ESMIZADEH, YALDA. Ph.D. The Disinformation Pandemic: Understanding, Identification, and Mitigation in COVID-19 Era. (2022)
Directed by Dr. Hamid Nemati. 152 pp.

In 2020 during COVID-19, in addition to the spread of coronavirus disease, we also observed a pandemic of disinformation about the disease. This pandemic of disinformation became known as *Infodemic* in the medical world. Just as coronavirus was infecting our bodies, *Infodemic* was infecting our information ecosystem and exasperating the fight against the COVID-19 pandemic. Disinformation can be produced by various sources including scientists, media personalities, and others and it can be disseminated by news media, webpages, and social media from one source to another. Additionally, disinformation can spread easily from web media to social media where it can spread even faster to a wider audience. Therefore, it is important that disinformation be detected before it has a chance to spread. However, the identification of disinformation is fraught with several challenges. This fact highlights the importance of studying and identifying disinformation both in the content of web pages and social media posts before it is allowed to spread. In this dissertation, I pursued a three-essay approach to understand, identify, and mitigate the disinformation pandemic. While manual fact-checking is difficult, time-consuming, and expensive, various automated detection solutions could speed up this process. Therefore, in my first essay, I explored whether Machine Learning (ML) techniques can be used to develop predictive models for automatic identification of disinformation. Computational linguistics methods are used to extract content-based, and sentiment-based features of selected webpage' articles to construct our study dataset. This dataset is used to train various ML algorithms to develop predictive models to identify disinformation. The results showed that there are significant differences among features of true and false information that can be used to identify disinformation. Since the spread of disinformation happens both on media pages and on social

media platforms, it is important to analyze disinformation at both levels. Moreover, the literature shows that disinformation spreads six times faster than true information on social media, demonstrating that users get more engaged with disinformation. Therefore, I extended my research to enhance the understanding of disinformation detection based on content-based features and its impact on users' engagement in social media posts. The findings of the second essay highlighted the critical role of linguistic structure, emotional tone, and the psychological load of social media posts on users' engagement that can be used to differentiate information from disinformation. The results of the first two essays confirmed that negative emotional tone was one of the most important factors in disinformation posts and was associated with a high engagement score. So, in the third essay, I explore the impact of negative emotional tones in developing users' perceptions regarding the accuracy of the content. Three separate experiments were developed to explore this. The results of experiments in the third essay highlighted the significant role of negative emotional tones on the believability of the content and their potential influence on behavioral change. My research findings allow for a better understanding and identification of disinformation by highlighting and identifying content-based features that are meant to mislead users to falsely perceive disinformation as information.

THE DISINFORMATION PANDEMIC: UNDERSTANDING, IDENTIFICATION, AND MITIGATION IN COVID-19 ERA

By

Yalda Esmizadeh

A Dissertation

Submitted to

the Faculty of The Graduate School at

The University of North Carolina at Greensboro

in Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

Greensboro

2022

Approved by

Dr. Hamid Nemati
Committee Chair

APPROVAL PAGE

This dissertation written by Yalda Esmizadeh has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair

Dr. Hamid Nemati

Committee Members

Dr. Indika Dissanayake

Dr. Nikhil Mehta

Dr. Minoo Modaresnezhad

Dr. Soheil Hooshangi

March 22, 2022

Date of Acceptance by Committee

March 15, 2022

Date of Final Oral Examination

ACKNOWLEDGEMENTS

During these years especially throughout the writing of my dissertation, I have received a great deal of support and assistance.

I would first like to thank my advisor, Professor Hamid Nemati, whose expertise was invaluable in developing my research topic and formatting my research questions. Your valuable feedback brought my work to a higher level.

I would like to express my deepest appreciation to my committee members for their collaboration in enhancing my research. Many thanks to Dr. Modaresnezhad for her valuable comment on essay I, Dr. Dissanayake and Dr. Mehta for their perceptive feedback on essay II, and Dr. Hooshangi for his invaluable support in designing the third essay experiment.

In addition, I would also like to extend my deepest gratitude to my family. Special thanks to my parents and my sister, Elnaz, for their emotional support. Finally, I could not have completed this dissertation without the full support of my soulmate, Eesa, who has always been there for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

Have you ever noticed that some topics on social media networks trend out of nowhere? Where do they come from? What could explain it? Could it be the content of the post, the emotions it evokes, or the social media network structure? Pew Research Center in 2016 reported that 62 % of U.S. adults receive their news on social media while 66 % of all Facebook users access the platform for news consumption (Gottfried and Shearer 2016). According to Pennycook et al. "Across a wide range of domains, it is critically important to correctly assess what is true and what is false: Accordingly, differentiating real from unreal is at the heart of society's constructs of rationality and sanity" (2018 p.1865). False information is one of the recent issues that has plagued social media platforms during, as well as after, the U.S. elections. False information can be defined as stories that have been fabricated but are presented in a legitimate form and promoted on social media to deceive the public for ideological and/or financial gain (Lazer et al. 2018).

While most research is focused on social media platforms (Aldwairi and Alwahedi 2018; Allcott and Gentzkow 2017; Loeb et al. 2020), less attention has been paid to the sources of these posts and the published articles on webpages. Nowadays, people prefer to search for health-related articles online rather than accessing traditional sources due to the easy access to massive amounts of available information (Ghenai 2017). According to a study by Pew Internet and American Life Project, weather-related, national events-related, and health-related news searches are the three most common online news topics explored through search engines (Ghenai, 2017; Purcell et al., 2010). Considering the importance of having access to true sources of information, how can we, as information systems (IS) researchers, help users to identify the false content on the internet? The massive spread of false information in these platforms could lead to numerous social problems such as breaking the validity balance of the news ecosystem, enhancing biased and false beliefs, and impacts the way that people respond to authentic information (Shao et al. 2018; Shu, Sliva, et al. 2017; Shu, Wang, et al. 2019). So, what are the applicable approaches in false information identification? How information systems related theories could enhance our understanding of false information detection? Can exposure to various content change users' decisions? In this thesis, I will try to clarify these questions by answering the following research question separately:

Essay I: The first essay focuses on understanding the efficacy of using Machine Learning (ML) techniques for identifying disinformation. I attempt to answer the following two research questions: "Can ML be used to identify disinformation?" and "Which characteristics of the content are important in the disinformation identification?

Essay II: The second essay focuses on understanding the characteristics of content that help with its dissemination. I attempt to answer the following two research questions: "What content factors differentiate information from disinformation on social media?" and "Which content dimensions of social media posts can influence the engagement score on social media?

Essay III: The third essay focuses on understanding the role of emotional tone, one of the content-based features, on the believability of the content and its potential effect of behavior change. I attempt to answer the following two research questions: "How the change in negative emotional tone will affect the believability of the disinformation content?" and "How the change in negative emotional tone would influence the behavior?"

In the *first* essay, I observe that during the COVID-19 pandemic in 2020, the popularity of several websites reputed for spreading disinformation grew significantly, resulting in citizens' distrust of public policies meant to fight the pandemic. These websites would argue that mandatory mask-wearing would encroach on citizens' freedoms. Additionally, anti-maskers, and anti-vaxxers tried to sow doubt in the efficacy of the vaccines by amplifying the uncertainty surrounding the origins of COVID-19 and the procedure and policies employed to develop and administer the vaccines (Das et al. 2021; Kanozia and Arya 2021). Many influencers try to propagate conspiracy theories or use fear tactics to gain support and increase their followers. This explosion of false information around the COVID-19 topic calls for novel approaches to detect false information to mitigate the catastrophic effects that disinformation may have on society. By collecting and integrating data from various known disinformation producers as well as those known for publishing authentic and accurate information during the COVID-19 pandemic, I attempted to investigate and shed light on the critical features that could distinguish disinformation from authentic ones. In this dissertation, I use a three-essay approach to achieve the goals of my dissertation. In the first essay, I seek to explore whether Machine Learning techniques can be used to develop models for automatic classification of false vs authentic information. I report on the details of using Machine Learning techniques employed on a dataset of 1795 health-related disinformation and information articles published over the 2020 year. I used computational linguistics to extract content-based, and sentiment-based features of the selected articles in our database to train my Machine Learning predictive models. The results showed that there are significant differences among features of true information and disinformation. My analysis illustrates that true articles have a lower word frequency average in comparison to disinformation articles, moreover, they use less fear, anger, and generally negative tones in the content. These findings highlight the importance of computational linguistics in disinformation detection procedures which could be particularly useful in designing disinformation recognition tools in flagging false content.

In the *second* essay, I present a theory-driven framework to understand how disinformation producers on social media produce content to engage and attract more users. Based on the uses and gratifications theory (UGT), users generate content on social media to satisfy their cognitive, affective, personal, social, and tension release needs (West et al. 2010). Furthermore, emotional broadcaster theory (EBT) represents that emotionally motivated disclosures often contain important information for listeners which will predict how far the stories travel across the social networks. In this study, I propose that disinformation producers are aware of these facts and attempt to produce content that responds to these needs to engage more users. To be able to measure UGT and EBT constructs through content, I collected a dataset of 6000 Facebook posts from well-known disinformation and information producers over the 2020 year. After collecting data, various applications and packages have been implemented to extract content features. I need to evaluate the various linguistic structures, emotional tones, and cognitive needs tone of textual messages. To do so, I applied a text analysis application called Linguistic Inquiry and Word Count (LIWC) and Syuzhet package in R as a scientific method to determine the tone of messages. The findings of the second essay highlight the importance of the theory-driven approaches in detecting disinformation and their effect on engaging audiences.

In the *third* essay, I expand the results of the second essay to design an experiment and evaluate the importance of my findings in the believability of the content and potential influence of disinformation on behavior change. Understanding the impact of the emotional tone of content on the believability of disinformation and users' decision-making will lead to a great contribution to disinformation awareness. A better understanding of this mechanism would be useful to provide meaningful strategies for social media networks to enhance public fact-checking behavior. According to the first essay, machine learning algorithms were able to detect disinformation from trustworthy ones with significant accuracy based on document term matrix and sentiment discovery analysis. The findings of the first essay highlighted the importance of linguistic structure and emotional tone in disinformation/information identification. Even though the second essay provided evidence for a significant relationship between content-based features and social media engagement, it did not establish a directional causality. In the literature, the strength of emotions evoked by information has been known as a powerful predictor of information spread (Cotter 2008; Izawa 2010; Wang et al. 2020) and highly intense emotional content is more likely to be shared (Rimé et al. 2002; Xiong and Lv 2021). In the third essay, I investigated the role of emotional tone in the believability of the content on social media and its potential effect on decision-making. To do so, I conducted a series of experiments to expose participants to different disinformation FB posts with high and low levels of emotional tone. After that, they were asked to score their perception regarding trustfulness of the message and the potential influence of the message on their future behavior. To be able to generalize the findings, I ran a series of experiments with three different types of emotions. The findings of essay III illustrated that increasing the negative emotional tone of content including fear, sadness, and anger would significantly increase the believability of the content. Moreover, participants of the experiment were estimated to be influenced more with high negative emotional tone content in comparison to low negative emotional content. In the following subsections, a detailed explanation of each essay has been provided.

The rest of my dissertation is organized as follows: the second chapter presents a theoretical background of false information taxonomy and false information detection classification in literature. After providing a holistic view of false information and its detection techniques, I will present each essay separately. Chapter three will present essay I which is entitled “Disinformation Detection by ML algorithms”. In the fourth chapter, I will present essay II which is entitled “Disinformation engagement on social media from a theoretical perspective”. Chapter 5 will represent a brief explanation of essay three which is entitled “Negative Emotion effect on believability and change of behavior”. Finally, the conclusion of my research is provided in chapter 6.

CHAPTER II: THEORETICAL BACKGROUND

Today's digital world provides people a platform to easily get access to any information, share it, and create their content. Nowadays people from all over the world are overloaded with information. The main problem of this overloading is that there is no way to guarantee the information they receive is true. It could be biased with respect to different perspectives or even untruthful to spread disinformation and user preference manipulation (Vysotska and Vysotska n.d.).

Pew Research Center in 2016 reported that 62 % of US adults receive their news on social media while 66% of all Facebook users use the platform for news consumption (Gottfried and Shearer 2016). Users' access and reliance on social media as a source of news have created new challenges (Allcott and Gentzkow 2017). Given that human beings are not very good at detecting deceptions (Rubin 2010; Twitchell et al. 2004), and manual fact-checking is difficult, time-consuming, and expensive, various automated detection solutions would speed up this process (Kula et al. 2020). False information detection faces several challenges as content is intentionally created to mislead readers (Pasławska and Popielska-Borys 2018) and users' social interaction with false information leads to big, incomplete, and noisy data (Tang et al. 2014).

During the COVID-19 pandemic, infodemic in the medical publishing world was observed besides the spread of coronavirus disease. An Infodemic or outbreak of false information can be defined as "a contagious disease infecting our information culture" (Solomon et al. 2020, P1806). Eysenbach (2020) classified the current infodemic into four levels: science, policy and practice, news media, and social media. Knowledge translation occurs by translating the information from one audience to another. Therefore, false information from news media can easily spread on social media via its users. As false information passes from the producer to the consumer on social media, various factors play a role in how it disperses and impacts users' ideologies, beliefs, and responses. Considering the importance of producer, consumer, and message content itself, it is critical to examine false information production and its diffusion from multiple theoretical perspectives to gain a comprehensive understanding of the phenomenon.

Authenticity and intent are two main factors in defining false information in the literature. The spread of false information is one of the recent issues that has plagued social media platforms during, as well as after, the U.S. elections. In recent years concerns regarding fake information in the text has gained attention by increasing the negative effect of false information in various fields including political polarization (Hameleers and van der Meer 2020), climate change debates (Treen et al. 2020), and public health-related topics (Southwell et al. 2019). The first step of studying false information and its detection techniques is providing false information definition in literature.

False information: Based on the literature, false information can be divided into five categories known as (1) rumors, (2) fake news, (3) misinformation, (4) hoax (Habib et al. 2019), and (5) disinformation. Rumor is a piece of information that the source is unreliable, the truthfulness of the content is in doubt, and probably has been produced under emergencies to create public panics and disturb the social order, to diminish government credibility, and even to threaten national security (Liang et al. 2015). The main consequence of the rumor is the injection of fear and anxiety to the people to the level that people will not be able to make rational decisions as they otherwise do (Jin et al. 2016).

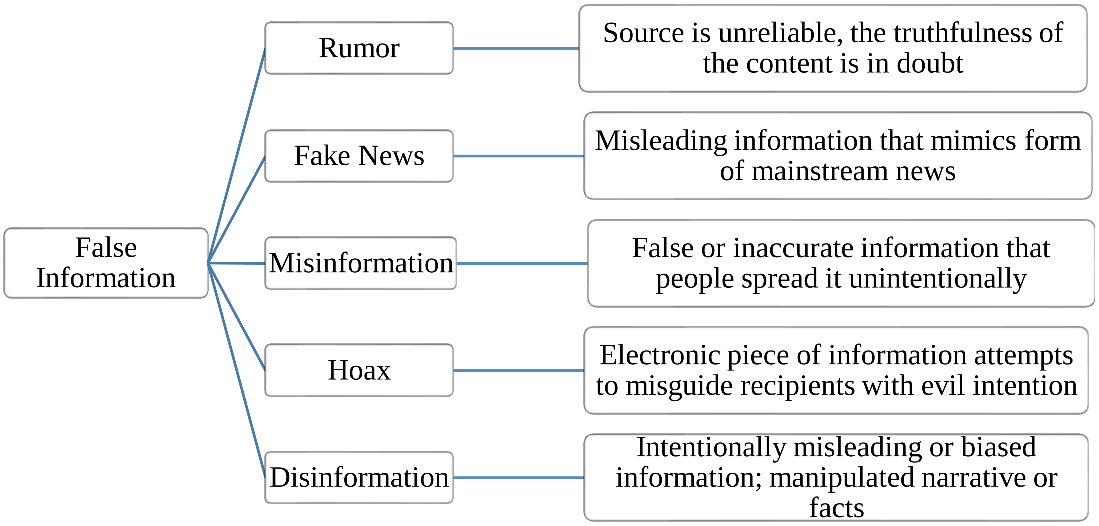
The second category is fake news. Fake news can be defined as a piece of information that mimics a form of mainstream news (Zimdars and McLeod 2020). Fake news has been produced to mislead its audiences to gain the readers' attention with financial or political intentions. Fake News is usually created by shocking, exaggerated, or false headlines (Allcott and Gentzkow 2017).

The consequence of fake news covers the range from being merely annoying to influencing and misleading societies (De Beer and Matthee 2020). Misinformation is the third type of false information on social media. Misinformation can be defined as a piece of false or inaccurate information that people spread intentionally or unintentionally in an attempt to present it as being true (Jia 2020). The main consequence of believing in misinformation is its effect on decision-making abilities and making wrong decisions (Karduni et al. 2018).

The fourth group of false information is a hoax, which can be defined as an electronic piece of information that attempts to misguide recipients consisting of audio, text, and multimedia content with evil intention. Hoaxes mainly spread due to the lack of media literacy with political interest, where people are unable to recognize true news and hoaxes on social media (Shu, Sliva, et al. 2017). Our literature review survey illustrates that disinformation is another group of false information. Based on the Merriam-Webster dictionary, disinformation can be defined as a piece of intentionally misleading or biased information; manipulated narrative or facts[1] that attempts to create deceptive and false statements to convince someone of untruth (Nimmo 2016). Moreover, in IS literature, disinformation can be defined as the act of creating and sharing false and manipulated information to deceive and mislead the audiences to gain political, personal, and financial benefits (Gelfert 2018; Tandoc Jr et al. 2018). Figure 1 illustrates the summary of these definitions.

[1] https://www.dictionary.com/browse/disinformation

Figure 1. False information taxonomy


False Information
Rumor
Source is unreliable, the truthfulness of the content is in doubt
Fake News
Misleading information that mimics form of mainstream news
Misinformation
False or inaccurate information that people spread it unintentionally
Hoax
Electronic piece of information attempts to misguide recipients with evil intention
Disinformation
Intentionally misleading or biased information; manipulated narrative or facts

Although these definitions look remarkably like each other, the following statements would help recognize different types of false information. If the false information is verified as false by the authorized sources, it should be considered as a rumor (Wu et al. 2015). If the false information contradicts the ground-truth information and makes people believe in the false version of the information, it should be labeled as fake news (Pomerantsev and Weiss 2014). If the false information has been spread usually on social media via tweets, blogs, and articles without deceptive intention it should be labeled as misinformation, and if it has been spread with deceptive intentions should be labeled as disinformation (Nimmo 2016). Finally, if the false information has been spread through the news outlet to deceive the truth, it should be considered a hoax (Hunt 2016).

Studying the reasons why users fall prey to false information is one of the major branches in this field. While some researchers believe that people are more likely to share false information when it is viewed as a novel (Vosoughi et al. 2018), others believe that it is rooted in personal differences (Vishwanath 2015). Moreover, some researchers look for the roots in content, for instance, Wang (Wang et al. 2015) represented users who are more likely to be engaged by Twitter posts by re-tweeting posts that have happiness emotions than anger emotion. Automated deception detection methods in textual information are of interest in information science (Rubin et al. 2015; Rubin and Lukoianova 2015) due to their applications in enhancing human abilities in recognizing true information from false information (Purian et al. 2020; Rasool et al. 2019), assessing information qualities (Osatuyi and Hughes 2018; Rubin and Lukoianova 2013), and assessing source credibility (Shao et al. 2018).

The vast amount of unstructured text data sets and automated deception detection are challenging tasks (Rubin and Chen 2012; Shu, Sliva, et al. 2017) that can be overcome with big data analytics techniques such as natural language processing (NLP) and machine learning (ML) techniques (Karimi et al. 2018; Rubin and Lukoianova 2015; Torabi Asr and Taboada 2019). ML techniques give the system the ability to automatically learn and improve itself from data with minimum interaction with humans. This feature lets researchers remove or eliminate human bias from the system. Machine learning algorithms are classified as supervised learning methods, unsupervised learning methods, and semi-supervised learning methods. While the prediction in the supervised learning method is made based on the labeled dataset, unsupervised learning algorithms use unlabeled datasets for classification, and the semi-supervised method uses both labeled and unlabeled datasets for learning (Habib et al. 2019).

In this section, a systematic literature review has been done to be able to find the research gaps and propose solutions for false information detection. From a data mining perspective, Shu (2017) divided false information detection on social media into two categories based on feature selection: a) false information detection based on content, b) false information detection based on social context. This classification has been provided since research in detecting false information in literature takes two approaches. While one group attempts to extract content features such as source, headline, body text, and image or video to construct machine learning models, the second group of researchers focus on the social context features such as user-based, post-based, and network-based features to construct their machine learning models. In the following, I will discuss them in detail.

Content-based False Information Detection: Content-based False information detection methods utilize content features to describe the meta-information related to a piece of information (Shu, Sliva, et al. 2017). In this approach, researchers evaluate linguistic and visual features of the content. Chen (2015) studied misleading online content to recognize clickbait as false news. The researcher believed that as false information is intentionally created for financial and political gain, they often contain biased, incentive, and shocker language which led to the rapid spread of rumors and misinformation online. They believed that recognizing textual and non-textual clickbaiting cues is essential in detecting clickbait as a form of deception. The finding of the study highlighted the importance of extracting linguistics features and dramatic headlines in detecting false information. Linguistic features are extractable at lexical level (character level and word level) and syntactic level (sentence level, punctuation level, and document level). On the other hand, visual features of the content focus on sensational and fake images to provoke emotional responses of the audiences (Qi et al. 2019).

After extracting features, model construction is the next step in machine learning methods of false information. Knowledge-based, style-based, and visual models are three major models that are constructed based on content-related features. One of the best ways in detecting false claims is to check the truthfulness of the claim. This approach is known as a knowledge-based strategy, where researchers use a third external source to evaluate and fact-check the claims in news content to assign a true or false tag to a claim (Thorne and Vlachos 2018; Vlachos and Riedel 2014). Expert-oriented, crowdsourcing-oriented, and computational-oriented are three major approaches to fact-checking. While the two first categories rely on human judgments and group judgments, the computational-oriented one is built on providing an automatic scalable system to classify true and false claims based on identifying check-worthy claims and distinguishing the validity of the claims. In other words, computational-oriented detection is achieved by the comparison of "the relational textual knowledge extracted from to-be-verified news articles and that of knowledge graphs representing facts or ground truth" (Zafarani et al. 2019, P.3208). Therefore, the construction of knowledge graphs is an active research area that directly evaluates news authenticity. Knowledge-based detection has been applied in both social media and news content for decision-making. The literature review illustrates the strengths and weaknesses of this approach. As an instance, Pan's study (2018) illustrates that computational-oriented fact-checking is not comprehensive enough to cover all the relations needed for false information detection and it is challenging to validate the correctness of the extracted check-worthy claims from news articles.

The second model of false information detection is a style-based model where detection techniques have been applied to identify the differences in writing styles between false information stories and true stories. Style-based approaches aim to detect false information by capturing the content manipulators through the writing style. To be able to recognize and capture these differences, NLP techniques within a machine learning framework are the best tools in automated detection (Mukherjee and Bala 2017). The written style extracted from the news can be used to evaluate the intention of the creator of the news (Zhou et al. 2020). Style-based detection is usually done via deception-oriented or objectivity-oriented approaches. In the first approach, researchers try to evaluate the syntax and rhetorical structure of the content to capture deception. As an instance, while Mukherjee and Bala (2017) used style-based NLP techniques to distinguish sarcasm in customer tweets, Litvinova et al. (2017) used style-based NLP techniques to detect deception in Russian text. While Mukherjee and Balam used supervised learning algorithms to differentiate features between a sarcastic and non-sarcastic sentence on customers' tweets, Litvinova's group used linguistic inquiry and word count software to identify the differences between a deceptive and true piece of textual content. In an objectivity-oriented approach, researchers aim to identify "style signals that can indicate a decreased objectivity of news content and thus the potential to mislead consumers, such as hyperpartisan styles and yellow-journalism" (Shu, Sliva, et al. 2017, P.28). Based on the literature, hyperpartisan styles which are detectable via linguistic features are defined as extreme behavior in favor of a particular political party, which usually correlates with strong motivation to create false information (Cruz et al. 2019; Potthast et al. 2017). News titles are one of the main sources that can be used to evaluate hyperpartisan; therefore, deceptive clickbait titles may work as a potential indicator for recognizing false information (Chen et al. 2015). Horne et al. (2019) studied the hyperpartisan news classifications and illustrated that content-based features such as writing styles are reasonably robust to changes in the news cycle. Therefore, the content-based detection method is one of the robust approaches in false information detection studies.

The third type of false information detection model, which is a more recent approach, is visual-based model where detection techniques intend to utilize graphics to evaluate false information content (Manzoor and Singla 2019). In continues to development of multimedia technologies, Cao et al state that "with the development of multimedia technology, false information attempts to utilize multimedia content with images or videos to attract and mislead consumers for rapid dissemination, which makes visual content an important part of fake news" (Cao et al. 2020, P.141). As an instance, Stahl's study (2018) illustrates that false information posts use significantly more graphics in content including morphed images, doctored video, or a combination of both. Giachanou et al. (2020) proposed a multimodal system to differentiate between fake and real posts based on a neural network and a combination of textual, visual, and semantic information. In this study, researchers proposed that estimating the similarities between the text and the image is essential in false information detection where false information may contain images that are not relevant to the text. As it is clear, content-based detection is not restricted to textual content and it can be extended to images, videos, and audio content as well. Moreover, it illustrated that the combination of the various content such as similarity of the image and text could be a useful predictor in the detection of false information (Zlatkova et al. 2019).

Social Context-based False Information Detection: The second stream of false information detection research focuses on social context features. Social context features can be obtained from the user-driven social engagement behavior on social media platforms. Social engagements indicate the news spread process over time, which could provide useful information to understand the truthfulness of news articles. Based on the literature, there are three major types of social media context features: users, generated posts, and networks (Shu, Wang, et al. 2020). From users' perspective, understanding users' behavior and capturing users' profiles information can be a useful source of information for false information detection (Shu et al. 2018). False information is likely to be created and spread by non-human accounts including social bots which can pose serious cyber threats to society and public opinion (Heidari et al. 2020).

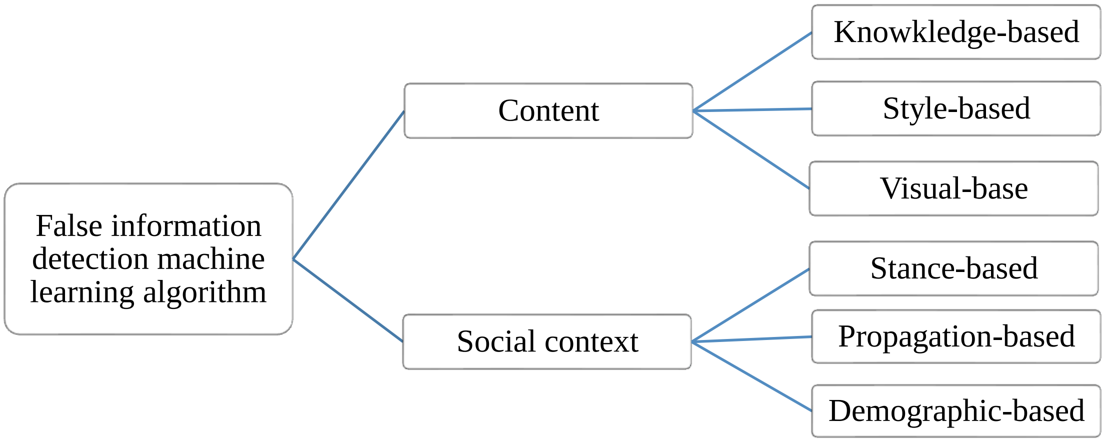
From a generated posts perspective, people express their emotions and opinions through social media posts which can be used as a factor to distinguish potential false information via public reactions to shared posts on social media (Sabeeh et al. 2020; Shu, Mahudeswaran, et al. 2020; Shu, Wang, et al. 2020). From the network perspective, false information propagation usually starts from an echo chamber cycle, where people seek out information that reinforces their existing views regarding specific topics without facing opposing views. The main effect of the echo chamber is on the users' network size and its diversity where individuals tend to preferentially connect to the people within a cluster and form a community with similar ideologies (Bruns 2017). However, social content features are not restricted to users' engagement, as researchers explore the correlation of publisher bias, news stance, and relevant user engagements to propose a Tri-Relationship false information detection framework (Shu, Wang, et al. 2017).

Social context-based detection models are categorized into three major groups: stance-based, social network-based, and user demographic-based (Shu, Sliva, et al. 2017). While stance-based models focus on individuals' viewpoints from relevant post content (Scarselli et al. 2008; Shu, Sliva, et al. 2017), social network models focus on forms of connection between users such as follower/followee relationships (Shu, Bernard, et al. 2019) and demographic-based models concentrate on demographic information of individual such as age, gender, education, and political affiliation (Long 2017). Based on the literature, automating stance evaluation is a valuable step towards assisting human fact-checkers to detect false claims. The main goal of stance-based models is to evaluate individual viewpoints from a post and classify the user attitude toward an event into 'agree', 'disagree', 'discuss', or 'unrelated' regarding some specific events (Riedel et al. 2017). Stance-classifiers rely on linguistic features to predict stances. Topic modeling methods such as structural topic modeling (Canziani et al. 2021; Esmizadeh et al. 2020) and latent Dirichlet allocation (Mohammad et al. 2017), and bag-of-words methods such as term frequency and term frequency-inverse document frequency are some of the well-known stance predictors (Appiah Otoo et al. 2020; Riedel et al. 2017). Researchers believe that by using these methods and identifying the stance values of relevant posts, machine learning algorithms can infer the news veracity.

On the other hand, propagation-classifiers rely on the fact that false and true news follow different propagation patterns and spread differently on social media (Monti et al. 2019). In this study, Monti applied a geometric deep learning algorithm to detect false information based on social network structure and propagation patterns. Their study indicates that propagation-based approaches have advantages over content-based methods due to their language independence. Moreover, demographic detection techniques rely on demographic information of speakers' profiles to differentiate false information from real ones. Long's (2017) study illustrates that by adding speaker profiles such as party affiliation, speaker title, location, and credit history, the detection model works significantly better in accuracy.

By considering the above literature, I can summarize false information detection strategies from ML perspective as Figure 2:

Figure 2. False information detection strategies



False information detection on social media is a newly emerging research area with several research directions from a data mining perspective. By considering the literature review from a data mining perspective, false information detection fields can be grouped as feature-oriented and model-oriented areas. In feature-oriented studies, researchers focus on finding the best features and best approaches to extract features from content, social context, and propagation structure. Furthermore, in model-oriented studies, researchers focus on applying the best supervised, semi-supervised and unsupervised machine learning algorithm to have a good detection tool with higher accuracy. However, from information system literature, the false information detection field is deficient in theory-driven detection methods. Although false information detection research has shown great progress in recent years, it is proven that interdisciplinary research on false information detection approaches by relying on scientific contributions from various disciplines including social science and engineering would provide higher interpretability in this field (Zhou and Zafarani 2019). False information detection can be done by evaluating writing style, propagation structure, and false information spreaders' intentions. While all these aspects can help to distinguish false information from the truth, without supportive theories, it would be difficult to interpret the latent features and the analysis would provide limited insight. Applying appropriate theories enables researchers to represent news content, propagation structure, and users' intentions based on interpretable theory-driven features to detect false information (Zhou and Zafarani 2019). The lack of theory-driven false information detection methods calls for a new stream of studies in this field.

Applicable Theories on False information detection

"Fundamental human cognition and behavior theories developed across various disciplines, such as social sciences and economics, provide invaluable insights for false information analysis"(Zhou and Zafarani 2020, P.5). Not only do these theories provide new opportunities for false information data studies, but they can also facilitate building well-justified and explainable models for false information detection and intervention as well (Miller et al. 2017; Zhou et al. 2020). In this section, some of the well-known interdisciplinary theories will be discussed.

Emotional broadcaster theory: In 2005 Harber and Cohen (2005) proposed the significant role of emotional disclosures on social media sharing behavior based on Emotional broadcaster theory (EBT). EBT is based on four central propositions: "(a) people who have experienced major events are emotionally compelled to communicate their experiences, (b) disclosures are most therapeutic when they succeed as reportage, (c) emotionally motivated disclosures often contain important information for listeners, and (d) the potency of tellers' own emotional experience will predict how far their stories travel across their social networks" (Harber and Cohen 2005, P.383). EBT predicts that stories that carry greater emotional impact are more likely to be shared across social media (Duffy et al. 2020; Rimé 2009; Valenzuela et al. 2017). By relying on the discrepancy theories of emotional arousal, EBT proposed that by drawing attention to event or belief discrepancies, emotions may prompt people to examine disrupting events more closely (Harber et al. 2014). Based on EBT, I believe that extracting emotional scores of the false content could be useful in detecting false information. Sentiment analysis is an NLP technique that determines whether textual data is positive, negative, or neutral. Studies on false information sentiment analysis on social media such as Twitter and Facebook demonstrate that hidden bias regarding concepts can be reflected using sentiment analysis (Bhutani et al. 2019; Dey et al. 2018; Hirlekar and Kumar 2020). Therefore, EBT has been used in Essay I to highlights the importance of sentiment analysis in applying ML techniques in disinformation detection.

Information Manipulation Theory: Based on Information Manipulation Theory (IMT) message deceptiveness is a message property that can be defined as a situation where a sender might assemble information to give a false impression to the readers (McCornack 1988). In this situation, the sender selects certain facts in the message from an available amount of information to get someone to believe that something untrue is true. Based on the information manipulation theory (IMT) quality, quantity, relation, and manner are the four main dimensions that can be used to determine the degree of deceptiveness in the literature (Anjalin and ZAMAN 2018; Jacobs et al. 1996; Zhou and Zafarani 2020).

In this situation, the sender selects certain facts in the message from an available amount of information to get someone to believe something untrue is true (Grice 1975) provided following conversational maxims (CM) as four main dimensions to determine the degree of deceptiveness:

- “Quantity: Make your contribution as informative (1) as is required (for the current purpose of the exchange). Do not make your contribution more informative than is required.

- Quality: Do not say what you believe to be false. Do not say that for which you lack adequate evidence.
- Manner: Avoid obscurity of expression. Avoid ambiguity. Be brief (avoid unnecessary prolixity). Be orderly.” (Thomas 1997, P.389)

McCornack in his second version of MIT proposed that extreme information quantity often exists in deception (McCornack et al. 2014). McCornack in his IMT theory presented that “deceptive messages function deceptively because they covertly violate the principles that govern conversational exchanges. Given that conversational interactants possess assumptions regarding the quantity, quality, manner, and relevance of information that should be presented, it is possible for speakers to exploit any or all of these assumptions by manipulating the information that they possess so as to mislead listeners” (McCornack 1992, P.1). IMT argued that deceptive messages are distinctive from others in a manner that they involve deviation from rational and cooperative conversational behavior. Grice demonstrates that (1989) deceptiveness can originate from the amount of disclosed information, misrepresentation of the information, the use of ambiguous phrases, and changing the relevance of the information.

Computational linguistics is an interdisciplinary field that tries to model the natural language based on computational approaches (Grishman 1986). Based on the study that has been done on the word frequency in computational linguistics, Kao and Jurafsky (Kao and Jurafsky 2012, P.9) stated that “poems written by professional poets may contain more difficult words and lower average word frequencies than poems written by amateur poets.” Just as word frequency has been used to evaluate professional poems from unprofessional ones, I believe that it is important to extend this approach to detect true public health-related articles and false ones that are produced by unprofessional producers. Therefore, in essay I, I proposed that evaluating the content based on the IMT construct could be extremely useful in detecting false information.

Uses and Gratification Theory: Uses and gratification theory (UGT) proposes an audience-centered perspective to identify why people engage with special media. UGT implies that individuals actively choose media content that complies with their needs and desires (Katz et al. 1973). In this way, studies have been done to understand users' engagement stimulus on social media networks as well. As an instance, Saho (2009) extended the UGT to investigate an individual's purpose for consuming, sharing, or creating content through internet-enabled media platforms. This study highlighted users engage with online media for self-expression, establishing social interactions, and entertainment. Moreover, Leung (2013) studied individuals' motivations and intentions to generate content on the internet. This study categorized users' needs satisfied through the process of online user content generation (UCG) into five groups: cognitive needs, affective needs, personal integrative needs, social integrative needs, and tension release needs. Cognitive needs highlight the fact that online-generated content broadens users' thinking process and learning of the world and community. Social integrative needs and affective needs indicate how people fulfill their social, communicational, and emotional needs by sharing their views, thoughts, and experiences and getting satisfaction. Personal integrative needs reveal the fact that individuals engage in an online content generation that provides credibility, respect, status, confidence, and stability. Finally, tension release needs indicate that online users engage in an online content generation to meet their entertainment needs as they see these activities to be fun, entertaining, and escape from tensions (West et al. 2010).

Other intentions including monetary benefits and political affiliations also affect online user content generation behaviors as well (Liu 2017; Liu et al. 2020). Online false information content creators engage in such behavior to satisfy their specific needs. I believe that just like ordinary online content generators, false information producers also look for ways to create content that meets their needs. In the same direction, in essay II, I proposed that disinformation producers are aware of these needs and produce content that meet audiences' five major cognitive needs. I believe that extracting UGT construct scores from generated content by dictionary-based software might be a powerful approach to recognize false information from true ones.

The provided theories in this section, mostly focus on the content features of information and how they can be significant on users' behavior. As this study focuses on detecting disinformation based on content and its effect on making decisions regarding the truthfulness of the message, I believe these theories could be useful in interpreting my research analysis and provide a broader area to implement the findings.

CHAPTER III: DISINFORMATION DETECTION BY ML ALGORITHM (ESSAY I)

Introduction

"Across a wide range of domains, it is critically important to correctly assess what is true and what is false: Accordingly, differentiating real from unreal is at the heart of society's constructs of rationality and sanity" (Pennycook et al. 2018 p.1865). The spread of false information is one of the recent issues that has plagued social media platforms during, as well as after, the U.S. elections. While most of the studies in this area have focused on the spread of false information on social media networks, including Twitter and Facebook, detecting false information on websites content is rare. People in their social media network will share their opinions about a topic by resharing or retweeting from others or by conveying the information on their own.

Nowadays, people prefer to search for health-related articles online rather than accessing traditional approaches due to the easy access to the massive amount of available information (Ghenai 2017). The massive spread of false information on these platforms could lead to social problems such as losing trust, increasing bias and false beliefs, and affecting the way people respond to real news (Shao et al. 2018; Shu, Wang, et al. 2019). As the COVID-19 pandemic has unfolded, the popularity of websites claiming to provide information about COVID-19 has grown significantly. The uncertainty inherent with issues surrounding the virus also gave rise to the number of websites spreading false information. The popularity of some of these websites spreading false information can also be understood due to the distrust of the official websites that were perceived to be promoting public policies that encroached on citizens' freedoms. For example, anti-vaxxers, who have strong organizational structures and an existing following, harnessed public distrust to amplify the uncertainty surrounding COVID-19 related policies (Balog-Way and McComas 2020).

Researchers in the false information detection field have primarily focused on formulating false information diffusion similar to disease epidemics with various infectious probabilities based on features including interaction frequency and geographic proximity (Ghoshal et al. 2020; Wu et al. 2016). In this study, I focus on disinformation detection, a specific type of false information (Figure 1). Based on the literature, disinformation can be defined as the act of creating and sharing false and manipulated information to deceive and mislead the audiences to gain political, personal, and financial benefits (Gelfert 2018; Tandoc Jr et al. 2018). The information manipulation theory (IMT) proposes that quality, quantity, relation, and manner are the four main dimensions that can be used to determine the degree of deceptiveness in the literature (Jacobs et al. 1996; Zhou et al. 2020). I believe detecting and labeling the disinformation content at the source before it reaches the public is an essential mission in mitigating its consequences. In this study, I attempt to extend IMT in detecting disinformation.

While manual fact-checking is difficult, time-consuming, and expensive, various automated detection solutions could speed up this process (Kula et al. 2020). Disinformation detection faces several challenges since it is intentionally created to mislead readers (Pasławska and Popielska-Borys 2018); therefore, it is interesting to provide an applicable classification algorithm to detect false narratives. False information detection methods have been dramatically limited by the lack of proper labeled benchmarking datasets (Wang 2017). To combat this weakness, presenting a systematic approach to detect false information websites is critical. This study provides a systematic approach to collect a comprehensive dataset for false and true textual information related to the COVID-19 pandemic over 2020 years to fill this gap. Specifically, I aim to answer the following research questions: “Can ML be used to identify disinformation?” and “Which characteristics of the content are important in the identification of disinformation? To answer the research questions, the current study utilizes IMT as its main theoretical framework to develop an analytical model for detecting systematic differences between the misleading and truthful text in terms of their structure to support the proposed claims. By focusing on content features, the finding of the present research illustrates the importance of evaluating the word repetition pattern as quantity metric and sentiment scores as manner metric in detecting disinformation.

Literature Review

In recent years concerns regarding false information in textual information has gained attention by increasing the negative effect of false information in various fields, including political polarization (Hameleers and van der Meer 2020), climate change debates (Treen et al. 2020), and public health-related topics (Southwell et al. 2019). In information science literature, automated deception detection methods in textual information are of interest (Rubin and Lukoianova 2015). These methods can be applied in enhancing human abilities in recognizing true information from misinformation (Purian et al. 2020; Rasool et al. 2019), assessing information qualities (Osatuyi and Hughes 2018; Rubin and Lukoianova 2013), and assessing source credibility (Shao et al. 2018).

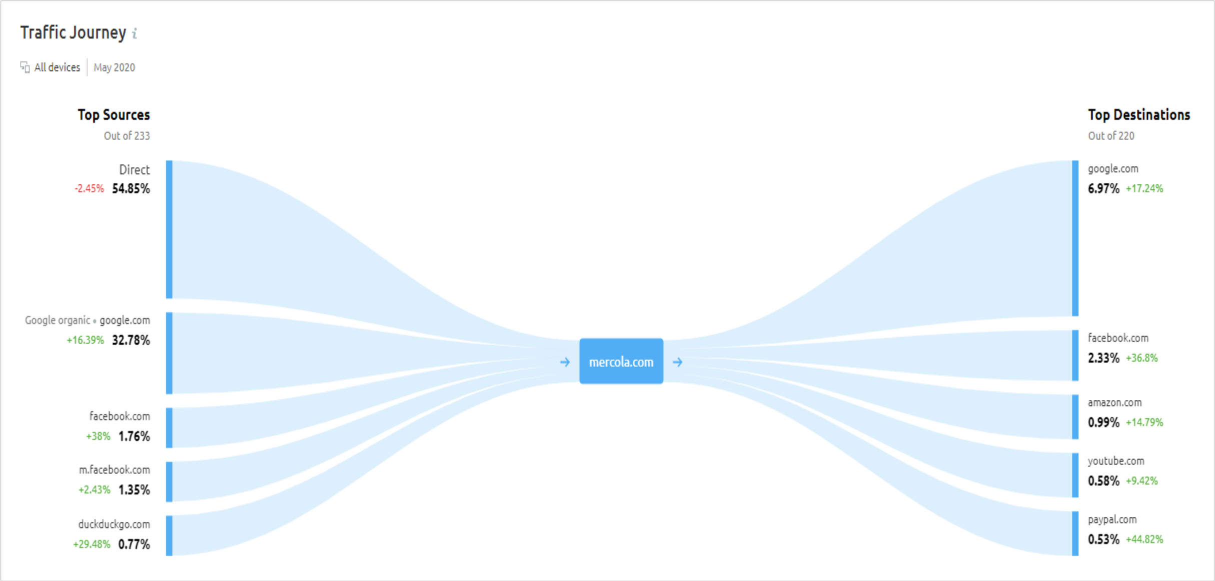
Moreover, the accessibility of online information over cyberspace and the expansion in computer-mediated communication have created challenges for every society by offering the internet as one of the main resources to persuade individuals to make decisions (Wall and Kaye 2018). Manipulated and deceptive information can mislead users and negatively affect decision-making procedures (Habib et al. 2019). Advanced deceptive information detection methods provide potential opportunities to help online information seekers and decision-makers by alerting them about the accuracy and reliability of the textual information resources. During the COVID-

19 pandemic, an infodemic in the medical publishing world was observed besides the spread of coronavirus disease. An infodemic or outbreak of false information can be defined as “a contagious disease infecting our information culture”(Solomon et al. 2020, P1806). On February 15, 2020, a high-profile group of authors associated with the World Health Organization (WHO) collected and organized global ideas to fight the coronavirus disease infodemic (Tangcharoensathien et al. 2020). In the current study, I use information related to the spread of the virus, online interactions, and public-generated content on social media to understand false information detection and mitigation. The vast amount of unstructured text data sets and automated deception detection present challenging tasks (Rubin and Chen 2012; Shu, Sliva, et al. 2017) that can be overcome with big data analytics techniques such as natural language processing (NLP) and machine learning (ML) techniques (Karimi et al. 2018; Rubin and Lukoianova 2015; Torabi Asr and Taboada 2019).

Eysenbach (2020) classified the current infodemic into four levels: science, policy and practice, news media, and social media. Knowledge translation occurs by shifting information from one audience to another. Therefore, false information from news media can easily spread on social media via its users. I believe that detecting false information at the media news level would mitigate the misinformation spread on social media platforms as well. For example, on Feb.18, 2021, U.S Food and Drug Administration (FDA) publicly released a warning letter informing Joseph Mercola (a well-known individual whose website frequently shares information that may be considered false medical information) that “the FDA has observed that your website offers “Liposomal Vitamin C,” “Liposomal Vitamin D3,” and “Quercetin and Pterostilbene Advanced” products for sale in the United States and that these products are intended to mitigate, prevent, treat, diagnose, or cure COVID-19 in people. Based on our review, these products are unapproved new drugs sold in violation of section 505(a) of the Federal Food, Drug, and Cosmetic Act (FD&C

Act), 2pt1 U.S.C. § 355(a).[2]" Although Facebook and Twitter try to restrict such accounts, traffic analysis of the website illustrates that more than 50% of the people reach their websites through direct links that could be sent to users' email accounts (Figure 3). Mercola supplements are also available on the Amazon website for purchase, and Figure 3 represented that in May 2020, amazon.com was the website's third direct destination, which illustrates the significant effect of the provided information on users' decision-making behavior to buy introduced products. The traffic analysis demonstrates that besides focusing on social media, researchers also need to investigate the content of the web pages to identify their information manipulation techniques and provide proper disinformation detection tools.

Figure 3. Traffic Journey for Mercola.com(https://www.semrush.com/dashboard/)



[2] https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/warning-letters/mercolacom-llc-607133-02182021

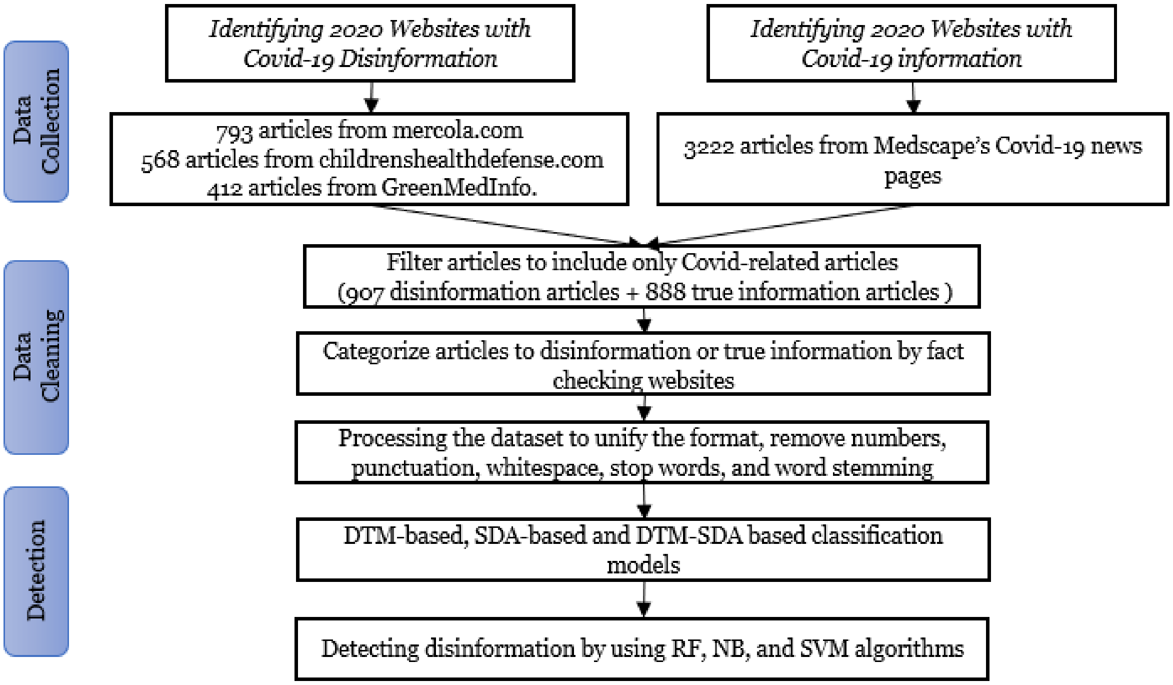
Based on the literature, certain lexical items could be predictive linguistic signals to recognize disinformation stories from real ones (Porter and Yuille 1996). Existing psycholinguistic lexicons techniques (Mihalcea and Strapparava 2009) and standard classification algorithms like decision trees and logistic regression (Fuller et al. 2009) for detecting deception in text reached an average classifier with 70% accuracy rate at the level of Lexico-semantic analysis (Rubin et al. 2015). Sentiment analysis is another feature extraction method that can be used in ML and NLP techniques to automatically classify text as false information or true piece of information (Mesquita et al. 2020).

The role of information manipulation in producing deceptive information, the lack of study on news media sources of false information, and the accuracy rate of the existing method call for highly accurate methods to evaluate disinformation and recognize it from trustworthy information sources. By considering these elements, I applied three well-known ML algorithms to our dataset and compared their performance in detecting articles containing false claims.

Methodology

The main goal of this paper is to classify the disinformation health-related articles from real articles. Using 1768 health-related published articles on different websites, I applied several machine learning algorithms to classify these articles, and then I evaluated these algorithms based on their prediction accuracy. Our proposed model for disinformation articles detection is based on computational linguistic terms, including the articles' word frequency and sentiment scores. To visualize the methodology's pattern, Figure 4 depicts the proposed method diagram that classifies the articles based on the word frequency and sentiment score patterns in each article.

Figure 4. Visual scheme of applied methodology


Data Collection
Data Cleaning
Detection
Identifying 2020 Websites with Covid-19 Disinformation
Identifying 2020 Websites with Covid-19 information
793 articles from mercola.com
568 articles from childrenshealthdefense.com
412 articles from GreenMedInfo.
3222 articles from Medscape's Covid-19 news pages
Filter articles to include only Covid-related articles
(907 disinformation articles + 888 true information articles)
Categorize articles to disinformation or true information by fact checking websites
Processing the dataset to unify the format, remove numbers, punctuation, whitespace, stop words, and word stemming
DTM-based, SDA-based and DTM-SDA based classification models
Detecting disinformation by using RF, NB, and SVM algorithms

Data collection

Identifying Websites with COVID-19 False information: For the purpose of the current study, it was necessary to collect data and create a dataset including a sample of COVID-19 public health-related published articles in 2020, as there was not a preexisted dataset. I gathered articles from various sources.

After a careful review of the literature regarding false online information and its impact on the COVID-19 pandemic, a survey of popular alternative health websites was undertaken. While many websites contain a mixture of political and health-related materials, for this study, only websites that focused nearly exclusively on health topics were considered. Lists found on various websites that describe alternative health websites were consulted. Some of these lists included

Quackwatch[3] and various pages posted on NewsGuard on the Coronavirus Misinformation Tracking Center page (detailed lists and data were formally publicly shared, although they are no longer available except by request). The Center for Countering Digital Hate reports were also consulted (counterhate.com). In March 2021, CCDH released an additional study which named Joseph Mercola, Robert F. Kennedy, and Sayer Ji as among a list that they identify as the "disinformation dozen", a group that they find is responsible for most of the current vaccine false information being spread on social media (The Disinformation Dozen | Center for Countering Digital Hate).

Using Massmine[4] which is a command-line tool designed for researchers to simplify the collection and use of data from online sources, a collection of tweets was compiled using common hashtags associated with COVID-19 false information (#vaccineskill, #nomaskonme, #scamdemic, #COVIDhoax, #vaccineinjury, #nomaskmandates, #Vaxxed2, #nomorelockdowns, #masksdontwork, #vaxxed, #nochip, #learntherisk, #antimask, #BillGates, #plandemic, #stopbillgates, #iwillnotcomply, #Lockdownskill, #nogates, #NoCOVIDVax, #burnyourmask, #COVIDisOver, #nomasks, #endlockdownsnow, #COVIDiots, #nomask, #HerdImmunity). By following links listed on Twitter posts and investigating the number of responses and followers frequently mentioned websites were identified. Ultimately, the resulting data was composed to create a database of 38 websites that included additional information about the individuals named as authors of the sites and key players associated with the websites.

[3] https://quackwatch.org/consumer-protection/COVID-19-consumer-protection/

[4] https://www.massmine.org/index.html

Each website in the database was investigated to gauge the number of available articles on the site and the type of content available. Some websites (like NVIC) were eliminated simply because few articles existed on the website, while other websites (like HealthNut News by Erin Elizabeth) were eliminated because a larger portion of the articles on the website were reprints from Mercola's website. Finally, other websites were eliminated because their content consisted mostly of videos or "courses" that individuals could enroll in to learn more about the information shared on the website. After applying these criteria, three popular websites with a reputation for publishing COVID-19 false information were chosen for scraping and data gathering: mercola.com, childrenshealthdefense.org, and greenmedinfo.com.

Identifying Websites with COVID-19 Information: Once the three websites were identified to collect articles that would represent COVID-19 false information, comparable websites were sought out to collect articles that would represent COVID-19 information the scientific and medical community would generally accept. Interestingly, few examples of websites counterparts with reliable medical information fit the genre of health false information websites. Few websites with large archives of reliable health-related articles written, edited, or collated by a single individual or health celebrity exist. As an alternative to the false information website, medscape.com was utilized to collect articles about COVID-19 that the scientific and medical community would generally accept. Medscape.com is a medical news website that contains a large volume of archives of articles with a variety of medical topics, including an archive of articles related to COVID-19 from a variety of news organizations, including WebMD, Associated Press, Reuters, Kaiser Health News, and Medscape’s journalists. These articles may be found in the Latest News section of their COVID-19 resource page (medscape.com/resource/coronavirus).

NewGuard has assigned medscape.com a rating of 92.5 out of 100 for adhering to credibility and transparency standards[5].

Final Dataset: Archived articles published by Mercola, CHD, and GreenMedInfo in 2020 were scraped using both RSelenium and rvest packages. URLs were identified, collected, and utilized to access articles. Some of the URLs identified had broken links and were eliminated. Additionally, URLs that were not associated with health articles or contained little text (either recipe, exercise illustrations, or videos) were also omitted. Once these URLs were eliminated, the text body of the article, article date, and article title were scraped and imported into CSV files. Because many articles had inconsistencies in the way in which author information was coded, this data was not collected. The resulting data contained a total of 793 articles from mercola.com, 568 articles from childrenshealthdefense.com, and 412 articles from GreenMedInfo. The same technique was used to scrape text from Medscape's COVID-19 news pages resulting in a total of 4995 articles.

Data Cleaning

Articles from mercola.com, childrenshealthdefense.org, and greenmedinfo.com were then filtered to include only articles that had been identified and labeled as containing COVID-19 false information. The resulting dataset of false information articles represented 907 articles from the three websites. An equivalent number of observations was randomly selected from the medscape.com data distributed evenly by month, and all the observations were combined into a single dataset, including 1795 observations.

[5] https://api.newsguardtech.com/label/medscape.com?cid=1a184c47-f65b-403a-af0b-824b5b5b9bbe

Articles from mercola.com, childrenshealthdefense.org, and greenmedinfo.com were labeled to identify those which contained mostly information regarding COVID-19 and then were investigated to identify whether the information shared could be categorized as false information. A variety of resources were used to identify common false information themes found in typical COVID-19 false information, including the fact-checking website Snopes.com's Coronavirus Collection[6], NewsGuard's collect of top COVID-19 myths[7], the Poynter's Institute's fact-checking website for COVID-19 information database[8], and the World Health Organization's COVID-19 Mythbusters' website[9]. Each article was rated and categorized to false information or true information by two individuals independently. Records where there was no full agreement in the ratings were either omitted or revisited and discussed until consensus was reached.

Since the data set was collected by data scraping from various websites, it was important to clean up and preprocess the data before applying ML algorithms. A variety of r packages were utilized to clean the resulting data scraped from the four websites. The data cleaning and data preprocessing of our paper include the following steps:

- Transforming strings to lower case to unify the format
- Removing numbers
- Removing punctuations
- Removing white spaces

[6] https://www.snopes.com/collections/new-coronavirus-collection/
[7] https://www.newsguardtech.com/coronavirus-misinformation-tracking-center/
[8] https://www.newsguardtech.com/coronavirus-misinformation-tracking-center/
[9] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters

- Removing English stop words to reduce words
- Word stemming to reduce words to unify across the document

Machine Learning Algorithm

To be able to detect false articles and the true ones, I adopted three different ML algorithms, including Random Forest (RF), support vector machine (SVM), and Naïve Bayes (NB) algorithms in R Studio to demonstrate the efficiency of the classification performance on our dataset.

Support vector machine is a classification algorithm that has been designed for binary classification problems by identifying the best plane that separates the data points of two classes with maximum margin (Ahmad et al. 2020). SVM classification algorithm has been used in several false content detection problems, including deep fake detection on fake videos (Yang et al. 2019), fake image detection (Kunbaz et al. 2019), and fake news detection (Gravanis et al. 2019). Text data normally contains tens of thousands of terms, which means that text data is very high-dimensional, and SVM can take care of it (Deng et al. 2019). The application of SVM in disinformation detection and its ability to manage the high-dimensional problem make SVM an excellent option for this study. The random forest algorithm is the second algorithm for this study. RF is an extended format of decision tree algorithm which consists of a large number of decision trees working individually to predict an outcome of a class. In this algorithm, the final prediction of classes is made based on a class that received majority votes therefore, it has a lower classification error in binary classification (Kirasich et al. 2018). The low classification error in binary classification makes RF another applicable option for this study. Our last selected algorithm is NB. NB classifier is another classification technique based on Bayes' Theorem, which works quite well in text classification (Xu 2018) by checking the frequency of each word category in the training document set (Deng et al. 2019).

Feature selection

After cleaning the data set, I need to extract features from the text. As what I have discussed earlier, the lack of theory-driven disinformation detection methods calls for novel studies, it this section I investigated the selection of content-based features based on IMT and repetition theories. Based on theory of the repetition, word frequency will be higher if the focus of the writer is on changing audience's worldview (Groppe 1984). To be able to evaluate the word frequency of the selected article the count vector extracted the word frequency in each article. Furthermore, from IMT perspective, quantity dimension measures the amount of the provided information in content. In this study word frequency and repetition also has been used to evaluate the quantity dimension of the article as well. To do so, I convert the corpus of data to a document-term matrix (DTM). Document-term matrix is a mathematical matrix that represents the frequency of the words in documents. The third step of data preparation is removing sparse terms from the text file to clean extremely sparse terms. Sparse value is a numeric for the maximal allowed sparsity in the range from bigger zero to smaller one. Reducing the sparse value means reducing some of the sparse terms.

Manner is another dimension of the IMT that can be extracted by text analytics tools from textual content. Manner dimension measures the way that message have been conveyed and which is about the way the message has been said (McCornack 1992). I proposed that sentiment analysis could be able to evaluate the way that message has been said by evaluating the emotional tone of the content. Therefore, in the second step, I extract the matrix of sentiment scores for our dataset. Sentiment discovery and analysis (SDA) describes underlying attitudes, sentiments, and subjectivity (Han et al. 2020). Syuzhet package in R has been used successfully across a variety of disciplines to identify emotions: anger, anticipation, disgust, fear, joy, sadness, surprise, as well as general negativity/positivity sentiments (Arun et al. 2017; Jockers 2017; Liu and Lei 2018; Valdivia et al. 2017). I divided emotion scores by article word length to compare proportionalities of each emotion across sample subgroups. Then the document term matrix was combined with sentiment discovery output for further analysis.

I split our dataset into 75% train data and 25% test data in the third step. To evaluate the ML algorithms accuracy, three types of analysis have been conducted: prediction based on DTM, prediction based on SDA, and prediction based on both features. The next section will represent the performance of the algorithms.

Results

ML Results

In this study, I have compared three predictive models on three different classification algorithms. I have considered sensitivity, specificity, positive prediction value, negative prediction value, prevalence, detection rate, detection prevalence, balanced accuracy, and 95% confidence interval for accuracy as our major comparison metrics, and the averaged performance of the algorithms are compared based on 10 runs. Table 1 illustrates the formulation that has been used to calculate each metric.

Table 1. The definition of major comparison metrics

Classification Metric	***Formula***
Sensitivity	TP/(TP+FN)
Specificity	TN/(TN+FP)
Positive prediction value	TP/(TP+FP)
Negative prediction value	TN/(TN+FN)
Prevalence	(FN+TP)/ (TP+FP+TN+FN)

Detection rate	TP/(TP+FP+TN+FN)
Detection prevalence	(FP+TP)/ (TP+FP+TN+FN)
Balanced accuracy	(Sensitivity + Specificity)/2

* TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative

DTM-based model

The first model included document-term matrices as predictors of our classification model. The DTM-based model detects disinformation articles based on the word repetition pattern and occurrence of the words in the written text. Firstly, the Random Forest algorithm was evaluated and then Naïve Bayes and SVM algorithms were run on the DTM based model. Among these three algorithms, the random forest algorithm has the highest accuracy rate (0.85) among the rest (Table 2Table 2). Based on the RF analysis, in each split, 8 variables have been used. "vaccine", "information", "people", "world", "public" and "covid" are a group of stem words that have been used repeatedly in detecting disinformation articles. Table 3 represents the Gini coefficients of the predictors. Gini coefficient is a metric to calculate the amount of probability of a specific feature that is classified incorrectly when selected randomly[10]. The output of the RF algorithm illustrates that these sets of words have been repeated more through disinformation articles rather than true articles. Table 2 demonstrates the performance of these algorithms in our DTM-based prediction model. The accuracy level of the DTM model highlights the importance of word frequency in false information detection.

[10] https://medium.com/analytics-steps

Table 2. Performance comparison in DTM-based classification model

Classification Metrics	*Random Forest*	*Naïve Bayes*	*SVM*
Sensitivity	0.86	0.89	0.89
Specificity	0.84	0.60	0.60
Positive prediction value	0.86	0.73	0.74
Negative prediction value	0.84	0.82	0.82
Prevalence	0.52	0.55	0.55
Detection rate	0.45	0.49	0.49
Detection prevalence	0.52	0.67	0.67
Balanced accuracy	0.85	0.75	0.75
95% CI	(0.82,0.88)	(0.72,0.80)	(0.72,0.80)

Table 3. Gini coefficients of DTM model

Predictor	**Gini Coefficients**
Vaccine	68.47
Information	41.10
People	25.35
World	24.58
Public	21.28
Covid	18.41
Patient	14.95
Corona Virus	12.97

SDA-based model

The SDA model is the second predictive model in this study. The SDA-based model attempts to recognize disinformation articles based on the tone or sentiment pattern and their occurrence in the written text. Just like the previous model, Random Forest, Naïve Bayes, and SVM algorithms were run on the SDA-based model. Among these three algorithms, the RF algorithm has the highest balanced accuracy rate (0.75) among the rest. Table 4 shows the performance of these algorithms in the SDA-based prediction model. Based on the RF analysis, “anger”, “sadness”, and “disgust” tones are a group of sentiments that has been used repeatedly in detecting disinformation articles. Table 5 represents the Gini coefficients of the predictors. The accuracy level of the SDA model highlights the importance of tone analysis in false information detection.

Table 4. Performance comparison in SDA-based classification model

Classification Metrics	***Random Forest***	***Naïve Bayes***	***SVM***
Sensitivity	0.76	0.64	0.75
Specificity	0.62	0.66	0.74
Positive prediction value	0.68	0.69	0.78
Negative prediction value	0.72	0.61	0.72
Prevalence	0.52	0.54	0.53
Detection rate	0.40	0.35	0.41
Detection prevalence	0.57	0.50	0.52
Balanced accuracy	0.75	0.65	0.70
95% CI	(0.71,0.79)	(0.61.0.70)	(0.65,0.69)

Table 5. Gini coefficients of SDA model

Predictor	**Gini Coefficients**
Anger emotional tone	90.61
Sadness emotional tone	79.40
Disgust emotional tone	70.91
Joy emotional tone	68.84
surprise emotional tone	61.48

DTM-SDA-based model

The last predictive model used both DTM and SDA variables to classify test articles as disinformation or true. It was assumed that the use of both variables would improve the classification model. Table 6 represents the performance of applied algorithms in the DTM-SDA-based prediction model. As it is clear, the combination of SDA variables and DTM variables resulted in better classification. The last model improved the accuracy level by 4% in all the classification algorithms. Likewise, in the previous models, RF has the highest accuracy rate of 88% with (an 84%-91%) confidence interval. The algorithms use both SDA and DTM variables to detect disinformation articles. Table 6 reveals the performance of these algorithms in the DTM-SDA-based prediction model. Table 7 represents the predictors with Gini coefficients of the predictors greater than 11 percentage.

Table 6. Performance comparison in DTM-SDA-based classification model

Classification Metrics	***Random Forest***	***Naïve Bayes***	***SVM***
Sensitivity	0.89	0.86	0.88

Specificity	0.86	0.61	0.84
Positive prediction value	0.88	0.71	0.86
Negative prediction value	0.87	0.80	0.86
Prevalence	0.54	0.53	0.52
Detection rate	0.49	0.46	0.46
Detection prevalence	0.54	0.64	0.52
Balanced accuracy	0.88	0.74	0.79
95% CI	(0.84,0.91)	(0.70,0.79)	(0.82,0.89)

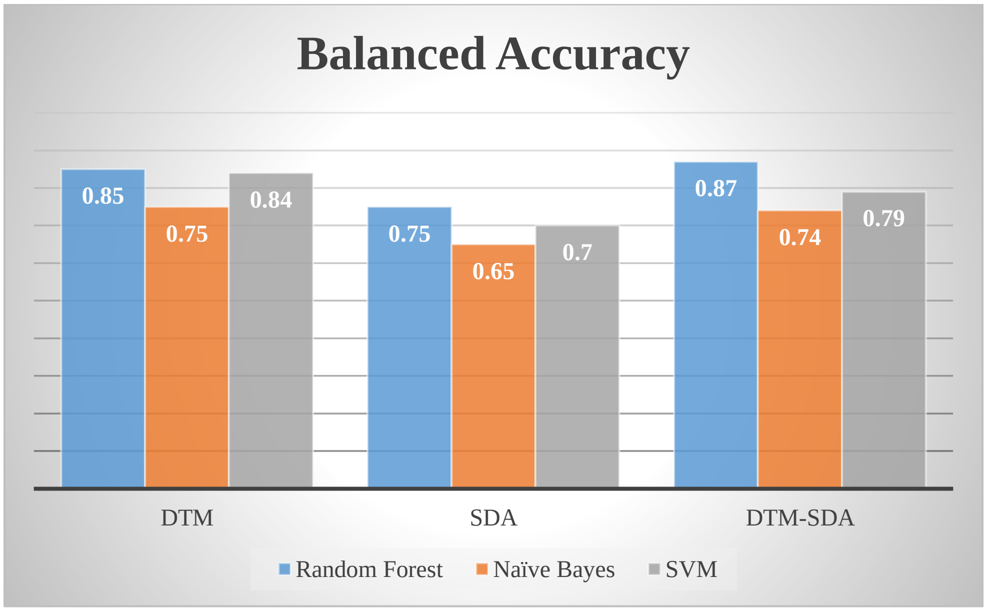
Table 7. Gini Coefficients of DTM-SDA model

Predictor	**Gini Coefficients**
Vaccine	64.61
information	22.61
Sadness emotional tone	20.14
Anger emotional tone	16.96
Disgust emotional tone	16.36
Public	16.16
virus	15.01
patient	13.44
Corona Virus	11.45

Comparative analysis

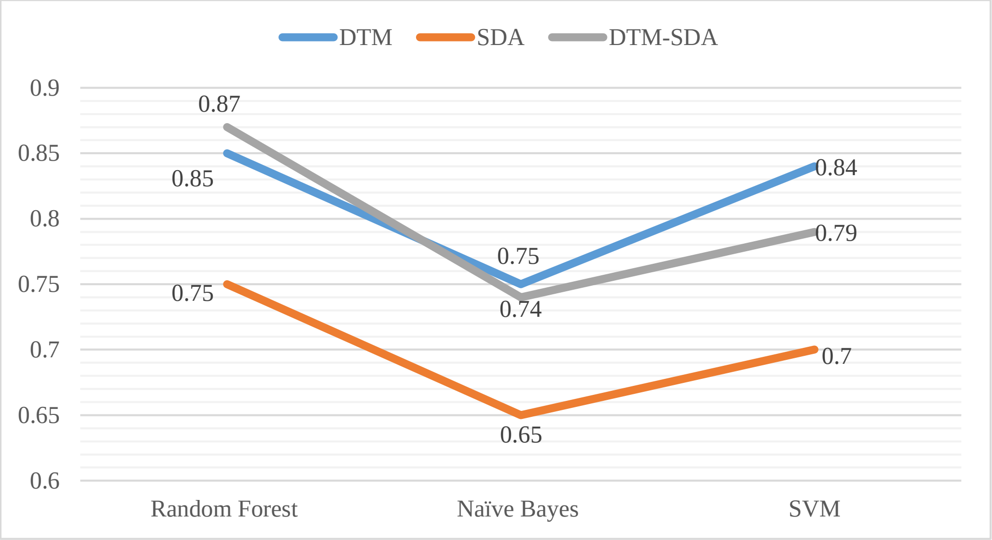
The comparative analysis of all the classification models illustrates that the Random Forest algorithm works the best among all and SVM works as the second-best model. Figure 5 depicts the performance of applied ML algorithms on our proposed classification models. In this graph, the blue columns, which represent the accuracy rate of RF algorithms, have the highest score in all models, and the orange ones, which represent NB algorithm performance, have the lowest performance in all the models.

Figure 5. ML algorithms performance based on classification models



Moreover, among the classification models, while the SDA-based model has the least accuracy to classify the disinformation articles correctly, its combination with the DTM-based model enhances the accuracy level significantly in all algorithms. Figure 6 illustrates that the gray line (DTM-SDA model) has better performance in comparison to the other two models in all algorithms. This comparison graph also highlights the higher importance of word selection and word frequency that emotional tones in detecting disinformation articles from the real ones too.

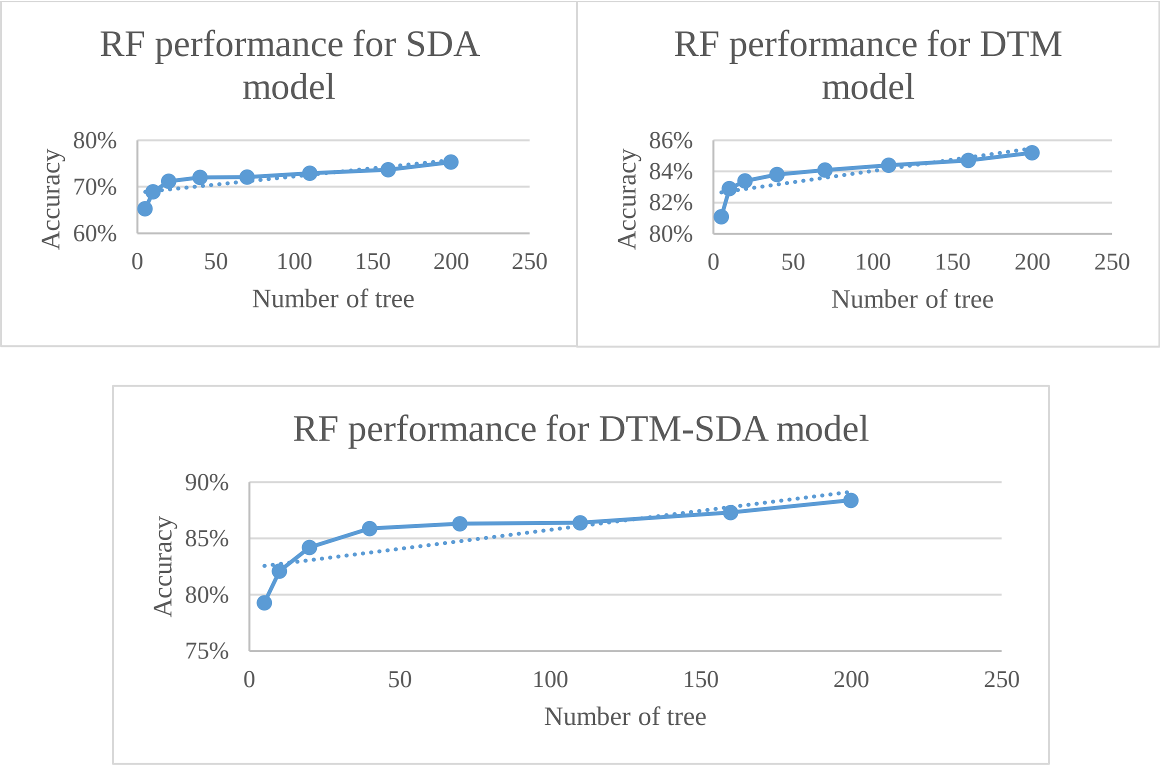
Figure 6. Model performance based on various algorithms.



Sensitivity analysis

In this section, I compare the performance of the RF algorithm and changes in its initial setting on the balanced accuracy. One of the main parameters that influence RF performance is the number of trees in the algorithm (Badawi et al. 2018). To do so, 5 independent runs have been performed and averaged to reflect the balanced accuracy based on per model and a various number of trees. Based on Figure 7, by increasing the number of trees in all models the averaged balanced accuracy gets better and then stays stable or with negligible changes after 200 trees. The improvement at the initial steps is significantly larger and later will gradually decrease as the number of trees increases. In this problem, the running time of the algorithm did not increase significantly by increasing the number of trees, therefore 200 is the optimum number of trees here. Figure 7 illustrates the influence of the number of trees on our problem.

Figure 7. The impact of the number of trees on the balanced accuracy of the RF algorithm.



Discussion

Summary of Findings

As has been discussed in previous sections, computational linguistics tries to model natural languages based on computational approaches. The analysis of word count illustrates that there is a significant difference between the word choice of disinformation articles and true ones. As an instance, Coronavirus, COVID, health phrases in disinformation articles have been repeated 2 to 3 times more than the true articles on average. This study extends the Kao and Jurafsky (2012) research on professional poets to the false information detection field. As Kao and Jurafsky state that poems written by professional poets may contain lower average word frequencies, the analysis of our dataset also confirms this fact is also supported in the true health-related article written by professionals. Based on the literature disinformation can be defined as the act of creating and sharing false and manipulated information to deceive and mislead the audiences to gain political, personal, and financial benefits (Gelfert 2018; Tandoc Jr et al. 2018). Groppe in his theory of repetition article stated that " if the focus of the writer or speaker is on the audience and the action the audience should take to change or maintain a world view, then repetition will be frequent" (1984. P.167). Considering the abovementioned definition, and the theory of repetition could be concluded that repetition would be one of the major tricks of the false information producers to influence their audiences influence worldview, where the results of both DTM-based and DTM-SDA-based models highlighted this fact. On the other hand, like social media such as Twitter and Facebook content, sentiment analysis of our study demonstrates that sentiment scores can reveal hidden bias regarding concepts to recognize disinformation articles from real articles. In this study, sentiment analysis illustrates that while COVID-19 disinformation articles have higher scores in anger, disgust, and fear tones, true articles have higher scores in anticipate, sadness, and trust tones. Both groups of articles have almost the same sentiment scores in surprise and positive tones.

Therefore, the results of classification algorithms on all three models illustrate the importance of computational linguistics in the false information detection field. However, the comparison between the performance of the DTM model and the SDA model illustrates the higher importance of word selection and word repetition in false information detection models. However, the third model also illustrates a better performance of our classification model when it includes both document-based and sentiment-based variables as predictors. It seems like COVID-19 false information producers try to influence their audience to change or maintain a worldview by repetition techniques. Based on sentiment features, our results demonstrate that COVID-19 false information producers try to influence their audience by injecting anger, disgust, and fear tones to encourage them to buy their supplement or participate in anti-mask or anti-vaccine groups.

Contribution

Based on what I discussed earlier, human beings are not particularly good at detecting deceptions, and manual fact-checking is difficult, time-consuming, and expensive, various automated detection solutions would speed up this process. The current study attempted to apply ML algorithms to help this procedure. The results of the algorithm and their accuracy rate highlight the good performance of ML techniques in detecting disinformation. Therefore, our deception detection models would make several important contributions to both research and practice. From an academic perspective, one major theoretical contribution of this study would be extending the theory of repetition in detecting health-related disinformation article detection research by applying a DTM-based model in the disinformation article detection algorithm. Based on the analysis, the linguistic structure including word repetition was one of the significant factors in detection disinformation which illustrates the importance of linguistic structure in the deception detection field. Furthermore, based on the information manipulation theory (IMT) manner is one of the dimensions that can be used to determine the degree of deceptiveness in the literature. the SDA-based model results represent that sentiment analysis and especially negative emotional tone is the key attribute in detection disinformation. The analysis demonstrates that disinformers produced negative emotion-embedded content in comparison to true articles. This change of manner in representing information can extend the application of IMT in medical-related articles and be also used to detect deception on webpages. I believe that the presented models can provide insight into the disinformation article content, writing structure, and sentiment tones. Moreover, it highlights the importance of computational linguistics in disinformation detection procedures.

From a practical perspective, these findings could be valuable for disinformation recognition tools providers to enhance their accuracy level by including both word repetition and sentiment tones as significant predictors in flagging false information. Disinformation detection is not restricted to political topics and could be used in various filed. As an instance, damaging the brands' reputation is another area that disinformation can affect drastically (Castellani and Berton 2017; Jahng 2021; Mills and Robson 2019). I believe these findings could also be valuable for business managers in evaluating disinformation reviews regarding their brands and products to mitigate their catastrophic financial damages. Recognizing disinformation reviews and studying their content could be useful to make strategic commercial decisions in facing and resolving the issues.

Limitation and Future Research

Like all other studies, this study has several limitations that might be addressed in further research studies. First, I believe that environmental factors and demographic information about the false information producers can also be critical in detecting disinformation articles. These variables need to be included as potential directions to improve the proposed model or to be addressed in future studies. I limit our dataset to COVID-19 related articles. This limitation makes it easier to classify articles as disinformation or true, however, restricts the generalizability of the model. A potential future direction could be the comparison of our model and algorithms in different content to understand how topics and content affect detection procedures. Moreover, another potential future direction could be in the cognitive bias study field, to understand how human behavior changes and how DTM and SDA features affect human decision-making procedures in presence of disinformation. Understanding cognitive procedures would be critical in understanding partisan polarization too which has been known as the primary psychological motivation behind political false information (Osmundsen et al. 2020).

CHAPTER IV: DISINFORMATION ENGAGEMENT ON SOCIAL MEDIA FROM A THEORETICAL PERSPECTIVE (ESSAY II)

Introduction

Have you ever noticed that some topics on social media networks trend out of nowhere? Where do they come from? What could explain it? Could it be the content of the post, the emotions it evokes, or the social media network structure? Pew Research Center in 2016 reported that 62 % of U.S. adults receive their news on social media while 66 % of all Facebook users access the platform for news consumption (Gottfried and Shearer 2016). According to Pennycook et al. "Across a wide range of domains, it is critically important to correctly assess what is true and what is false: Accordingly, differentiating real from unreal is at the heart of society's constructs of rationality and sanity" (2018 p.1865). False information is one of the recent issues that has plagued social media platforms during, as well as after, the U.S. elections. Based on the literature disinformation is one type of false information that can be defined as a piece of intentionally misleading or biased information; manipulated narrative or facts[11] that attempts to create deceptive and false statements to convince someone of untruth (Nimmo 2016).

The massive spread of the disinformation could lead to numerous social problems such as breaking the validity balance of the news ecosystem, enhancing biased and false beliefs, and impacts the way that people respond to authentic information (Shao et al. 2018; Shu, Sliva, et al. 2017; Shu, Wang, et al. 2019). During the COVID-19 pandemic, numerous users on social media accounts and their blogs produced intentionally misleading or biased information to convince their

[11] https://www.dictionary.com/browse/disinformation

audience that their product can mitigate or cure COVID-19[12]. As an instance, on Feb.18, 2021, U.S Food and Drug Administration (FDA) publicly released a warning letter informing Joseph Mercola (a well-known individual whose website frequently shares information that may be considered medical false information) that “the FDA has observed that your website offers “Liposomal Vitamin C,” “Liposomal Vitamin D3,” and “Quercetin and Pterostilbene Advanced” products for sale in the United States and that these products are intended to mitigate, prevent, treat, diagnose, or cure COVID-19 in people. Based on our review, these products are unapproved new drugs sold in violation of section 505(a) of the Federal Food, Drug, and Cosmetic Act (FD&C Act), 2pt1 U.S.C. § 355(a).[13]” However, the range of disinformation was not restricted to selling products. Additionally, anti-maskers, anti-vaxxers tried to sow doubt in the efficacy of the vaccines by amplifying the uncertainty surrounding the origins of COVID-19 and the procedure and policies employed to develop and administer the vaccines. Many influencers try to propagate conspiracy theories or use fear tactics to gain support and increase their followers. This explosion of disinformation around the COVID-19 topic calls for novel approaches to investigate the nature of disinformation to mitigate the catastrophic effects of disinformation on society. The spread of disinformation over social networks and the released letter highlighted the significance of the disinformation during COVID-19 ear as well.

To be able to investigate the nature of disinformation and the ways that have been used to engage users, this study attempted to identify the various emotional, cognitional, and structural components in disinformation written posts and highlight the way disinformation producers

[12] https://www.fda.gov/consumers/health-fraud-scams/fraudulent-coronavirus-disease-2019-COVID-19-products

[13] https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/warning-letters/mercolacom-llc-607133-02182021

enhance their users' engagement. Therefore, by focusing on understanding the characteristics of content, I will answer the following two research questions: "What content factors differentiate information from disinformation on social media?" and "Which content dimensions of news content can influence the engagement score of the posts on social media?

Studying the reasons why users fall prey to disinformation is one of the major branches in the disinformation field. The literature illustrates that while some researchers believe that people are more likely to share disinformation when it is viewed as novel (Vosoughi et al. 2018), some believe it is rooted in personal differences (Vishwanath 2015) and some researchers look for the roots in content. For instance, Wang (Wang et al. 2015) represented users are more likely to be engaged by Twitter posts by re-tweeting posts that have happiness emotions than anger emotion. Although disinformation detection research has shown great progress in recent years, it is proven that interdisciplinary research on disinformation detection approaches by relying on scientific contributions from various disciplines including social science and engineering would provide higher interpretability in this field (Zhou and Zafarani 2019). Disinformation detection can be done by evaluating writing style, propagation structure, and false information spreaders' intentions. While all these aspects can help to distinguish the untrue messages from the truth, without supportive theories, it would be difficult to interpret the latent features and would provide limited insight. Applying appropriate theories enables researchers to represent news content, propagation structure, and users' intentions based on interpretable theory-driven features to detect disinformation (Zhou and Zafarani 2019).

Uses and gratification theory(UGT) has been proposed to explain why users engage with special media and how media content fulfills their certain needs (Katz et al. 1973). Based on UGT, users generate content on social media to satisfy their cognitive, affective, personal, social, and entertainment needs (West et al. 2010). Furthermore, emotional broadcaster theory (EBT) represented that emotionally motivated disclosures often contain important information for listeners, and the potency of tellers' own emotional experience will predict how far their stories travel across their social networks. While most of the studies in IS literature focus on understanding users' behavior from the UGT perspective, there is a lack of study to implement this perspective to recognize engagement in disinformation spread and a metric to detect disinformation and real information. In this study, I try to extend the uses and gratifications theory and emotionally broadcaster theory to evaluate the audiences' engagement with disinformation stories and a theoretical approach to recognize disinformation. I propose that disinformation producers by relying on UGT constructs and producing emotional content would try to attract more users to their social media posts (Harber and Cohen 2005). By considering that disinformation producers are looking for methods to engage more people on their published posts, I need to have a better understanding of the social media engagement techniques. In other words, I need to study social media content to understand the mechanism that disinformation producers use to engage more people with their posts and accounts. To have a better idea, I also need to evaluate the produced content, compare both information and disinformation producers' content, and highlight the differences among them.

To be able to investigate this fact, I need to evaluate various emotional, cognitive, and structural components and determine the tone of disinformation producers' verbal messages. To do so, I will apply a text analysis application called Linguistic Inquiry and Word Count (LIWC) as a scientific method to determine these factors. The finding of this study can be used in several directions. From a cyber security perspective, it can be used as a metric to alert about the trustworthiness of the posts based on the content evaluation. From a software engineering perspective, it can be used to provide disinformation detection tools to warn about the emotional arousing content or logical arousing one and highlight the gaining attraction tricks on FB posts. Moreover, from an educational perspective, it could be used through information literacy education in increasing the students' awareness and their ability to identify disinformation from information on social media.

Literature review

Based on the Merriam-Webster dictionary, disinformation can be defined as a piece of intentionally misleading or biased information; manipulated narrative or facts[14] that attempts to create deceptive and false statements to convince someone (Nimmo 2016). In IS literature, disinformation can be defined as the act of creating and sharing false and manipulated information to deceive and mislead the audiences to gain political, personal, and financial benefits (Gelfert 2018; Tandoc Jr et al. 2018). During the COVID-19 pandemic, on February 15, 2020, a high-profile group of authors associated with the World Health Organization (WHO) collected and organized global ideas to fight the coronavirus disease infodemic (Tangcharoensathien et al. 2020). Based on this study, information related to the spread of the virus, online interactions, and public-generated content on social media can be useful sources in understanding disinformation detection and mitigation. Studying the reasons why users fall prey to false information is one of the major branches in this field. Sharing false information when it is viewed as novel (Vosoughi et al. 2018), personal differences (Vishwanath 2015), and content differences such as emotional tones (Wang et al. 2015) are known as some of the reasons for sharing false information on social media. Automated deception detection methods in textual information are of interest in information

[14] https://www.dictionary.com/browse/disinformation

science (Rubin et al. 2015; Rubin and Lukoianova 2015) due to their applications in enhancing human abilities in recognizing true information from disinformation (Purian et al. 2020; Rasool et al. 2019), assessing information qualities (Osatuyi and Hughes 2018; Rubin and Lukoianova 2013), and assessing source credibility (Shao et al. 2018). The vast amount of unstructured text data sets and automated deception detection are challenging tasks (Rubin and Chen 2012; Shu, Sliva, et al. 2017) that can be overcome with big data analytics techniques such as natural language processing (NLP) and machine learning (ML) techniques (Karimi et al. 2018; Rubin and Lukoianova 2015; Torabi Asr and Taboada 2019).

From a data mining perspective, Shu (2017) divided false information detection on social media into two categories based on feature selection: a) false information detection based on content, b) false information detection based on social context. While the first group attempts to extract content features such as source, headline, body text, and image or video to construct machine learning models, the second group of researchers focuses on the social context features such as user-based, post-based, and network-based features to construct their machine learning models.

I believe that to increase the awareness regarding disinformation through informational literacy, researchers need to 1) understand the differences between disinformation and trustful information, 2) identify the way disinformation producers manipulate the data, and 3) investigate the approaches that disinformation producers use to increase users' engagement. Although false information detection research has shown great progress in recent years, it is proven that interdisciplinary research on detection approaches by relying on scientific contributions from various disciplines including social science and engineering would provide higher interpretability in this field (Zhou and Zafarani 2019). False information detection can be done by evaluating writing style, propagation structure, and false information spreaders’ intentions. While all these aspects can help to distinguish false news from trustworthy ones, without supportive theories, it would be difficult to interpret the latent features and would provide limited insight.

Uses and gratification theory (UGT) has been proposed to explain why users engage with special media and how media content fulfills their certain needs (Katz et al. 1973). Later, various studies have been done to understand users’ behavior on social media based on UGT. As an instance, while Shao examined an individual’s purpose of using, sharing content on social media (Shao 2009b), (Poon and Leung 2013) focused on the identification of the gratifications sought by the Net-generation while producing user-generated content (UGC) on the internet. Based on UGT, users consume and generate UGC to satisfy five major needs including 1) cognitive needs to acquire information, knowledge, and understanding, 2) affective needs to satisfy Emotion, pleasure, feelings, 3) personal integrative needs to provide them credibility, respect, status, confidence, and stability, 4) social integrative needs to enhance the connection to family and friends, and 5) tension release needs to escape from routines or daily problems (West et al. 2010). While cognitive needs focus on the role of online content generation on learning of the world and community, social needs concentrate on expressing and sharing feelings, thoughts, and personal experiences. As disinformation content producers convey their stories in a way to meets their needs, they will produce content that entices audiences' engagement (Meier et al. 2018). Likewise, I propose that disinformation content producers create content to educate their audiences about world issues and fulfill cognitive needs or to meet their social needs and express their opinion to their audiences. Moreover, while individuals consume generated content to satisfy their needs (Shao 2009b), users act on a disinformation story to meet their needs by ignoring, sharing, or responding to the information they receive. While most of the studies on UGT have been focused on understanding the online content producer’s needs, there is a lack of literature on the role of these features to identify disinformation from the information. In this study, I will study the UGT constructs’ effect in textual content to evaluate their role in increasing users’ engagement and their role in recognizing produced content of disinformation producers (dis-informers) and information producers (informers). To do so I will pursue our analysis in two streams. First, the text analytic tools will be applied to extract the applicable UGT constructs scores from content and evaluate the change of their load on users’ engagement scores. Second, I will investigate the significance of these content features on the credibility of the posts to be used in the classification algorithm.

Harber and Cohen (2005) proposed the significant role of emotional disclosures on social media sharing behavior based on Emotional broadcaster theory (EBT). EBT is based on four central propositions: “(a) people who have experienced major events are emotionally compelled to communicate their experiences, (b) disclosures are most therapeutic when they succeed as reportage, (c) emotionally motivated disclosures often contain important information for listeners, and (d) the potency of tellers’ own emotional experience will predict how far their stories travel across their social networks” (Harber and Cohen 2005, P.383). EBT predicts that stories that carry greater emotional impact are more likely to be shared across social media (Duffy et al. 2020; Rimé 2009; Valenzuela et al. 2017). By relying on the discrepancy theories of emotional arousal, EBT proposed that by drawing attention to event or belief discrepancies, emotions may prompt people to examine disrupting events more closely (Harber et al. 2014). Similar to what was proposed in EBT, other studies illustrate that users are more likely to draw on affective processes when facing social media content, rather than relying on cognitive processes (Huang et al. 2016; Shiau et al. 2021). Therefore, grounding on EBT assumptions, I are going to investigate if emotionally motivated disclosures are used as a mechanism to prompt social engagement and consequently an engine to create disinformation. From this framework, several testable predictions can be derived, which determine when engagement will occur and how these elements have been used in disinformation posts. Essays II extends the results of essay I to propose a model to explain the differences between disinformation and trustworthy information Facebook posts. As a human, the producer holds her opinion, which might be polarized, on a specific issue. Also, communicative intentions form in her mind, which drive her to communicate some idea to her followers online; these intentions are conveyed by a set of speech acts she chooses to use. I believe that these factors could be essential in understanding the nature of the disinformation content.

Conceptual Model

In this section, I propose our conceptual model based on EBT and UGT theories. Based on EBT and by considering the fact that stories that carry greater emotional impact are more likely to be shared across social media (Duffy et al. 2020; Rimé 2009; Valenzuela et al. 2017), I propose that disinformation producers use emotional tone and emotionally motivated disclosures as an engine to create disinformation that prompts social engagement. Moreover, based on UGT and by considering the fact that users consume user-generated content to satisfy five major needs (West et al. 2010), I purpose that disinformation producers actively consider people's psychological needs in producing posts and stories on social media to attract more people and heighten audiences' engagement. I also plan to extend these features to evaluate the credibility of the content. Finally, the results of the first essay demonstrate the importance of linguistic structure and word repetition in disinformation detection. Therefore, I extend this finding to disinformers' social media posts and propose that disinformers manipulate the linguistic structure of content as a method to gain the attention of their audience.

By considering above mentioned assumptions, three attributes were defined to represent content features. The first component is linguistic structure (LS) which evaluates how linguistic structures including the number of words in the sentence, the number of unique words, and the total number of words in the Facebook posts change engagement. The second component is called the psychological process (PP) captures the UGT element and evaluates how psychological needs including cognitive, social, personal, affective dimensions differ engagement scores on Facebook posts. The third component of the model is the emotional tone that reflects the positive emotional tone (PET) and negative emotional tone (NET) of Facebook posts and evaluates how the change of emotional tone results in a change in engagement scores. The fourth component is engagement scores that are collected from Facebook directly which included the number of likes, number of shares, the number of comments, and number of reactions. The last component of the model is credibility, which is coded by 0 and 1, where 0 shows information and 1 shows disinformation post. The abovementioned constructs are presented in Table 8.

Table 8. Conceptual model concepts and their definitions

Concept	Definition
Linguistic Structure (LS)	The score that indicates linguistic features such as word count and word length (Pennebaker et al. 2015)
Psychological Process (PP)	A measure of the degree to which data contains words categories tapping psychological constructs (Pennebaker et al. 2015)
Emotional Tone ***(NET: negative emotional tone*** ***PET: positive emotional tone)***	A measure that represents the emotional tone of the piece of information (Jockers 2017)
Engagement	The level of participation in FB posts through the number of likes, comments, sharing, and emotional reactions
Credibility	The credibility of the posts is the level of trustworthiness of the FB post where it could be true or false.

Based on what I discussed in the literature review, content-based false information detection methods highlight the importance of content features to describe the meta information related to a piece of information (Shu, Sliva, et al. 2017). In this approach, researchers evaluate linguistic and visual features of the content. Based on the literature, various linguistic features could be differing between disinformation and information. Literature and analysis in the first essay illustrated those linguistic features such as word count and word repetition, which are critical factors in the detection of false articles from true ones. By extending the findings of essay I, I propose that disinformation producers use linguistic features to enhance their users' engagement rate as well. Therefore, the first set of hypotheses of essay II based on linguistic features are as follow:

H1a: *Linguistic features positively affect users' engagement scores*.

H1b: *Linguistic features play a significant role in detecting the credibility of the posts.*

Freud in his book entitled "The Psychopathology of Everyday Life" provided that a person's hidden intentions would reveal themselves in apparent linguistic mistakes (Freud et al. 1978). From a psychological perspective, researchers believe that specific words -style words- can represent how people are communicating more than conveying what they are saying (Tausczik and Pennebaker 2010). Studying the role of psychometric properties of language in consulting, Newman et al. (2017) studied the psychometric properties of language and proposed a positive link between psychometric properties of language and consultant perceptions of dyadic collaboration.

In this study I extend this effect to evaluate the role of psychometric properties of language on engagement score of the audiences. These words and psychometric properties of language can be used as a measure of people's social and psychological worlds and linguistic inquiry and word count (LIWC) can be used as a tool to count words in psychologically meaningful categories.

Therefore, by being able to extract psychological scores of the content, I propose the second set of hypotheses as follow:

H2a: *Psychological features of content positively affect users' engagement scores.*

H2b: *Psychological features of content play a significant role in detecting the credibility of the posts.*

Based on the first essay, sentiment analysis is another technique of NLP that attempts to determine whether textual data is positive, negative, or neutral. Studies on false information sentiment analysis on social media such as Twitter and Facebook demonstrate that hidden bias regarding concepts can be reflected using sentiment analysis. By considering this fact, I propose that emotional tone is a critical role in enhancing engagement rate and the third set of hypotheses will be defined as follow:

H3a: *Negative/Positive emotional tone positively/negatively affects users' engagement scores.*

H3b: *Negative/Positive Emotional tone of content play a significant role in detecting the credibility of the posts.*

To the best of our knowledge, in literature usually, the main effect of content features has been studied in the disinformation detection field. However, the interaction effect among these features is probable. As an instance, it can be expected that a positive emotional tone would moderate the effect of negative emotional tone on engagement scores, such that negative emotions would have a stronger effect on engagement score when the positive emotional tone is low than when the positive emotional tone is high. Therefore, the last set of H4 hypothesis can be defined as interaction effect on engagement score and credibility as follow:

H4a: The linguistic structure will moderate the effect of psychological processes on users' engagement score and credibility of the posts.

H4b: The linguistic structure will moderate the effect of emotional tone on users' engagement score and credibility of the posts.

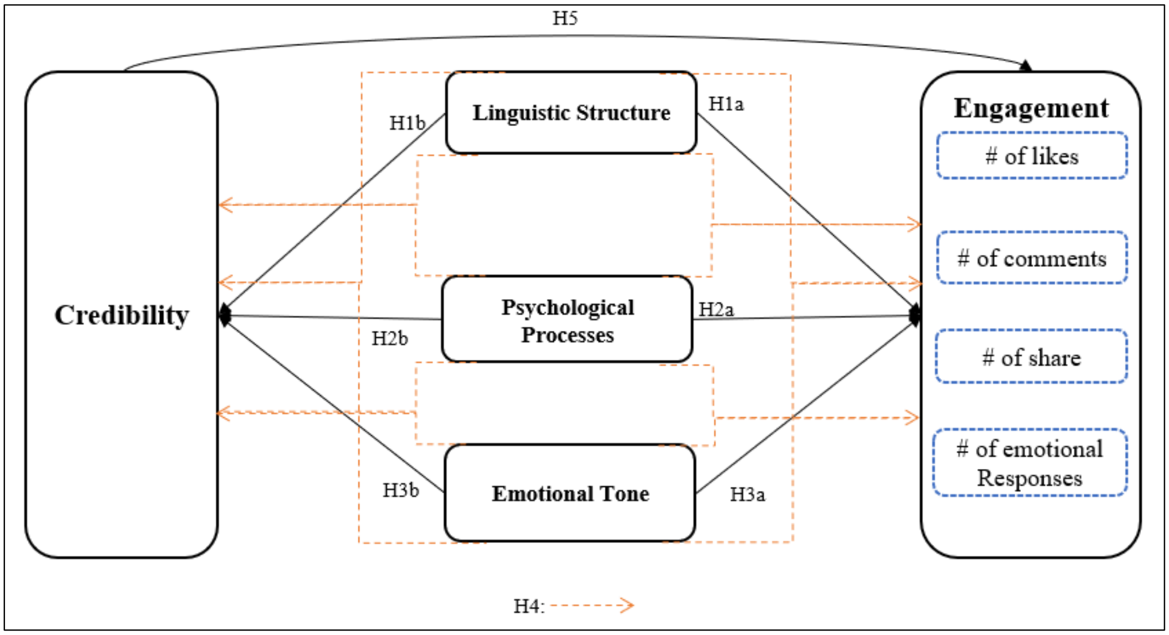
H4c: The emotional tone will moderate the effect of psychological process on users' engagement score and credibility of the posts

Literature also shows that falsehood usually spread faster than true information by human (Meyer 2018) to the extent that Green et. al. (2021) asserted that false claims on social media spread six times faster than accurate news. Pathak and Srihari (2019) in their study and by emphasizing the importance of false information detection mentioned that "Fake news is a widespread menace which has resulted into protests and violence around the globe" P.357. Moreover, the faster, farther, deeper, and more broadly spread of falsehood is not limited to one area and it shows itself in all categories of information (Vosoughi et al. 2018). Therefore, it can be proposed that the credibility of the posts or its level of truthfulness will significantly affect users' engagement behavior as well, therefore the fifth hypothesis is proposed as follow:

H5: *Credibility (level of trustfulness) significantly affect users' engagement score.*

Considering above mentioned hypothesis, Figure 8 illustrates my proposed framework for this study.

Figure 8. Proposed engagement/credibility model



Methodology

Data collection

According to a 2019 survey by Pew Research, around 14% of Americans have changed their mind about an issue because of something they saw on social media[15]. Based on this research, while 29 percent of men between the ages of 18 and 29 admitted to changing their minds about a political or social view, only 6 percent of over 65 years old users changed their minds (Bialik, 2018). While most of the research articles focused on individual preferences (Phua et al. 2017; Wang and Song 2017) and user-based factors, fewer studies have been done to evaluate the content of the posted content (Evans et al. 2017). The content-based study demonstrates that emotions can impact the

[15] https://www.pewresearch.org/fact-tank/2018/08/15/14-of-americans-have-changed-their-mind-about-an-issue-because-of-something-they-saw-on-social-media/

retweet frequency where people are more likely to re-tweet posts that have emotions associated with happiness than those that contain anger (Wang et al. 2015). Since Facebook recently added emotional responses, features including *love, care, wow, angry, and haha*, I decided to use FB as our main social media platform. As there was not a pre-existing dataset for disinformation and information COVID-related FB posts, first I needed to identify main disinformation and information producers on FB.

Identifying COVID-19 Disinformation Producers: After a careful review of the literature regarding online disinformation and its impact on the COVID-19 pandemic, a survey of popular alternative health websites was undertaken. While many websites contain a mixture of political and health-related materials, for this study, only websites that focused nearly exclusively on health topics were considered. Lists found on various websites that describe alternative health websites were consulted. Some of these lists included Quackwatch[16], and various pages posted on NewsGuard on the Coronavirus Misinformation Tracking Center page (detailed lists and data were formally publicly shared although they are no longer available except by request). Reports published by the Center for Countering Digital Hate were also consulted (counterhate.com). In March 2021 CCDH released an additional study which named Joseph Mercola, Robert F. Kennedy, and Sayer Ji as among a list that they identify as the "disinformation dozen", a group that they find is responsible for most of the current vaccine disinformation being spread on social media (The Disinformation Dozen | Center for Countering Digital Hate). After identifying these promoters of COVID-19 disinformation, I needed to collect their published FB posts and their

[16] https://quackwatch.org/consumer-protection/COVID-19-consumer-protection/

attributes such as number of likes, number of comments, number of sharing, and number of emotional responses. To do that, I focused on the year 2020 published posts that were mainly related to the COVID-19 pandemic. Their published FB posts have been collected by using the CrowdTangle platform to analyze and report on what is happening across social media[17]. By removing posts that did not have any textual content, I ended up with 2610 posts from the selected disinformation producers.

Identifying COVID-19 information producers: As an alternative to the disinformation website, medscape.com was utilized to collect articles about COVID-19 that would generally be accepted by the scientific and medical community. Medscape.com is a medical news website that contains a large volume of archives of articles with a variety of medical topics including an archive of articles related to COVID-19 from a variety of news organizations including WebMD, Associated Press, Reuters, Kaiser Health News, and Medscape's journalists. These articles may be found in the Latest News section of their COVID-19 resource page (medscape.com/resource/coronavirus). NewGuard has assigned medscape.com a rating of 92.5 out of 100 for adhering to credibility and transparency standards[18]. Moreover, these websites published links to their articles on their FB account too, which made it a good option to collect COVID-19 related FB posts. The first stage of information collection ended up in more than 9000 posts.

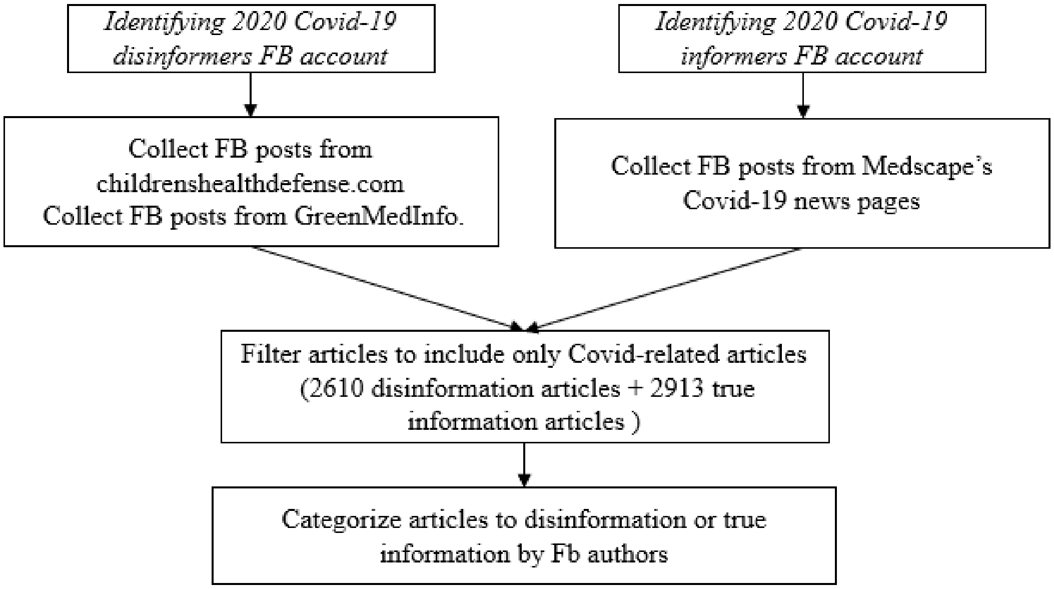
Final dataset: After identifying both disinformation and information producers' FB accounts, the FB posts were collected by the Crowdtangle platform. However, to have a more balanced data set, 32913 FB posts have been selected randomly out of the initial 9000 posts.

[17] https://www.crowdtangle.com/

[18] https://api.newsguardtech.com/label/medscape.com?cid=1a184c47-f65b-403a-af0b-824b5b5b9bbe

Finally, both disinformation and information datasets were combined for future analysis. To be able to analyze the disinformation content and understand their structure, the collected posts from disinformation producers were labeled as False and the rest were labeled as True information. Figure 9 summarizes the data collection procedure.

Figure 9. Data collection summary


Identifying 2020 Covid-19 disinformers FB account
Identifying 2020 Covid-19 informers FB account
Collect FB posts from childrenshealthdefense.com Collect FB posts from GreenMedInfo.
Collect FB posts from Medscape's Covid-19 news pages
Filter articles to include only Covid-related articles (2610 disinformation articles + 2913 true information articles)
Categorize articles to disinformation or true information by Fb authors

Data Analysis

Structural equation modeling (SEM) is known as the second-generation data analysis technique that has been used in research to evaluate the extent to which IS research meets standards for high-quality statistical analysis. SEM tools are increasingly being used in behavioral science research for the causal modeling of complex, multivariate data sets in which the researcher gathers multiple measures of proposed constructs. In contrast to linear regression, preliminary factor and reliability analysis are not necessary for SEM because the testing of measurement properties is simultaneous with the testing of the hypothesis. Partial-least-squares-based model and covariancebased SEM are the two most widely used SEM models in the IS literature (Gefen et al. 2000). This study uses the lavaan package in R-studio for the Structural Equation Modeling (SEM) analysis, and the fit indices adopted were Likelihood-ratio c2, standardized root mean square residual (SRMR), comparative fit index (CFI), and root mean square error of approximation (RMSEA). It was determined that a good fit would be proven by non-significant chi-square test results, a value less than .10 for SRMR, a value greater than .80 for CFI, and a value less than .10 for RMSEA (Kline 2015; Shi et al. 2019). Besides SEM, hierarchical regression using standardized weighted means for the construct scores to assess my hypothesis and understand their significance level.

Computational Result

Exploratory data analysis

Dependent variable: Since I are interested in the potential impacts of linguistic, tonal, and psychological features of text content on predicting the trustworthiness of the message and engagement rate, I defined a binary variable as trustworthiness which is equal to False or True. Trustworthiness of the selected content from dis-informers has been categorized as *False* and Trustworthiness of the selected content from informers have been categorized as *True*. The engagement rate of the users with Facebook posts is calculated based on the number of likes, number of comments, number of shares, and the total number of emotional responses. In our selected dataset, 47% of the selected posts are in the False category and 53% of them are in the True category. While the GreenmedInfo account has about 15k followers, the ChildrenHealthDefense account has 167K followers, the Medscape account which publishes posts from authorized sources has 759k followers[19]. Although the true informer account has a higher number of followers, the engagement rate of their audiences is surprisingly low. Table 9 illustrates the average of interaction on disinformers' and informers' Facebook posts. Moreover, Table 10 highlights the basic statistical analysis of dependent variables separately for disinformation and information groups.

Table 9. Average of engagement rate with respect to interaction factors

Group	Likes	Comments	Shares	Love	Wow	Haha	Sad	Angry	Care
Dis-informer	228.39	47.54	167.43	42.10	17.39	7.41	13.52	21.75	0.92
Informer	94.17	8.45	64.66	3.35	6.88	2.53	8.62	1.69	0.32
Welch Two Sample t-test (p-value)	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	1.968e-10	6.766e-15	< 2.2e-16	8.535e-11

Based on our observation, user engagement on disinformation posts is significantly higher than information posts which supports our H5 claim, however, this fact will also be studied in regression analysis in the next sections.

Table 10. Statistical analysis of dependent variables

	Likes	Comments	Shares	Love	Wow	Haha	Sad	Angry	Care
Disinformation									
Min	3	0	0	0	0	0	0	0	0
1st Qu	64	5	31	2	1	0	0	0	0
Median	141	17	78	8	4	0	1	1	0
Mean	228.4	47.54	167.4	42.10	17.39	7.47	13.52	21.75	0.92
3rd Qu	281	49	177	32.25	18	1	10	10.25	1
Max	4722	1887	7233	5562	509	1046	1074	981	123

[19] The data captured in August 2021

Information									
Min	0	0	0	0	0	0	0	0	0
1st Qu	13	0	3	0	0	0	0	0	0
Median	32	1	11	1	0	0	0	0	0
Mean	72.28	6.08	52.6	2.65	4.812	1.67	5.45	1.12	0.28
3rd Qu	71	4	35	2	2	0	1	0	0
Max	1974	517	5235	281	641	827	839	342	65

Independent variables: To be able to evaluate the proposed model, I need to find a way to assess the constructs, and to assess a construct I need to measure it first. To be able to measure these constructs, the following variables have been extracted from the Facebook message by using two different NLP tools known as the Syuzhet package and LIWC application.

Syuzhet package in R has been used successfully across a variety of disciplines (Liu and Lei 2018); (Valdivia et al. 2017) to identify emotions: anger, anticipation, disgust, fear, joy, sadness, surprise, as well as general negativity/positivity sentiments (Jockers 2017). Syuzhet algorithm extracts sentiment and sentiment-derived plot arcs from text using a variety of sentiment dictionaries conveniently packaged for consumption by R users which implements various dictionaries including *syuzhet*, *afinn* (Nielsen 2011), *bing* (Liu et al. 2005), and *nrc* (Mohammad and Turney 2010) which are developed by various researchers[20]. I used the Syuzhet default dictionary to extract the sentiment features as it is the best lexicon under a number of words and resolution criteria (Naldi 2019). I divided emotion scores by posts word length to be able to compare proportionalities of each emotion across samples. In this study sadness, fear, and anger

[20] https://cran.r-project.org/web/packages/syuzhet/syuzhet.pdf

dimensions have been used to measure the negative emotional tone (NET) and joy, trust, and anticipate dimensions have been used to measure positive emotional tone (PET).

To be able to extract the psychological aspect of content, I applied LIWC 15 application. This software is provided by Pennebaker et al. (Pennebaker et al. 2015), one of the well-known researchers in the natural language process field. The authors and developers believed that the ways people use words in their daily lives can provide rich information about their beliefs, fears, thinking patterns, social relationships, and personalities. “From the time of Freud’s writings about slips of the tongue to the early days of computer-based text analysis, researchers began amassing increasingly compelling evidence that the words I use have tremendous psychological value” (Pennebaker et al. 2015, P.1). Linguistic Inquiry and Word Count (LIWC) is a text analysis application that has been developed to provide an efficient and effective method for studying the various emotional, cognitive, and structural components in individuals’ verbal and written samples. LIWC provides 90 output variables including word count, summary language variables, general descriptor categories, standard linguistic dimensions, word categories tapping psychological constructs (e.g., affective processes, social processes, cognitive processes, perceptual processes, and biological processes), personal concern categories, informal language markers, and punctuation categories. In this study word count, number of unique words, and number of word per sentence dimensions have been used to measure the linguistic structure (LS) and affective processes, social processes, cognitive processes, and personal drives dimensions have been used to measure Psychological process construct (PP).

Computational linguistics is an interdisciplinary field concerned with the computational modeling of natural language to respond to linguistic research questions. Computational linguistics has been founded based on linguistics, computer science, artificial intelligence, mathematics, logic, philosophy, cognitive science, cognitive psychology, psycholinguistics, anthropology, and neuroscience, to answer linguistic research questions. I believe that these variables extracted from the above three tools would provide rich information regarding textual content and extend the computational linguistic studies in false information detection as well.

The below tables illustrate emotional, tonal, psychological, and structural differences among false and true Facebook posts of our dataset. Table 11 illustrates that even though literature highlights the fact that false information has more negative emotions (Ajao et al. 2019; Martel et al. 2020), our dataset illustrates that in the COVID-19 case while all authorized sources attempted to alert people about the COVID-19 severity to encourage people to wear a mask, keep social distance and get vaccinated, dis-informers attempted to encourage people to hesitate mask and vaccine, by expressing more positive emotions. Our study illustrates that emotional tone can differ from situation to situation. In other words, disinformers try to attack the real situation by producing opposite emotions or creating surprising posts. In a rumor spreading model based on information entropy, Wang et al. (2017) represented the role of memory, conformity effects, and differences in the subjective propensity in producing distortions and changing the degree of trust among people. Moreover, Table 12 represents the statistical analysis of emotional tone variables based on information groups. The represented value and t-test analysis illustrate the significant differences in the emotional tone between disinformation and information posts.

Table 11. Averaged emotional tone percentages among disinformation and information.

Group	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
Disinformation	8.4%	14.7%	6.1%	14.5%	8%	8.5%	5.5%	19.7%
Information	2.8%	4.4%	2.4%	6.4%	1.8%	4.7%	1.5%	5%
Welch Two Sample t-test (p-value)	< 2.2e-16	2.008e-12	0.135	1.583e-07	< 2.2e-16	< 2.2e-16	3.546e-07	< 2.2e-16

Table 12. Statistical analysis of emotional tone variables

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
Disinformation								
Min	0	0	0	0	0	0	0	0
1st Qu	0	0	0	0	0	0	0	0
Median	0	10%	0	10%	0	0	0	10%
Mean	8.4%	14.7%	6%	14%	8%	8.5%	5.5%	19%
3rd Qu	10%	20%	10%	20%	10%	10%	10%	30%
Max	15%	33.33%	17%	23%	20%	21%	18%	33%
Information								
Min	0	0	0	0	0	0	0	0
1st Qu	0	0	0	0	0	0	0	0
Median	0	0	0	0	0	0	0	0
Mean	2.8%	4.4%	2.4%	6.4%	1.8%	4.7%	1.5%	5%
3rd Qu	8%	10%	12%	10%	9%	10%	15%	10%
Max	23%	40%	30%	50%	40%	30%	30%	30%

Finally, Table 13 represents the averaged structural and psychological scores over disinformation and information groups. As it is clear, disinformers focus more on the psychological aspects of their post compared to informers. Therefore, I believe that structural and psychological variables are also important variables in increasing the engagement rate and an important factor in disinformation detection as well. Furthermore, Table 14 illustrates the statistical analysis of structural and psychological scores of selected contents based on disinformation and information groups.

Table 13. Averaged structural and psychological score over disinformation and information posts

Group	WC	Unique. Words/WC	WPS	affect	social	cog process	Personal drives
Disinformation	53.84	0.81	19.42	4.53	7.71	7.77	9.03
Information	23.29	0.90	16.86	3.52	3.76	6.23	5.78
Welch Two Sample t-test (p-value)	< 2.2e-16	< 2.2e-16	< 2.2e-16	2.677e-13	< 2.2e-16	6.194e-15	< 2.2e-16

Table 14. Statistical analysis of structural and psychological variables

	WC	Unique. Words/WC	WPS	affect	social	cog process	Personal drives
Disinformation							
Min	1	0.30	1	0	0	0	0
1st Qu	23	0.72	13	0	3.57	3.57	4.49
Median	37	0.82	18	3.7	7.14	7.06	8.23
Mean	53.84	0.81	19.42	4.52	7.71	7.76	9.03
3rd Qu	65	0.90	24.50	6.59	10.99	11.11	12.50
Max	757	1	172	50	50	100	50
Information							
Min	3	0.41	1.33	0	0	0	0
1st Qu	12	0.84	10	0	0	0	0
Median	18	0.92	14	0	0	2.86	3.33
Mean	23.29	0.90	16.86	3.52	3.76	6.23	5.78
3rd Qu	34	1	20.50	5.88	5.88	10	9.09
Max	90	1	68	33.33	41.67	50	66.67

Structural equation Modelling (SEM)

In this study, SEM analysis has been used to assess the relationships among the latent variables. Besides verifying the structure of the model, SEM helps to verify how well the structure has been capable of representing the data (Chatterjee and Bhattacharjee 2020). To do so, the *lavaan* package in R studio has been applied. This package is a free open-source, but commercial-quality package for latent variable modeling. The researcher can use lavaan to estimate various multivariate statistical models, including path analysis, confirmatory factor analysis, and structural equation modeling[21].

This study adopted fit indices such as Likelihood-ratio c2, standardized root mean square residual (SRMR), comparative fit index (CFI), and root mean square error of approximation (RMSEA). It was determined that a good fit would be proven by non-significant chi-square test results, a value less than .10 for SRMR, a value greater than .90 for CFI and TLI, and a value less than .050 for RMSEA (Kline 2015; Shi et al. 2019). The output of the SEM analysis is provided in Table 15.

Table 15. Model fit index summary

Fit index	Recommended value based on literature	Extracted value from our model
Comparative Fit Index (CFI)	>=0.90	0.948
Tucker-Lewis Index (TLI)	>=0.90	0.927
RMSEA	<= 0.05	0.06
Standardized Root Mean Square Residual	<= 0.10	0.042

[21] https://lavaan.ugent.be/

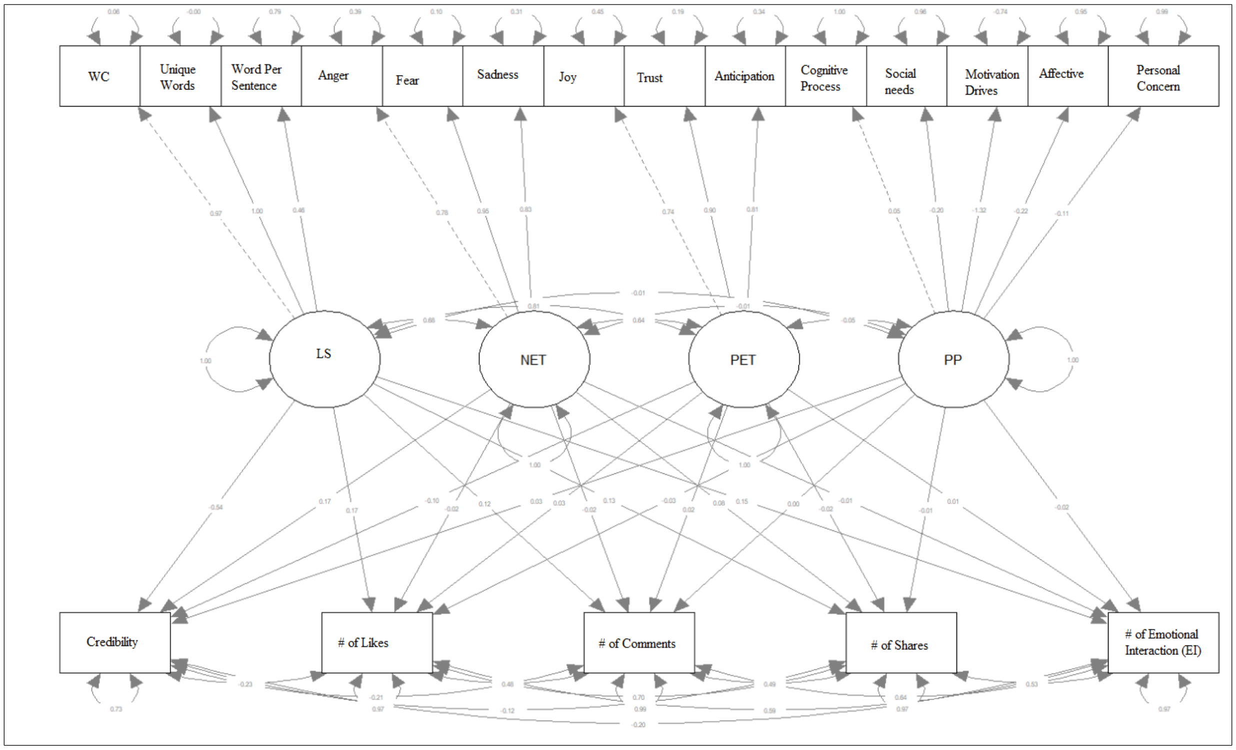
This table illustrates that the parameters are within standard acceptable limits which indicates the adequacy of the model fit. Moreover, provided P-values in Table 16 indicate the importance of the latent variants and their loadings in defining constructs. These loadings will be used to measure the constructs in regression models. The structural model with path weights and level of significance is represented in Figure 10.

Table 16. Summary of SEM loadings

	Estimate	**P-value**
LS=~		
Word count	1	
Unique words	1.3	$p < 0.001$
Words per sentence	0.4	$p < 0.001$
NET=~	1	
Anger	1.92	$p < 0.001$
Fear	1.23	$p < 0.001$
Sadness		
PET=~	1	
Joy	2.28	$p < 0.001$
Trust	1.56	$p < 0.001$
Anticipate		$p < 0.001$
PP=~	1	
Cognitive process	-3.56	$p < 0.001$
Social	-3.2	$p < 0.001$
Affective	-26	$p < 0.001$

The result of this study illustrates that there are three main hypotheses that can be divided into several sub-hypothesis. The analysis highlights that there are significant differences among linguistic structure (H1b), psychological processes (H2b), and emotional scores (H3b) of disinformation and true information posts. To be able to pursue the analysis, disinformation and information posts have been coded as binary variables where 0 reflects disinformation and 1 reflects information posts. It seems like disinformation posts carry higher linguistic structure scores and negative emotional scores as well. In contrast to the literature, COVID-related posts that include disinformation, have higher positive emotional scores and their posts have higher psychological process scores.

Figure 10. Structural Model with Path Weights



This study also analyzed the effect of content features on users' engagement scores. To do so, the number of likes, comments, shares, and emotional responses have been extracted from posts. The analysis highlights the fact that posts with higher linguistic structure scores – means a higher number of words, longer sentences, and unique words- have higher engagement scores than posts with low linguistic scores. Moreover, the analysis clarifies that emotional tone has a significant impact on engagement scores, where negative tones engage more users than positive tones.

Econometric Model:

In this study, I also used hierarchical regression using standardized weighted means for the construct scores to assess my hypothesis and understand their significance level. Based on (Aiken et al. 1991; Serrano and Karahanna 2016) multiple regression can be used to test the hypothesis by applying a series of regressions in each subsequent model until a full set of hypotheses are tested. The result of regression analysis is presented in Table 17 to Table 21.

Table 17 to Table 20 represents 3 hierarchical models that evaluate *H1a*, *H2a, H3a*, and *credibility effect*. The independent model of each table is one measure of engagement score in the dataset including the number of likes, number of comments, number of shares, and number of emotional interactions. In all models of these tables, the relationship between the credibility and engagement score is significant which confirms the significant differences of information and disinformation engagement scores that support the *credibility effect* of the proposed model. The second model of each table evaluates the main effects of content features on engagement score in presence of credibility variable. These models highlight the fact that linguistic structure, psychological process, and emotional tone of posts have a significant role in engagement scores which supports *H1a*, *H2a*, and *H3a* in our proposed model. The third model of each table also assesses the interaction effect between content features. In all these models, adding interaction items does not improve the regression models and their effects are insignificant. Therefore, models without interaction terms work best to evaluate content features on engagement scores.

Table 17. Results of Hierarchical Regression Analysis for the first measure of engagement (# of likes)

	Likes					
Variable	Model 1.1		Model 1.2		Model 1.3	
	Beta	(S.E.)	Beta	(S.E.)	Beta	(S.E.)
Credibility(t)	-1.55***	0.04	-1.47***	0.05	-1.31***	0.05
Main Effects						
LS			0.07	0.06	2.32***	1.05
PP			0.02*	0.007	-0.015	0.021
NET			0.15**	0.010	-0.5**	0.021
PET			-0.03	0.013	-1.01	1.2
Interaction effect						
LS* PP					1.60**	0.04
NET* LS					0.12*	0.005
NET* PP					-0.028	0.002
PET* LS					-0.017	0.004
PET*PP					-0.067	0.003
PET*NET					0.055	0.013
Sig. F Change	< 2.2e-16		0.0005		0.2	
R^2	0.16		0.17		0.08	

Standardized coefficients (standard errors), n = 5607; ***p < 0.001; **p;<;0.01; *p < 0.05.

Table 18. Results of Hierarchical Regression Analysis for the second measure of engagement (# of comments)

	Comments					
Variable	Model 2.1		Model 2.2		Model 2.3	
	Beta	(S.E.)	Beta	(S.E.)	Beta	(S.E.)
Credibility(t)	-1.56***	0.037	-1.50***	0.043	-1.31***	0.043
Main Effects						
LS			0.014***	0.021	-0.49**	0.01
PP			0.017*	0.006	-0.56	0.02

NET			0.13**	0.432	-1.08	1.6
PET			-0.14**	0.008	1.87	1.5
Interaction effect						
LS* PP					0.016***	0.004
NET* LS					0.012*	0.003
NET* PP					-0.023	0.002
PET* LS					-0.017	0.028
PET*PP					-0.067	0.002
PET*NET					0.006	0.071
Sig. F Change	< 2.2e-16		0.0005		0.06	
R^2	0.238		0.25		0.23	

*Standardized coefficients (standard errors), n = 5607; ***p < 0.001; **p;<;0.01; *p < 0.05.*

Table 19.Results of Hierarchical Regression Analysis for the third measure of engagement (# of shares)

	Shares					
Variable	Model 3.1		Model 3.2		Model 3.3	
	Beta	(S.E.)	Beta	(S.E.)	Beta	(S.E.)
Credibility(t)	-1.68***	0.051	-1.57***	0.061	-1.62***	0.067
Main Effects						
LS			0.013*	0.037	0.051	0.009
PP			0.03*	0.008	-0.001	0.001
NET			0.41***	0.013	0.352*	0.022
PET			-0.19**	0.015	-0.114	0.023
Interaction effect						
LP* PP					0.04	0.001
NET* LS					0.06	0.038
NET* PP					-0.02	0.031
PET* LS					-0.05	0.047
PET*PP					-0.015	0.004
PET*NET					-0.02	0.014
Sig. F Change	< 2.2e-16		5.363e-06		0.54	
R^2	0.15		0.17		0.16	

*Standardized coefficients (standard errors), n = 5607; ***p < 0.001; **p;<;0.01; *p < 0.05.*

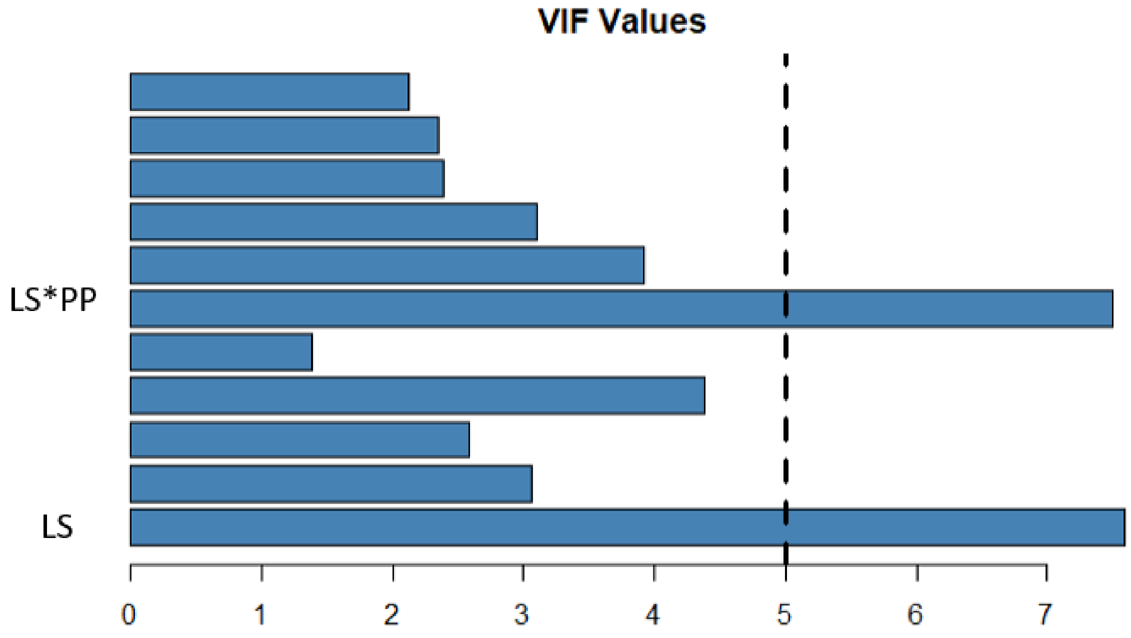
Table 20. Results of Hierarchical Regression Analysis for the fourth measure of engagement (# of emotional interaction)

	EI					
Variable	Model 4.1		Model 4.2		Model 4.3	
	Beta	(S.E.)	Beta	(S.E.)	Beta	(S.E.)
Credibility(t)	-2.011***	0.043	-1.957***	0.052	-1.978***	0.057
Main Effects						
LS			0.099*	0.031	-0.039*	0.007
PP			0.082	0.007	-0.021	0.001
NET			0.022***	0.091	-0.026*	0.572
PET			-0.013**	0.121	0.007	0.193
Interaction effect						
LS* PP					0.098**	0.051
NET* LS					0.003	0.003
NET* PP					0.001	0.002
PET* LS					0.004*	0.004
PET*PP					-0.004	0.003
PET*NET					-0.007	0.012
Sig. F Change	< 2.2e-16		8.793e-09		0.29	
R^2	0.27		0.29		0.28	

*Standardized coefficients (standard errors), n = 5607; ***p < 0.001; **p;<;0.01; *p < 0.05.*

Post hoc analysis: Due to the possibility of multicollinearity between the interaction effect and the fact that adding interaction effects changed main effects status to insignificant, I also conduct variance inflation factor (VIF) analysis on the models. As it is visualized in Figure 11 linguistic structure factor and its interaction with the psychological process factor have higher values than the acceptable threshold and exceed 5 (Craney and Surles 2002). This analysis helps to remove predictor variables with high VIF values and evaluates their effect on the r-squared value of the model (Miles 2014). As the below graph represents LS variable is not significant in the engagement regression models, however, the elimination of LS did not improve the r-squared value and did not change the significance level of other predictors in models.

Figure 11. VIF values for each predictor variable of engagement models



On the other hand, Table 21 highlights how content features change from information to disinformation. To be able to evaluate this part of the model (Hb), credibility has been analyzed as a factorial variable. The evaluation in model 5.1 illustrates the main effect of linguistic structure, psychological process, and emotional tone scores on credibility. All the main effects are significant and support Hb. While disinformation posts have higher linguistic scores and higher negative emotion scores, information posts have a higher psychological and positive emotional score. Model 5.2 added interaction effect on credibility as well. The output of the model illustrates the significance of the interaction effect on model performance.

Table 21. Results of Hierarchical Regression Analysis for credibility (# information VS. disinformation)

	Credibility			
Variable	Model 5.1		Model 5.2	
	Beta	(S.E.)	Beta	(S.E.)
Main Effects				
LS	0.03^{***}	0.116	0.218^{***}	0.118
PP	0.01^{***}	0.007	0.078^{*}	0.033
NET	0.11^{***}	0.019	0.427^{***}	0.048
PET	-0.06^{***}	0.024	-0.251	0.051
Interaction effect				
LS* PP			0.001^{***}	0.002
NET* LS			0.0025^{***}	0.008
NET* PP			0.0029^{***}	0.006
PET* LS			0.002^{***}	0.001
PET*PP			-0.005^{***}	0.007
PET*NET			$-0.006^{\ ***}$	0.003
Sig. F Change	< 2.2e-16		< 2.2e-16	
R^2	0.24		0.34	

*Standardized coefficients (standard errors), n = 5607; ***p < 0.001; **p;<;0.01; *p < 0.05.*

Discussion

Summary of Findings

In essay II I attempted to understand the disinformation mechanism based on content features. To do so, content features including psychological processing load, emotional load, and linguistic structure of the content have been studied. The regression analysis on the relationship between content-based features and the credibility of the FB posts illustrated how content-based factors can be used to distinguish disinformation from information that supports H1b, H2b, H3b. Based on the results, from a linguistic structure point of view, while disinformation posts have longer sentences and use a greater number of words in their content, their posts include a smaller number of unique words. In other words, they include more repetitive words in their social media posts. This fact was also the same in the first essay results which means that disinformers use the same linguistic pattern both in their social media posts and their published content on webpages. From a psychological point of view, the results highlight the fact that disinformers attempt to produce social media posts that have higher psychological processing loads, which means that they focus on the topics that discuss more about UGT mentioned major cognitive needs in comparison to the posts published by informers. From an emotional tone perspective, it seems like disinformers concentrate to produce content that evokes more negative emotions in comparison to positive emotions. Our study also supports the fact that, disinformation producers that look for more engagement use more negative emotions. Therefore, the emotional tone could be another critical factor in deception detection.

Besides the use of content-based features to evaluate the falsehood of the content, I also used them to analyze users' engagement scores. The findings of this study imply that the above-mentioned content-based features are also important contributing factors of users' engagement on social media which supports H1a, H2a, and H3a. Results illustrate that disinformers use a different mechanism to produce eye-catching content. This result highlights the significant differences between disinformation content and information content. Therefore, content-based features can play a critical role in distinguishing disinformation from information and the understanding of these facts can be useful in providing more accurate disinformation detection tools.

Contributions

In my second essay, I tried to understand the content-based mechanism that differs COVID-19 related disinformation FB posts from true information. Therefore, this research provides several contributions to the disinformation detection research field. The first expected contribution is providing a theory-based disinformation detection technique. As I discussed in the literature review, the false information detection field requires theory-based detection methodologies and without supportive theories, it would be difficult to interpret the latent features and would provide limited insight. Uses and gratification theory seeks the answers that why people use social media. But in this study, I attempt to look at the UGT from the disinformers' perspective. Here I proposed that disinformers use content that matches users' five major cognitive needs to gain their attention. Based on the analysis, disinformers use content with a higher psychological need load and high negative emotional tone. The extension of uses and gratification theory and emotional broadcaster theory to this area would enhance our understanding of disinformation created on social media and provide opportunities for false information detection.

Additionally, literature proposes that people are likely to spread information that evokes disgust, fear emotions on social media regardless of the information truth level (Peters et al. 2009; Wang et al. 2020). Based on the statistician analysis of the second essay there is a significant difference in the negative tone of disinformation and information posts. The outputs of regression analysis also highlight that higher engagement scores of social media posts have been associated with higher negative scores of the posts.

The contribution of this study is not limited to providing a theory-based deception detection framework. The finding could be useful in practical solutions as well, in other words, the analysis reveals the potential use of social media engagement theories in developing deception detection tools. The content-based mechanism of disinformation recognition can be valuable for cyber security researchers and educational researchers as well. From a cyber security perspective, it can be used as a metric to alert about the trustworthiness of the posts based on the content evaluation.

From a software engineering perspective, it can be used to provide disinformation detection tools to warn about the emotional arousing content or logical arousing one and highlight the gaining attraction tricks on FB posts. Auberry (2018) proposed that offering an information literacy module, which could be embedded within a course's Learning Management System (LMS), would provide a great opportunity for students to increase their ability to identify disinformation through information literacy education and content management systems. I believe that besides providing a theory-based framework for potential disinformation alerting systems, our proposed research model could be a good template in defining information literacy syllabus in enhancing disinformation awareness level. Therefore, from an educational perspective, it could be used through information literacy education in increasing the students' awareness and their ability to identify disinformation from information on social media.

Limitation and Future Study

Like all other studies, this study has limitations that might be addressed in further research studies, however, I tried to overcome the limitation by applying both structural equation modeling and hierarchical regression modeling as my main methodologies. In social media research, I believe that behavioral and personal factors can also be critical in improving disinformation detection and social media engagement. As the data has been collected from Facebook, there was no demographic data available. Therefore, providing comparative analysis was not achievable. These variables need to be studied as potential direction to improve the study in the content of the experiment or be addressed in future studies. I will try to overcome this limitation in essay III. The other limitation of this study would be extending the finding of this study to different research topics and evaluating the change of topic on model performance. As an instance, this study clarifies that a negative emotional tone may not always play a critical role in identifying disinformation from a true one. In other words, the content and discussed topic may play a more important role in using emotional tones in detecting disinformation. Conducting multi-topic analysis and comparing the results would generalize the findings of this study for future application.

CHAPTER V: AN EXAMINATION OF THE IMPACT OF NEGATIVE EMOTIONS ON BELIEVABILITY OF DISINFORMATION AND THE RESULTING BEHAVIORAL CHANGE (ESSAY III)

Introduction

Understanding the impact of the emotional tone of content on the believability of disinformation and users' decision-making will lead to a great contribution to disinformation awareness. A better understanding of this mechanism would be useful to provide meaningful strategies for social media networks to enhance public fact-checking behavior. Based on the literature, viral false information on Twitter tends to spread faster and reach more people than trustworthy news (Duffy et al. 2020; Vosoughi et al. 2018). According to the first essay, machine learning algorithms were able to detect disinformation from trustworthy information with significant accuracy based on document term matrix and sentiment discovery analysis. The findings of the second essay highlighted the importance of linguistic structure and emotional tone in disinformation/information identification. In the literature, the strength of emotions evoked by information has been known as a powerful predictor of information spread (Cotter 2008; Izawa 2010; Wang et al. 2020) and highly intense emotional content is more likely to be shared (Rimé et al. 2002; Xiong and Lv 2021). Even though the second essay provided evidence for a significant relationship between the emotional content of social media posts and users' engagement, it did not establish a directional causality.

The increasing importance of the role of social media in the consumption of public news has resulted in a surge of academic discussions on the generation and dispersion of false information in social media (Grinberg et al. 2019; Pennycook et al. 2018; Wardle and Derakhshan 2017). Literature review on false information sharing behavior illustrates high trust in social media sources, social media fatigue (Khan and Idris 2019; Laato et al. 2020), socialization, and status-seeking gratification (Apuke and Omar 2021; A. K. M. N. Islam et al. 2020; Thompson et al. 2019), and users' cognitive motivators (Quan-Haase and Young 2010; Xu 2019) as three main reasons for engaging in false information spread on social media. Vosoughi et al. (Vosoughi et al. 2018) showed that false stories were 70% more likely to be spread in comparison to trustworthy stories. A few recent studies (for example Dennis et al. 2021; Jain and Pradhan 2020; Kim and Dennis 2019) have examined critical factors that affect audiences' judgment of the believability of the content and its potential influence on their behavior. During the COVID-19 pandemic and besides the spread of coronavirus disease, an infodemic in the medical publishing world was observed which worked like a contagious disease that infected our information culture (Solomon et al. 2020). The massive spread of the disinformation could lead to numerous social problems such as breaking the validity balance of the news ecosystem, enhancing biased and false beliefs, and impacts the way that people respond to authentic information (Shao et al. 2018; Shu, Sliva, et al. 2017; Shu, Wang, et al. 2019). Therefore, investigating the determinants that affect the believability of disinformation content on social media platforms would provide a wider understanding of social media audiences' perceptions regarding disinformation.

Emotions are known to have a critical impact on a recipient's perception, information processing, and diffusion of the content on social media (Brady et al. 2017; Clore et al. 2014; Vlasceanu et al. 2020). Disinformation studies represented that, false information content contains stronger negative emotions than positive emotions compared to trustworthy content (Horne and Adali 2017; Paschen 2019). Moreover, social media posts with a negative emotional tone generally diffuse faster and deeper than content with a positive emotional tone (Firdaus et al. 2021; Naveed et al. 2011; Zhang and Wang 2021). Therefore, studying negative emotional tone on social media and its influence on health-related beliefs and behaviors is essential in IS research. Persuasion is the attempt to influence an individual's beliefs, attitude, intentions, motivation, and behaviors (Gass and Seiter 2018). Existing literature on persuasion theory can be categorized into three groups (Miller and Levine 2019): 1) research studies that focus on the pragmatic issue of isolating factors that enhance persuasions like studying the role of typography and word-driven advertising in increasing the persuasive power of advertisement (McCarthy and Mothersbaugh 2002), 2) research studies which focus on explaining why persuasive messages are persuasive like understanding the factors that affect persuasion from source effect, message effect, and recipient characteristics perspective in celebrity-endorsed advertisements among consumers (Azab 2011), and 3) research studies which investigate the generation of persuasive messages, such as examining the role of emotional factors in persuasion and its significant influence on the change of attitudes (Petty et al. 2003).

Appraisal tendency (Lerner and Keltner 2000) is a theoretical framework to evaluate the role of emotions on judgment. Appraisal tendency theory suggests that distinct emotions can cause cognitive appraisals that can influence subjects' behavior and their judgment (Lerner and Keltner 2000). Furthermore, this framework has been used as an important basis to understand the psychological construction of emotion and its consecutive effects (Eysenck and Keane 2015; Lerner and Keltner 2001; Russell 2003). As an instance, Lerner and Keltner (2001) propose that emotions trigger changes in cognition, physiology, and action that often persist beyond the eliciting situation(Lerner and Keltner 2001, P.146). Based on the importance of emotions in literature and considering the importance of negative emotions on information spread and the findings of essay

II, in the current essay, I will examine the impact of emotional tone in COVID-19 disinformation textual posts on readers' tendency to believe the (dis)information and change their behaviors. In essay III, I run a series of experiments to examine and establish the causal relationship between the negative emotional content of written disinformation and the believability of the content and its ability to influence readers' behavior. These experiments reveal the micro-behavioral foundations of the impact of emotions on the believability of disinformation in written posts. By conducting this study, I will answer the following research questions: "How the change in negative emotional tone will affect the believability of the disinformation content?" and "How the change in negative emotional tone would influence the behavior?"

Investigating the role of emotional tone on the persuasive power of disinformation puts our study within the persuasion theory research category which concerns isolating factors that enhance persuasion. The finding of this study can be used in several theoretical and empirical directions. These findings could be valuable for the disinformation detection research field to understand disinformation believability determinants. Moreover, for social media network providers, it would highlight the importance of user-generated content and its effect on users' behavior. The findings of the study also can be used to increase the social media users' awareness regarding content that can mislead their judgment and deception detection ability.

The rest of this paper is organized as follows: the second section presents a theoretical background of disinformation and emotional tone effect on the judgment in IS literature. In the third section, the methodology of the study is discussed. The fourth section provides a detailed description of the data analysis and statistical results of the study. Finally, the discussion and conclusion are provided in sections five and six respectively.

Literature review

Today’s digital world provides a publicly accessible platform to easily get access to a huge amount of information, share it, modify it, or add to it. Nowadays people from all over the world are overloaded with information. Information overload makes it increasingly difficult for information recipients to verify the accuracy of the information and ascertain it is not unintentionally biased or intentionally manipulated to spread disinformation. People are susceptible to error and they are unaware of how introspection, emotional processing, and experience offer poor guidance in making various decisions from wrong investment decisions and wrong hiring decisions to government inefficiency and social injustice (Bazerman and Moore 2012; Hooshangi and Loewenstein 2018). The appraisal tendency framework investigates the critical role of emotions on poor guidance and its effect on changes in cognition, judgment, and action (Lerner and Keltner 2000). In a study on feeling and believing, Dunn et al. (2005) examined the effect of emotional state on trust and found a significant impact of emotions on information processing.

Communication is the main procedure of forming relationships and a critical factor in acquiring information to make decisions (Veil et al. 2008). From social media behavior perspective, Wang et al. (2021) studied the effect of emotional tone on the virality of food safety messages over social media. Their study provided evidence for a positive association between increasing negative emotional tone (such as anger or sadness emotions) in social media posts and increased virality of the posts on social media.

In information system research, the emotional tone has recently observed a surge in research attention, particularly in studies on false information. In image processing, Karduni et al

(2021) proposed that highly emotional images from sources of false information can greatly influence our judgments where angry facial emotions impact users' perceived content bias and source credibility. In another study, Bakir et al. (2018) studied the use of emotionally targeted news and asserted that exposure to a particular type of emotional content in users’ news feeds stimulates the posting behavior of the social media users. Parikh et al. (2019) also in a study of the tone of fake news proposed that fake news stories are intended to be filled with negative emotions to change the opinion of the users.

Literature on studying the believability of the content mostly focuses on political topics (Ali and Zain-ul-abdin 2021; Peixoto 2019) and non-content based factors such as reliability of the source of the message (Trivedi et al. 2021), virality metrics of posts including the number of retweets, likes, and replies (Kim 2018), social norms (Gimpel et al. 2021) and source rating (Kim and Dennis 2019). However, this study moves forward and focuses on the content-based features of medical-related disinformation posts to investigate the effect of negative tones in social media posts on human perception regarding the accuracy of the content and the potential influence on behavior.

In this study, I provide a theoretical explanation and experimental evidence to answer whether social media disinformers use techniques that employ emotions to persuade and mislead users’ perceptions regarding the accuracy of the content. Persuasion is defined as a procedure intentionally designed to change an individual’s or a group’s attitude and belief toward some event, idea, or object by using written, spoken words or some visual techniques to convey information, feelings, and reasoning (Peleckis and Peleckiene 2015). The literature demonstrates that the study of persuasion has generally conjured up images of manipulation, deceit, or brainwashing (Gass and Seiter 2018). As an instance, Boush et al. (2015) studied the deception in the marketplace and attempted to investigate what makes persuasive communication misleading. In their study rhetorical deception was introduced as one of the main deception tactics. Rhetoric is known as the art of persuasion which aims to study the techniques that speakers utilize to inform and persuade their audiences (Walton 2007).

Today's social media networks provide people access to richer information than ever before, to the extent that many people pick social channels over traditional sources to access updated news. Because any user of social media can, and many of them generate news, it is challenging for organizations that aim to publish factual news to find the news that contradicts the facts, and even more challenging to come up with a plan to counteract them. Nevertheless, given the volume and speed of information diffusion in social media, it is imperative to identify the factors that prevent people from verifying their consumed content on social media (Tambuscio et al. 2015). In essay 1, I demonstrated the importance of linguistic structure and emotional tone in disinformation detection; and in essay II, I highlighted the significant association between linguistic structure and negative tone with the engagement score of the social media audiences. Considering the higher load of negative emotional tone in disinformation content and its association with higher engagement scores, it is important to investigate the impact of emotional tone on individuals' decisions about the believability of content and its potential influence on their behavior.

Existing research on false information in social media can be categorized into three streams which examine: 1) use of machine learning algorithm as a falsehood detection tool to detect deception such as the use of textual characteristics of news title to detect fake news (Shrestha and Spezzano 2021), 2) effect of content-based features on social media engagement such as investigating the human-generated content on audience engagement (Mallipeddi et al. 2017), and

3) study the content effect on behavior such as studying the firm generated content in social media on customer behavior (Kumar et al. 2016). The current study contributes to the third stream by examining the impact of various negative emotional tones on social media posts' believability and the effect of emotions on change of behavior. In this research, I use an experimental design in three experiments to study the influence of anger, fear, and sadness tone of disinformation posts on the believability of the content and readers' tendency to be behaviorally influenced by the content. This study proposes negative embedded content as a persuasive technique of disinformation posts to misguide audiences' deception detection ability. My research will investigate the application of persuasion theory in interpreting the effect of negative emotional tone on the believability of disinformation in social media posts. I will argue that understanding the reasons and associated factors that induce people to believe the accuracy of the disinformation posts will contribute to various fields including political, public health, and social media networks studies.

Methodology

In this study, I used an incentive-compatible one-way independent sample experimental design to assess the effect of emotional tone on the believability of the post and evaluate their consecutive effect on the change of behavior.

Experiment procedure

In this research, a series of experiments will be performed to evaluate the effect of 3 different negative emotions -anger, fear, and sadness- in the textual content on participants' judgments and decisions. In essay II, I used the Syuzhet package (Jockers 2017) to create emotional content scores for 2700 Facebook posts. In this study, I selected 10 posts for each of the anger, fear, and sadness emotion. I adopted a few criteria to ensure that selected posts will be appropriate for the experiment. Specifically, I limited the length of posts to 80 to 160 words to avoid noticeably short or exceptionally long posts. For each emotion, a total of 10 posts were randomly selected and used in the experiment. Out of 10 posts for each emotion, five posts were randomly selected from the emotions' top quantile (the quantile which had the highest emotional tone score), and another five posts were randomly selected from the emotion's bottom quantile (the quantile which had the lowest emotional tone score). Each participant was requested to read two separate texts, one at a time. One of the texts had a high emotional score and the other had a low emotional score. The order of presenting the texts was evenly randomized to eliminate the possible impact of the order of exposure on the experimental results. Each participant was asked to answer two slider-type questions following reading a post. The first question asked participants whether they believe the post provides accurate information to the readers. The second question asked participants whether they believe the post has the potential to influence the behavior of people who read it. Participants could provide an answer ranging from 0% and 100% using a slider which is illustrated in

Figure **12** and Figure 13.

Figure 12. Accuracy measuring tool in survey

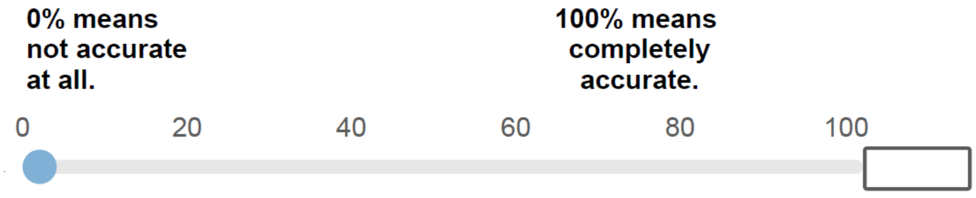
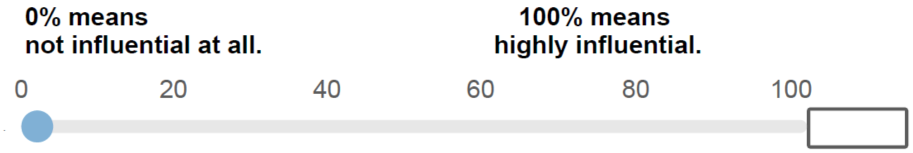


Figure 13. Influential measuring tool in survey



The experiment uses a novel incentive-compatible experimental design which ensures participants have the incentive to provide the best objectively correct answer that they can think of. Specifically, participants were rewarded in the form Amazon.com electronic gift card based on the average accuracy of their response to the two questions they were asked to answer. Each participant's reward amount depended on how his/her ratings were compared to the average rating provided by all participants. The more his/her rating deviated from all participants' average rating, the less his/her monetary reward became, as illustrated in Table 22. Participants had full knowledge of how the payoff mechanism worked. Therefore, they knew it was in their best interest to provide ratings that represent their true belief about how the post will be perceived by readers.

Table 22. The experiment's incentive-compatible payoff table

Deviation from participants' average rating	Less than or equal to 5%	More than 5% but less than or equal to 10%	More than 10% but less than or equal to 25%	More than 25% but less than or equal to 40%	More than 40%
Your reward	$5	$4	$3	$2	$1

All parts of the experiment were run online using Qualtrics. Participants were recruited from the students at a large business school in North Carolina. The study was first announced in class and students were requested to share their emails if they were interested in participating in the experiment. Following the registration, a total of 511 invitation emails were sent, and 151 participants completed the experiment which represents a 29% response rate. All participants completed the experiment voluntarily. Information such as gender, education level, the purpose of use of social media, perception regarding their ability in disinformation detection, and their preferred social media platform was collected from the participants after they completed the experiment.

Data Analysis and Computational Result

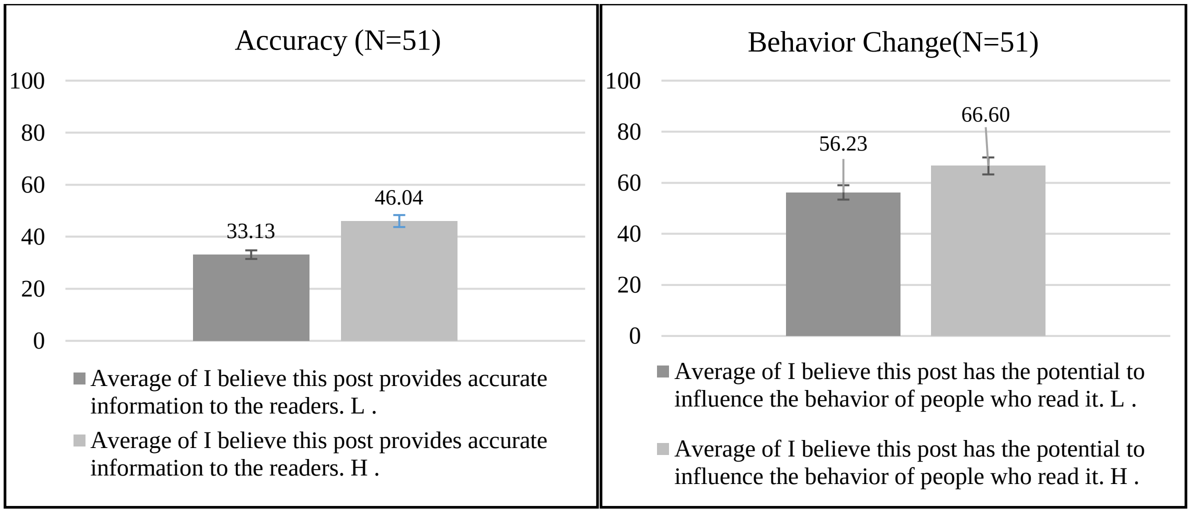
As I have several emotions in this study, here I will present the finding of each emotion separately.

Fear: The first emotion that I studied is fear.

In this test number of participants is 51. On average, participants evaluated disinformation posts with high anger tone as more accurate than text with a low fear tone. The average of participants' perceived accuracy of high and low fear tone posts was %46 and %33 respectively. Readers rated high emotional content posts as 39% more accurate than low emotional content posts, and this increase was significant at 10% level (p=0.071). The dark grey and light grey bars in Figure 14 represent how participants evaluated the accuracy of the posts with low and high fear tone respectively. Participants considered posts with high fear to convey more accurate information and posts with low fear tone to convey less accurate information. It seems like that, using more fear tone could mislead audiences' disinformation recognition abilities.

Participants' evaluation of the ability of high and low fear tone posts to change readers' behavior revealed a significant difference where the influence score of high and low fear tone was %66 and %56 respectively. Readers rated the potential behavior change in high emotional content posts as 18% more accurate than low emotional content posts, and this increase was significant at 10% level (p=0.079). The *behavior-change* of Figure 14 represents how participants estimated the change of behavior after reading low and high fear tone respectively. As it is illustrated, participants estimated that posts with high fear tone are more capable to influence the behavior of people who read them. It seems like the use of more fear tone could be more influential in the change of behavior.

Figure 14. Accuracy: Participants performance regarding low and high fear tone. Behavior change: Estimated potential change of behavior after reading posts with different fear tone



As this study ran three separate experiments, I will provide the sociodemographic analysis separately for each emotion. Figure 15 and Figure 16 represent the sociodemographic determinants of change in believability and change of behavior upon exposure to disinformation about COVID-19 with high and low fear tone. The *SM* graph of Figure 15 and Figure 16 depict the preferred social media (SM) platform among participants. In this group, the majority use Instagram (35%) as their preferred social media platform, after that TikTok, Facebook, Twitter, and YouTube each with ~16% were the most popular platforms. Participants who use Twitter tend to have the lowest average accuracy ratings in the low condition and show the most pronounced difference in accuracy ratings between low and high emotional content conditions. Participants who use Twitter rated high emotional content posts almost 228% more accurate than low emotional content posts.

Instagram is the social media that has the highest number of users among participants. Instagram users rated high emotional content posts 64% more accurate than low emotional content posts, and this difference is significant at 10% level (p=0.024).

The *Education* graph of Figure 15 and Figure 16 display the accuracy and change of behavior among different educational levels where the majority have some college degree (47%), 27% of the group a college degree, and 26% have a graduate degree. Based on the *Education* graph of Figure 15 , participants with graduate degrees are inclined to evaluate disinformation as more accurate. Participants who have some college education rated high emotional content posts almost 84% more accurate than low emotional content posts, and this difference is significant at 10% level (p=0.02).

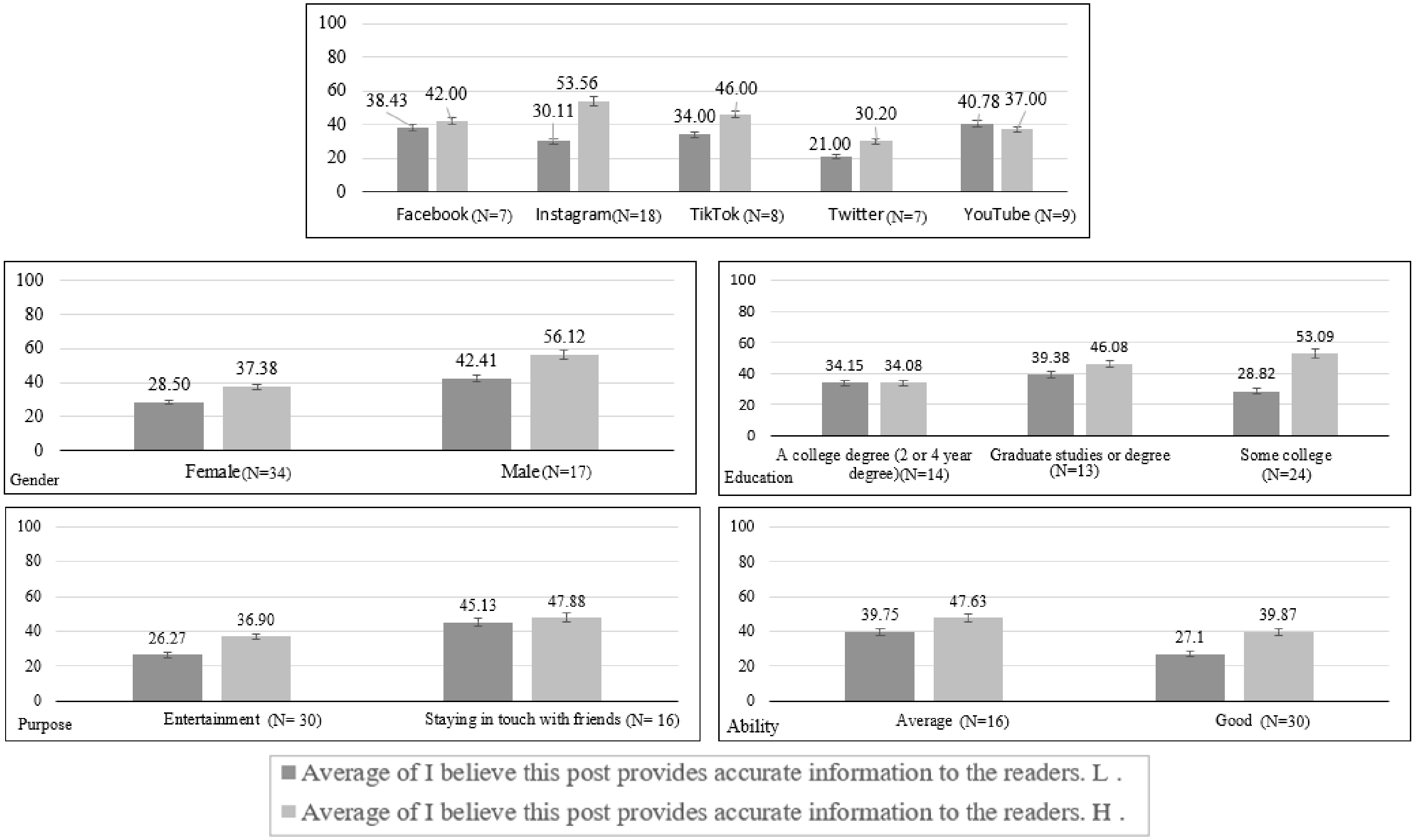
The *Purpose* graph of Figure 15 and Figure 16 exhibits how the purpose of using social media platforms changes participants' responses. The majority in this experiment use social media for entertainment (70%) and 18% of them use them to be connected to their friends. Based on the *Purpose* graph of Figure 15, users that employ social media as an entrainment platform tended to evaluate posts as less accurate than people that use social media for connecting with their friends. Plus, both groups believed that both posts are cable to influence their behavior. Participants who use social media to be entertained are evaluated to be influenced by high emotional content posts almost 25% more than low emotional content posts, and this difference is significant at 10% level (p=0.04).

The last part of Figure 15 and Figure 16 is the *ability*, which asked participants to evaluate their ability in disinformation detection. While 59% of the participants classified themselves as good in deception detection, 31% of them classified their ability as average. As it clarifies, participants in the average group evaluated disinformation less accurate than the other two groups.

Participants who evaluate themselves as good rated high emotional content posts almost 96% more accurate than low emotional content posts, and this difference is significant at 10% level (p=0.08). Moreover, this group evaluated to be influenced by high emotional content posts almost 26% more than low emotional content posts, and this difference is significant at 10% level (p=0.06).

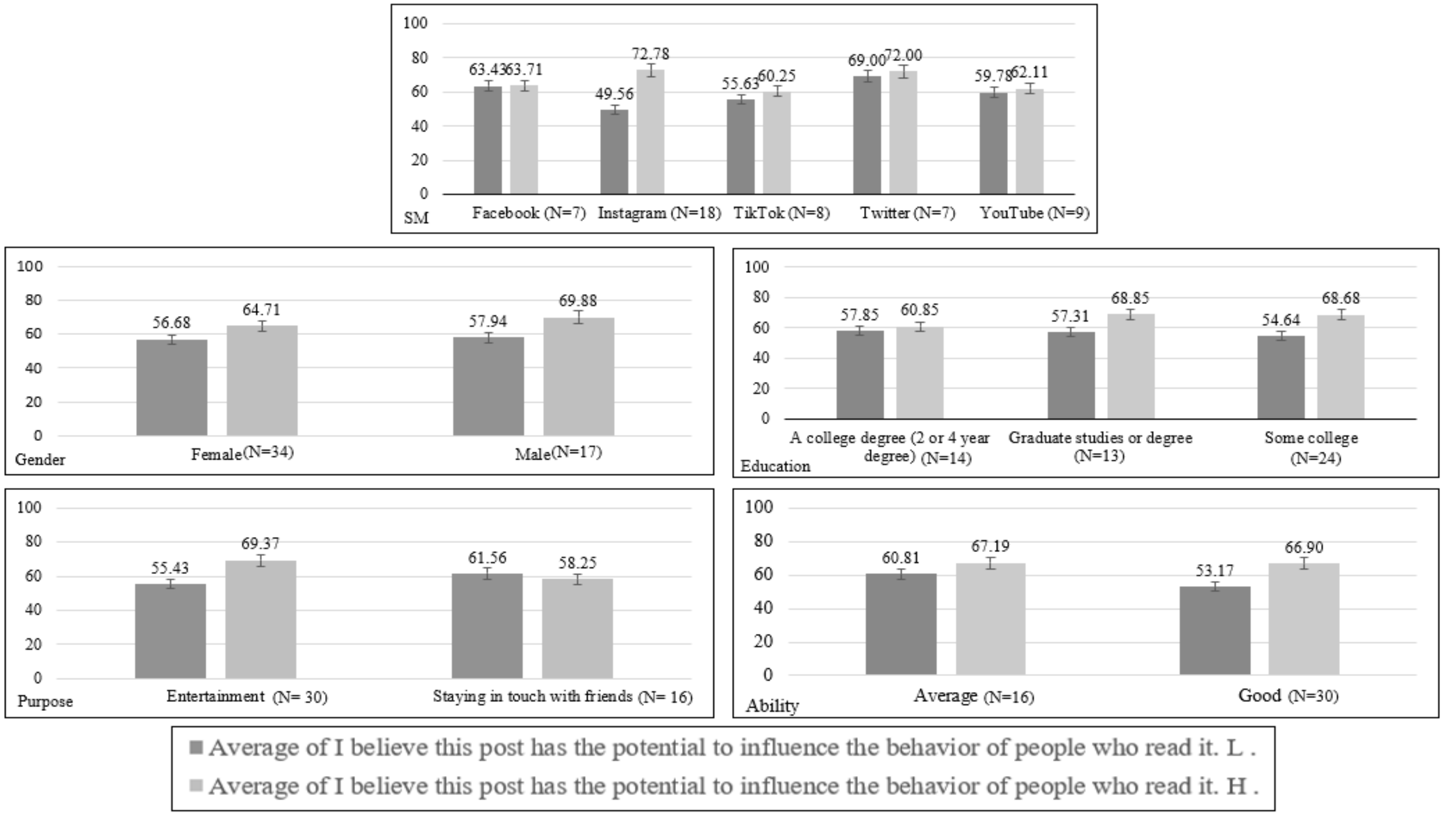
In the next subsections, the sociodemographic distribution of participants in sadness and anger evaluation will be provided. The comparison of the performance among all groups would provide a generalizable result from the study.

Figure 15. Fear tone sociodemographic determinants of the accuracy of the posts.[1]



[1] Groups that have less than 5 participants (N<5) are not reflected in graphs.

Figure 16. Fear tone sociodemographic determinants of the potential influence of the posts.[1]



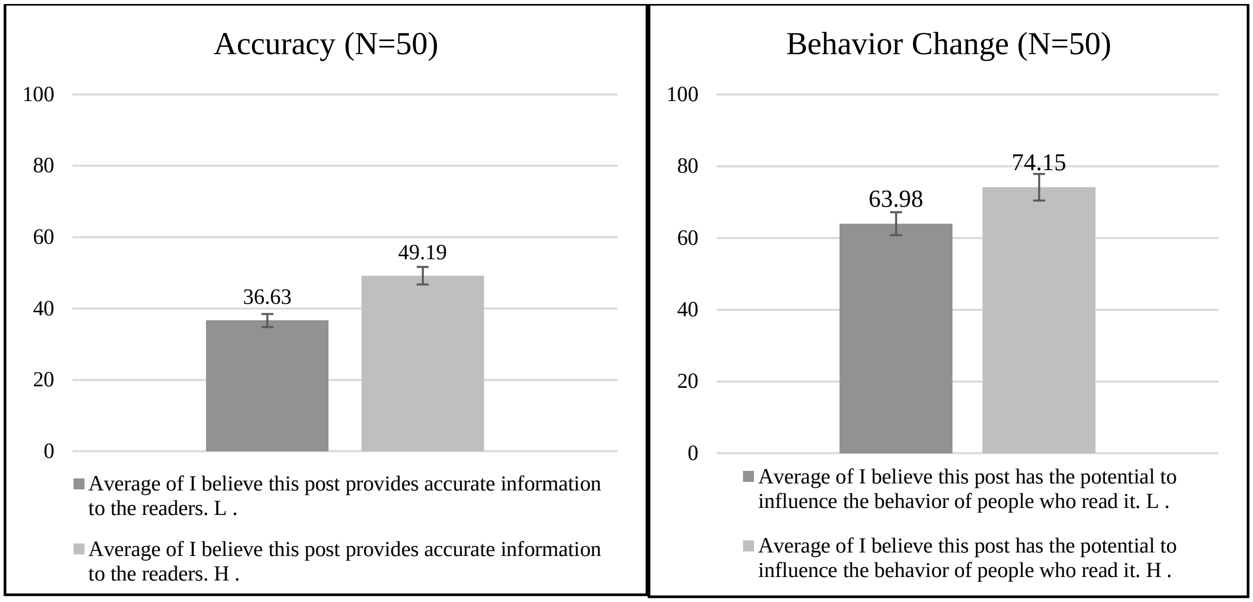
[1] Groups that have less than 5 participants (N<5) are not reflected in graphs.

Sadness: The second emotion that I studied is sadness.

In this test number of participants is fifty. Like other emotions, the results illustrate a significant difference in deception detection between low and high sadness tone. Participants evaluated posts with high sadness tone to convey more accurate than posts with low sadness tone, where the accuracy of the high and low sadness tone are averaged to be %49 and %37 respectively. Readers rated high emotional content posts as 34% more accurate than low emotional content posts, and this increase was significant at 10% level (p=0.068). In other words, participants considered posts with high sadness tone to convey more accurate information and posts with low sadness tone to convey less accurate information. It seems like that, using high sadness tone also could mislead audiences' disinformation recognition abilities.

Moreover, the influence power of the high sadness tone has been estimated higher than low sadness tone posts, where the influence score of high and low fear tone were %74 and %64 respectively. Readers rated the potential behavior change in high emotional content posts as 16% more accurate than low emotional content posts, and this increase was significant at 10% level (p=0.03). Figure 17 represent how participants evaluated posts accuracy and estimated the change of behavior after reading low and high sadness tone respectively. As it is illustrated, participants estimated that posts with high sadness tone are more capable to influence the behavior of people who read them. It seems like the use of high sadness tone could be more influential in a change of behavior.

Figure 17. Accuracy: Participants performance regarding low and high sadness tone. Behavior change: Estimated potential change of behavior after reading posts with different sadness tone.



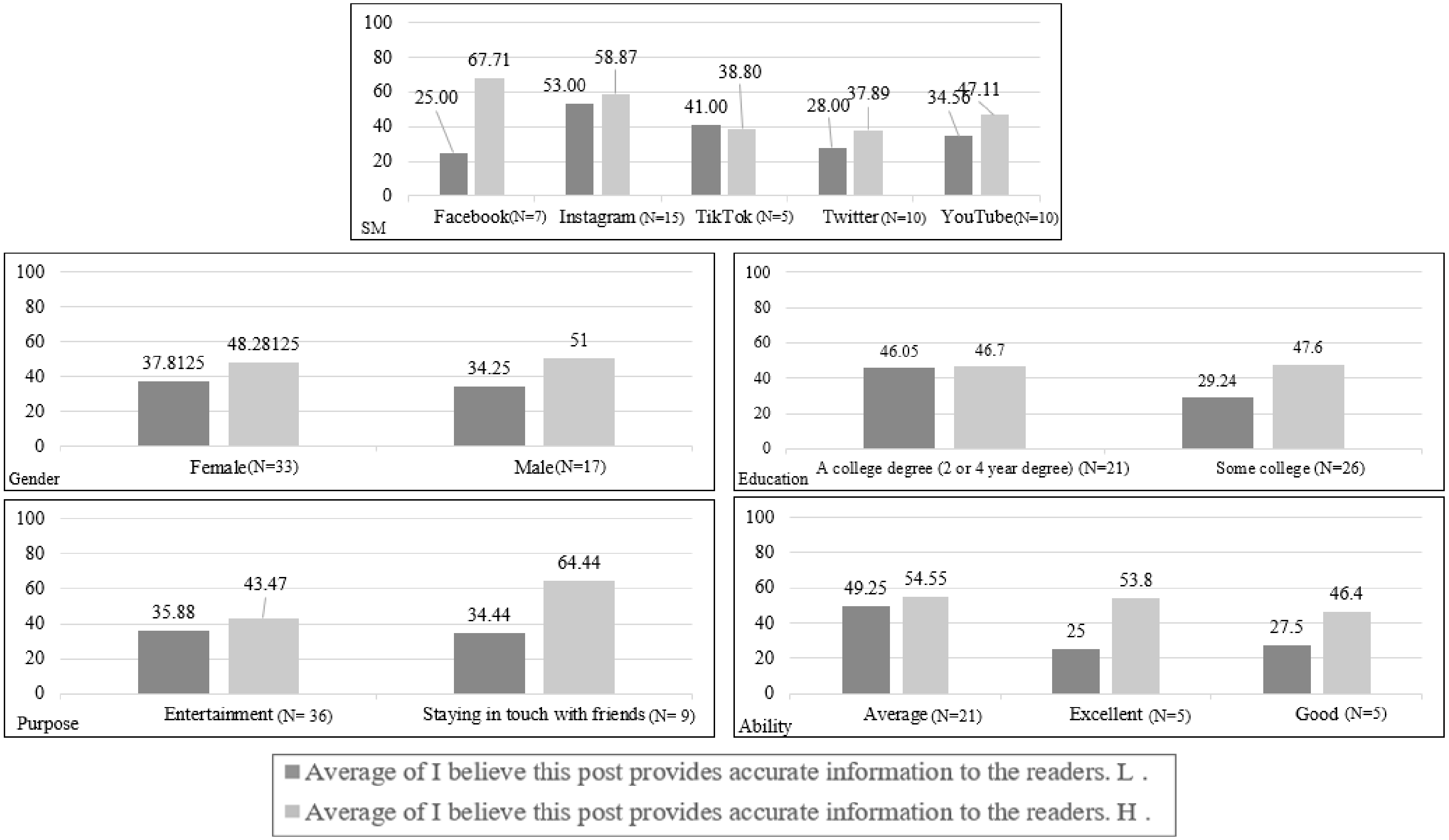
The sociodemographic graph (Figure 18 and Figure 19) illustrates the distribution among the sadness experiment participants. Regarding the *SM* graph of Figure 18 and Figure 19, the majority of students use Instagram (30%) as their preferred social media platform, after that Twitter, and YouTube each with 20%, Facebook with 14%, and TikTok with 10% were the most popular ones. Among all the SM platforms, Facebook and Twitter users evaluated posts with low sadness tone to be less accurate, and all groups of participants evaluate posts with high sadness tone to be more accurate. Participants who use Twitter tend to have the lowest average accuracy ratings in the low condition and show the most pronounced difference in accuracy ratings between low and high emotional content conditions. Participants who use Facebook rated high emotional content posts almost 168% more accurate than low emotional content posts, and this difference is significant at 10% level (p=0.020). Among social media platforms, TikTok users estimated to be less influenced by the posts. The *Gender* graph of Figure 18 also does not represent any significant differences between males and females. In this experiment, female genders where 66% of the participants were female and 34% were males. Male participants evaluate to be influenced by high emotional content posts almost 18% more than low emotional content posts, and this difference is significant at 10% level (p=0.077).   The *Education* graph of Figure 18 and Figure 19 display that the majority have some college degree (52%) and 40% of the group a college degree. Participants who have some college education rated high emotional content posts almost 62% more accurate than low emotional content posts, and this difference is significant at 10% level (p=0.03). Moreover, this group evaluated to be influenced by high emotional content posts almost 13% more than low emotional content posts, and this difference is significant at 10% level (p=0.05).

The *Purpose* graph of Figure 18 and Figure 19 exhibits that participants that use social media for entertainment (72%) on average evaluate posts to convey less accurate information than those who use social media to be connected to their friends (18%). Participants who use social media to be in touch with their friends rated high emotional content posts almost 63% more accurate than low emotional content posts, and this difference is significant at 10% level (p=0.038). Based on the *Purpose* graph of Figure 19, users that employ social media as an entrainment platform believed people could be less influenced by provided posts in comparison to the other group. Participants who use social media to be entertained evaluated to be influenced by high emotional content posts almost 21% more than low emotional content posts, and this difference is significant at 10% level (p=0.09).

The last part of Figure 18 and Figure 19 is the *ability*, which asked participants to evaluate their ability in disinformation detection. While 42% of the participants classified themselves as average in deception detection, 10% of them classified their ability as good and 10% as excellent.

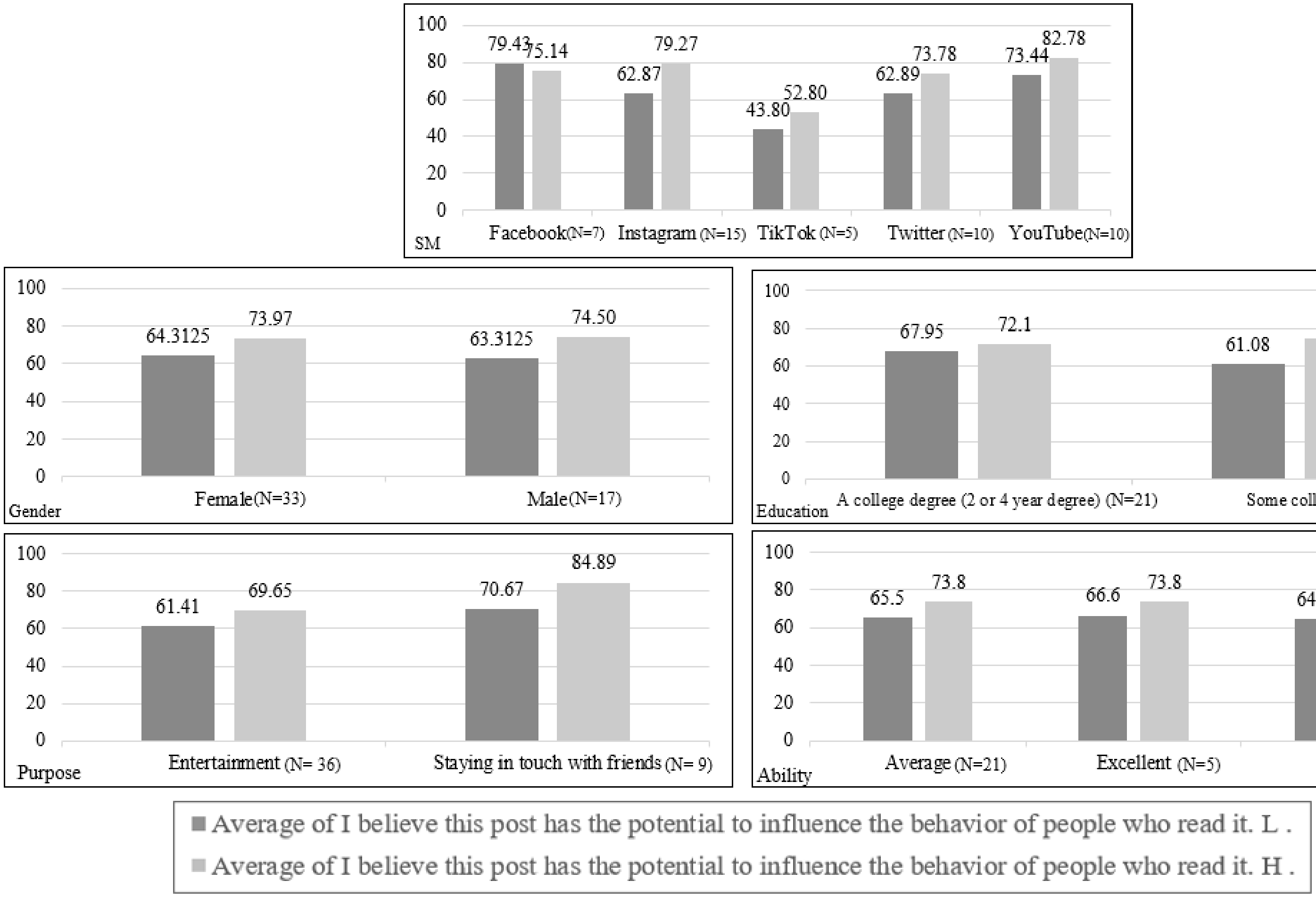
As it clarifies, participants in excellent and good groups evaluate the disinformation posts to convey less accurate information. Moreover, the change of perception regarding the accuracy of the disinformation by change of emotional tone level is interesting. The graph represents participants who rated their ability to identify the truthfulness of something they read on social media as excellent were most misled by the increase in the emotional content of the posts. Based on the *ability graph of* Figure 18 the perception of the excellent group regarding the accuracy of the post for low sadness tone was 25% and for high sadness tone was 53.8%. Readers in this group rated high emotional content posts as 115% more accurate than low emotional content posts.

Figure 18. Sadness tone sociodemographic determinants of the accuracy of the posts[1].



[1] Groups that have less than 5 participants (N<5) are not reflected in graphs.

Figure 19. Sadness tone sociodemographic determinants of the potential influence of the posts[1].



[1] Groups that have less than 5 participants (N<5) are not reflected in graphs.

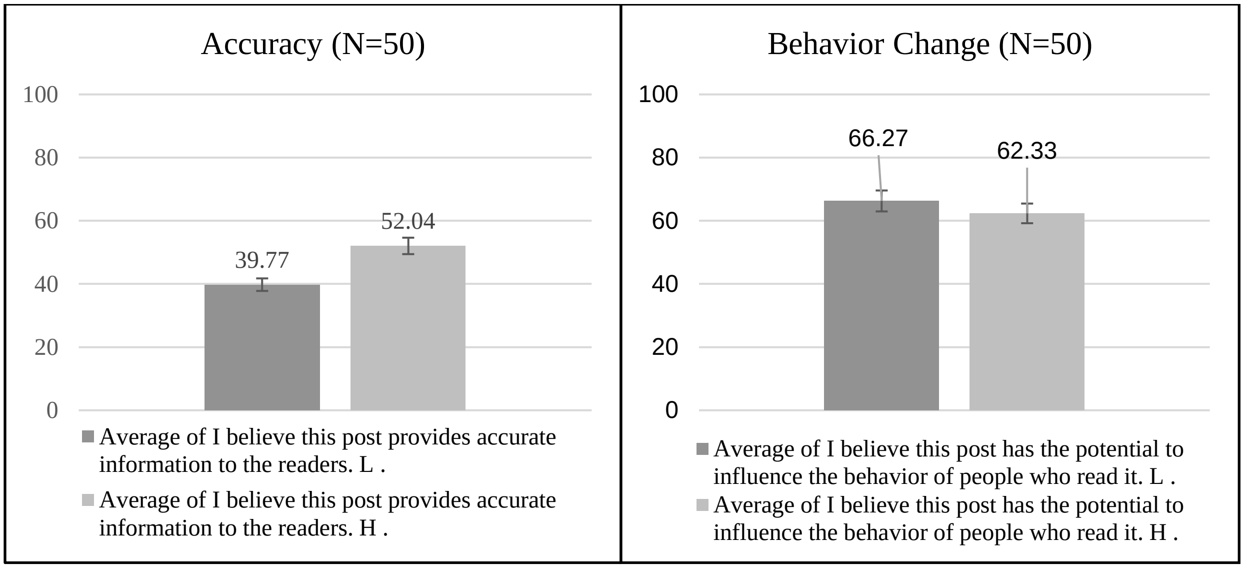
Anger: The third emotion that I studied is anger.

In this test number of participants is fifty. On average, participants evaluated posts with high anger tone as more accurate than posts with low anger tone. The average of readers' perceived accuracy of high and low anger tone posts was %52 and %40 respectively. Readers rated high emotional content posts as 30% more accurate than low emotional content posts, and this increase was significant at 10% level (p=0.02).

The dark grey bar and light grey bar in Figure 20 (*Accuracy*) represent how participants evaluated the accuracy of the posts with low anger tone and high anger tone respectively. In this analysis, higher scores mean participants evaluate the post as information and lower scores mean post has been evaluated to contain disinformation rather than information. As it is depicted, even though all posts were disinformation, participants considered posts with high anger tone to convey more accurate information and posts with low anger tone to convey less accurate information. These findings provide evidence that using high anger tone misled audiences' disinformation recognition abilities by making them increasingly inclined to believe misinformation as accurate information.

Participants' evaluation of the ability of high tone and low tone posts to change readers' behavior did not reveal a significant difference (p-value=0.4).

Figure 20. Accuracy: Participants performance regarding low and high anger tone. Behavior change: Estimated potential change of behavior after reading posts with different anger tone.



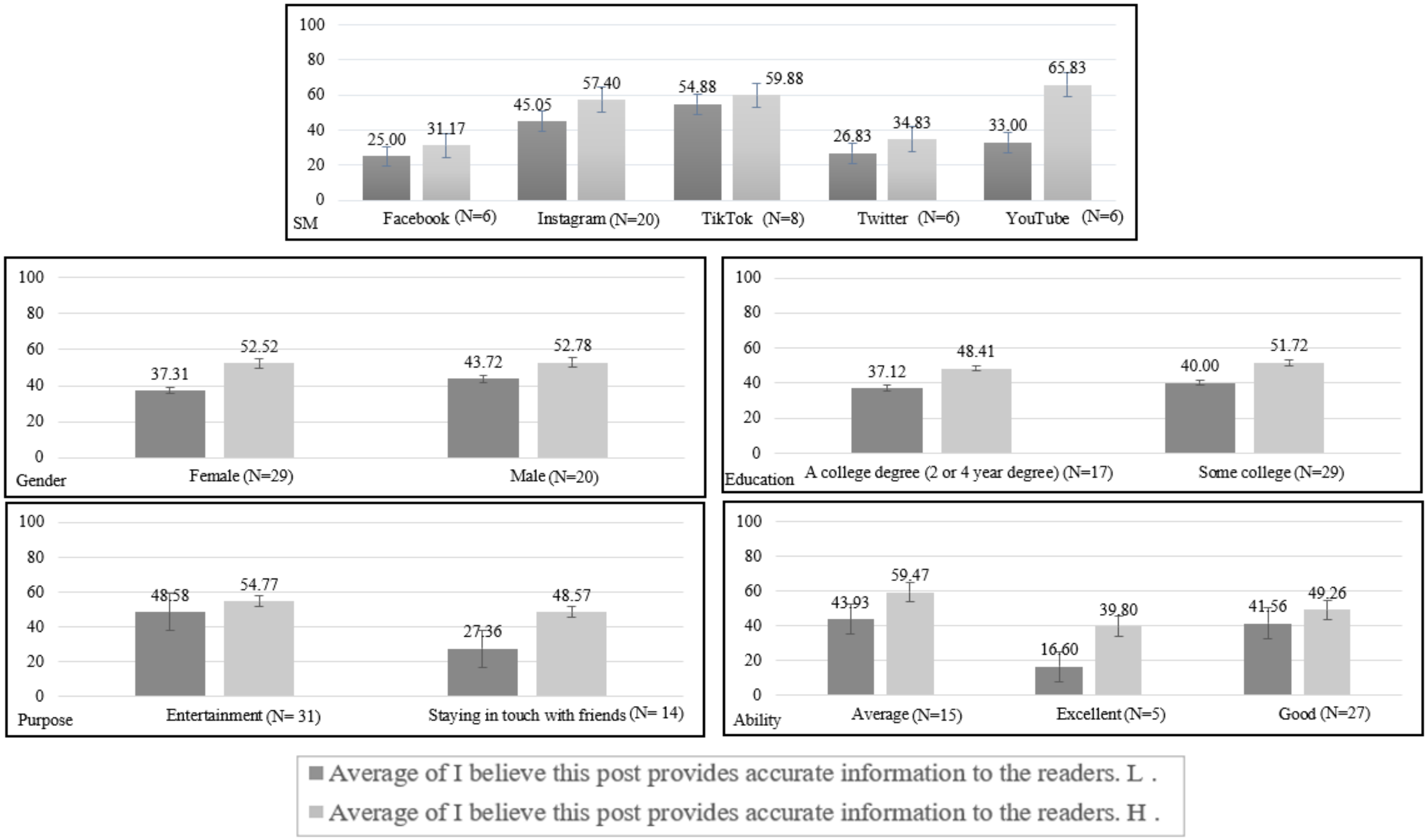
Besides comparing the average values, I also study the difference of responses among various sociodemographic factors. Figure 21 represents the sociodemographic determinants of change in believability upon exposure to disinformation about COVID-19.

The *SM* graph of Figure 21 depicts the preferred social media (SM) platform among participants. Participants who use Facebook tend to have the lowest average accuracy ratings across all conditions, and participants who use YouTube show the most pronounced difference in accuracy ratings between low and high emotional content conditions. Participants who use YouTube rated high emotional content posts almost 100% more accurate than low emotional content posts. The number of participants is small for Facebook and YouTube, therefore the differences are condition, even though practically large, are not statistically significant. Instagram is the social media that has the highest number of users among participants. Instagram users rated high emotional content posts 27% more accurate than low emotional content posts.

The *Gender* graph of Figure 21 depicts the differences among males and females where 58% of participants were female and 40% were male. Based on the results, females tend to evaluate posts as more accurate in comparison to male participants. Based on this graph females evaluated the accuracy of the post for low sadness tone 37% and for high sadness tone 52%. Readers in the females group rated high emotional content posts as 40% more accurate than low emotional content posts, and this increase was significant at 10% level (p=0.041). The *Education* graph (Figure 21) displays the accuracy of posts among different educational levels where the majority have some college degree (58%) and 34% of the group a college degree. Based on the *Education* graph of Figure 21, participants with college degrees in average evaluated both high and low anger tone evaluated post as more accurate in comparison to the other group.

The *Purpose* graph of Figure 21 exhibits how the purpose of using social media platforms changes participants' responses. The majority in this experiment use social media for entertainment (62%) and 28% of them use them to be connected to their friends. Based on this graph, users that employ social media as an entrainment platform tend to evaluate posts to have more accurate information (*Purpose* graph of Figure 21). In contrast with this group, people that use social media for connecting with their friends tend to evaluate the posts as less accurate. Participants who use social media to be in touch with their friends rated high emotional content posts almost 77% more accurate than low emotional content posts, and this difference is significant at 10% level (p=0.042). It is an important finding of the study that reflects how the purpose of people in using social media and their perspective regarding social media platform usage can affect their judgment.

Figure 21. Anger tone sociodemographic determinants of the accuracy of the posts.[1]



[1] Groups that have less than 5 participants (N<5) are not reflected in graphs

The last part of Figure 21 is the *ability*, which asked participants to evaluate their ability in disinformation detection. While 54% of the participants classified themselves as good in deception detection, 30% of them classified their ability as average and 10% as excellent. Participants who use rate themself as excellent tend to have the lowest average accuracy ratings across all conditions. However, this group rated high emotional content posts almost 140% more accurate than low emotional content posts. The graph represents participants who rated their ability to identify the truthfulness of something they read on social media as excellent, were most misled by the increase in the emotional content of the posts.

Additionally, I did not find the predicted difference between high and low post ratings for behavior change for anger tone, therefore, I do not further investigate it. In the next section, the overall findings of sociodemographic factors will be summarized.

Discussion

Summary of Findings

The analysis of essay II demonstrated that disinformation posts which contain high negative tone were more successful than information posts with low negative tone in engaging their audiences based on the number of likes, comments, shares, and emotional interactions. Based on the second essay, I predicted that using a higher negative emotional tone in social media posts would be used to mislead users' ability in recognizing disinformation. Considering the findings of essay II, I proposed running experiments to evaluate the role of negative emotional tone on people's perception of the accuracy of the textual content and its potential influence on decision making and change of behavior. Consistent with my predictions, participants rate disinformation with high negative emotional content more accurate than written disinformation with low negative emotional content. This finding was consistent among three experiments that were used to generalize the fact that an increase in negative tone would mislead the deception detection ability of the readers. Literature on the relationship between emotional content and the ability to detect deception has illustrated the unique role of emotional processing in susceptibility to false information Martel et al. (2020). Our findings also aligned with the literature and proved that participants' ability in recognizing disinformation decreased significantly by increasing the negative emotional tone in COVID-19 related disinformation. Therefore, the results show that using high negative emotional tone structure which is a method frequently used by disinformation-propagating social media content producers, which was also supported by essay II, is indeed effective in increasing the perceived accuracy of, and thus believing the (dis)informational content.

The analysis of essay II represented a significant association between the negative emotional tone and the post engagement score. Loomba et al. stated that "The public's willingness to accept a vaccine is therefore not static; it is highly responsive to current information and sentiment around a COVID-19 vaccine, as well as the state of the epidemic and perceived risk of contracting the disease"(Loomba et al. 2021, P.337). Higher engagement with disinformation posts during the COVID-19 pandemic caused serious changes in behavior such as reducing the vaccine acceptance rate (Loomba et al. 2021; Solís Arce et al. 2021). Therefore, the study of the change of behavior is an inseparable part of the disinformation research study. In this experiment, participants also rated the potential change of behavior after reading posts. The analysis of change of behavior illustrated that participants have been estimated that posts with high fear and sadness emotional tone would significantly change their future behavior. By considering both emotions, on average the participants' evaluation showed that their behavior might get influenced 17% more in high negative emotional content. Therefore, it can be concluded that disinformation posts that contain a higher negative tone would have higher influential power to change the behavior of their audiences.

Summing up my observation over sociodemographic factors would provide a better idea about the groups' performance on average. By considering all three experiments, I can sort our findings regarding sociodemographic factors as follow. On average females evaluated disinformation to convey less accurate information than male participants in both high and low emotional tone. However, the analysis reveals that on average female participants rated high emotional content posts almost 73% more accurate than low emotional content posts while the number is equal to 54% among male participants. This finding reveals that although females, on average, perform better than males in detecting disinformation, an increase in the emotional content of the posts misleads females more than males. Comparing the pattern of results for social media platforms, I found that, on average, participants who use Twitter tend to have the lowest average accuracy ratings in low emotional tone condition, and participants who use Facebook show the most pronounced difference in accuracy ratings between low and high emotional content conditions. Participants who use Facebook rated high emotional content posts almost 150% more accurate than low emotional content posts. Twitter is another social media that has a high difference in accuracy rating. Twitter users rated high emotional content posts 156% more accurate than low emotional content posts. The analysis uncovers that these two platforms' users are more vulnerable to high emotion-embedded content. Participants' perception about their ability in recognizing disinformation is another important factor in accuracy analysis. Based on all three experiments, although people with excellent ability rate posts with low negative emotional tone to be less informative, this group rate high emotional tone posts 156% more accurate than low emotional tone posts. These findings reveal that participants who rate their ability at the highest level, turn out to be the ones who are least capable in keeping their judgment unbiased when they are exposed to highly emotional content. The patterns of self-rating and realized performance for this group of participants manifest overconfidence in their abilities because they rate themselves excellent while they are the group that is most highly misled by change of emotional content of the posts.

Limitations

As with all research, this study is not without limitations. To reduce the limitation of the research several procedures have been considered. Our sample of posts included disinformation about COVID-19, which mostly included negative emotions. This limited our ability to investigate a possible impact of positive emotions on readers' judgment of posts' accuracy and ability to change behavior. Because of our relatively small number of participants per experiment, a few outliers or ignorant participants can drastically bias the results. This possibly explains the insignificant results for the ability to change behavior in Anger. Our participants consist entirely of students of one university. Although this can be considered a limitation, prior research provided evidence that student participants generate reliable results that are no different than randomly selected participants from a diversified population.

Additionally, in this study, I selected social media posts that contain disinformation to be able to evaluate the effect of negative tone on the believability of the content. The lack of true information in the survey list could be the potential limitation of this study. This limitation can be eliminated and be directed as a future study by running the 2 by 2 factorial design setting for a sample of true information disinformation FB posts at varying emotional tone levels. Comparative analysis among disinformation posts results and true information posts results could enhance our understanding of the relationship between negative emotional tone and believability of the content.

Contributions

This experimental study contributes to the theory and practice in several ways. From a theoretical perspective, the findings of this study extended our understanding regarding the believability of disinformation in social media and integrates the emotional tone of content to persuasion theory in the field of social media. Our analysis illustrates the significant role of negative emotional tone on the perceived accuracy of social media posts and their potential to influence behavior. As I discussed earlier studying the believability of the content mostly focuses on political topics based on non-content-based factors. Although the impact of content-based factors on the believability of the content has been examined in political science literature, the disinformation pandemic underlined the importance of believability study in other fields such as health and medical sciences. This study moves forward and investigates the content-based feature to fill this gap. The results of conducted experiments discovered how a change of negative emotional tone can mislead human perception regarding the accuracy of the content. The analysis disclosed that the disinformation posts with higher negative emotional tone have been evaluated to convey more accurate information.

Another observation from essay II was the significant differences in engagement scores of disinformation and information posts. This difference motivated me to study the change of behavior after exposure to disinformation. Therefore, from a change of behavior perspective, this essay contributes to the literature of social media by analyzing the effect of disinformation tone on change of behavior. The results of essay II also revealed that social media posts audiences get more engaged with disinformation posts that have a higher negative emotional tone. The current research provided a better understanding of the effect of negative emotional tone on change of behavior. The findings of my research illustrate that the estimated change of behavior after reading disinformation posts is associated with the human perception regarding the accuracy of the content. Based on our analysis, participants estimated that high negative emotion embedded posts that have been evaluated to convey more accurate information would cause more change of behavior in comparison to the low negative emotional tone posts. In other words, considering the findings of the studies, I can conclude that higher negative emotional tone content has higher believability power which could lead to higher engagement scores.

Furthermore, the current study investigated the role of various negative emotions to provide detailed insight regarding different emotions. From a holistic perspective, a negative emotional tone is a significant factor in the believability power of the content. By looking closer to the analysis, my research shows that while readers rated high anger emotional content posts as 30% more accurate than low emotional content posts, readers rated high sadness emotional content posts as 34% more accurate than low emotional content posts, and readers rated high fear emotional content posts as 39% more accurate than low emotional content posts. These results represent that fear tone and sadness tone might have a stronger effect than anger tone in misleading their audiences. These findings demonstrate the importance of a content-based feature in social media research.

My findings have several implications for practice. First, social media providers can use the findings to navigate and alert their audiences about the accuracy of the content based on emotional scoring tools. I believe the study of the emotional tone on the believability of the content can be used to provide an intervention strategy to mitigate the cognitive effect of negative tone on perceptions. Providing negative and positive emotion rates of social media posts can be studied as a future direction to study the cognitive awareness on the believability of the content. These findings also could be important for political and medical studies to understand the nature of disinformation in a different field to inform social media audiences about the strategies that disinformers use to mislead people's judgment. Moreover, the findings of this study can be used to provide guidelines in the information literacy field. Information literacy can be defined as an ability to find information needed to address a given problem, evaluate information, organize them, use and communicate information (Klusek and Bornstein 2006; Webber and Johnston 2000). Agosto (2018) in her "*Information Literacy and Libraries in the age of Fakes News"* book highlights the importance of information evaluation and mentioned that librarians need to teach users how to think critically and how to evaluate information. In this book, the author believes that this is librarians' responsibility to teach their communities how to determine whether the information they encounter online is accurate, reliable, and worthy of being shared. The analysis of the essay I illustrated that word repetition and negative emotion were key factors in identifying disinformation from true information. Here, the results illustrated the significant role of negative emotional tone on human perception regarding the accuracy of the content and its potential influence on the change of behavior. I believe that content-based features especially the role of negative emotional tone in deception can be useful in providing a guideline for evaluating information accuracy to enhance information literacy skills. Librarians need to emphasize and warn users regarding content-based features while reading text. As an instance, if the textual content conveys unnecessary negative emotion regarding the topic, it could be a red flag for deceptions and users need to verify the reference of the claims. Groppe in his theory of repetition article stated that " if the focus of the writer or speaker is on the audience and the action the audience should take to change or maintain a world view, then repetition will be frequent" (1984. P.167). It could be concluded that repetition would be one of the major tricks of the disinformation producers to influence their audiences influence worldview. Again, readers need to be educated regarding the word repetition and use it as a red flag for verifying the authenticity of the content.

CHAPTER VI: CONCLUSION AND CONTRIBUTION

In this research, I tried to understand the nature of disinformation textual content both on published articles on disinformers' websites and on their social media accounts to evaluate how content-based features be used to identify disinformation and true information based on machine learning algorithms. Moreover, to investigate content-based features effect on engagement score with social media posts, people deception detection ability, and its potential influence on people's future decisions. Disinformation can be defined as the act of creating and sharing false and manipulated information to deceive and mislead the audiences to gain political, personal, and financial benefits (Gelfert 2018; Tandoc Jr et al. 2018). To be able to conduct this research, 3 separate steps have been applied.

In the first step, I attempted to understand the structure of COVID-19 related disinformation textual content on websites to be able to provide an applicable approach to identify disinformation from true information by machine learning algorithms. Using the published article in 2020 on three well-known COVID-19 disinformers, this study reveals that word repetition is one of the major techniques that have been used in creating disinformation textual content. Groppe in his theory of repetition article stated that " if the focus of the writer or speaker is on the audience and the action the audience should take to change or maintain a world view, then repetition will be frequent" (1984. P.167). Considering the abovementioned definition, and the theory of repetition could be concluded that repetition would be one of their major tricks of the false information producers to influence their audiences influence worldview, where the results of both DTM-based and DTM-SDA-based models highlighted this fact.

I also have explored the sentimental score of the articles as well. The result of the analysis illustrates the significant differences between negative and positive emotions of the disinformation and true information content. In this study, sentiment analysis illustrates that while COVID-19 disinformation articles have higher scores in anger, disgust, and fear tones, true articles have higher scores in anticipate, sadness, and trust tones. Furthermore, the findings of my research proposed that the combination of word repletion and sentiment scores would lead to a better performance of the ML algorithms. Based on the results, the random forest algorithm shows the best performance and Naïve Bayes showed the weakest performance. This study proposes that using word repetition patterns and sentiment scores together can empower ML-based disinformation detection techniques to flag disinformation content.

In my second essay, I tried to extend my finding to social media content. Nowadays, social media platforms are an inevitable part of daily lives that people use to create a social network, enjoy the connection, and share content. However, the existence of the deceptive content can mislead users' ability in recognizing disinformation and fall into believing disinformation. Disinformation detection can be defined as an observation that differs significantly from other observations which can arouse suspicion that it is generated with a different mechanism. In recent years by the growth of social media platforms and the growth of disinformation in social media networks, disinformation detection gained attention as one of the newly emerging research areas (M. R. Islam et al. 2020). In the social media research field, uses and gratification theory proposes that people use social media to fulfill their major needs: 1) cognitive needs to acquire information, knowledge, and understanding, 2) affective needs to satisfy Emotion, pleasure, feelings, 3) personal integrative needs to provide them credibility, respect, status, confidence, and stability, 4) social integrative needs to enhance the connection to family and friends, and 5) tension release needs to escape from routines or daily problems (West et al. 2010). To the best of my knowledge, literature uses UGT to understand users' reasons, however, in this study, I attempt to look at the UGT from disinformers' perspective. In my second essay, I proposed that disinformers use content that matches users' five major cognitive needs to gain their attention. In other words, they use it as a mechanism to create more attractive content and mislead audiences' deception detection skills. Based on the analysis, disinformers use content with a higher psychological need load and high negative emotional tone. The extension of uses and gratification theory and emotional broadcaster theory to this area would enhance our understanding of disinformation created on social media and provide opportunities for false information detection. Additionally, literature proposes that people are likely to spread information that evokes disgust, fear emotions on social media regardless of the information truth level (Peters et al. 2009; Wang et al. 2020). Based on this fact, Martel et al. (2020) illustrate the unique role of emotional processing in susceptibility to false information. Based on the statistician analysis of the second essay there is a significant difference in the negative tone of disinformation and information posts. The outputs of regression analysis also highlight that higher engagement scores of social media posts have been associated with higher negative scores of the posts. The results of the second study also support the fact that disinformation producers that look for more engagement use more negative emotions and increase psychological processes load to have higher engagement scores.

The analysis of essay II demonstrated that disinformation posts which contain a high negative tone were more successful in engaging their audiences based on the number of likes, comments, shares, and emotional interactions. Based on the second essay, I predicted that using a higher negative emotional tone in social media posts would be used to mislead users' ability in recognizing disinformation. By considering the essay II findings, in essay III, I proposed running a series of experiments to evaluate the role of negative emotional tone on people's perception of the accuracy of the textual content and its potential influence on decision making and change of behavior. Literature on studying the believability of the content mostly focuses on political topics (Ali and Zain-ul-abdin 2021; Peixoto 2019) and non-content based factors such as reliability of the source of the message (Trivedi et al. 2021), virality metrics of posts (Kim 2018), social norms (Gimpel et al. 2021) and source rating (Kim and Dennis 2019). However, this study moves forward and investigates the content-based feature on medial related disinformation posts to fill this gap and show the effect of negative tones in social media posts on human perception regarding the accuracy of the content and the potential influence to change in behavior. Consistent with my predictions, participants rate disinformation posts with high negative emotional content more accurate than written disinformation with low negative emotional content. My analysis illustrates the significant role of negative emotional tone on human perception about the accuracy of social media posts and its potential influence on the change of behavior. Considering the findings of these studies, I can conclude that higher negative emotional tone content has higher believability power which could lead to higher engagement scores. The findings of this study improve the understanding of the role of emotional tone on persuasion regarding the accuracy of the content, provide a potential contribution to the deception detection research field, and contribute to providing guidelines to enhance information literacy in society. Information literacy can be defined as an ability to find information needed to address a given problem, evaluate information, organize them, use and communicate information (Klusek and Bornstein 2006; Webber and Johnston 2000). Librarians have a great opportunity and responsibility to teach their communities how to determine whether the information they encounter online is accurate, reliable, and worthy of being shared (Agosto 2018). I believe that content-based features especially the role of negative emotional tone in deception can be useful to provide a guideline for evaluating information accuracy to enhance information literacy skills.

REFERENCES

Agosto, D. E. 2018. *Information Literacy and Libraries in the Age of Fake News,* ABC-CLIO.

Ahmad, I., Yousaf, M., Yousaf, S., and Ahmad, M. O. 2020. "Fake News Detection Using Machine Learning Ensemble Methods," *Complexity* (2020), Hindawi.

Aiken, L. S., West, S. G., and Reno, R. R. 1991. *Multiple Regression: Testing and Interpreting Interactions,* sage.

Ajao, O., Bhowmik, D., and Zargari, S. 2019. "Sentiment Aware Fake News Detection on Online Social Networks," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* IEEE, pp. 2507–2511.

Aldwairi, M., and Alwahedi, A. 2018. "Detecting Fake News in Social Media Networks," *Procedia Computer Science* (141), Elsevier, pp. 215–222.

Ali, K., and Zain-ul-abdin, K. 2021. "Post-Truth Propaganda: Heuristic Processing of Political Fake News on Facebook during the 2016 US Presidential Election," *Journal of Applied Communication Research* (49:1), Taylor & Francis, pp. 109–128.

Allcott, H., and Gentzkow, M. 2017. "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives* (31:2), pp. 211–236.

Anjalin, U., and ZAMAN, A. S. M. S.-U. 2018. "An Organization of Theories According to a Causal Chain Framework in Relation to Social Media Constructs for Future Research," *Journal of Economic & Management Perspectives* (12:2), pp. 711–718.

Appiah Otoo, B., Esmizadeh, Y., and Palvia, P. 2020. *" Like Talking To a Wall": A Study of Multinational Customers' Online Shopping Experiences*.

Apuke, O. D., and Omar, B. 2021. "Fake News and COVID-19: Modelling the Predictors of Fake News Sharing among Social Media Users," *Telematics and Informatics* (56), Elsevier, p. 101475.

Arun, K., Srinagesh, A., and Ramesh, M. 2017. "Twitter Sentiment Analysis on Demonetization Tweets in India Using R Language," *International Journal of Computer Engineering In Research Trends* (4:6), pp. 252–258.

Auberry, K. 2018. "Increasing Students' Ability to Identify Fake News through Information Literacy Education and Content Management Systems," *The Reference Librarian* (59:4), Taylor & Francis, pp. 179–187.

Azab, M. D. E. D. 2011. *Perceptions of Effectiveness of Celebrity Endorsed Advertisements among Egyptian Consumers*.

Badawi, A. A., Al-Kabbany, A., and Shaban, H. 2018. "Multimodal Human Activity Recognition from Wearable Inertial Sensors Using Machine Learning," in *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, IEEE, pp. 402–407.

Bakir, V., and McStay, A. 2018. "Fake News and the Economy of Emotions: Problems, Causes, Solutions," *Digital Journalism* (6:2), Taylor & Francis, pp. 154–175.

Balog-Way, D. H. P., and McComas, K. A. 2020. "COVID-19: Reflections on Trust, Tradeoffs, and Preparedness," *Journal of Risk Research* (23:7–8), Taylor & Francis, pp. 838–848.

Bazerman, M. H., and Moore, D. A. 2012. *Judgment in Managerial Decision Making*, John Wiley & Sons.

De Beer, D., and Matthee, M. 2020. "Approaches to Identify Fake News: A Systematic Literature Review," in *International Conference on Integrated Science*, Springer, pp. 13–22.

Bhutani, B., Rastogi, N., Sehgal, P., and Purwar, A. 2019. "Fake News Detection Using Sentiment Analysis," in *2019 Twelfth International Conference on Contemporary Computing (IC3)*, IEEE, pp. 1–5.

Boush, D. M., Friestad, M., and Wright, P. 2015. *Deception in the Marketplace: The Psychology of Deceptive Persuasion and Consumer Self-Protection*, Routledge.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., and Van Bavel, J. J. 2017. "Emotion Shapes the Diffusion of Moralized Content in Social Networks," *Proceedings of the National Academy of Sciences* (114:28), National Acad Sciences, pp. 7313–7318.

Bruns, A. 2017. "Echo Chamber? What Echo Chamber? Reviewing the Evidence," in *6th Biennial Future of Journalism Conference (FOJ17)*.

Canziani, B. F., Esmizadeh, Y., and Nemati, H. R. 2021. "Student Engagement with Global Issues: The Influence of Gender, Race/Ethnicity, and Major on Topic Choice," *Teaching in Higher Education*, Taylor & Francis, pp. 1–22.

Cao, J., Qi, P., Sheng, Q., Yang, T., Guo, J., and Li, J. 2020. "Exploring the Role of Visual Content in Fake News Detection," *Disinformation, Misinformation, and Fake News in Social Media*, Springer, pp. 141–161.

Castellani, P., and Berton, M. 2017. "Fake News and Corporate Reputation: What Strategies Do Companies Adopt against False Information in the Media?," in *Toulon-Verona Conference" Excellence in Services"*.

Chatterjee, S., and Bhattacharjee, K. K. 2020. "Adoption of Artificial Intelligence in Higher Education: A Quantitative Analysis Using Structural Equation Modelling," *Education and Information Technologies* (25:5), Springer, pp. 3443–3463.

Chen, Y., Conroy, N. J., and Rubin, V. L. 2015. “Misleading Online Content: Recognizing Clickbait as" False News",” in *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*, pp. 15–19.

Clore, G. L., Schwarz, N., and Conway, M. 2014. “Affective Causes and Consequences of Social Information Processing,” *Handbook of Social Cognition: Volume 1: Basic Processes*, Psychology Press, p. 323.

Cotter, E. M. 2008. “Influence of Emotional Content and Perceived Relevance on Spread of Urban Legends: A Pilot Study,” *Psychological Reports* (102:2), SAGE Publications Sage CA: Los Angeles, CA, pp. 623–629.

Craney, T. A., and Surles, J. G. 2002. “Model-Dependent Variance Inflation Factor Cutoff Values,” *Quality Engineering* (14:3), Taylor & Francis, pp. 391–403.

Cruz, A., Rocha, G., Sousa-Silva, R., and Cardoso, H. L. 2019. “Team Fernando-Pessa at Semeval-2019 Task 4: Back to Basics in Hyperpartisan News Detection,” in *Proceedings of the 13th International Workshop on Semantic Evaluation*, pp. 999–1003.

Das, A., Ghai, G., Alam, M., Pardeshi, G., and Kishore, J. 2021. “COVID Appropriate Behaviour Compliance and Vaccine Hesitancy: Findings from a COVID-19 Health Education Campaign in a Government Tertiary Care Hospital in Delhi, India,” *Disaster Medicine and Public Health Preparedness*, Cambridge University Press, pp. 1–17.

Deng, X., Li, Y., Weng, J., and Zhang, J. 2019. “Feature Selection for Text Classification: A Review.,” *Multimedia Tools & Applications* (78:3).

Dennis, A. R., Galletta, D. F., and Webster, J. 2021. “Fake News on the Internet,” *Journal of Management Information Systems* (38:4), Taylor & Francis, pp. 893–897.

Dey, A., Rafi, R. Z., Parash, S. H., Arko, S. K., and Chakrabarty, A. 2018. “Fake News Pattern Recognition Using Linguistic Analysis,” in *2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (IcIVPR)*, IEEE, pp. 305–309.

Duffy, A., Tandoc, E., and Ling, R. 2020. “Too Good to Be True, Too Good Not to Share: The Social Utility of Fake News,” *Information, Communication & Society* (23:13), Taylor & Francis, pp. 1965–1979.

Dunn, J. R., and Schweitzer, M. E. 2005. “Feeling and Believing: The Influence of Emotion on Trust.,” *Journal of Personality and Social Psychology* (88:5), American Psychological Association, p. 736.

Esmizadeh, Y., Canziani, B., Nemati, H. R., and Mondaresnezhad, M. 2020. *Sharing Economy: Application of Structural Topic Models*.

Evans, N. J., Phua, J., Lim, J., and Jun, H. 2017. “Disclosing Instagram Influencer Advertising: The Effects of Disclosure Language on Advertising Recognition, Attitudes, and Behavioral Intent,” *Journal of Interactive Advertising* (17:2), Taylor & Francis, pp. 138–149.

Eysenbach, G. 2020. “How to Fight an Infodemic: The Four Pillars of Infodemic Management,” *Journal of Medical Internet Research* (22:6), JMIR Publications Inc., Toronto, Canada, p. e21820.

Eysenck, M. W., and Keane, M. T. 2015. *Cognitive Psychology: A Student’Handbook*, Psychology press.

Firdaus, S. N., Ding, C., and Sadeghian, A. 2021. “Retweet Prediction Based on Topic, Emotion and Personality,” *Online Social Networks and Media* (25), Elsevier, p. 100165.

Freud, S., Strachey, J., and Freud, A. 1978. *The Psychopathology of Everyday Life:(1901),* Hogarth Press: Institute of Psycho-Analysis.

Fuller, C. M., Biros, D. P., and Wilson, R. L. 2009. “Decision Support for Determining Veracity via Linguistic-Based Cues,” *Decision Support Systems* (46:3), Elsevier, pp. 695–703.

Gass, R. H., and Seiter, J. S. 2018. *Persuasion: Social Influence and Compliance Gaining,* Routledge.

Gefen, D., Straub, D., and Boudreau, M.-C. 2000. “Structural Equation Modeling and Regression: Guidelines for Research Practice,” *Communications of the Association for Information Systems* (4:1), p. 7.

Gelfert, A. 2018. “Fake News: A Definition,” *Informal Logic* (38:1), Informal Logic, pp. 84–117.

Ghenai, A. 2017. “Health Misinformation in Search and Social Media,” in *Proceedings of the 2017 International Conference on Digital Health,* pp. 235–236.

Ghoshal, A. K., Das, N., and Das, S. 2020. “Influence of Community Structure on Misinformation Containment in Online Social Networks,” *Knowledge-Based Systems,* Elsevier, p. 106693.

Giachanou, A., Zhang, G., and Rosso, P. 2020. “Multimodal Fake News Detection with Textual, Visual and Semantic Information,” in *International Conference on Text, Speech, and Dialogue,* Springer, pp. 30–38.

Gimpel, H., Heger, S., Olenberger, C., and Utz, L. 2021. “The Effectiveness of Social Norms in Fighting Fake News on Social Media,” *Journal of Management Information Systems* (38:1), Taylor & Francis, pp. 196–221.

Gottfried, J., and Shearer, E. 2016. *News Use Across Social Medial Platforms 2016*, Pew Research Center.

Gravanis, G., Vakali, A., Diamantaras, K., and Karadais, P. 2019. “Behind the Cues: A Benchmarking Study for Fake News Detection,” *Expert Systems with Applications* (128), Elsevier, pp. 201–213.

Green, L. W., Fielding, J. E., and Brownson, R. C. 2021. “More on Fake News, Disinformation, and Countering These with Science,” *Annual Review of Public Health* (42), Annual Reviews, v–vi.

Grice, H. P. 1975. “Logic and Conversation,” in *Speech Acts*, Brill, pp. 41–58.

Grice, P. 1989. *Studies in the Way of Words*, Harvard University Press.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., and Lazer, D. 2019. “Fake News on Twitter during the 2016 US Presidential Election,” *Science* (363:6425), American Association for the Advancement of Science, pp. 374–378.

Grishman, R. 1986. *Computational Linguistics: An Introduction*, Cambridge University Press.

Groppe, J. D. 1984. “Reality as Enchantment—a Theory of Repetition,” *Rhetoric Review* (2:2), Taylor & Francis, pp. 165–174.

Habib, Ammara, Asghar, M. Z., Khan, Adil, Habib, Anam, and Khan, Aurangzeb. 2019. “False Information Detection in Online Content and Its Role in Decision Making: A Systematic Literature Review,” *Social Network Analysis and Mining* (9:1), Springer, p. 50.

Hameleers, M., and van der Meer, T. G. L. A. 2020. “Misinformation and Polarization in a High-Choice Media Environment: How Effective Are Political Fact-Checkers?,” *Communication Research* (47:2), SAGE Publications Sage CA: Los Angeles, CA, pp. 227–250.

Han, Z., Wu, J., Huang, C., Huang, Q., and Zhao, M. 2020. “A Review on Sentiment Discovery and Analysis of Educational Big-data,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (10:1), Wiley Online Library, p. e1328.

Harber, K. D., and Cohen, D. J. 2005. “The Emotional Broadcaster Theory of Social Sharing,” *Journal of Language and Social Psychology* (24:4), Sage Publications Sage CA: Thousand Oaks, CA, pp. 382–400.

Harber, K. D., Podolski, P., and Dyer, L. 2014. “Hearing Stories That Violate Expectations Leads to Emotional Broadcasting,” *Journal of Language and Social Psychology* (33:1), Sage Publications Sage CA: Los Angeles, CA, pp. 5–28.

Heidari, M., Jones, J. H., and Uzuner, O. 2020. “Deep Contextualized Word Embedding for Text-Based Online User Profiling to Detect Social Bots on Twitter,” in *2020 International Conference on Data Mining Workshops (ICDMW)*, IEEE, pp. 480–487.

Hirlekar, V. V., and Kumar, A. 2020. “Natural Language Processing Based Online Fake News Detection Challenges–A Detailed Review,” in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, pp. 748–754.

Hooshangi, S., and Loewenstein, G. 2018. “The Impact of Idea Generation and Potential Appropriation on Entrepreneurship: An Experimental Study,” *Management Science* (64:1), INFORMS, pp. 64–82.

Horne, B., and Adali, S. 2017. “This Just in: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News,” in *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 11), pp. 759–766.

Horne, B. D., Nørregaard, J., and Adali, S. 2019. “Robust Fake News Detection over Time and Attack,” *ACM Transactions on Intelligent Systems and Technology (TIST)* (11:1), ACM

New York, NY, USA, pp. 1–23.

Huang, N., Hong, Y., and Burtch, G. 2016. “Social Network Integration and User Content Generation: Evidence from Natural Experiments,” *MIS Quarterly (Forthcoming), Fox School of Business Research Paper* (17–001).

Hunt, E. 2016. “What Is Fake News? How to Spot It and What You Can Do to Stop It,” *The Guardian* (17), p. 12.

Islam, A. K. M. N., Laato, S., Talukder, S., and Sutinen, E. 2020. “Misinformation Sharing and Social Media Fatigue during COVID-19: An Affordance and Cognitive Load Perspective,” *Technological Forecasting and Social Change* (159), Elsevier, p. 120201.

Islam, M. R., Liu, S., Wang, X., and Xu, G. 2020. “Deep Learning for Misinformation Detection on Online Social Networks: A Survey and New Perspectives,” *Social Network Analysis and Mining* (10:1), Springer, pp. 1–20.

Izawa, M. 2010. *What Makes Viral Videos Viral?: Roles of Emotion, Impression, Utility, and Social Ties in Online Sharing Behavior,* Citeseer.

Jacobs, S., Dawson, E. J., and Brashers, D. 1996. “Information Manipulation Theory: A Replication and Assessment,” *Communications Monographs* (63:1), Taylor & Francis, pp. 70–82.

Jahng, M. R. 2021. “Is Fake News the New Social Media Crisis? Examining the Public Evaluation of Crisis Management for Corporate Organizations Targeted in Fake News,” *International Journal of Strategic Communication*, Taylor & Francis, pp. 1–19.

Jain, T., and Pradhan, V. 2020. “Role of Socio-Demographic on Social Media User Believability and Attitude towards Veracity of News about Covid-19 Pandemic,” *International Journal of Modern Agriculture* (9:3), pp. 1185–1202.

Jia, F. 2020. “Misinformation Literature Review: Definitions, Taxonomy, and Models,” *International Journal of Social Science and Education Research* (3:12), Boya Century Publishing Limited, pp. 85–90.

Jin, Z., Cao, J., Zhang, Y., Zhou, J., and Tian, Q. 2016. “Novel Visual and Statistical Image Features for Microblogs News Verification,” *IEEE Transactions on Multimedia* (19:3), IEEE, pp. 598–608.

Jockers, M. 2017. “Package ‘Syuzhet,’” *URL: Https://Cran. r-Project. Org/Web/Packages/Syuzhet*.

Kanozia, R., and Arya, R. 2021. “‘Fake News’, Religion, and COVID-19 Vaccine Hesitancy in India, Pakistan, and Bangladesh,” *Media Asia* (48:4), Taylor & Francis, pp. 313–321.

Kao, J., and Jurafsky, D. 2012. “A Computational Analysis of Style, Affect, and Imagery in Contemporary Poetry,” in *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature,* pp. 8–17.

Karduni, A., Wesslen, R., Markant, D., and Dou, W. 2021. “Images, Emotions, and Credibility: Effect of Emotional Facial Images on Perceptions of News Content Bias and Source Credibility in Social Media,” *ArXiv Preprint ArXiv:2102.13167*.

Karduni, A., Wesslen, R., Santhanam, S., Cho, I., Volkova, S., Arendt, D., Shaikh, S., and Dou, W. 2018. “Can You Verifi This? Studying Uncertainty and Decision-Making about Misinformation Using Visual Analytics,” in *Twelfth International AAAI Conference on Web and Social Media*.

Karimi, H., Tang, J., and Li, Y. 2018. “Toward End-to-End Deception Detection in Videos,” in *2018 IEEE International Conference on Big Data (Big Data)*, IEEE, pp. 1278–1283.

Katz, E., Blumler, J. G., and Gurevitch, M. 1973. “Uses and Gratifications Research,” *The Public Opinion Quarterly* (37:4), JSTOR, pp. 509–523.

Khan, M. L., and Idris, I. K. 2019. “Recognise Misinformation and Verify before Sharing: A Reasoned Action and Information Literacy Perspective,” *Behaviour & Information Technology* (38:12), Taylor & Francis, pp. 1194–1212.

Kim, A., and Dennis, A. R. 2019. “Says Who? The Effects of Presentation Format and Source Rating on Fake News in Social Media,” *MIS Quarterly* (43:3), pp. 1025–1039.

Kim, J. W. 2018. “Rumor Has It: The Effects of Virality Metrics on Rumor Believability and Transmission on Twitter,” *New Media & Society* (20:12), SAGE Publications Sage UK: London, England, pp. 4807–4825.

Kirasich, K., Smith, T., and Sadler, B. 2018. “Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets,” *SMU Data Science Review* (1:3), p. 9.

Kline, R. B. 2015. *Principles and Practice of Structural Equation Modeling*, Guilford publications.

Klusek, L., and Bornstein, J. 2006. “Information Literacy Skills for Business Careers: Matching Skills to the Workplace,” *Journal of Business & Finance Librarianship* (11:4), Taylor & Francis, pp. 3–21.

Kula, S., Choraś, M., Kozik, R., Ksieniewicz, P., and Woźniak, M. 2020. “Sentiment Analysis for Fake News Detection by Means of Neural Networks,” in *International Conference on Computational Science*, Springer, pp. 653–666.

Kumar, A., Bezawada, R., Rishika, R., Janakiraman, R., and Kannan, P. K. 2016. “From Social to Sale: The Effects of Firm-Generated Content in Social Media on Customer Behavior,” *Journal of Marketing* (80:1), SAGE Publications Sage CA: Los Angeles, CA, pp. 7–25.

Kunbaz, A., Saghir, S., Arar, M., and Sönmez, E. B. 2019. "Fake Image Detection Using DCT and Local Binary Pattern," in *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, IEEE, pp. 1–6.

Laato, S., Islam, A. K. M. N., Islam, M. N., and Whelan, E. 2020. "What Drives Unverified Information Sharing and Cyberchondria during the COVID-19 Pandemic?," *European Journal of Information Systems* (29:3), Taylor & Francis, pp. 288–305.

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., and Rothschild, D. 2018. "The Science of Fake News," *Science* (359:6380), American Association for the Advancement of Science, pp. 1094–1096.

Lerner, J. S., and Keltner, D. 2000. "Beyond Valence: Toward a Model of Emotion-Specific Influences on Judgement and Choice," *Cognition & Emotion* (14:4), Taylor & Francis, pp. 473–493.

Lerner, J. S., and Keltner, D. 2001. "Fear, Anger, and Risk.," *Journal of Personality and Social Psychology* (81:1), American Psychological Association, p. 146.

Leung, L. 2013. "Generational Differences in Content Generation in Social Media: The Roles of the Gratifications Sought and of Narcissism," *Computers in Human Behavior* (29:3), Elsevier, pp. 997–1006.

Liang, G., He, W., Xu, C., Chen, L., and Zeng, J. 2015. "Rumor Identification in Microblogging Systems Based on Users' Behavior," *IEEE Transactions on Computational Social Systems* (2:3), IEEE, pp. 99–108.

Litvinova, O., Seredin, P., Litvinova, T., and Lyell, J. 2017. "Deception Detection in Russian Texts," in *Proceedings of the Student Research Workshop at the 15th Conference of the*

European Chapter of the Association for Computational Linguistics, pp. 43–52.

Liu, B., Hu, M., and Cheng, J. 2005. "Opinion Observer: Analyzing and Comparing Opinions on the Web," in *Proceedings of the 14th International Conference on World Wide Web*, pp. 342–351.

Liu, D., and Lei, L. 2018. "The Appeal to Political Sentiment: An Analysis of Donald Trump's and Hillary Clinton's Speech Themes and Discourse Strategies in the 2016 US Presidential Election," *Discourse, Context & Media* (25), Elsevier, pp. 143–152.

Liu, S. 2017. *A Study on the Effect of Financial Media in" Internet+" Environment——Based on "the Uses and Gratification Theory,"* Atlantis Press.

Liu, X., Min, Q., and Han, S. 2020. "Understanding Users' Continuous Content Contribution Behaviours on Microblogs: An Integrated Perspective of Uses and Gratification Theory and Social Influence Theory," *Behaviour & Information Technology* (39:5), Taylor & Francis, pp. 525–543.

Loeb, S., Taylor, J., Borin, J. F., Mihalcea, R., Perez-Rosas, V., Byrne, N., Chiang, A. L., and Langford, A. 2020. "Fake News: Spread of Misinformation about Urological Conditions on Social Media," *European Urology Focus* (6:3), Elsevier, pp. 437–439.

Long, Y. 2017. *Fake News Detection through Multi-Perspective Speaker Profiles*, Association for Computational Linguistics.

Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., and Larson, H. J. 2021. "Measuring the Impact of COVID-19 Vaccine Misinformation on Vaccination Intent in the UK and USA," *Nature Human Behaviour* (5:3), Nature Publishing Group, pp. 337–348.

Mallipeddi, R., Janakiraman, R., Kumar, S., and Gupta, S. 2017. "The Effects of Social Media Tone on Engagement: Evidence from Indian General Election 2014," *Information Systems*

Research, Forthcoming.

Manzoor, S. I., and Singla, J. 2019. “Fake News Detection Using Machine Learning Approaches: A Systematic Review,” in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, IEEE, pp. 230–234.

Martel, C., Pennycook, G., and Rand, D. G. 2020. “Reliance on Emotion Promotes Belief in Fake News,” *Cognitive Research: Principles and Implications* (5:1), Springer, pp. 1–20.

McCarthy, M. S., and Mothersbaugh, D. L. 2002. “Effects of Typographic Factors in Advertising-based Persuasion: A General Model and Initial Empirical Tests,” *Psychology & Marketing* (19:7-8), Wiley Online Library, pp. 663–691.

McCornack, S. A. 1988. “The Logic of Lying: A Rational Approach to the Production of Deceptive Messages,” in *Annual Meeting of the Speech Communication Association, New Orleans, LA.*

McCornack, S. A. 1992. “Information Manipulation Theory,” *Communications Monographs* (59:1), Taylor & Francis, pp. 1–16.

McCornack, S. A., Morrison, K., Paik, J. E., Wisner, A. M., and Zhu, X. 2014. “Information Manipulation Theory 2: A Propositional Theory of Deceptive Discourse Production,” *Journal of Language and Social Psychology* (33:4), Sage Publications Sage CA: Los Angeles, CA, pp. 348–377.

Meier, K., Kraus, D., and Michaeler, E. 2018. “Audience Engagement in a Post-Truth Age: What It Means and How to Learn the Activities Connected with It,” *Digital Journalism* (6:8), Taylor & Francis, pp. 1052–1063.

Mesquita, C. T., Oliveira, A., Seixas, F. L., and Paes, A. 2020. “Infodemia, Fake News and Medicine: Science and the Quest for Truth,” *International Journal of Cardiovascular*

Sciences (33:3), SciELO Brasil, pp. 203–205.

Meyer, R. 2018. “The Grim Conclusions of the Largest-Ever Study of Fake News,” *The Atlantic* (8), p. 2018.

Mihalcea, R., and Strapparava, C. 2009. “The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language,” in *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pp. 309–312.

Miles, J. 2014. “Tolerance and Variance Inflation Factor,” *Wiley StatsRef: Statistics Reference Online*, Wiley Online Library.

Miller, M. D., and Levine, T. R. 2019. “Persuasion,” in *An Integrated Approach to Communication Theory and Research*, Routledge, pp. 261–276.

Miller, T., Howe, P., and Sonenberg, L. 2017. “Explainable AI: Beware of Inmates Running the Asylum or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences,” *ArXiv Preprint ArXiv:1712.00547*.

Mills, A. J., and Robson, K. 2019. “Brand Management in the Era of Fake News: Narrative Response as a Strategy to Insulate Brand Value,” *Journal of Product & Brand Management*, Emerald Publishing Limited.

Mohammad, S. M., Sobhani, P., and Kiritchenko, S. 2017. “Stance and Sentiment in Tweets,” *ACM Transactions on Internet Technology (TOIT)* (17:3), ACM New York, NY, USA, pp. 1–23.

Mohammad, S., and Turney, P. 2010. “Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon,” in *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pp. 26–34.

Monti, F., Frasca, F., Eynard, D., Mannion, D., and Bronstein, M. M. 2019. "Fake News Detection on Social Media Using Geometric Deep Learning," *ArXiv Preprint ArXiv:1902.06673*.

Mukherjee, S., and Bala, P. K. 2017. "Detecting Sarcasm in Customer Tweets: An NLP Based Approach," *Industrial Management & Data Systems*, Emerald Publishing Limited.

Naldi, M. 2019. "A Review of Sentiment Computation Methods with R Packages," *ArXiv Preprint ArXiv:1901.08319*.

Naveed, N., Gottron, T., Kunegis, J., and Alhadi, A. C. 2011. "Bad News Travel Fast: A Content-Based Analysis of Interestingness on Twitter," in *Proceedings of the 3rd International Web Science Conference*, pp. 1–7.

Newman, D. S., Guiney, M. C., and Barrett, C. A. 2017. "Language in Consultation: The Effect of Affect and Verb Tense," *Psychology in the Schools* (54:6), Wiley Online Library, pp. 624–639.

Nielsen, F. Å. 2011. "A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs," *ArXiv Preprint ArXiv:1103.2903*.

Nimmo, B. 2016. "Identifying Disinformation: An ABC," *Institute for European Studies* (2016/1).

Osatuyi, B., and Hughes, J. 2018. "A Tale of Two Internet News Platforms-Real vs. Fake: An Elaboration Likelihood Model Perspective," in *Proceedings of the 51st Hawaii International Conference on System Sciences*.

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., and Petersen, M. B. 2020. *Partisan Polarization Is the Primary Psychological Motivation behind "Fake News" Sharing on Twitter*, PsyArXiv.

Pan, J. Z., Pavlova, S., Li, C., Li, N., Li, Y., and Liu, J. 2018. “Content Based Fake News Detection Using Knowledge Graphs,” in *International Semantic Web Conference*, Springer, pp. 669–683.

Parikh, S. B., Patil, V., and Atrey, P. K. 2019. “On the Origin, Proliferation and Tone of Fake News,” in *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, IEEE, pp. 135–140.

Paschen, J. 2019. “Investigating the Emotional Appeal of Fake News Using Artificial Intelligence and Human Contributions,” *Journal of Product & Brand Management*, Emerald Publishing Limited.

Pasławska, P., and Popielska-Borys, A. 2018. “Phenomenon of Fake News,” *Social Communication* (4:s1), Sciendo, pp. 136–140.

Pathak, A., and Srihari, R. K. 2019. “BREAKING! Presenting Fake News Corpus for Automated Fact Checking,” in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pp. 357–362.

Peixoto, F. V. 2019. *I Want to (Dis) Believe: Political Ideology and Misinformation in the Marketplace*.

Peleckis, K., and Peleckiene, V. 2015. “Persuasion in Business Negotiations: Strategic Orientations and Rhetorical Argumentation,” *Universal Journal of Management* (3:10), pp. 413–422.

Pennebaker, J. W., Boyd, R. L., Jordan, K., and Blackburn, K. 2015. “The Development and Psychometric Properties of LIWC2015.”

Pennycook, G., Cannon, T. D., and Rand, D. G. 2018. “Prior Exposure Increases Perceived Accuracy of Fake News.,” *Journal of Experimental Psychology: General* (147:12), American Psychological Association, p. 1865.

Peters, K., Kashima, Y., and Clark, A. 2009. “Talking about Others: Emotionality and the Dissemination of Social Information,” *European Journal of Social Psychology* (39:2), Wiley Online Library, pp. 207–222.

Petty, R. E., Fabrigar, L. R., and Wegener, D. T. 2003. *Emotional Factors in Attitudes and Persuasion.*, Oxford University Press.

Phua, J., Jin, S. V., and Kim, J. J. 2017. “Gratifications of Using Facebook, Twitter, Instagram, or Snapchat to Follow Brands: The Moderating Effect of Social Comparison, Trust, Tie Strength, and Network Homophily on Brand Identification, Brand Engagement, Brand Commitment, and Membership Intentio,” *Telematics and Informatics* (34:1), Elsevier, pp. 412–424.

Pomerantsev, P., and Weiss, M. 2014. *The Menace of Unreality: How the Kremlin Weaponizes Information, Culture and Money*, (Vol. 14), Institute of Modern Russia New York.

Poon, D. C. H., and Leung, L. 2013. “Effects of Narcissism, Leisure Boredom, and Gratifications Sought on User-Generated Content among Net-Generation Users,” in *Evolving Psychological and Educational Perspectives on Cyber Behavior*, IGI Global, pp. 49–63.

Porter, S., and Yuille, J. C. 1996. “The Language of Deceit: An Investigation of the Verbal Clues to Deception in the Interrogation Context,” *Law and Human Behavior* (20:4), Springer, pp. 443–458.

Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., and Stein, B. 2017. “A Stylometric Inquiry into Hyperpartisan and Fake News,” *ArXiv Preprint ArXiv:1702.05638*.

Purcell, K., Rainie, L., Mitchell, A., Rosenstiel, T., and Olmstead, K. 2010. “Understanding the Participatory News Consumer,” *Pew Internet and American Life Project* (1), pp. 19–21.

Purian, R., Ho, S. M., and Te'Eni, D. 2020. *Resilience of Society to Recognize Disinformation: Human and/or Machine Intelligence*.

Qi, P., Cao, J., Yang, T., Guo, J., and Li, J. 2019. "Exploiting Multi-Domain Visual Information for Fake News Detection," in *2019 IEEE International Conference on Data Mining (ICDM)*, IEEE, pp. 518–527.

Quan-Haase, A., and Young, A. L. 2010. "Uses and Gratifications of Social Media: A Comparison of Facebook and Instant Messaging," *Bulletin of Science, Technology & Society* (30:5), SAGE Publications Sage CA: Los Angeles, CA, pp. 350–361.

Rasool, T., Butt, W. H., Shaukat, A., and Akram, M. U. 2019. "Multi-Label Fake News Detection Using Multi-Layered Supervised Learning," in *Proceedings of the 2019 11th International Conference on Computer and Automation Engineering*, pp. 73–77.

Riedel, B., Augenstein, I., Spithourakis, G. P., and Riedel, S. 2017. "A Simple but Tough-to-Beat Baseline for the Fake News Challenge Stance Detection Task," *ArXiv Preprint ArXiv:1707.03264*.

Rimé, B. 2009. "Emotion Elicits the Social Sharing of Emotion: Theory and Empirical Review," *Emotion Review* (1:1), Sage Publications Sage UK: London, England, pp. 60–85.

Rimé, B., Corsini, S., and Herbette, G. 2002. "Emotion, Verbal Expression, and the Social Sharing of Emotion," in *The Verbal Communication of Emotions*, Psychology Press, pp. 193–216.

Rubin, V. L. 2010. "On Deception and Deception Detection: Content Analysis of Computer-Mediated Stated Beliefs," *Proceedings of the American Society for Information Science and Technology* (47:1), John Wiley & Sons, Ltd, pp. 1–10. (https://doi.org/10.1002/meet.14504701124).

Rubin, V. L., and Chen, Y. 2012. "Information Manipulation Classification Theory for LIS and NLP," *Proceedings of the American Society for Information Science and Technology* (49:1), Wiley Online Library, pp. 1–5.

Rubin, V. L., Conroy, N. J., and Chen, Y. 2015. "Towards News Verification: Deception Detection Methods for News Discourse," in *Hawaii International Conference on System Sciences*, pp. 5–8.

Rubin, V. L., and Lukoianova, T. 2015. "Truth and Deception at the Rhetorical Structure Level," *Journal of the Association for Information Science and Technology* (66:5), Wiley Online Library, pp. 905–917.

Rubin, V., and Lukoianova, T. 2013. "Veracity Roadmap: Is Big Data Objective, Truthful and Credible?," *Advances in Classification Research Online* (24:1), p. 4.

Russell, J. A. 2003. "Core Affect and the Psychological Construction of Emotion.," *Psychological Review* (110:1), American Psychological Association, p. 145.

Sabeeh, V., Zohdy, M., Mollah, A., and Al Bashaireh, R. 2020. "Fake News Detection on Social Media Using Deep Learning and Semantic Knowledge Sources," *International Journal of Computer Science and Information Security (IJCSIS)* (18:2).

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. 2008. "The Graph Neural Network Model," *IEEE Transactions on Neural Networks* (20:1), IEEE, pp. 61–80.

Serrano, C., and Karahanna, E. 2016. "The Compensatory Interaction between User Capabilities and Technology Capabilities in Influencing Task Performance: An Empirical Assessment in Telemedicine Consultations," *Mis Quarterly* (40:3), Society for Information Management and The Management Information Systems …, pp. 597–621.

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., and Menczer, F. 2018. "The Spread of Low-Credibility Content by Social Bots," *Nature Communications* (9:1), Nature Publishing Group, pp. 1–9.

Shao, G. 2009a. "Understanding the Appeal of User-Generated Media: A Uses and Gratification Perspective," *Internet Research*, pp. 7–25. (https://doi.org/10.1108/10662240910927795).

Shao, G. 2009b. "Understanding the Appeal of User-generated Media: A Uses and Gratification Perspective," *Internet Research*, Emerald Group Publishing Limited.

Shi, D., Lee, T., and Maydeu-Olivares, A. 2019. "Understanding the Model Size Effect on SEM Fit Indices," *Educational and Psychological Measurement* (79:2), Sage Publications Sage CA: Los Angeles, CA, pp. 310–334.

Shiau, W.-L., Shi, P., and Yuan, Y. 2021. "A Meta-Analysis of Emotion and Cognition in Information System," *International Journal of Enterprise Information Systems (IJEIS)* (17:1), IGI Global, pp. 125–143.

Shrestha, A., and Spezzano, F. 2021. "Textual Characteristics of News Title and Body to Detect Fake News: A Reproducibility Study," in *European Conference on Information Retrieval*, Springer, pp. 120–133.

Shu, K., Bernard, H. R., and Liu, H. 2019. "Studying Fake News via Network Analysis: Detection and Mitigation," in *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*, Springer, pp. 43–65.

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., and Liu, H. 2020. "Fakenewsnet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media," *Big Data* (8:3), Mary Ann Liebert, Inc., publishers 140 Huguenot Street, 3rd Floor New …, pp. 171–188.

Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H. 2017. “Fake News Detection on Social Media: A Data Mining Perspective,” *ACM SIGKDD Explorations Newsletter* (19:1), ACM New York, NY, USA, pp. 22–36.

Shu, K., Wang, S., Lee, D., and Liu, H. 2020. “Mining Disinformation and Fake News: Concepts, Methods, and Recent Advancements,” in *Disinformation, Misinformation, and Fake News in Social Media*, Springer, pp. 1–19.

Shu, K., Wang, S., and Liu, H. 2017. “Exploiting Tri-Relationship for Fake News Detection,” *ArXiv Preprint ArXiv:1712.07709* (8).

Shu, K., Wang, S., and Liu, H. 2018. “Understanding User Profiles on Social Media for Fake News Detection,” in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, IEEE, pp. 430–435.

Shu, K., Wang, S., and Liu, H. 2019. “Beyond News Contents: The Role of Social Context for Fake News Detection,” in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pp. 312–320.

Solís Arce, J. S., Warren, S. S., Meriggi, N. F., Scacco, A., McMurry, N., Voors, M., Syunyaev, G., Malik, A. A., Aboutajdine, S., and Adeojo, O. 2021. “COVID-19 Vaccine Acceptance and Hesitancy in Low-and Middle-Income Countries,” *Nature Medicine* (27:8), Nature Publishing Group, pp. 1385–1394.

Solomon, D. H., Bucala, R., Kaplan, M. J., and Nigrovic, P. A. 2020. *The “Infodemic” of COVID-19*, Wiley Online Library.

Southwell, B. G., Niederdeppe, J., Cappella, J. N., Gaysynsky, A., Kelley, D. E., Oh, A., Peterson, E. B., and Chou, W.-Y. S. 2019. “Misinformation as a Misunderstood Challenge to Public Health,” *American Journal of Preventive Medicine* (57:2), Elsevier, pp. 282–285.

Stahl, K. 2018. “Fake News Detection in Social Media,” *California State University Stanislaus* (6), pp. 4–15.

Tambuscio, M., Ruffo, G., Flammini, A., and Menczer, F. 2015. “Fact-Checking Effect on Viral Hoaxes: A Model of Misinformation Spread in Social Networks,” in *Proceedings of the 24th International Conference on World Wide Web*, pp. 977–982.

Tandoc Jr, E. C., Lim, Z. W., and Ling, R. 2018. “Defining ‘Fake News’ A Typology of Scholarly Definitions,” *Digital Journalism* (6:2), Taylor & Francis, pp. 137–153.

Tang, J., Chang, Y., and Liu, H. 2014. “Mining Social Media with Social Theories: A Survey,” *ACM Sigkdd Explorations Newsletter* (15:2), ACM New York, NY, USA, pp. 20–29.

Tangcharoensathien, V., Calleja, N., Nguyen, T., Purnat, T., D’Agostino, M., Garcia-Saiso, S., Landry, M., Rashidian, A., Hamilton, C., and AbdAllah, A. 2020. “Framework for Managing the COVID-19 Infodemic: Methods and Results of an Online, Crowdsourced WHO Technical Consultation,” *Journal of Medical Internet Research* (22:6), JMIR Publications Inc., Toronto, Canada, p. e19659.

Tausczik, Y. R., and Pennebaker, J. W. 2010. “The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods,” *Journal of Language and Social Psychology* (29:1), Sage Publications Sage CA: Los Angeles, CA, pp. 24–54.

Thomas, J. 1997. “Conversational Maxims,” *LAMARQUE, PV/ASHER, RE (1997)(Hrsgg.): Concise Encyclopedia of Philosophy of Language. New York: Pergamon*, pp. 388–393.

Thompson, N., Wang, X., and Daya, P. 2019. “Determinants of News Sharing Behavior on Social Media,” *Journal of Computer Information Systems*, Taylor & Francis.

Thorne, J., and Vlachos, A. 2018. “Automated Fact Checking: Task Formulations, Methods and Future Directions,” *ArXiv Preprint ArXiv:1806.07687*.

Torabi Asr, F., and Taboada, M. 2019. “Big Data and Quality Data for Fake News and Misinformation Detection,” *Big Data & Society* (6:1), SAGE Publications Sage UK: London, England, p. 2053951719843310.

Treen, K. M. d’I, Williams, H. T. P., and O’Neill, S. J. 2020. “Online Misinformation about Climate Change,” *Wiley Interdisciplinary Reviews: Climate Change* (11:5), Wiley Online Library, p. e665.

Trivedi, N., Lowry, M., Gaysynsky, A., and Chou, W.-Y. S. 2021. “Factors Associated with Cancer Message Believability: A Mixed Methods Study on Simulated Facebook Posts,” *Journal of Cancer Education*, Springer, pp. 1–9.

Twitchell, D. P., Nunamaker, J. F., and Burgoon, J. K. 2004. “Using Speech Act Profiling for Deception Detection,” *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (3073), Springer Verlag, pp. 403–410.

Valdivia, A., Luzón, M. V., and Herrera, F. 2017. “Sentiment Analysis in Tripadvisor,” *IEEE Intelligent Systems* (32:4), IEEE, pp. 72–77.

Valenzuela, S., Piña, M., and Ramírez, J. 2017. “Behavioral Effects of Framing on Social Media Users: How Conflict, Economic, Human Interest, and Morality Frames Drive News Sharing,” *Journal of Communication* (67:5), Oxford University Press, pp. 803–826.

Veil, S., Reynolds, B., Sellnow, T. L., and Seeger, M. W. 2008. “CERC as a Theoretical Framework for Research and Practice,” *Health Promotion Practice* (9:4_suppl), Sage Publications Sage CA: Los Angeles, CA, pp. 26S-34S.

Vishwanath, A. 2015. “Diffusion of Deception in Social Media: Social Contagion Effects and Its Antecedents,” *Information Systems Frontiers* (17:6), Springer, pp. 1353–1367.

Vlachos, A., and Riedel, S. 2014. “Fact Checking: Task Definition and Dataset Construction,” in *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pp. 18–22.

Vlasceanu, M., Goebel, J., and Coman, A. 2020. “The Emotion-Induced Belief-Amplification Effect.,” in *CogSci*.

Vosoughi, S., Roy, D., and Aral, S. 2018. “The Spread of True and False News Online,” *Science* (359:6380), American Association for the Advancement of Science, pp. 1146–1151.

Vysotska, O., and Vysotska, S. (n.d.). *Information Evaluation: Teaching Students to Detect Bias, Fake, and Manipulation Ukrainian Perspective*.

Wall, H. J., and Kaye, L. K. 2018. “Online Decision Making: Online Influence and Implications for Cyber Security,” in *Psychological and Behavioral Examinations in Cyber Security*, IGI Global, pp. 1–25.

Walton, D. 2007. *Media Argumentation: Dialectic, Persuasion and Rhetoric*, Cambridge University Press.

Wang, C., Tan, Z. X., Ye, Y., Wang, L., Cheong, K. H., and Xie, N. 2017. “A Rumor Spreading Model Based on Information Entropy,” *Scientific Reports* (7:1), Nature Publishing Group, pp. 1–14.

Wang, G., and Song, J. 2017. “The Relation of Perceived Benefits and Organizational Supports to User Satisfaction with Building Information Model (BIM),” *Computers in Human Behavior* (68), Elsevier, pp. 493–500.

Wang, Q., Lin, Z., Jin, Y., Cheng, S., and Yang, T. 2015. “ESIS: Emotion-Based Spreader–Ignorant–Stifler Model for Information Diffusion,” *Knowledge-Based Systems* (81), Elsevier, pp. 46–55.

Wang, R., He, Y., Xu, J., and Zhang, H. 2020. “Fake News or Bad News? Toward an Emotion-Driven Cognitive Dissonance Model of Misinformation Diffusion,” *Asian Journal of Communication* (30:5), Taylor & Francis, pp. 317–342.

Wang, W. Y. 2017. “‘ Liar, Liar Pants on Fire’: A New Benchmark Dataset for Fake News Detection,” *ArXiv Preprint ArXiv:1705.00648*.

Wang, X. R., Nan, X., and Stanley, S. J. 2021. “Emotion and Virality of Food Safety Risk Communication Messages on Social Media.,” *Journal of Applied Communications* (105:3), Agricultural Communicators in Education, pp. 1c-1c.

Wardle, C., and Derakhshan, H. 2017. “Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making,” *Council of Europe Report* (27).

Webber, S., and Johnston, B. 2000. “Conceptions of Information Literacy: New Perspectives and Implications,” *Journal of Information Science* (26:6), Sage Publications Sage CA: Thousand Oaks, CA, pp. 381–397.

West, R. L., Turner, L. H., and Zhao, G. 2010. *Introducing Communication Theory: Analysis and Application*, (Vol. 2), McGraw-Hill New York, NY.

Wu, K., Yang, S., and Zhu, K. Q. 2015. “False Rumors Detection on Sina Weibo by Propagation Structures,” in *2015 IEEE 31st International Conference on Data Engineering*, IEEE, pp. 651–662.

Wu, L., Morstatter, F., Hu, X., and Liu, H. 2016. “Mining Misinformation in Social Media,” *Big Data in Complex and Social Networks*, CRC Press, pp. 123–152.

Xiong, H., and Lv, S. 2021. “Factors Affecting Social Media Users’ Emotions Regarding Food Safety Issues: Content Analysis of a Debate among Chinese Weibo Users on Genetically Modified Food Security,” in *Healthcare* (Vol. 9), Multidisciplinary Digital Publishing

Institute, p. 113.

Xu, B. 2019. *Understanding Sticky News: Analyzing the Effect of Content Appeal and Social Engagement for Sharing Political News Online*, University of Maryland, College Park.

Yang, X., Li, Y., and Lyu, S. 2019. "Exposing Deep Fakes Using Inconsistent Head Poses," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 8261–8265.

Zafarani, R., Zhou, X., Shu, K., and Liu, H. 2019. "Fake News Research: Theories, Detection Strategies, and Open Problems," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 3207–3208.

Zhang, X., and Wang, G. 2021. "Application of Information Diffusion of Negative Emotion and Binomial Regression Method on Retweet Cascades during Covid-19," in *2021 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, IEEE, pp. 799–802.

Zhou, X., Jain, A., Phoha, V. V, and Zafarani, R. 2020. "Fake News Early Detection: A Theory-Driven Model," *Digital Threats: Research and Practice* (1:2), ACM New York, NY, USA, pp. 1–25.

Zhou, X., and Zafarani, R. 2019. "Fake News Detection: An Interdisciplinary Research," in *Companion Proceedings of The 2019 World Wide Web Conference*, p. 1292.

Zhou, X., and Zafarani, R. 2020. "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," *ACM Computing Surveys (CSUR)* (53:5), ACM New York, NY, USA, pp. 1–40.

Zimdars, M., and McLeod, K. 2020. *Fake News: Understanding Media and Misinformation in the Digital Age*, MIT Press.

Zlatkova, D., Nakov, P., and Koychev, I. 2019. “Fact-Checking Meets Fauxtography: Verifying Claims about Images,” *ArXiv Preprint ArXiv:1908.11722*.