400312. URL: <https://doi.org/10.1177/002200946900400312>.
- [3] Frank Ellis. *The Stalingrad Cauldron. Inside the Encirclement and Destruction of the 6th Army*. 2013.
- [4] Isabelle Kaminska. “A lesson in fake news from the info-wars of ancient Rome”. In: *Financial Times* January 17 (2017). URL: <https://www.ft.com/content/aaf2bb08-dca2-11e6-86ac-f253db7791c6>.
- [5] Jennifer Grygiel Xinyi Zhou Reza Zafarani Niraj Sitaula Chilukuri K. Mohan. “Credibility-based Fake News Detection”. In: *Arxiv* 1911.00643 (2019). URL: <https://arxiv.org/abs/1911.00643>.
- [6] Sinan Aral Soroush Vosoughi. “The spread of true and false news online”. In: *Science* 359.6380 (2018), pp. 1146–1151. DOI: 10.1126/science.aap9559. URL: <https://science.sciencemag.org/content/359/6380/1146>.
- [7] Usareza Zafrani Xinyi Zhou. “A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities”. In: *Arxiv* 1911.00643 (2020). URL: <https://arxiv.org/abs/1812.00315>.