Alerting Users to Fake News on Twitter

A Systematic Literature Review

A paper submitted in partial fulfilment of the Research Methods and Professional Practice module (RMPP_PCOM7E)

MSc Computer Science
University of Essex Online
April 2022

Table of Contents

Introduction

The topic of this study originates from the increasing number of reports on the influence of misinformation or ‘fake news’ in the sphere of political discourse and, in particular, its influence on the outcome of the electoral process in democratic nations. The ability to successfully identify false, misleading or malicious news allows voters to make truly informed choices and strengthens the democratic process.

This literature review is divided into three sections. The first section aims to provide an understanding of the need for this research by exploring the literature surrounding the sharing of political misinformation on Twitter and its effect on various national elections. It begins by providing a definition of misinformation and the various forms it can take and identifies a number of papers which have measured and compared the influence of misinformation in different electoral contexts. The second section addresses existing research into the efficacy of various machine learning algorithms and methods to identify political misinformation on Twitter and proposes a rationale for the undertaking of this research. In the third section, research on how users react to being notified their posts are unreliable.

Contextualising the Research

This section defines misinformation and its associated terms, and reviews the existing research into its effect on national and international elections over the past decade

Political Misinformation

This research defines misinformation as an umbrella term for a variety of false or inaccurate information created with a purpose to mislead, which is spread both intentionally by willing propagators and unintentionally by those duped by it (Tandoc et al, 2017). Figure 1, taken from Wu, Morsatter et. al's 2019 paper provides a list of the various types of misinformation. Fake news, a buzzword lionised by Donald Trump in the 2016 U.S. presidential elections, is the topic of this research and can be further defined as "news articles that are intentionally and verifiably false, and could mislead readers" (Allcott and Gentzkow 2017, 213).

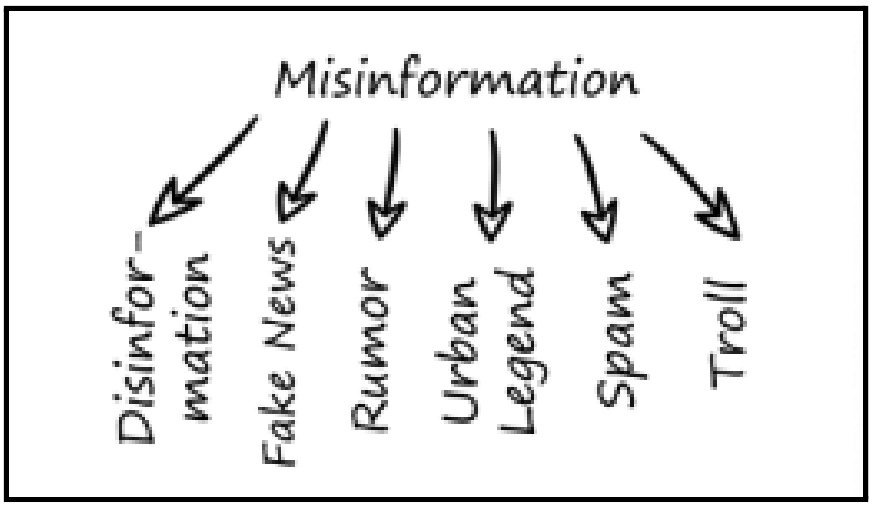

Figure 1: Key terms associated with misinformation (Wu, Morsatter et al., 2019, 80)

Twitter

Twitter is a large social media platform which enables users to communicate by posting short (no longer than 280 character) tweets, which may contain photos, videos, links, and text. The platform has 217 million active daily users and is most popular in the U.S.A., Japan and India (Aslam, 2022).

Influence of Fake News

A meta-analysis of 200 experiments conducted over 40 years has shown that humans are only 4% better than chance at detecting lies in writing (Bond & DePaulo, 2006). This implies that Twitter users cannot distinguish fake news from real news in text-based tweets, and are therefore susceptible to believing and spreading false news articles created with the intention to promote certain political biases. In fact, a recent study analysing 126,000 news stories shared and retweeted by 3 million users showed that political fake news spreads further, faster, and wider than the truth (Vosoughi et al., 2018: Fox, 2018: Meyer, 2018). This is despite the fact that most Americans mistakenly believe they can identify misleading content (Etkins, 2016).

Orchestrated campaigns of misinformation are inexpensive and can be used to try to sway public opinion at election time (Levin, 2017: Caldarelli et al., 2020). In the next subsection, recent studies into the effects of fake news on elections are discussed. As the country with the highest proportion of its population on Twitter (Aslam, 2022), we will begin with the U.S presidential elections in 2012 and 2016.

Fake News in Elections

The search terms “twitter misinformation election”, “twitter fake news election”, “social media fake news election” were used to identify the candidate papers. Papers about other social media platforms were disregarded, leaving the list of papers in Figure 2 overleaf.

Paper	Authors	Election	Sample
Political rumoring on Twitter during the 2012 US presidential election: Rumor diffusion and correction	Shin, J., et al.	2012 US presidential election	330,000 tweets
Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign	Badawy, A., et al.	2016 US presidential election	43 million tweets
Influence of fake news in Twitter during the 2016 US presidential election	Bovet, A. & Makse, H.A.	2016 US presidential election	30 million tweets, from 2.2 million users
Not All Lies Are Equal. A Study Into the Engineering of Political Misinformation in the 2016 US Presidential Election	Oehmichen, A., et al	2016 US presidential election	58 million tweets reduced to 9001 that achieved 1000 retweets
Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017	Bail, C. A. et al	2016 US presidential election	N/A
Political Knowledge and Misinformation in the Era of Social Media: Evidence From the 2015 UK Election	Munger, K. et al	2015 UK general election	N/A
Disinformation and Social Bot Operations in the run up to the 2017 French Presidential Election	Ferrara, E.	2017 French presidential election	17 million tweets
Junk News and Bots during the German Parliamentary Election: What are German Voters Sharing over Twitter?	Neudert, L. M. et al	2017 German general election	N/A
Political Bots and the Swedish General Election	Fernquist, J. et al	2018 Swedish general election	N/A
Misinformation on Twitter During the Danish National Election: A Case Study	Derczynski, L. et al	2019 Danish general election	1.5 million tweets
The limited reach of fake news on Twitter during 2019 European elections	Cinelli, M. et al	2019 European elections	400,000 tweets

Figure 2: Papers Concerning Election Fake News on Twitter

The single research paper concerning the 2012 U.S. elections investigated whether fake news and rumours were corrected as they were propagated across Twitter. Shin et al. found that Twitter functioned more as an echo chamber for like-minded partisans than a system where fake news was corrected, though their sample of tweets was much smaller than the other studies (Shin et al., 2016).

Fake news across a multitude of social media platforms was in the news in 2016, the year of the U.S. presidential election between Hillary Clinton and Donald Trump (Ritchie, 2016: Hotchkiss, 2017). In one study of 171 million tweets, Bovet & Makse found that almost 25% of the 30 million news-related retweets contained false or misleading content (Bovet & Makse, 2019). Much of this fake news was intended to polarise opinion and its propagation appeared co-ordinated and orchestrated (Oehmichen et al., 2019). The impact this had on the election outcome is difficult to measure, but may be somewhat limited as retweets by automated Russian accounts were mainly shared between users who were already similarly politically aligned (Badawy et al., 2018: Bail et al., 2018). Many of these accounts disappeared after the election, and were reactivated in the build-up to the 2017 French presidential election (Ferrara, 2017).

The spread of fake news is not limited to the U.S.A. A 2019 report by the UK Office of Communication showed that almost half of British adults get their news on social media platforms, one in five from Twitter (Ofcom, 2019). U.K. Twitter users tend to be better informed on political issues than those not on the platform, though often the information they are receiving and sharing is polarising and of questionable veracity (Munger et al., 2020). Organised clusters of automated twitter accounts were identified in Germany and Sweden sharing right wing materials during their 2017 and 2019 elections, respectively, though at a much lower level than that of the

U.S. (Fernquist, 2018: Neudert, 2017). It seems that the spread of fake election news is not at pandemic levels, though, as studies found little evidence of the phenomenon in the Danish elections of 2019, and the European parliament elections of the same year (Derczynski et al., 2019: Cinelli et al., 2020).

This organised propagation of misinformation and disinformation for political gain is increasing in some countries as partisan parties attempt to influence the results of elections. Though the success or failure of these campaigns is yet to be proven, identifying fake news in an attempt to prevent its distribution has become a hot topic for research. In the next subsection, recent research into the use of machine learning methods to detect political misinformation is identified and discussed.

Using machine learning techniques to identify misinformation

In recent years, a number of studies have been conducted on the efficacy of machine learning algorithms to detect fake news on social media (della Vedova et al., 2018: Lin et al., 2019: Khan et al., 2021). It is not a simple and straightforward proposition as misinformation is difficult to discern from trusted news as the articles often contain similar language, images and structure (Conroy et al., 2015).

Some studies have combined analysis of content with analysis of the network of sharing users in an attempt to improve accuracy of identification. In this section we discuss the most relevant of these, the various machine learning algorithms used on data from Twitter, though it should be noted that not all of the papers were based solely on political misinformation but also fake news in other domains.

Detection and Visualization of Misleading Content on Twitter

In this 2018 paper, the authors analysed both the content and the creator in an attempt to detect misinformation in multimedia tweets about a variety of events. Using Logistic Regression(LR) and Random Forest(RF) algorithms on a dataset of 6,225 real and 9,404 fake tweets, they report accuracy of 88% for LR (Boididou, 2018), though the limited size of the dataset calls into question the generalisability of this result.

Experiments on Detecting Fake News

Using a somewhat limited dataset of ~25,000 tweets on a variety of topics from Trump to Covid, the authors of this study used 4 different ML algorithms to detect untrustworthy news (Kudarvalli, H., & Fiaidhi, 2020). Logistic Regression and Support Vector Machine were found to be the most effective with over 90% accuracy, but the authors do not specify how they established which of the tweets were ‘fake’ nor whether the tweets were solely text-based or contained other media.

Machine Learning Algorithm based model for classification of fake news on Twitter

This 2020 paper uses a dataset of text-based political misinformation tweets about India. The authors collected an unspecified number of tweets about 4000 news articles, 40% of which contained fake news. They used TF-IDF classification methods and measured prediction of fake news using Naive Bayes and Passive Aggressive classifiers. They conclude that the Passive Aggressive classifier was the most successful predictor of fake news with a 78% accuracy (Nikam & Dalvi, 2020) but it is difficult to measure the significance of this finding without information about how many tweets were used in the training and testing of the models.

Detecting Fake News with Machine Learning Method

The Thai authors of this paper claim a 97% success rate at detecting tweets containing fake news using the Naive Bayes algorithm, significantly higher than the other studies found (Aphiwongsophon & Chongstitvatana, 2018). The results lack some important details about the dataset of over 300,000 tweets, and together with a lack of clarity in the writing, leave the credibility of these extraordinary numbers in doubt.

Multiclass Fake News Detection Using Ensemble Machine Learning

Using a dataset from Kaggle of approximately 50,000 headlines, the authors of this paper found that a Gradient Boosting algorithm was the most accurate predictor of fake news scoring 86% accuracy (Kaliyar et al., 2019). While this is an impressive result, and the paper is well written, containing sufficient details about the experiment, the experiment contains a mixture of different media and is not limited to tweets, and so cannot be presumed to be transferable.

Detecting Fake News, Then What?

Using machine learning to identify fake news seems to be possible to varying degrees of success but does its use reduce the sharing and spread of disinformation?

Seo et al. report on an experiment where users were given a warning when they were about to share dubious content. The warnings originated from two sources: a fact-checker, and a machine learning algorithm. The study found that although the machine learning algorithm increased a user's ability to spot fake news in future, they placed less trust in its findings than the fact-checking warning (Seo et

al., 2020). TrustyTweet is a plugin which notifies users when a tweet's content is dubious, though research to its effectiveness has only been conducted on a small sample of 27 people (Hartwig & Reuter, 2019). It seems, however, that notifying user's that their tweets included misinformation or fake news may not have the intended effect and could result in users sharing more partisan and biassed content (Mosleh et al., 2021).

Research identified in this review failed to identify a clear and obvious choice of machine learning algorithm to accurately identify fake news in tweets, and it is unclear how best to share this information with users when detected. It is the purpose of this research to fill this gap in the literature.

References

Allcott, H., Gentzkow, M. (2017) Social Media and Fake News in the 2016 Election. Journal of Economic Perspectives, Number 31, Volume 2, 211-36 https://doi.org/10.1257/jep.31.2.211

Aphiwongsophon, S. & Chongstitvatana, P. (2018) 'Detecting Fake News with Machine Learning Method' 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2018, pp. 528-531, https://doi.org/10.1109/ECTICon.2018.8620051

Aslam, S. (2022) Twitter by the Numbers: Stats, Demographics & Fun Facts. Available from: https://www.omnicoreagency.com/twitter-statistics/ [Accessed 30/3/2022].

Badawy, A., Ferrara, E. & Lerman, K. (2018) 'Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign,' 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2018, pp. 258-265. https://doi.org/10.1109/ASONAM.2018.8508646

Bail, C. A., Guay, B., Maloney, E., Combs, A., Hillygus, D. S., Merhout, F. (2020) Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017. Proceedings of the National Academy of Sciences, 117(1), 243–250.

Boididou, C., Papadopoulos, S., Zampoglou, M., Apostolidis, L., Papadopoulou, O., & Kompatsiaris, Y. (2018) Detection and visualization of misleading content on Twitter. Int J Multimed Info Retr 7, 71–86 (2018). https://doi.org/10.1007/s13735-017-0143-x

Bond, C. F. and DePaulo, B. M. (2006) Accuracy of Deception Judgments. Personality and Social Psychology Review, 10(3), pp. 214–234. https://doi.org/10.1207/s15327957pspr1003_2

Bovet, A. & Makse, H.A. (2019) Influence of fake news in Twitter during the 2016 US presidential election. Nat Commun 10, 7 (2019). https://doi.org/10.1038/s41467-018-07761-2

Caldarelli, G., De Nicola, R. & Del Vigna, F. (2020) The role of bot squads in political propaganda on Twitter. Commun Phys 3, 8. https://doi.org/10.1038/s42005-020-0340-4

Cinelli, M., Cresci, S, Galeazzi, A., Quattrociocchi, W. & Tesconi, M. (2020) The limited reach of fake news on Twitter during 2019 European elections. PLoS ONE 15(6): e0234689. https://doi.org/10.1371/journal.pone.0234689

Conroy, N.K., Rubin, V.L. and Chen, Y. (2015), Automatic deception detection: Methods for finding fake news. Proc. Assoc. Info. Sci. Tech., 52: 1-4. https://doi.org/10.1002/pra2.2015.145052010082

Della Vedova, M.L., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M. & de Alfaro, L. (2018) "Automatic Online Fake News Detection Combining Content and Social Signals," 22nd Conference of Open Innovations Association (FRUCT), 2018, pp. 272-279, https://doi.org/10.23919/FRUCT.2018.8468301

Derczynski, L., Albert-Lindqvist, T., Bendsen, M., Inie, N., Pedersen, V., & Pedersen, J. (2019). 'Misinformation on Twitter During the Danish National Election: A Case Study'. Conference: Conference for Truth and Trust Online 2019 https://doi.org/10.36370/tto.2019.16

Etkins, B. Forbes.com, (2016) Americans believe they can detect fake news. Studies show they can't. Available from: https://www.forbes.com/sites/brettedkins/2016/12/20/americans-believe-they-can-detect-fake-news-studies-show-they-cant/#1f4c85cf4022 [Accessed: 16/4/2022].

Fernquist, J., Kaati, L & Schroeder, R. (2018) Political Bots and the Swedish General Election. 2018 IEEE International Conference on Intelligence and Security Informatics (ISI), 2018, pp. 124-129, https://doi.org/10.1109/ISI.2018.8587347

Ferrara, E. (2017) Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election. First Monday. 22. https://doi.org/10.5210/fm.v22i8.8005

Fox, M. NBC News (2018). Want something to go viral? Make it fake news. Available from:

https://www.nbcnews.com/health/health-news/fake-news-lies-spread-faster-social-media-truth-does-n854896 [Accessed: 27/3/2022]

Hajli, N., Saeed, U., Tajvidi, M. and Shirazi, F. (2021) Social Bots and the Spread of Disinformation. Social Media: The Challenges of Artificial Intelligence. Brit J Manage. https://doi.org/10.1111/1467-8551.12554

Hartwig, K., & Reuter, C. (2019) TrustyTweet: An Indicator-based Browser-Plugin to Assist Users in Dealing with Fake News on Twitter. Wirtschaftsinformatik.

Hotchkiss, J. Huffington Post. (2017) Russia used fake news. This is how Russia used fake news on facebook to help elect Donald Trump. Available from: https://www.huffingtonpost.com/entry/this-is-how-russia-used-fake-News-on-facebook-to-help_us_59b60b64e4b0bef3378ce1be [Accessed: 22/4/2022].

Kaliyar, R.K., Goswami, A. & Narang, P. (2019) 'Multiclass Fake News Detection using Ensemble Machine Learning' 2019 IEEE 9th International Conference on Advanced Computing (IACC), 2019, pp. 103-107, https://doi.org/10.1109/IACC48062.2019.8971579

Khan, J. Y., Khondaker, M. T. I., Afroz, S., Uddin, G. & Iqbal, A. (2021) A benchmark study of machine learning models for online fake news detection. Machine Learning with Applications, Volume 4, 2021 https://doi.org/10.1016/j.mlwa.2021.100032

Kudarvalli, H., & Fiaidhi, (2020) Experiments on Detecting Fake News using Machine Learning Algorithms. International Journal of Reliable Information and Assurance. 8. 15-26. https://doi.org/10.21742/IJRIA.2020.8.1.03

Levin, S. The Guardian. (2017) Pay to sway: report reveals how easy it is to manipulate elections with fake news. Available from: https://www.theguardian.com/media/2017/jun/13/fake-news-manipulate-elections-paid-propaganda [Accessed: 4/4/2022].

Lin, J., Tremblay-Taylor, G., Mou, G., You, D. & Lee, K. (2019) 'Detecting Fake News Articles' 2019 IEEE International Conference on Big Data (Big Data), 2019, pp. 3021-3025, https://doi.org/10.1109/BigData47090.2019.9005980

Meyer, R., The Atlantic. (2018) The Grim Conclusions of the Largest-Ever Study of Fake News. Available from: https://www.theatlantic.com/technology/archive/2018/03/largest-study-ever-fake-news-mit-twitter/555104/ [Accessed: 27/3/2022].

Mosleh, M., Martel, C., Eckles, D. & Rand, D. (2021) 'Perverse Downstream Consequences of Debunking: Being Corrected by Another User for Posting False Political News Increases Subsequent Sharing of Low Quality, Partisan, and Toxic Content in a Twitter Field Experiment' CHI '21: the 2021 CHI Conference on Human Factors in Computing Systems, May 2021 Article No.: 182 Pages 1–13 https://doi.org/10.1145/3411764.3445642

Munger, K., Egan, P. J., Nagler, J., Ronen, J. and Tucker, J. (2020) Political Knowledge and Misinformation in the Era of Social Media: Evidence From the 2015 UK Election. British Journal of Political Science. Cambridge University Press, 52(1), pp. 107–127. https://doi.org/10.1017/S0007123420000198

Neudert, L. M., Kollanyi, B. & Howard, P. N. (2017) Junk news and bots during the german parliamentary election: What are german voters sharing over twitter? Data Memo, UK: Project on Computational Propaganda. comprop.oii.ox.ac.uk. Oxford, UK, 2017.

Nikam S. S. & Dalvi, R. (2020) "Machine Learning Algorithm based model for classification of fake news on Twitter," *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2020, pp. 1-4, https://doi.org/10.1109/I-SMAC49090.2020.9243385

Oehmichen, A., Hua, K., Amador Díaz López, J., Molina-Solana, M., Gómez-Romero, J. & Guo, Y. (2019) Not All Lies Are Equal. A Study Into the Engineering of Political Misinformation in the 2016 US Presidential Election. IEEE Access, vol. 7, pp. 126305-126314, 2019, https://doi.org/10.1109/ACCESS.2019.2938389

Ofcom (2019) Half of people now get their news from social media. Available from: https://www.ofcom.org.uk/about-ofcom/latest/features-and-news/half-of-people-get-news-from-social-media [Accessed on: 1/4/2020].

Ritchie, H. CNBC (2016) Read all about it: The biggest fake news stories of 2016 Available from: https://www.cnbc.com/2016/12/30/read-all-about-it-the-biggest-fake-news-stories-of-2016.html [Accessed 23/4/2022].

Shin, J., Jian, L., Driscoll, K. & Bar, F. (2016) Political rumoring on Twitter during the 2012 US presidential election: Rumor diffusion and correction. New Media & Society. 19. https://doi.org/10.1177/1461444816634054

Seo, H., Xiong, A., & Lee, D. (2019) 'Trust It or Not: Effects of Machine-Learning Warnings in Helping Individuals Mitigate Misinformation' Proceedings of the 10th ACM Conference on Web Science (WebSci '19). Association for Computing Machinery, New York, NY, USA, 265–274. https://doi.org/10.1145/3292522.3326012

Tandoc, E. C. Jr., Lim, Z. W. & Ling, R. (2018) Defining "Fake News", Digital Journalism, 6:2, 137-153, https://doi.org/10.1080/21670811.2017.1360143

Vosoughi, S., Roy, D., Aral, S. (2018) The spread of true and false news online. Science 359(6380), 1146–1151 (2018) https://doi.org/10.1126/science.aap9559

Wu, L., Morstatter, F., Carley, K. M., & Liu, H. (2019) Misinformation in Social Media: Definition, Manipulation, and Detection. SIGKDD Explor. Newsl. 21, 2 (December 2019), 80–90. https://doi.org/10.1145/3373464.3373475