Unmasking the News: Emotion and Content Analysis of Fake News

Daniel Vasilyev
Oliver Cheng
Renata Krivenko*
dmv4@stmarys-ca.edu
okc2@stmarys-ca.edu
rk22@stmarys-ca.edu
Saint Mary's College of California
Moraga, California, USA

Abstract

The rise of fake news has contributed to many negative and adverse effects on our society, such as increasing political polarization, spreading misinformation, and increasing the chances of public health risks. Research has suggested that false information spreads faster than real news, often appealing to emotions rather than logic, and being more persuasive when people examine the information using emotional intuition rather than critical analysis. It is also more persuasive to individuals who analyze it using emotional intuition rather than critical thinking. Building on this foundation, our study investigates the emotional differences between fake and real news headlines and articles and how the themes of fake news content evolve over time. By merging several Kaggle datasets, we gathered data on fake news articles from 2015 to 2018 and real news articles from 2016 to 2017. We then conducted an emotion analysis of fake and real news articles from 2016 to 2017 using a Transformers package that assessed seven emotions: anger, disgust, fear, joy, neutral, sadness, and surprise. The analysis revealed that fake news titles contained more anger and disgust than real news titles, while real news titles were more neutral. Additionally, fake news text showed higher levels of anger and fear, whereas real news text had more joy and neutral emotions. Following this, we applied thematic analysis using an LDA model to examine the themes in fake news from 2015 to 2018. Our findings revealed consistent topics, like Donald Trump, and more situational or year-based topics like Business and Health, indicating a response to trending public interests. In future research, it will be important to explore the role of emotional appeal in spreading fake news and its impact on public behavior. Including more years in the dataset, particularly notable periods like 2020 during the COVID-19 pandemic and 2022 amid the Ukraine War, could help uncover how emotional and thematic trends change over time.

*All authors contributed equally to this research.

Conference acronym '25, 978-1-4503-XXXX-X/2025/05

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

ACM Reference Format:
Daniel Vasilyev, Oliver Cheng, and Renata Krivenko. 2025. Unmasking the News: Emotion and Content Analysis of Fake News. In *Proceedings of (Conference acronym '25).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Fake news is a term that has been widely used by many people over the last nine or so years. It has been labeled as devastating to democracies and has been identified as one of the most powerful tools in creating division and polarization in the United States. But what is fake news? According to [2], this is a very broad term that encompasses news displayed as an honest and accurate source of information, but which is, in fact, deceptive and false.

Previous research has gone into the harmful effects that fake news can have on the population. For example, it has been shown to have a large prevalence during the 2016 election [1]. The study revealed that the amount of fake news increased during and surrounding the election, and it estimated that approximately 14% of people had fake news on social media as their primary source of news information. Studies have also extended beyond the prevalence of fake news in politically related domains and have demonstrated how it can transcend various contexts. Research has been conducted on health-related misinformation on social media. For example, one study found that there is health misinformation, like anti-vaccine claims, that leads to harmful behaviors and public health risks [10]. All of this research into how fake news can penetrate and affect different domains is united under a much larger problem. [8] found that lies spread faster than the truth. The study calculated that the top one percent of fake news was disseminated to 1000 to 100,000 people, whereas real news and the truth rarely reached more than 1000 people. This shows that not only is fake news damaging to our society, but the organization of our social systems, along with the psychological effects of fake news, make it extremely powerful in its ability to spread quickly.

This raises the question: Why is fake news so effective? The literature suggests that this is due to two main reasons. Firstly, studies have indicated that part of the reason people believe fake news is not because they have prior beliefs and attitudes that influence their biases, but instead that people do not consume the content with analytical thinking [6]. The study found that people often seem to believe misinformation, even when it contradicts their prior attitudes, and being resistant to fake news was more likely to result from critical thinking. The second reason fake news is so effective, which relates to the first, is that it excels when people evaluate the information using their emotional intuition [4, 5]. People, when they went with their "gut instinct", were much more susceptible to fake news.

While some research has been conducted into the psychological and thematic aspects of fake news, the majority of studies have focused on the methodological aspects, specifically how to identify fake news. Fake news detection started from traditional machine learning and has evolved into more complex approaches like deep learning, which can take into account multiple aspects of an article, like the text and images [3, 9]. While not yet 100% reliable, the fake news detection approaches have reached a reasonably high level of sophistication and accuracy. This has helped pave the way for different types of research into fake news. For example, the content. One study found that, before the pandemic, the most common themes in fake news included scientific, political, cultural, financial, and several other topics [7]. Our research aims to expand upon this by examining how the content and themes of fake news evolve, as well as comparing them to real news to identify their differences and similarities.

2 Methods

2.1 Data

The data was sourced from Kaggle. For fake news, two datasets were merged to cover the years 2015, 2016, 2017, and 2018. This dataset includes the article titles, full text, and year of publication. The real news data, also from Kaggle, covers the years 2016 and 2017 and includes the article titles, full text, subject, and date of publication. As a preliminary analysis, we did a word-level analysis of the twenty most common words and bigrams for all of the years (2015, 2016, 2017, and 2018) in the fake news dataset to gain an understanding of the data and preliminary insights into common topics and themes prominent in fake news. We then applied deep learning techniques to compare the sentiment between fake news and real news for the years 2016 and 2017. Finally, we did topic analysis on the fake news data to gain insight into how the dominant topics in fake news changed between 2015 and 2018.

2.2 Preliminary Analysis

For the preliminary analysis, we used SpaCy, a natural language processing (NLP) model, to determine which words and phrases were most commonly used in fake news from 2015 to 2018. This provided some insight into the common themes and topics in the news for the time period. We found several recurring themes like "trump", "hillary", "obama", "state", "president", etc. The bigrams provided more information as to how the words were related to each other for each year: "trump" went with "donald trump", "hillary" went with "clinton", "obama" went with "president" and "barack", and "state" went with "secretary".

2.3 Sentiment Analysis

To understand how sentiments change between fake news and real news, we analyzed the years 2016 and 2017 for fake news and real news, and used transformers, which are neural networks, to understand how words and phrases relate to various sentiments or emotions. The transformer package that we used categorized the emotions into seven different categories. The categories were "anger", "disgust", "fear", "joy", "neutral", "sadness", and "surprise". When running the analysis, we had to limit the analysis to 1000 topics for each type of news, because the run time was taking too long in Google Colab.

2.4 Topic Analysis

To understand how the topics of fake news changed between 2015 and 2018, we applied the Latent Dirichlet allocation (LDA) transformer model, which uses the Bayesian network to automatically extract topics in textual corpora. To analyze the topics, we extracted the 6 most popular topics in each year, and based on the words in each topic, we categorized them into categories that seemed to make the most sense. This was due to the fact that the model only applied numerical labels for each topic from zero to five, and therefore, to give more meaning to the model, we had to correlate the most common words in each topic to give a name for each topic. For example, for topic two in 2016, the most common words were "clinton", "fbi", "russian", "trump", "say", "investigation", "russia", "report", "intelligence", and "hillary". To tie these words together into a topic that made sense, we chose to name the topic "Hillary Scandal". We repeated this process of human input for all of the topics to convert the numerical topic labels into contextual topic labels.

3 Results

3.1 Word Frequency

Word Frequency analysis from 2015-2018 for fake news text revealed that there were spoke recurring themes, like Trump, but did not give too much insight into any more holistic themes or evolution/differences between the years. To expand on this analysis, we also did n-grams and found that Donald Trump was a very common theme among all of the years. While this showed us some interesting insight, we decided that to answer our research questions we would need to expand into more complex data science techniques.

3.2 Topic Analysis

The topic analysis identified six themes for each year of fake news coverage (as seen in figures 1 through 4), with keywords clustered based on semantic similarity. These clusters were reviewed by all of the authors and were interpreted towards overarching topics and themes (such as Donald Trump). The analysis demonstrated consistency across all the years and differences, showing that fake news content and themes are based on relevancy. We can see that Donald Trump was present from 2015 to 2018, which makes sense given that he was a presidential candidate and president of the United States all these years. Other big topics throughout the years feature some themes like cultural conflicts and foreign affairs, for example, the United States' involvement in the Middle East and the United States' relationship tensions with Russia.

3.3 Emotion Analysis

To better understand the differences between fake and real news, emotions were analyzed for the text and title of each article.

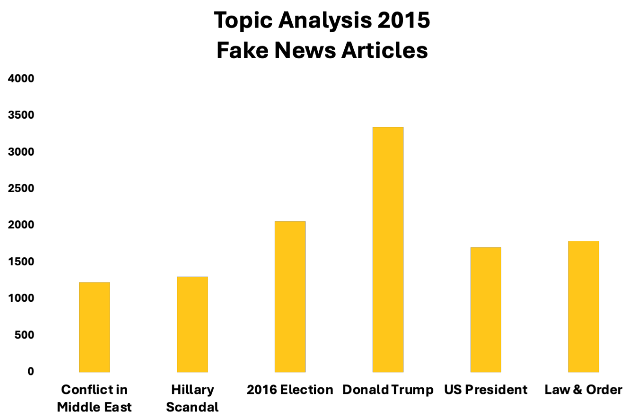

Figure 1

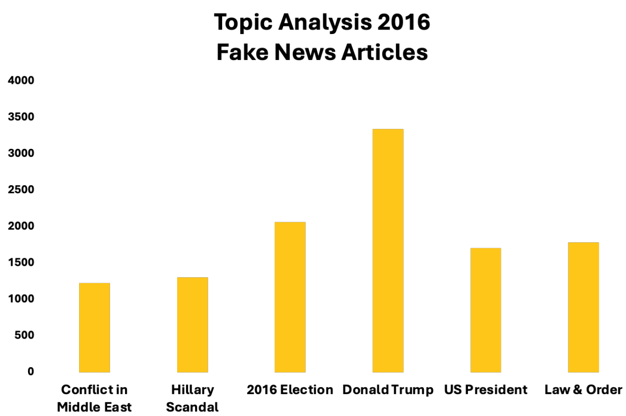

Figure 2

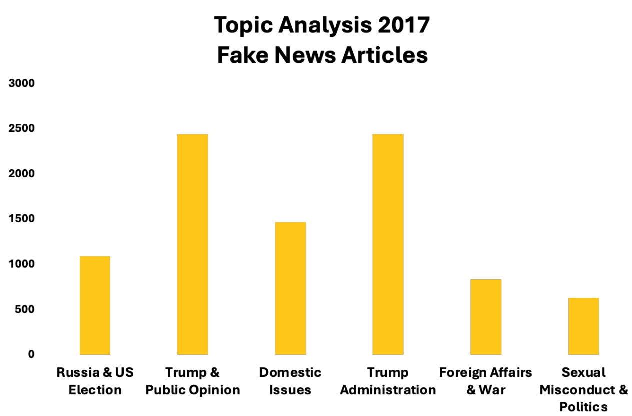

Figure 3

3.4 Title

The analysis of article titles showed that neutral was the most common emotion for both real and fake news titles (Figure 5).

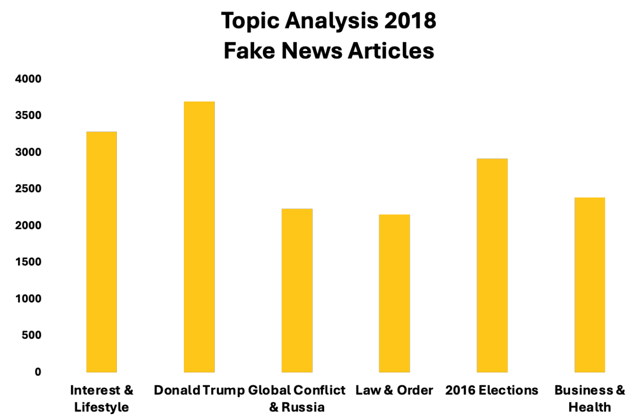

Figure 4

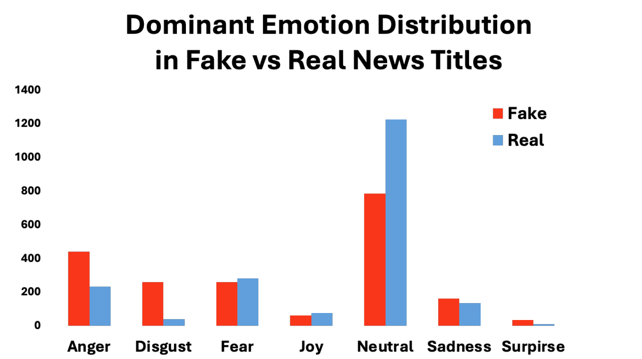

Figure 5

However, fake news headlines contained more anger and disgust than real news. In contrast, emotion expression in real news titles was more moderate, with no particular emotion (aside from neutral) standing out significantly. This suggests that both real and fake news rely on neutral framing to some extent, but phony news more frequently incorporates negatively emotionally charged language.

3.5 Text

Anger and fear emerged as the most dominant emotions for both fake and real news articles (Figure 6). This aligns with previous research that indicates a strong negative valence of emotions is common in news articles. However, there were notable differences in the presence of other emotions between real and fake news. Real news contained substantially more joy and neutral sentiment than fake news. This suggests that real news has a more even spread of emotional valence across positive and negative for the titles of articles, compared to fake news.

4 Discussion

This study has examined the similarities and differences between real and fake news and investigated how the topics of fake news evolve and change over time. To do this, we merged Kaggle sources to create a more comprehensive dataset with four years of fake

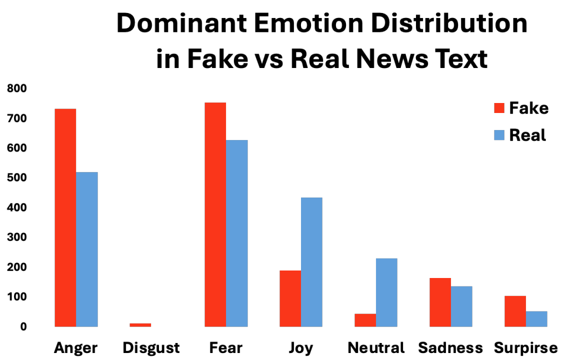

Figure 6

news data (2015-2018) and two years of real news data (2016-2017). Using a variety of Natural Language Processing techniques, such as sentiment analysis and topic analysis, we were able to draw some conclusions about the data.

We found that a key difference between real and fake news is the extent to which emotions are present in the articles. Specifically, we found fake news headlines had more emotions of disgust and anger, while real news headlines tended to have more neutral emotional tones. For the article text, we found that fake news has more negative emotions, while real news has more joy and neutral emotions. These findings align with previous research that indicates that fake news spreads faster than real news and that fake news plays more on people's emotions. The more emotion the article can evoke from a person, the less rational they will be in their analysis of the material, and the more likely they are to feel stronger about the article's topic. These findings help to contribute to the growing body of literature on fake news and help support and explain why fake news is so powerful.

In this research, we have helped expand on the fairly small literature surrounding thematic analysis of fake news content. While some previous research has examined the general themes that are prevalent, there hasn't been much if any analysis on how the themes evolve and change, or stay similar, over time. The results of our analysis were generally consistent with what we anticipated. There were some similarities and differences between the years. The similarities, like Donald Trump, seem to be ever constant as the topic of Donald Trump himself is ever constant during these years, as he was president of the United States. But there are also some topics that are unique to the years, demonstrating that fake news tries its best to stay relevant with whatever is going on, thereby making it more likely to be consumed by people. Combining relevancy with a powerful emotional valence makes fake news very dangerous and potent, reflecting its impact on the country for the last decade or so.

While this research has some interesting results and implications, there are some limitations that are important to address. Firstly, the data that we sourced was a combination of Kaggle datasets. While previous papers and studies have used the same datasets and have verified their validity, it is still important to note that we ourselves did not collect the article data and that it may not be as valid as some previous papers have claimed. Secondly, we used many Natural Language Processing techniques for our analysis. While these methods have been used by many data scientists for quite a long time, they are definitely prone to error and not perfectly reliable. It would be useful to run multiple different Natural Language Processing techniques to analyze the topics and emotions and see whether the results are consistent or change between the different approaches.

Looking ahead, there are several avenues of future directions towards which this research can be built. As mentioned earlier, increasing the number of different NLP techniques to see if the results are consistent. Additionally, it would be useful and interesting to see how the emotion differences specifically affect the audience of both real news and fake news and how different emotions lead to different reactions and behaviors to the article. Another promising avenue could be to see how the emotions connect to each topic that is prevalent within the articles and see whether certain topics are connected more with certain emotions and see what patterns unfold there. For example, more emotion of disgust for fake news surrounding health related topics. Exploring these future directions could lead to a better grasp of how fake news operates and spreads. With that understanding, we may be able to develop smarter ways to reduce its reach and lessen the harm it causes the United States.

References

[1] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31, 2 (2017), 211–236.
[2] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096.
[3] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting rumors from microblogs with recurrent neural networks. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. AAAI Press, 3818–3824.
[4] Cameron Martel and Gordon Pennycook. 2020. Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications* 5, 1 (2020), 47. doi:10.1186/s41235-020-00252-3
[5] Cameron Martel, Gordon Pennycook, and David G Rand. 2020. Go with your gut: How reliance on affective intuition promotes belief in fake news. *Unpublished manuscript* (2020).
[6] Gordon Pennycook and David G Rand. 2019. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* 188 (2019), 39–50.
[7] Marta Pérez-Escolar, Darren Lilleker, and Alejandro Tapia-Frade. 2023. A systematic literature review of the phenomenon of disinformation and misinformation. *Media and Communication* 11, 2 (2023). doi:10.17645/mac.v11i2.6453
[8] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151. doi:10.1126/science.aap9559
[9] Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Karan Jha, Lu Su, and Jing Gao. 2018. EANN: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 849–857.
[10] Yuxi Wang, Martin McKee, Aleksandra Torbica, and David Stuckler. 2019. Systematic literature review on health-related misinformation on social media. *Social Science & Medicine* 240 (2019), 112552.