Communications of the Association for Information Systems

Volume 53 Paper in press

2023

Combating Fake News on Social Media: A Framework, Review, and Future Opportunities

Mona Nasery
DeGroote School of Business McMaster University, naserym@mcmaster.ca

Ofir Turel
School of Computing and Information Systems University of Melbourne

Yufei Yuan
DeGroote School of Business McMaster University

Communications of the Association for Information Systems

Accepted Manuscript

Combating Fake News on Social Media: A Framework, Review, and Future Opportunities

Mona Nasery
DeGroote School of Business
McMaster University
Hamilton, Canada
naserym@mcmaster.ca

Ofir Turel
School of Computing and Information Systems
University of Melbourne
Melbourne, Australia

Yufei Yuan
DeGroote School of Business
McMaster University
Hamilton, Canada

Please cite this article as: Nasery, M., Turel, O., & Yuan, Y. (in press). Combating fake news on social media: A framework, review, and future opportunities. *Communications of the Association for Information Systems.*

Communications of the Association for Information Systems

Research Paper ISSN: 1529-3181

Combating Fake News on Social Media: A Framework, Review, and Future Opportunities

Mona Nasery
DeGroote School of Business
McMaster University
Hamilton, Canada
naserym@mcmaster.ca

Ofir Turel
School of Computing and Information Systems
University of Melbourne
Melbourne, Australia

Yufei Yuan
DeGroote School of Business
McMaster University
Hamilton, Canada

Abstract:

Social media platforms facilitate the sharing of a vast magnitude of information in split seconds among users. However, some false information is also widely spread, generally referred to as “fake news”. This can have major negative impacts on individuals and societies. Unfortunately, people are often not able to correctly identify fake news from truth. Therefore, there is an urgent need to find effective mechanisms to fight fake news on social media. To this end, this paper adapts the Straub Model of Security Action Cycle to the context of combating fake news on social media. It uses the adapted framework to classify the vast literature on fake news to action cycle phases (i.e., deterrence, prevention, detection, and mitigation/remedy). Based on a systematic and inter-disciplinary review of the relevant literature, we analyze the status and challenges in each stage of combating fake news, followed by introducing future research directions. These efforts allow the development of a holistic view of the research frontier on fighting fake news online. We conclude that this is a multidisciplinary issue; and as such, a collaborative effort from different fields is needed to effectively address this problem.

Keywords: Fake News, Social Media, Deterrence, Prevention, Detection, Mitigation.

[Department statements, if appropriate, will be added by the editors. Teaching cases and panel reports will have a statement, which is also added by the editors.]

[Note: this page has no footnotes.]

This manuscript underwent [editorial/peer] review. It was received xx/xx/20xx and was with the authors for XX months for XX revisions. [firstname lastname] served as Associate Editor.] **or** The Associate Editor chose to remain anonymous.]

1 Introduction

Social media platforms such as Twitter and Facebook provide an easier, cheaper, and faster way for individuals to consume and share news. About two-thirds of U.S. adults (68%) got news on social media in 2018 [1]. However, these benefits come at a cost, namely a large volume of fake news on social media platforms. Fake news are news items that are false, regardless of the intentions of the news originator (Zhou & Zafarani, 2020). As such, they include misinformation (false or misleading information with no intention to deceive) and disinformation (false information that is purposely spread to deceive people) (Lazer et al., 2018).

The spread of fake news on social media can have severe negative impacts on individuals and societies. For example, in the context of Covid-19, fake news about ingesting fish-tank cleaning products, alcohol, or injecting bleach to treat the virus can pose a serious threat to people's lives. The harmful impacts of fake news have been shown in other various contexts such as politics (Allcott & Gentzkow, 2017), economy (Kogan et al., 2019), and people responses to natural disasters (Gupta et al., 2013). Thus, there is an acute need for effective mechanisms to stop or limit the harmful consequences of fake news. Indeed, giant tech companies such as Google, Microsoft, Facebook, and Twitter issued a joint statement to combat fake news about Covid-19[2]. In response, scholars have proposed numerous approaches to combat fake news. However, such approaches, as we discuss later, primarily focused on one action, namely detection, and overlooked tackling the problem through different stages of fake news dissemination.

Several barriers for combating fake news online exist. First, fake news on social media spread faster, farther, and deeper than true news (Vosoughi et al., 2018). That is, fake news can spread exponentially fast at early stages and pose harmful impacts in a very short time. For example, a false tweet about Barack Obama being injured in a White House explosion, although debunked quickly, wiped out $130 billion in stock market in a matter of seconds. Second, in many cases it is difficult to identify whether the news is fake or not. Manual fact-checking and debunking fake news cannot keep up with the large volume and fast spread of fake news on social media. To address this, a large body of research focused on automated fake news detection. However, regardless of the type of the algorithm for fake news detection (text-based, propagation-based, etc.), they are still not very effective.

Thus, it is important to devise strategies to stop fake news not only after its spread, but also before its spread and even before its creation. Here, we aim to examine this broad landscape by focusing on all lifecycle stages of fake news dissemination. We specifically seek to provide a comprehensive picture of combating fake news on social media. This holistic view affords considering synergies among approaches and more careful and hopefully effective plans to tackle the problem. To this end, we adapt the Straub Model of Security Action Cycle to the context of combating fake news on social media. This model comprises four steps (countermeasures) to address security threats: deterrence, prevention, detection, and mitigation/remedy. Notably, the Straub Model of Security is rooted in criminology and is hence not limited to the security context. It can be applied to any undesirable behavior. Since creating or spreading fake news on social media is an undesirable behavior with destructive impacts on individuals and societies, we propose similar steps to combat fake news on social media. Based on a thorough investigation of the relevant literature, we use this model to classify the vast literature on fake news. We believe that this framework helps readers to grasp the whole picture of the research frontier.

We note that in recent years there have been several attempts to review the literature on fake news from different perspectives. Table 10 in the appendix B summarizes the various literature review papers on fake news, their combat stage, classification criteria, and the type of false information addressed in their review. Based on this review, we conclude that most existing reviews focus on fake news detection approaches (Bondielli & Marcelloni, 2019; Sharma et al., 2019; K. Shu et al., 2017; K. Shu, Bernard, et al., 2019; Zhang & Ghorbani, 2020; Zhou & Zafarani, 2020; Zubiaga et al., 2018). Thus, there is a need to consider and review also other approaches, such as deterrence and to conduct a systematic and multidisciplinary review on the full lifecycle of combating fake news on social media.

In doing so, we make the following contributions. First, we provide an overview of fake news definitions and several related terms in the literature. This will help scholars to have a better understanding of the

[1] https://www.pewresearch.org/
[2] https://twitter.com/microsoft/status/1239703041109942272?lang=en

term “fake news” and other relevant terms that are often used interchangeably in the literature. Second, we provide an inter-disciplinary approach to combat fake news on social media. To this end, we suggest deterrence, prevention, detection, and mitigation/remedy as action areas for reducing the dissemination of fake news. Third, we conduct a comprehensive and systematic review of 164 articles related to the four countermeasures of the framework. We also provide some descriptive statistics of the reviewed papers. We hope this analysis can better depict the current status of fake news combating research. Finally, we use the adapted framework to discuss the approaches to combat fake news on social media, challenges involved, limitations of the current approaches, and directions for future research. These should allow the IS community to take a more systematic and active role in combating fake news, and not just in fake news detection.

The organization of this paper is as follows: In section 2, we review different definitions of the term “fake news” and related terms. Section 3 describes the Straub Model of Security Action cycle and proposes similar steps to combat fake news on social media. We also explain the rationale behind applying this framework to the context of fake news and compare it to the security context from several perspectives. In section 4, we describe the methodology for our review process along with some descriptive statistics about the reviewed articles. Section 5 provides a review of the approaches to combat fake news on social media, identify several research gaps and future opportunities in each stage of the fake news combat cycle. We provide further discussion on the limitations of this research and possible future research directions in section 6. Conclusions are provided in section 7.

2 Overview of Fake News and Related Terms

2.1 Fake News Definitions

The term “fake news” has gained widespread attention mainly after the 2016 U.S. Presidential Campaign. There has been no overall agreement on the definition of fake news. This is because the term "fake news" covers a wide range of (with or without intention) false or inaccurate information such as deceptive stories, rumors, satires, and conspiracy theories. Therefore, in this section we aim to provide an overview of the ways that the term “fake news” has been used and defined in the literature.

Allcott & Gentzkow (2017) define fake news as *“a news article that is intentionally false and is verifiable”.* Several other studies (e.g., Bondielli & Marcelloni, 2019; Kim et al., 2019; Shu et al., 2017) adapted this definition. This is, however, a narrow definition of fake news, which emphasize on both authenticity and intention of the information. There are also broader definitions of fake news, which do not restrict the intention of the information/news. For example, Zhou & Zafarani, (2020) broadly defined fake news as false news. Table 1 shows different definitions of fake news. In this paper, we purposefully adopt the broad definition of fake news provided by Sharma et al., (2019): “*a news article or message published and propagated through media, carrying false information regardless the means and motives behind it*”. The broad definition of fake news allows us to cover different types of fake news and related terms, such as rumors, misleading news, and conspiracies.

Table 1. Various Definitions of Fake News in the Literature

Fake News Definition	Reference(s)
Fake news is false news (broad definition)	
A news article or message published and propagated through media, carrying false information regardless of the means and motives behind it (broad definition).	(Sharma et al., 2019)
News article that is intentionally and verifiably false (narrow definition)	(Allcott & Gentzkow, 2017), (Bondielli & Marcelloni, 2019), (A. Kim et al., 2019a)
Fabricated information that mimics news media content in form but not in organizational process or intent	(Lazer, et al., 2018)
False stories disguised as a credible news source for political or financial gain	(Shin et al., 2018), (Silverman, 2017)
Information presented as a news story that is factually incorrect and designed to deceive the consumer into believing it is true	(Golbeck et al., 2018)

2.2 Different Types of False Information

There are several terms and concepts linked to fake news that have been frequently used in the literature. For example, Tandoc Jr et al., (2018) identified six ways that the term “fake news” has been used in the literature: satire, parody, fabrication, manipulation, propaganda, and advertising. A good distinction between fake news and different terms related to fake news is provided by (Zhou & Zafarani, 2020) which is based on three characteristics: *intention to deceive* or mislead others, *authenticity* (whether it includes non-factual information), and *whether the information is news or not.* For example, based on intention, false information can be divided into two broad categories of **misinformation** and **disinformation**. Misinformation refers to “inadvertent sharing of false information” (there is no intention). Disinformation, on the other hand, refers to the “deliberate creation and sharing of false information” (S. Kumar & Shah, 2018; Wardle, 2017). Rubin et al., (2015) identified three types of fake news as: serious fabrications (tabloids and yellow journalism), large-scale hoaxes (deliberate falsification causing harm), and humorous fakes (satire and parody). Table 2 presents different types of fake news and the associated definitions. It differentiates between different types of fake news based on two main dimensions: (1) the authenticity or facticity of the news stories (does it rely on facts? Is it based on factual or non-factual statement?), and (2) intention to deceive or mislead readers/users.

Table 2. Different Types of False Info. with Definitions and Classification Based on Truthfulness and Intention

Truthfulness	Intention	Relevant Terms & Definitions
False	*Malicious*	**Disinformation**: False information with the intention to deceive (S. Kumar & Shah, 2018; Wardle, 2017)
		Hoax: Reports of false information disguised as proper news (Bondielli & Marcelloni, 2019; Rubin et al., 2015). A false story used to masquerade the truth, originating from the verb hocus, meaning "to cheat" (Nares 1822)
		Serious Fabrication: Prototypical form of fake news, i.e. articles with a malicious intent that often become viral through social media (Bondielli & Marcelloni, 2019; Rubin et al., 2015)
		Propaganda: News stories which are created by a political entity to influence public perceptions (Tandoc et al., 2018)
	No Malicious intention	**Misinformation**: False or misleading information without the intention to deceive (S. Kumar & Shah, 2018; Wardle, 2017)
		Parody: Use of non-factual and fabricated content to inject humor (Tandoc et al., 2018)
		Satire[3]: News stories that are factually incorrect, but the intent is not to deceive but rather to call out, ridicule, or expose behavior that is shameful, corrupt, or otherwise "bad (Golbeck et al., 2018). Mock news programs, which typically use humor or exaggeration to present audiences with news updates (Tandoc et al., 2018)
True	*Malicious*	**Misleading Content**: Misleading use of information to frame an issue (Sharma et al., 2019)
		Irrelevant Context: Using true information in an unrelated context to mislead people
	No Malicious intention	**Real News or True Information**: genuine news or true information based on facts.
The True or False is unknown	*Malicious*	**Conspiracy Theory (CT)**: A proposed explanation of some events in terms of the significant causal agency of a relatively small group of persons acting in secret (Keeley, 1999). Causal narratives of an event as a covert plan orchestrated by a secret cabal of people (or organizations) instead of a random or natural happening (Banas & Miller, 2013; Douglas & Sutton, 2008)
		Clickbait: Use of misleading headlines to entice readers to click on links under false pretenses (Ireton & Posetti, 2018). Article titles or social media posts whose aim is to attract readers to follow a link to the actual article page (Bondielli & Marcelloni, 2019; Y. Chen et al., 2015)
	No Malicious intention	**Rumor*:*** Stories whose truthfulness is ambiguous or never confirmed (Zannettou et al., 2019). Circulating story of questionable veracity, which is apparently credible but hard to verify, and produces sufficient skepticism and/or anxiety (Zubiaga et al., 2018)

[3] The truthfulness (facticity) of satire depends on which definition we adopt. For example, Tandoc et al., (2018) considered satire as facts and stated that "their being fake only refers to their format", while Golbeck et al., (2018) considered satires as "factually incorrect" stories. In this paper, we adapted the latter.

3 A Framework to Combat Fake News on Social Media

In this section, we describe a framework to combat fake news on social media. The adopted framework (shown in Figure 1) is inspired by the Straub Model of Security Action Cycle (Straub & Welke, 1998). According to the model, the first step to address the system risks is to use "deterrents" such as administrative policies or employee training. Deterrents are passive countermeasures to discourage individuals from engaging in illicit behavior or committing a crime. Deterrence is applicable in the stage where the adversaries have intention but have not yet taken any action to launch security attacks. If deterrents fail, the next step is to use "preventives". These are active countermeasures to impede or stop individuals from criminal activities or illegal behavior. This means that prevention may happen when an abuser has taken an action, but the system will stop them. If an abuser overcomes the first two stages and engaged in the undesirable behavior, then detection approaches should be used. Detection refers to the process of monitoring and identifying the undesirable behavior. Finally, an effective IS system should be able to mitigate or remedy the destructive impacts of undesirable behavior. Remedy refers to the post-attack process or activities that reduce the negative impacts of undesirable behavior.

In this paper, we apply the Straub model of Security Action Cycle to the context of fake news on social media and propose similar steps to combat fake news on social media platforms. The rationale behind this is twofold. First, similar to Information security threats that have negative impact on individuals, organizations, and society, creating or spreading fake news is also an undesirable phenomenon, which can negatively affect many different entities such as individuals, organizations, political parties, and financial markets. Research shows the destructive and far-reaching impacts of fake news on many aspects of our lives, including but not limited to politics (Allcott & Gentzkow, 2017), businesses (Bakir & McStay, 2018; Petratos, 2021), healthcare (Carrieri et al., 2019), or people's responses to natural disasters such as Hurricane Sandy (Gupta et al., 2013). Thus, both security and fake news represent undesirable behaviors that can be deterred, prevented detected and remedied. Second, fake news can sometimes (and certainly not always) represent a security threat, which makes the application of models from the security domain to fake news (Botha & Pieterse, 2020). In some cases, the alluring nature of click bytes can be used for spreading malicious software (Zeng et al., 2020).

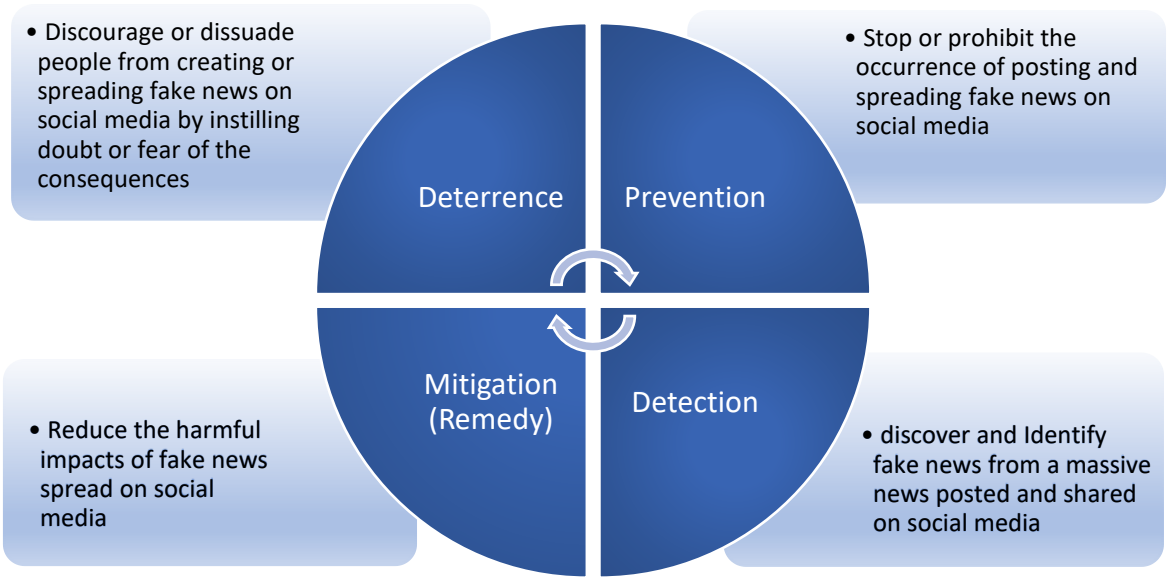

Figure 1. A Framework to Combat Fake News on Social Media (Stages and Definitions)

Importantly, fake news and security issues do not always have the same attributes. We therefore outline the similarities and differences between fake news and information security threats in Table 3 and Table 4. The tables demonstrate nuanced differences between security issues and fake news, but also point to key similarities, namely in the undesirability of the behavior, the problems it causes, and the potency of deterrence, prevention, detection, and remedy to reduce the behavior or its adverse outcomes. Given

such similarities, and the possibility to apply the stages in Table 3 to fake news, we view the application of the Straub model to fighting fake news as reasonable.

Table 3. Example Application of the Framework in Security and Fake News Contexts.

Combat Stage	Description	Examples in Security	Examples in Fake News
Deterrence	The first step to cope with system risks (in case of this research, to combat fake news) is to use deterrents. Deterrents are passive countermeasures to discourage individuals from engaging in illicit behavior or committing a crime. Deterrents are passive in that in that they have no inherent provision for enforcement and depend on the willingness of users (Straub & Welke, 1998).	• Policies and guidelines for proper system use • Educate users (e.g., Security awareness programs) about the risks and threats in organizational environment and to emphasize the certainty and severity of sanctions for violation	• Establish laws, policies, and regulations by government, authorities, and social media platforms. • Educating users and increase their awareness about fake news and its destructive impacts. • Information literacy, media literacy, and other training programs
Prevention	Preventives are “active countermeasures with inherent capabilities to enforce policy and ward off illegitimate use” (Gopal & Sanders, 1997; Straub & Welke, 1998).	• Locks on computers • Password access control	• Block or suspend malicious accounts. • Block or remove known fake content
Detection	If deterrents and preventives don’t work and the abuser penetrate the system (in our case, when fake news is already published and disseminated), the next step is to identify and detect misuse (in our case detecting fake news)	• System Audits to monitor computer use activities. • Transaction log reports • Virus scanning	• Fact-checking (Manual, Crowd-sourced, Automated) • Algorithmic Solutions (Machine learning, and other approaches)
Mitigation (Remedy)	The last stage is to mitigate or reduce the harmful effects of abuse (in our case, reducing the negative impacts of fake news)	• Software recovery • Prosecution of perpetrators • Legal actions such as criminal and civil suits	• Minimize the spread of fake news by blocking certain nodes in the network (e.g., influential nodes) • Spreading true information • Platform interventions (account-level, and content-level) to stop or limit the spread of fake news

Table 4. Comparison of Fake News and Security Context

	Fake News	Security Attacks
Creators (Who)	• Bots • Malicious/fake accounts • Politicians, or governments, etc. • Normal people	• Hackers • Corporate spies • Terrorist groups • People with security knowledge (in contrast to fake news that can be propagated by any individual, security attacks can be done only by people who have relevant knowledge)
Motives (Why)	• Monetary motives (e.g., increase revenue or web traffic in case of clickbait), • Ideological motives, • Political motives (e.g., during elections)	• Financial/Monetary motives • Access data • Political motives (Hacktivism)
Intention	Anyone with or without malicious intent may spread fake news (e.g., many individuals may share fake news and misinformation without knowing it is false)	Often with malicious intent (however, sometimes security threats can occur because of carelessness, or compromised credentials
Why (why people fall for it)	• Ideological beliefs, Confirmation Bias, Naïve realism (people tend to believe they have the "true" perception of reality and those who disagree with them must be uninformed, irrational, or biased), • Social Normative Theory (the influence of other people that leads us to conform in order to be liked and accepted by them), • Intuitive or emotional response and lack of analytical thinking (Dual Process Theory), • Familiarity with the topic, • Social validation Echo-chambers (because of personalized contents, people are primarily exposed to contents that agree with their beliefs), etc.	• Lack of enough security measures (e.g., weakness in security policies) • System weaknesses (e.g., weakness in computer technologies such as network protocols (TCP/IP) or operating systems' weaknesses) • Individuals' sloppiness or negligence • Lack of knowledge
Where	Social media, messaging apps, peer-to-peer, …	Organizations, firms
Targets (Who) & Impacts	• **Individuals** (increase panic, distrust, conflict, radicalization/extremism), • **Societies** (echo-chambers, polarization, voting patterns), • **Organizations** (impact on the relationship between companies and consumers, destroy brand reputation).	• Often on **organizations** (e.g., economic loss, loss of customer and stakeholder trust, destroy brand reputation) • **Societies** (e.g., shortage of products or services, panic buying, etc.) • Sometimes **individuals** are targets too (e.g., because of weak passwords, or storing their personal information on devices while using unsecure public networks).
Example Impacts	Fake news can be used to **manipulate public opinion**, **reducing trust** in governments, institutions, or experts. For example, in the context of Covid-19, fake news reduced trust in medical experts and doctors. Another example is Macedonian teenagers who were targeting Trump supporters in the 2016 US presidential election, although their motivation was financial (for advertising revenue). In some cases, such as the "Pizzagate" incident (Fisher et al., 2016), fake news resulted in **physical violence**.	Security attacks often impact organizations. For example, Microsoft Exchange Servers data breach in 2021 was one of the biggest cyberattacks of US history, which affected more than 30,000 US companies. Security attacks can also impact individuals and societies. For example, in case of the Colonial Pipeline ransomware attack in May 2021, millions of people experienced fuel shortages, and many airlines had to cancel or change flights due to jet fuel shortage.

4 Review Process Methodology

To find the relevant literature, we used two major online scientific databases, namely Google Scholar and Scopus. Google Scholar was linked to major online libraries and databases such as Web of Science, EBSCOhost, ProQuest, ACM Digital Library, and IEEE Xplore. We set six criteria to include or exclude

articles in the literature review: 1) we selected articles written in English, 2) we included journal publications, conference papers, as well as the grey literature to expand scholarly efforts and gain more practical insights about the fake news phenomenon (Adams et al., 2017), 3) since fake news research is a multidisciplinary topic, we included studies from various disciplines such as Information Systems (IS), Computer Science (CS), Information Security, Psychology, Social Science, etc. 4) we selected articles that focus on combating fake news on social media, conceptual papers about fake news, relevant literature review papers, and a few studies from the security literature (our theoretical foundation is based on a model from the security literature), 5) We also excluded studies about fake news propagation, echo-chambers, filter bubbles, and polarization, 6) Finally, we did not limit our search to any specific time range.

To obtain more effective search results, we used the following keywords in our search query: ("fake news" OR "misinformation" OR "disinformation" OR “Rumor" OR “false information" AND ("combat" OR "fight") OR "deter fake news" OR “prevent fake news” OR "detect fake news" OR “mitigate fake news”. We searched documents titles, abstracts, and keywords. This search strategy and selection criteria identified 1640 articles in Google Scholar and 925 articles in Scopus. After eliminating the overlapping materials and reading and skimming the abstracts, 245 papers were selected for further screening and reading the full text. Screening the full text also led to elimination of 81 more papers. The final number of papers included in this review was 164 articles. We note that our literature search was by no means exhaustive, rather we tried to provide a representative summary of the relevant research to combat fake news on social media. Figure 2 shows the flow diagram for our literature review process.

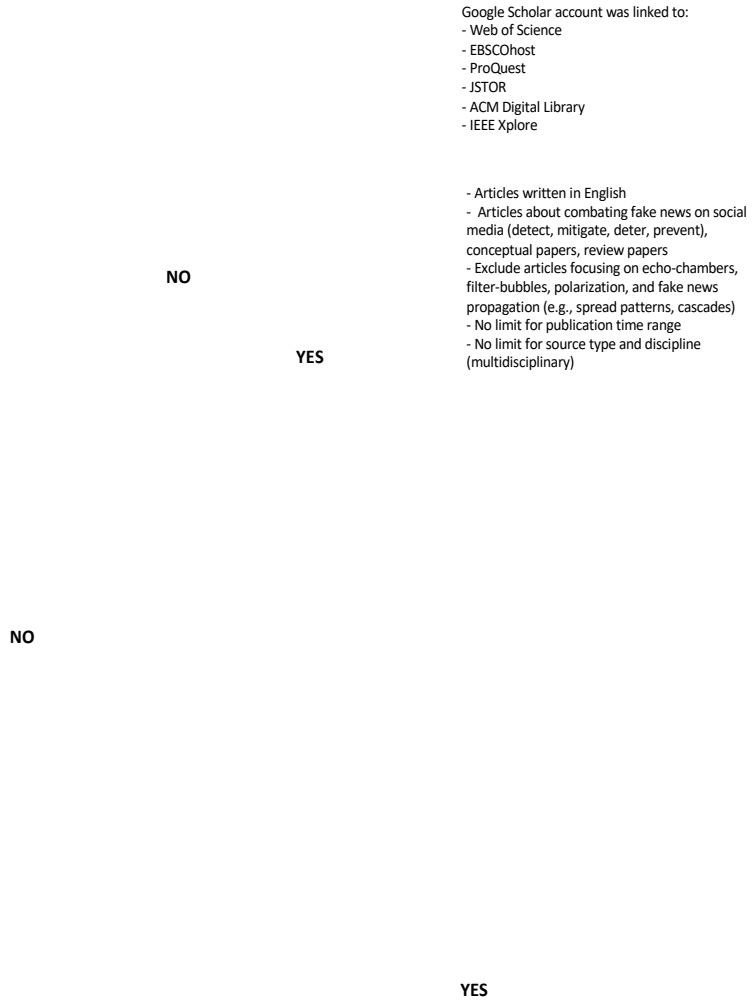

Figure 2. Flow Diagram for the Literature Review Process

4.1 Descriptive Statistics of the Articles

Figure 3 illustrates descriptive statistics about the articles reviewed in this paper. First, Figure 3(a) shows year-wise distribution of articles reviewed in this research. This figure shows an increasing trend in the number of publications about fake news, which shows a growing interest in this topic, especially after the year 2016. This is largely due to the proliferation of fake news during the 2016 US. presidential election (Allcott & Gentzkow, 2017). Second, the figure on the top right (Figure 3 (c)), shows the number of reviewed articles by the publisher, where the "ACM Digital Library", "Taylor & Francis", "Elsevier", "IEEE", and "Springer" are among the top 5 publishers. Third, Figure 3(b) on the bottom left shows the distribution of reviewed papers by discipline. This figure shows that the reviewed articles about fake news come from a range of disciplines. The majority of the contribution comes from the Computer Science (36%) field, followed by Information Systems (16%) field. Finally, we can see that most of the work on combating fake news on social media is focused on "detection", while "deterrent" strategies have gained less attention from academic scholars (Figure 3(d)). We note that for this figure, we only included articles focusing on combating fake news and excluded other papers such as review papers, and theoretical papers from Information Security literature. Also, if a paper focused on more than one stage, for example all four stages, it is presented in all the stages of the pie chart. The reason for doing this is because if we considered a separate part in the pie chart for all combinations (e.g., deter & prevent, deter & detect, …), each part would have been very small (there are 14 possible combinations). Also, our goal is to show number of studies (portion of the research) for each combat stage. For example, if a paper addressed both detection and mitigation, it should appear in both "detection" and "mitigation" slices of the pie chart. However, based on our review, only few papers focused on more than one stage.

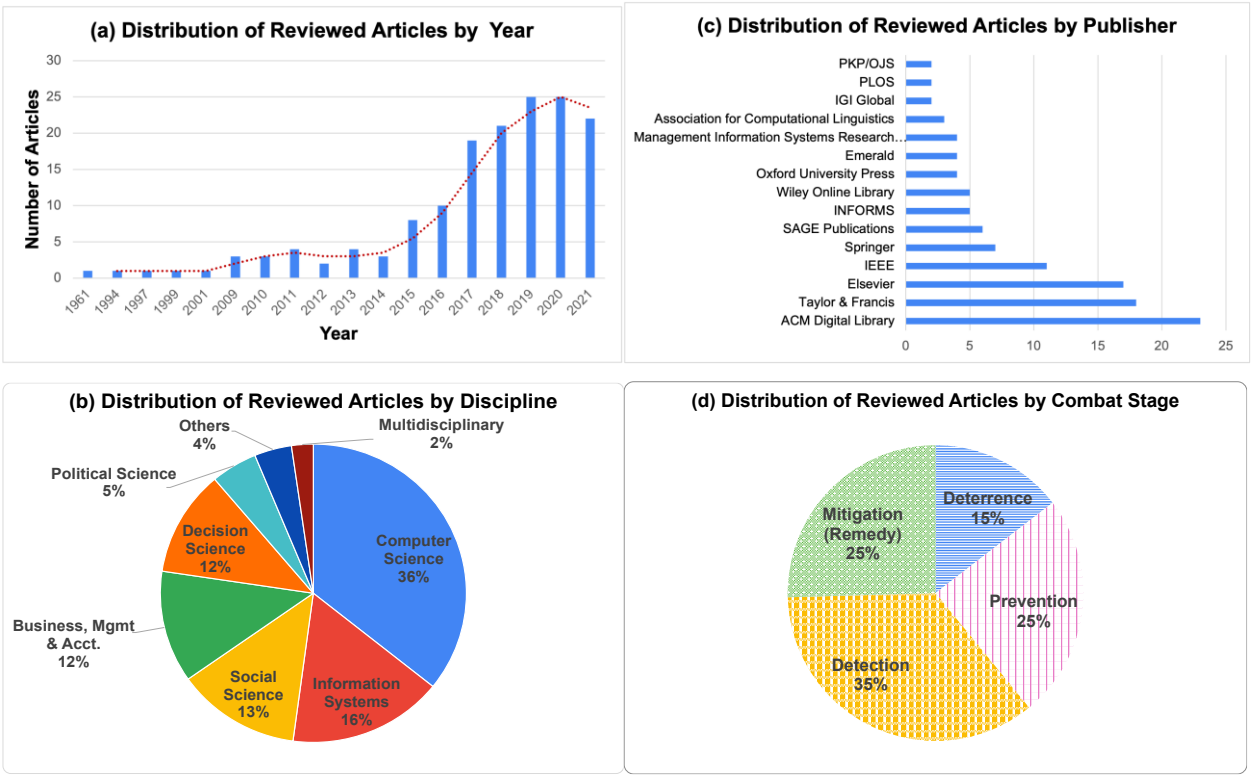

Figure 3. Descriptive Statistics of Reviewed Papers

As mentioned earlier, fake news is a multidisciplinary field and the articles reviewed in this research come from a variety of disciplines. However, the contribution of different fields varies across different stages of combating fake news. As depicted in Figure 4 (a), most of the reviewed articles related to fake news "deterrence" come from the field of "Social Science" (27%). In terms of fake news "prevention" (Figure 4 (b)), almost half of the articles belong to the "Social Science" and "Psychology" fields (29% and 19%

respectively). While the Social science discipline has the highest contribution in fake news "deterrence" and "prevention" research, there are very few research (only 3%) in fake news detection. Figure 4 (c) shows that the "Computer Science" discipline plays the dominant role in fake news detection studies (47%). Interestingly, more than 70% of reviewed papers in fake news detection are from "Computer Science" and "Information Systems". This is probably due to the technical nature of fake news detection on social media platforms. Finally, as shown in Figure 4 (d), the research on fake news mitigation is mainly covered in "Computer Science" (42%). In the following section, we will further explain each stage of combating fake news on social media.

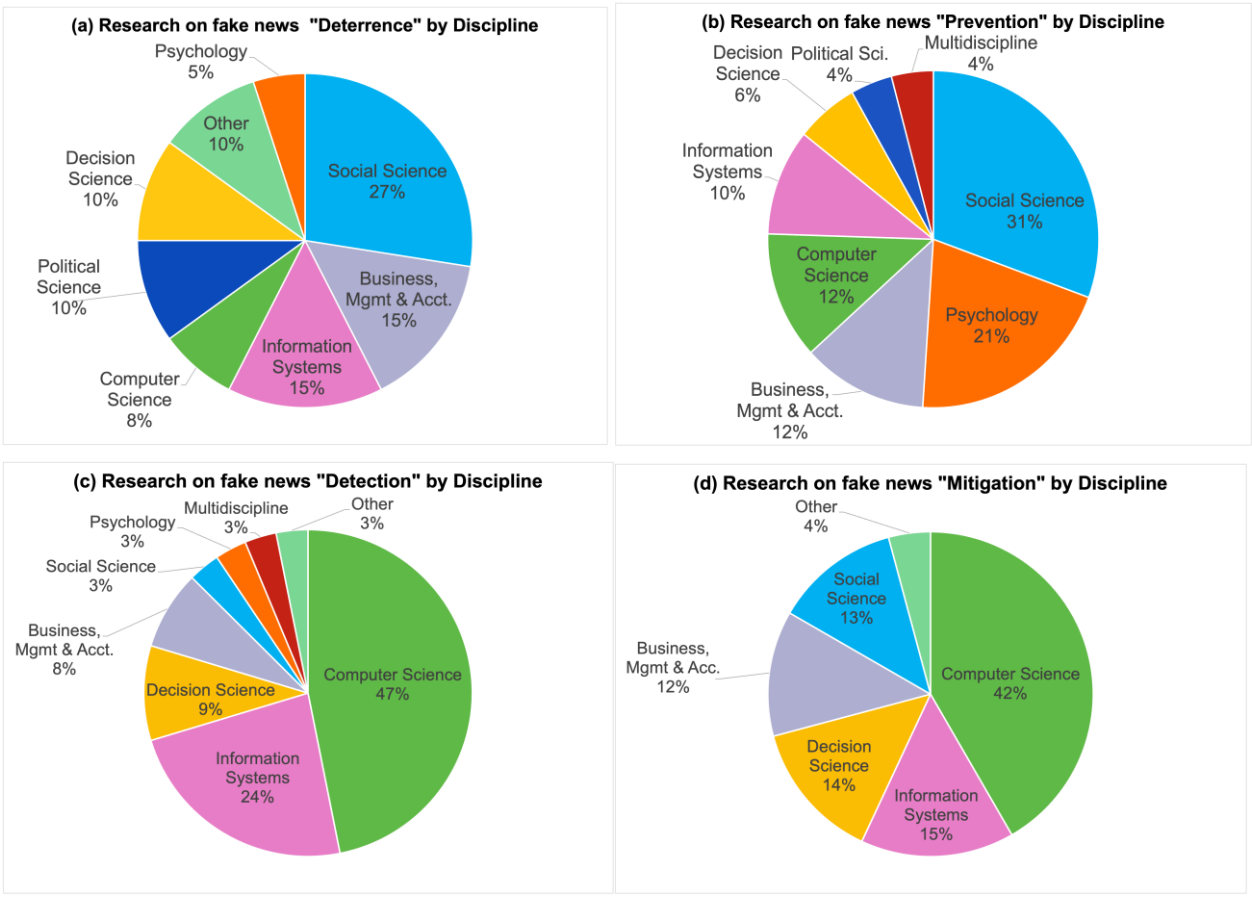

Figure 4. Distribution of the Reviewed Articles on Fake News Combat Stages across Disciplines

5 Combating Fake News on Social Media

In this section, we further discuss each stage of the fake news combat cycle in our framework namely, deterrence, prevention, detection, and remedy (mitigation). For each stage of the fake news combat cycle, we provide our definition for that stage, the challenges that exist to implement that stage, the existing approaches, limitations of current approaches, and directions for future research. Table 5 provides a summary of the challenges, approaches, limitations, and future directions for each stage of combating fake news on social media. In addition, Table 9 in the appendix provides a complete list of reviewed articles classified by fake news combat stage (please note that it only contains papers relevant to combating fake news, and excludes review papers, conceptual papers, etc.).

5.1 Deterrence

The first stage to combat fake news is deterrence defined as discouraging or dissuading people from creating or spreading fake news on social media by instilling doubt or fear of the consequences. Importantly, deterrents dissuade people from action through the threat of force and not the actual use of

force. Since deterrence is about demotivating people, we first need to understand the motives behind creating fake news.

Table 5. Fake News Combat Stages, Challenges, Approaches, Limitations, and Future Opportunities

	Challenges	Approaches	Limitations & Future Research
Deterrence	• Several motivations for fake news creation and propagation • Difficult to discourage people from creating or posting fake news especially when it is politically or ideologically motivated. • Social media companies lack incentives to police their platforms.	• Establish laws, policies, and regulations on fake news by governments, authorities, and social media platforms. • Educate users to increase the awareness of regulations	• Fake news has not been legally treated as a crime and no agreement on which criteria to consider a fake news as a crime. • Regulation may be viewed as restriction of freedom of speech. • Why laws and regulations are less effective to deter fake news.
Prevention	• Difficult to apply preventives due to the nature of free information exchange. • Fake news prevention could be interpreted as censorship to against freedom of speech. • Governments and authorities may misuse fake news prevention against opposition for political purpose.	• Block and suspend malicious accounts on social media platforms. • Block or filter the known fake news on social media platforms. • Prebunking (inoculation against fake news by e.g., preemptive warnings)	• How to effectively prevent wide and fast spreading of fake news in social media? • How to distinguish and balance the fake news prevention and freedom of speech? • How to prevent true information to be mistakenly blocked • How to combat people's ideology biases in relation to fake news?
Detection	• Fake news is masqueraded as true news and humans are often unable to identify fake news. • People like to receive and share the news they like without considering if they are true or fake. • Social media facilitate the spread of massive news, and it is difficult to check every news piece.	• Manual detection (either by experts or through crowdsourcing). • Automated detection (computational fact-checking, algorithmic solutions using ML, propagation pattern, etc.) • Guidelines for fake news detection	• Manual detection is difficult and time consuming. • There are needs to further improve the effectiveness and applicability of algorithmic solutions (semi-supervised and unsupervised models, fake audio and video detection, the use of social contexts features) • Educate people to detect fake news
Remedy (Mitigation)	• Fake news causes significant damage to the individuals trust believe and the justice of democratic society. • It is difficult to make people disbelieve fake news and change behavior accordingly. • Continued Influence Effect (CEI), i.e., when discredited information (e.g., flagged fake news) continues to affect behavior and beliefs.	• Minimize the influence of fake news propagation. • Spreading truth through both social media and public media to discredit fake news. • Platform interventions to clean up fake news. • Execute legal sanctions against those who caused significant damage by creating and spreading fake news.	• Anti-fake news actions can backfire and increase the spread of fake news. • Platform interventions have also some limitations, e.g., people may perceive unflagged content as true • What is appropriate rule of multiple stakeholders such as governments, political parties, social media providers, organizations, and individuals to maintain healthy social media environment.

5.1.1 Deterrence Challenges

One of the main challenges of this stage is that there are different motivations to create or spread fake news on social media: 1) Political motives to influence public opinion, to advance a preferred candidate and political party, or to damage opponents, especially during election periods (Allcott & Gentzkow, 2017). 2) Economic/Financial motives to generate revenue and monetary profit. A common example is using

clickbait headlines which entice/attract users to click through and subsequently generate revenues through increasing page traffic. 3) Ideological motives to promote ideological views. For example, the ISIS terrorist group uses social media platforms to promote their opinions through spreading propaganda (Zannettou et al., 2019), 4) Other Individual motives: These include malicious intents (to hurt others in various ways), influence (to get power or to manipulate public opinion), sow discord (confusion), and fun (Zannettou et al., 2019).

However, there are insufficient discouragement mechanisms to demotivate people from creating and spreading fake news on social media. This is in part because there is no clear governing body. Social media platforms as the main actor in this space have little incentives to deter the production of fake news. At the same time, governments struggle to restrict freedom of speech and create perceptions of effective deterrence. The challenge lies in deterrents dependency on users' free will, and it is difficult to restrict or control it without effective "carrots and sticks". In the following subsections, we discuss the approaches to demotivate or deter users from creating and spreading fake news on social media.

5.1.2 Deterrence Approaches (Deterrents)

A common deterrent approach to fight against fake news is to establish laws, regulations and policies that clearly define sanctions and consequences for those who create and/or spread fake news on social media. According to General Deterrence Theory, perceived certainty and severity of sanctions deter individuals from engaging in illegal behavior or committing a crime (in criminology) or IS misuse intention (in IS security). The idea behind this is that people will avoid abusive behavior (e.g., create or spread fake news) if they believe that cost of their actions is higher than the benefits. Therefore, establishing laws, regulations, and policies is an important deterrent to dissuade people from creating or spreading fake news on social media. Although such attempts conflict with free speech ideas and ideals, some level of restriction on free speech is inevitable to discourage the creation and spread of fake news, rather than just preventing its spread (Helm & Nasu, 2021).

In recent years, there have been some attempts by governments, policymakers, legislators, and social media platforms to address the fake news problem. For example, Malaysia's government was one of the first to establish a law to combat fake news by penalizing offenders with a 10-year jail sentence, a fine up to (£90,000) or both[4]. In 2018, the German parliament established a law, known as NetzDG, which oblige large social media companies to remove fake news and hate speech content within a 24-hour deadline or pay the penalty of up to 50 million euros[5]. In Italy, the anti-trust chief Giovanni Pitruzzella has called for the EU to establish rules to consider the penalty for companies that spread false content (Morgan, 2018). Following claims of Russia's meddling in the 2017 French presidential election, president Emmanuel Macron promised anti-fake news laws in 2018 to stop fake news (Nugent, 2018). A comprehensive list of anti-misinformation actions around the world is provided in (Funke & Flamini, 2022).

Another (non-legislative) deterrence approach is to use educational and training programs. Such programs dissuade users from illicit behaviors (create or spread false content in the context of fake news) by increasing awareness about regulations and policies, and the penalties associated with violating the laws. In security literature, it has been shown that the best way to ensure the viability of a security policy is to educate users about it to make sure they understand it and accept necessary precautions (Whitman et al., 2001). IS research found that user's awareness of security policies and SETA (Security Education, Training, and Awareness) program deter IS misuse (D'Arcy et al., 2009). A similar study found that employees can better manage cybersecurity tasks when they are aware of their company's information security policy (Li et al., 2019). In the context of online fake news, governments in several countries took some steps to increase users' awareness about fake news through training and media literacy initiatives. For example, in 2019, federal government of Canada announced it was giving $7 million to projects aimed at increasing public awareness of online fake news[6]. In the same year, the Netherlands government launched a public awareness campaign to inform their citizens about the spread of fake news online.

5.1.3 Deterrence Limitations and Future Opportunities

There are several limitations in effectively implementing deterrence strategies, especially in the context of fake news. First, establishing laws and regulations to deter users from creating or spreading fake news in

[4] https://www.theguardian.com/world/2018/mar/26/malaysia-accused-of-muzzling-critics-with-jail-term-for-fake-news
[5] https://www.reuters.com/article/us-singapore-politics-fakenews-factbox-idUSKCN1RE0XN
[6] https://www.canada.ca/en/canadian-heritage/services/online-disinformation.html

the context of fake news is more difficult and complex compared to security or criminology contexts. One limitation is that it is not easy to recognize fake news as a crime because there is not even an overall agreement on how to define fake news, or when to consider it as a crime. For example, in the context of politics, a content that the left party consider as true news may be considered as fake by the right party.

In general, there are not enough deterrent mechanisms against fake news on social media. There should be more effective laws, regulations, and policies by governments, authorities, and social media platforms to discourage users from creating/spreading fake news. Sanctions and penalties against fake news should be certain and severe to be effective as deterrents. However, the laws and regulations established by governments can be viewed against freedom of speech, especially by people who don't trust their governments and those who think these laws increase the corruption and prevent their right of free speech. Research shows that regulations are not the preferred choice of the public to combat fake news on social media as people may view regulations as a restriction to freedom of speech. Most people, even when they perceive fake news harmful to society, if they have a choice, they prefer non-regulatory solutions such as education over regulations (Jang & Kim, 2018). The authors explain that most people prefer education over regulations because they "do not want to sacrifice their freedom of speech to protect other's vulnerability".

Ultimately, more research is needed to understand *why* anti-fake news laws and regulations are less effective, *how* differences between laws affect the motivation and ability to generate fake news, and how, why, and when people respond differently to deterrence measures against fake news generation and spread. This line of work should also examine interactions of legislation and other means. As pointed by Haciyakupoglu et al., (2018), legislations should be complemented by other means such as pre-emptive inoculation, immediate measures (e.g., fact-checking), and long-term measures (e.g., education and media literacy). We discuss all these measures and more in the remainder of this paper.

5.2 Prevention

If people choose to ignore the deterrents, the next stage is to use preventive actions, defined as *"active countermeasures with inherent capabilities to enforce policy and ward off illegitimate use"* (Gopal & Sanders, 1997; Straub & Welke, 1998). Applied to fake news, preventive actions are active countermeasures to prevent individuals from creating or spreading fake news on social media. In the context of fake news, blocking fake accounts or blocking fake content are examples of preventive countermeasures (users may create the fake account, but it will be blocked or removed). We further explain the prevention stage in the remainder of this section.

5.2.1 Prevention Challenges

Implementing preventive measures in the context of fake news is more challenging compared to the security context. One challenge is the debate over censorship and freedom of speech, which can be a potential explanation for the weakness of social media platforms in implementing effective preventive countermeasures. For example, preventive measures such as blocking or suspending social media accounts can be misinterpreted as a censorship or as conflicting with freedom of speech ideals. The laws against fake news established by governments can especially be questioned by people who do not trust their governments and those who think these laws increase the corruption and prevent their right to free speech. In fact, in some cases, governments and authorities may use preventive measures to censor the opposing views and further spread the information aligned with their own views and benefits. In addition, prevention mechanisms vary based on the countries in which they are implemented. For example, some countries have taken stronger preventive measures and have more control over the information their people consume online. However, as mentioned earlier, there is a concern that the governments use it to further spread fake news. In the remainder of this section, we further explain and review the current preventive approaches to combat fake news on social media. Based on our review, we also discuss the research gaps and future opportunities for this stage of fake news combat cycle.

5.2.2 Prevention Approaches (Preventives)

In recent years, there has been growing concerns about the role of social media in facilitating the spread of fake news and several studies called for actions by social media platforms to fight against fake news (Flew et al., 2019; Hartley & Vu, 2020; Hemphill, 2019; Smyth, 2019). In response, social media platforms have taken some steps to prevent the spread of fake news by e.g., blocking fake and malicious accounts and updating their algorithms to remove incentives for users who promote false information. In terms of

preventive measures, Facebook updated its recidivism policy to stop people who repeatedly violate its Community Standards from being able to create new pages or groups[7]. Following the 2020 presidential election campaign in the United States, Facebook banned deepfake media (manipulated videos or photos)[8] from its platform. Twitter has accelerated its combat against fake accounts by suspending millions of fake and suspicious accounts in 2018 (over 70 million only in May and June). Twitter's growing campaign against bots and trolls was driven by political pressure from the U.S congress following reports of manipulation by Russian disinformation during the 2016 presidential election (Timberg & Dwoskin, 2018). In addition, Twitter announced a COVID-19 misinformation policy in response to a large volume of false and misleading information related to COVID-19. Depending on the severity of the violation, the consequences of violating this policy may include tweet deletion, labelling the tweet, and even account locks and permanent suspension of the accounts for severe or repeated violations of this policy[9].

In academia, several studies focused on platform interventions to fight fake news on social media. A type of platform intervention that restricts the accounts from publishing fake news is *"account-level intervention"*. Several attempts have been proposed in this direction such as algorithms to identify bots and malicious accounts (Sharma et al., 2019), and network monitoring which leverages a set of nodes to filter the information they receive and block what they identify as fake news (Amoruso et al., 2020; Kimura et al., 2009; Zhang et al., 2016). More recently, Ng et al., (2021) examined "fake news flags" as content-level and "forwarding restriction" as an account-level intervention to combat fake news. They found that the two types of interventions have different effects on fake new: flagging fake news leads to the more centralized and less dispersed spread of fake news while forwarding restriction leads to less direct and more indirect forwarding of fake news, compared with true news.

Another preventive approach is Prebunking Fake News by Inoculation. According to the Inoculation Theory (Papageorgis & McGuire, 1961), people can be inoculated against persuasion by being exposed to a refuted version of a counterargument beforehand. Just like vaccines, a sufficiently weakened does of counterargument triggers the production of "mental antibodies", immunizing people to unwanted persuasion (Compton, 2013). Inoculation involves two elements: (a) forewarning – a warning of a forthcoming threat, designed to motivate resistance and defend one's attitudes, and (b) a pre-emptive refutation (or prebunking) of the persuasive arguments. Several studies have shown inoculation as an effective strategy to confer resistance against fake news on social media. For example, inoculation, based on logical communication and facts, reduces the influence of conspiracy persuasion by increasing the degree of skepticism towards conspiratorial claims (Banas & Miller, 2013). In the context of climate change, inoculation has been shown to neutralize the influence of misinformation on a perceived consensus about climate change (Cook et al., 2017). Similarly, preemptive warnings help protect (inoculate) public attitudes about the scientific consensus against misinformation (Van der Linden et al., 2017). In the context of COVID-19, the theory of inoculation is shown to be an effective strategy to confer resistance against fake news (van Der Linden et al., 2020). Research shows that inoculation or prebunking fake news is more effective than debunking it, and preexposure warnings have a stronger effect than corrections (King et al., 2021). In other words, prevention is better than cure. For example, Jolley & Douglas, (2017) found that anti-conspiracy arguments that were present prior to conspiracy theories improved vaccination intention, but they were not effective if they came afterwards (once established, conspiracy theories become resistant to correction).

5.2.3 Prevention Limitations and Future Opportunities

There has been insufficient mechanisms and strategies to prevent the creation or spread of fake news on social media. Unlike the security context where using password or lock on computers can be used as a preventive measure, implementing preventives in the context of fake news is not that easy. As mentioned earlier, one issue is that preventives such as blocking social media accounts can be interpreted as censorship and against the freedom of speech. However, the harmful impacts of fake news may outweigh the benefits of free speech (Helm & Nasu, 2021). Therefore, one important direction for future research is to investigate the balance between freedom of speech and preventive measures against the creation or spread of fake news on social media. Another concern with preventive measures such as blocking accounts on social media is that it might unintentionally prevent the spread of truth if mistakenly blocks legitimate accounts. Moreover, only limited number of malicious accounts can be blocked compared to the

[7] https://about.fb.com/news/2020/09/keeping-facebook-groups-safe/
[8] https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/
[9] https://help.twitter.com/en/rules-and-policies/medical-misinformation-policy

large volume of fake news on social media. In addition, there has been a few studies on inoculation and education to prepare users to fight against fake news, mostly in the context of climate change (Lewandowsky & Van Der Linden, 2021; Van der Linden et al., 2017). Finally, prior studies mainly focused on passive inoculation where people are inoculated against the same information to which they will be exposed later. However, recent research shows that "active inoculation" where people are exposed to similar, but not the same information is more effective in creating resistance against fake news (Roozenbeek & Van Der Linden, 2019).

Ultimately, more research is needed on preventive measures: *when* they work, *why* they work, *for what types* of fake news or in what contexts they work, and for what *types of people* they work best. Findings from such studies can help social media providers apply effective restrictions. One more question that is relevant in this context is: how to prevent true information to be mistakenly blocked?

5.3 Detection

If fake news cannot be stopped at the first two stages, which means fake news is already spread on social media, the next stage is to detect fake news. We define detection as discovering and identifying fake news from a massive news posted and shared on social media.

5.3.1 Detection Challenges

Detecting fake news on social media is a challenging task. First, fake news is always decorated as true news which makes its detection difficult. As pointed by George et al. (2021):

> *FN is created with truth-subversive language, designed to play on emotion and connect with recipients by signaling authenticity and homophilic characteristics on the part of the originator. The objective of such strategies is to seed FN content effectively, and to increase the propagation of FN messages through social networks (p. 6)*

At the same time, people's ability to identify fake news is only slightly better than chance (Kumar et al., 2016; Ott et al., 2011; Rubin, 2010). More importantly, the term fake news has been highly polarized and misused, especially by politicians who label any piece of content that is not aligned with their view as "fake news" (Vosoughi et al., 2018).

Fake news detection is especially challenging in the context of social media where everyone can post any content, real or fake, with no cost or friction, resulting in a massive amount of news posted every day. It is difficult to monitor and detect all the fake news posted on social media. In general, people like to receive and share the news they like and believe what they like without considering if the content is true or fake (Moravec et al., 2019). In addition, social media platforms facilitate the spread of fake news through personalized recommendations which leads to the formation of "echo-chambers". Echo-chamber (Sunstein, 1999), refers to an effect when users in social media form groups with like-minded individuals where they are largely exposed to the information that confirm their own opinions (Shore et al., 2018). Echo chambers facilitate the spread of fake news, which can be explained through two psychological factors: social credibility (people tend to perceive a source as credible if others perceive it is credible) and frequency heuristic (when processing information, people favor information they have seen more frequently, even if it is fake) (Shu et al., 2017). In the remainder of this section, we review the existing fake news detection approaches and discuss the limitations and future research opportunities.

5.3.2 Detection Approaches

Fact-checking: One of the main approaches to detect fake news on social media is through fact-checking. Fact-checking is the process of evaluating the authenticity of news by comparing the knowledge extracted from a to-be-checked content with facts. There are three types of fact-checking. First, *"Expert-based fact-checking"* uses credible fact-checkers to manually assess the accuracy of the news. In recent years, several fact-checking organizations such as PolitiFact[10], and Snopes[11] have emerged to verify the veracity of information. For example, the PolitiFact's Truth-O-Meter provides six ratings including true, mostly true, half true, mostly false, and false to reflect the accuracy of a claim. The website also provides a "scorecard" to show the accuracy of statements based on the mentioned ratings. The Snopes website also has a similar rating scale with a few more labels such as unproven, miscaptioned, scam, etc. Second,

[10] https://www.politifact.com/
[11] https://www.snopes.com

"Crowdsource based Fact-checking" uses a group of regular individuals to evaluate the accuracy of information. For example, Fiskkit[12] is a crowd-based fact-checking website where users can apply tags to judge the article's accuracy and view how others evaluated the article. The International Fact-Checking Network (IFCN) [13] has launched a huge crowdsourcing project (The "CoronaVirusFacts" Alliance database) that unites more than 100 fact-checking organizations worldwide to fight the COVID-19 infodemic. Recently, Twitter introduced "Birdwatch", a crowdsourced fact-checking pilot that allows people to flag Tweets they perceive as misleading and write notes to provide additional context for why it may be misleading. Finally, *"Computational (Automated) Fact-checking"* uses computational solutions such as ML and NLP to automatically fact-check fake news. Two well-known examples are Truthy (Ratkiewicz et al., 2011) which track political memes in Twitter and help detect misinformation, and Hoaxy (Shao et al., 2016), a platform for automatic tracking of fake news diffusion and its competition with fact-checking efforts on Twitter. Some other examples include Factmata, an AI project by Google (Dale, 2017), ClaimBuster (Hassan et al., 2017), and ClaimRank (Gencheva et al., 2017) that use machine learning approaches for fact-checking. Kim et al., (2018) used both the crowd and expert knowledge to detect and prevent the spread of fake news. They developed CURB, a scalable online algorithm to decide which stories to send for fact-checking and when to do so. Table 6 shows a comparison of fact-checking approaches.

Table 6. Comparison of Fact-checking Approaches

Fact-checking	Advantage(s)	Drawback(s)
Expert-based (Manual)	• High accuracy (because it use experts)) • Expert-based fact-checking websites can be used as a public data repository for fake news research, e.g., LIAR (Wang, 2017) and FakeNewsNet (Shu et al., 2020)	• Slow • Costly • Low scalability (they cannot keep up with the large volume and rapid spread of fake news on social media)
Crowdsource-based (Manual)	• Faster than expert-based fact-checking • More scalable than expert-based	• Low accuracy (because it relies on regular people for verification) • Vulnerable to manipulation and misuse by adversaries • Less scalable than computational (automated) fact-checking
Computational (Automated)	• Faster than both expert-based and crowdsourced-based fact-checking • High scalability	• Less accurate than expert-based fact-checking

Automated Algorithmic Solutions: In recent years, there has been several survey papers reviewed the literature on fake news detection on social media and classified the approaches to detect fake news from different perspectives such as fake news component (content, user, context), methodology, etc. *From a data mining perspective,* fake news detection methods are classified into *knowledge-based* and *style-based* methods (based on content features) and *stance-based* and *propagation-based* approaches (based on social context features) (Shu et al., 2017; Zhou & Zafarani, 2020). *From a methodology perspective*, there has been several categorizations. For example, fake news detection approaches can be broadly divided into: *classification* (ML and DL), and *other approaches* (propagation pattern, retweet behavior, etc.) (Bondielli & Marcelloni, 2019). Another categorization based on methodology is: *machine learning, systems* (systems that inform users about detected fake news), and *other models/algorithms* such as epidemiological models, Howkes processes, etc. (Zannettou et al., 2019). Fake news detection approaches have also been categorized *based on fake news components* (content, user, context). For example, fake news detection methods can be divided into three types: *Content-based* (identify fake news based on the content of the information), *Feedback-based* (based on user responses on social media), and the *Intervention-based* (actively identify and contain the spread of fake news and mitigate their impacts) (Sharma et al., 2019). Finally, a comprehensive review of fake news detection approaches is provided in (Zhang & Ghorbani, 2020), where authors provided three different perspectives to classify fake news detection approaches: *Component-based* (creator/user, content, social context), *Data mining-based* (supervised, unsupervised), *and Implementation-based* (online, offline). Please note that, in this paper, we have not provided a new classification of fake news detection approaches because this was previously

[12] http://fiskkit.com
[13] https://www.poynter.org/ifcn/

done. However, we provide example references for different fake news detection approaches (classified based on methodology) in Table 8. In addition, Table 7 provides a summary of review papers on fake news detection, their classification criteria, and the type(s) of fake news they addressed in their study.

Table 7. Review Papers on Fake News Detection, Classification Criteria, and Type of Fake News Studied

Fake News Review Papers	Classification Criteria for Fake News Detection Approaches	Type of False Information
(Shu et al., 2017)	• Content Models: knowledge-based, style-based • Context Models: stance-based, propagation-based	Fake News
(S. Kumar & Shah, 2018)	Based on algorithms: • Feature-based • Graph-based • Model-based (Temporal, Propagation models)	Fake News, Fake Reviews, Hoaxes
(Zubiaga et al., 2018)	No specific classification for rumour detection	Rumours
(Shu, Bernard, et al., 2019)*	Based on Network: • Interaction network embedding • Temporal diffusion • Friendship network embedding • Knowledge network matching	Fake News
(Zannettou et al., 2019)*	• Machine learning • Systems • Other Models/Algorithms	Rumours, Hoaxes, Conspiracy Theories, Satire, Clickbait, Fabricated
(Sharma et al., 2019)*	• Content-based • Feedback-based (based on user responses) • Intervention-based (detection and mitigation)	Fake News, Rumour
(Bondielli & Marcelloni, 2019)	• Classification approaches (ML, DL) • Other approaches (Crowdsourcing, Diffusion patterns, etc.)	Fake News, Rumour
(X. Zhang & Ghorbani, 2020)	• Component-based (Creator analysis, Content analysis, Context analysis) • Data mining-based (Supervised learning, Unsupervised learning) • Implementation-based (Online/Real-time, Offline detection)	Fake News, Fake Review, Rumour or Satire
(Zhou & Zafarani, 2020)	• Knowledge-based (Manual fact-checking, Automated fact-checking) • Style-based (based on content) • Propagation-based (using News Cascades, Propagation Graphs) • Credibility-based (source credibility)	Fake News
(Collins et al., 2021)	Classified fake news detection into 8 categories: Experts/Fact-check approach, Crowdsourced, Hybrid (Expert-crowdsource, Human-Machine), ML, DL, NLP, Graph-based methods, Recommender Systems	Fake News (Clickbait, Propaganda, Satire & Parody, Hoax, other)
(Khan et al., 2021)	• Knowledge-based • Feature-based • Network Propagation • Hybrid Approach	Fake News (including Rumor & Clickbait detection)

Guidelines for fake news detection: It is also important to help users improve their ability to detect fake news. In recent years, numerous workshops, training programs, and courses have been developed to help people recognize fake news from true news. A common approach is to provide guidelines for people to detect fake news. These guides often suggest a checklist for evaluating a news source. The CAARP (currency, authority, accuracy, relevance, and purpose) test, SMART (source, motive, authority, review, two-source test), or SMELL (source, motive, evidence, logic and left-out) are just a few examples (Lim, 2020). Other examples include but are not limited to a research guide on “Fake News, Misinformation, and

Propaganda" by Harvard University library, two research guides offered by the University of Toronto library, and the "LibGuide", a popular library guide offered by librarians at Indiana University to help students in evaluating the credibility of information (Banks, 2017).

5.3.3 Detection Limitations and Future Opportunities

A large body of research have focused on fake news detection technologies, especially through algorithmic solutions. However, there are still many limitations. First, there is a lack of large-scale publicly available datasets on fake news that can be used as a benchmark to compare different algorithms. Such datasets are helpful in building and evaluating models in a situation similar to the real world. In recent years, some public datasets have been developed (Shu et al., 2020; Wang, 2017). However, there is stuffiest research on comparing different categories of algorithms on these datasets. Second, most existing detection algorithms use supervised learning, based on labeled datasets for training and validation. In real world scenarios, most data are either unlabeled or only a few labels are available in which cases unsupervised or semi-supervised models should be applied. Also, unsupervised models can better handle large amount of data in real time, which is especially useful in the context of social media where a large volume of information is created and disseminated every day. Third, prior research in fake news detection have mainly focused on the content. However, the context can help to identify if the content is true or false. For instance, the person described in the news could not be in the place at the time as mentioned. Although there has been some recent works using contextual features (Atanasova et al., 2019; Nguyen et al., 2020; Shu et al., 2019), social context features need to be further investigated for fake news detection. Finally, information on social media platforms comes in various formats such as text, audio, video, etc. Usually, pictures or video recording can be used as the evidence of truth. However, with the advances in information technology, especially artificial intelligence in recent years, it is easy to use photo editing or deepfake technology to make fake images or videos that appear authentic but are practically indistinguishable by humans (Westerlund, 2019). It is important to develop methods that can detect not only fake text, but also fake audio or video (Yu et al., 2021).

Overall, most detection efforts, especially the algorithmic solutions are in computer science. Although fake news research have gained more attention among IS scholars in recent years, they mainly focused on user behavior and the psychological and cognitive factors in sharing fake news (Kim & Dennis, 2019; Moravec et al., 2019, 2022; Turel & Osatuyi, 2021). Fake news is a multidisciplinary research field in nature, and we believe that there is an opportunity for IS scholars to further contribute to solving this problem. For instance, questions around *why* and *when* people believe algorithmic screening should be examined. There is also an opportunity to examine human-bot interactions in the process of screening fake news, and whether such approaches are superior to using just bots or just humans.

5.4 Remedy (Mitigation)

The remedy (mitigation) stage aims at reducing the destructive impacts of fake news diffusion on social media. In this paper we use the words remedy and mitigation interchangeably.

5.4.1 Remedy/Mitigation Challenges

Fake news causes significant damage to trust and beliefs of individuals (Ognyanova et al., 2020). It has also significant negative impacts on global issues faced by human society such as fighting COVID-19 pandemics (Shirish et al., 2021), or the recent war between Russia and Ukraine (e.g., deepfake videos of Putin or Zelenskyy circulating on social media amid the conflict[14]).

To reduce the negative impact of fake news, it is important to know why people believe fake news even when they were told it is fake. People's ideology and pre-existing beliefs play an important role in believability and spread of fake news. In fact, people believe what they want to believe, even when it makes no sense at all (Moravec et al., 2019). From a theoretical perspective, several theories explain this. First, the theory of *Confirmation Bias* (Nickerson, 1998) posits that people tend to believe what confirms their pre-existing beliefs. Second, according to the theory of *Naïve Realism* (Ross & Ward, 1996) people tend to believe they have the "true" perception of reality and those who disagree with them must be uninformed, irrational, or biased. Finally, people are also influenced by their peers, and they tend to share information that is more aligned with their peers' beliefs to gain social acceptance and affirmation,

[14] https://www.dw.com/en/fact-check-the-deepfakes-in-the-disinformation-war-between-russia-and-ukraine/a-61166433

regardless of the veracity of that information (*Social Normative Theory*) (Deutsch & Gerard, 1955). In political contexts, partisanship and political ideology of individuals are common explanations for why people believe fake news, i.e., people perceive fake news as accurate if it is consistent with their political ideology (Turel & Osatuyi, 2021).

5.4.2 Remedy/Mitigation Approaches

A common mitigation strategy is to minimize the influence of fake news by limiting the scope of its spread, e.g., by blocking certain nodes or links in the network. The goal is to minimize the impact of fake news spread on social media. The impact of fake news on social media can be assessed by the number of people that are affected by fake news. Blocking the flow of information from influential users in the network can significantly reduce the impact of fake news spread as these users have many followers. Indeed, finding a minimum subset of individuals who are neighbors with the rumour community can help in limiting the spread of the rumour to the rest of the network (Fan et al., 2013). Tong et al., (2017) addressed the rumour blocking problem in online social networks by using a random-based approach. They evaluated their randomized algorithm on both real and synthetic social networks (Power2500, Wiki, Epinion, and YouTube) and showed that their algorithm outperforms the state-of-the-art rumour blocking algorithms such as greedy algorithm with the Monte Carlo simulation in terms of running time. Another example is the DRIMUX model (dynamic rumour influence minimization with user experience), minimized the influence of rumours by blocking a subset of nodes while considering users' experience (a time threshold that a particular node is willing to wait while being blocked) (Wang et al., 2017).

Another approach to mitigate the impacts of fake news is through increasing the spread of true information (Shu, Bernard, et al., 2019). To this end, most prior research used competing cascades which contain true information, to compete with the fake news cascade as the falsehood begins to spread through the network rather than after its diffusion. The goal is to make sure that true news reach users who are exposed to fake news, to reduce the chance of believing fake news, and to make social media more reliable source of information. Several models have been proposed in this direction. For example, Budak et al., (2011) models the spread of two cascades evolving simultaneously: "bad campaign" spreading bad information (fake news) and "good campaign" to counteract the effects of fake news. They identified a subset of individuals (k influential users) to spread true information with the objective of minimizing the number of users who at the end of the propagation process adopt the bad campaign. One limitation of the approach used in Budak et al., 2011 is that their model assumes if a user is exposed to a piece of news, then they will also shares the news. In a similar notion, Nguyen et al., (2012) proposed a model which finds a small set of influential nodes (users) to spread "good information" to contain misinformation. Their findings depict that when the number of required nodes to spread true information is small, it is most effective to select influential nodes in large communities. However, when more nodes are required, selecting influential nodes from smaller communities is more effective in limiting the fake news spread. Wang et al., (2014) developed two strategies to select the smallest set of influential nodes decontaminated with true information to effectively contain the spread of fake news. Their experimental results using three datasets from Twitter, Friendster, and a random synthetic network proved the performance benefits of their proposed strategies.

In the IS, there has been a growing interest in platform interventions to fight fake news on social media, either through content-level interventions (interventions that only target a piece of content) or account-level interventions (interventions that target the account that post fake news) (Ng et al., 2021). We discussed the account-level interventions in the prevention section. Content-level interventions reduce the impact of fake news by triggering users' cognition, e.g., through flagging fake news or highlighting the source of the article. A common example is using "fake news flags". In IS, scholars mainly studied the effectiveness of flagging on changing users' beliefs and limiting the spread of fake news. In this vein, two different approaches to implement a fake news flag was examined; one designed to trigger system 1 ("automatic cognition" or "fast-thinking") and the other to trigger system 2 ("deliberate cognition", or "slow-thinking"). (Moravec et al., 2020). Both approaches are shown effective in reducing the believability of fake news and combining both approaches was about twice as effective. To understand whether some type of flagging is more effective than others, three flagging strategies were examined: fact-checker flags, peer-generated flags, and publishers' self-identified humor flags (Garrett & Poulsen, 2019). They found that publishers' self-identified flags were the most effective strategy in reducing people's beliefs and sharing intentions of fake news. In addition to fake news flags, "highlighting the source of article" and "source rating" are other forms of content-level interventions proposed in the literature (Kim & Dennis, 2019). The authors showed that both changing the interface to highlight the source of the article, and source rating

(showing low ratings for the source) can nudge users to be more skeptical of fake news and less likely to believe and spread any article. Finally, different rating mechanisms (*experts rating, users article rating*, and *users' source rating*) influence user beliefs in news articles (Kim et al., 2019). Author found that users perceive expert ratings as more cognitive and user ratings more emotional.

5.4.3 Remedy/Mitigation Limitations and Future Opportunities

Unfortunately, corrective information does not necessarily change people's beliefs and can actually have the opposite effect (Flynn et al., 2017). In politics, not only correction may fail to reduce misperceptions, but it can also backfire and strengthens misperception among ideological subgroup holding those misperceptions (Nyhan & Reifler, 2010). In line with this, King et al., (2021) used Twitter data to examine the dynamic interaction between true and fake news and found that information correction does not reduce the spread of fake news. Instead, it backfires and increases the propagation of fake news on social media. These findings are in line with prior research that shows that any attempt to debunk fake news by confronting falsehood and truths facilitates the acceptance of fake news (Pennycook et al., 2018). This is because frequent exposure (in this case repeating fake news) increases familiarity, which in turn increases the chance of accepting fake news. More research is needed to clarify these contradictory findings. Timing of information correction is also important and different methods may be useful in different phases of fake news propagation. For example, He et al., (2015) proposed an optimization approach that combines two methods (*blocking rumours* at influential users and *spreading the truth* to clarify rumours). They showed that the method of *"spreading truth"* should play a dominant role in the start of rumour containment, whereas the method of "rumour *blocking"* should be used extensively when approaching the end of rumour restraining phase. This is because the exposure to fake news increases as time passes. The more fake news is circulated and repeated, it increases users' familiarity and acceptance. As a result, the "spreading truth" method may be less effective after longer exposure to fake news.

In terms of content-level interventions, most prior research studied their effectiveness in terms of psychological and cognitive aspects such as believability. Believability is an important factor in studying fake news on social media and prior research found the strong effect of believability on users actions such as read, like, share, and comment (Kim et al., 2019a). However, there are contradictory findings about the effectiveness of content-level interventions (e.g., flagging fake news) on reducing users' belief in fake news, and there are several factors (e.g., prior beliefs or source reputation) that can weaken the effectiveness of such interventions. For example, although using fake news flags triggers more cognitive activity, it is shown that it cannot overcome the role of confirmation bias and users continue to believe what they want to believe, regardless of the truth of a news article (Moravec et al., 2019). Also, a trusted source with high reputation can lower the impact of flags on reducing the believability of fake news (Figl et al., 2019). Finally, using fake news flags may cause an implied truth effect, meaning that it may lead people to believe that unflagged contents are trustworthy (Pennycook, Bear, et al., 2020).

Even though prior research on the effectiveness of content-level platform interventions is inconclusive, such interventions are still helpful in combating fake news because they trigger users' cognition and nudge them to think more deeply before sharing contents on social media (Moravec et al., 2022). However, there are many more opportunities for IS scholar to understand when, how and why people resist the temptation to spread fake news and have a stronger motivation to check news items before they share them.

Table 8. Approaches to Combat Fake News and Example References for Each Stage

	Combat Approaches		Sample Articles (References)
Deterrence	Establish Laws, Regulations, and Policies/Increase Public Awareness about Policies		(Batchelor, 2017; Delellis & Rubin, 2018; Helm & Nasu, 2021; Jones-Jang et al., 2021), (Haciyakupoglu et al., 2018), (Hartley & Vu, 2020), (Morgan, 2018), (Nugent, 2018), (Jang & Kim, 2018), (Kreiss & McGregor, 2019) (Flew et al., 2019), (Hemphill, 2019), (Hensel & Kacprzak, 2021), (Smyth, 2019), (D'Arcy et al., 2009), (Li et al., 2019), (Whitman et al., 2001)
Prevention	Inoculation (Prebunking)		(Banas & Miller, 2013), (Cook et al., 2017), , (Jolley & Douglas, 2017),(Bolsen & Druckman, 2015), (Roozenbeek & Van Der Linden, 2019), (Roozenbeek & van der Linden, 2019), (Van der Linden et al., 2017), (Basol et al., 2020), (Lewandowsky & Van Der Linden, 2021)
	Inoculating through Education/Misconception-based Learning		(De Paor & Heravi, 2020), (McCuin et al., 2014), (Cook et al., 2014),(Cook, 2022), (Mihailidis & Viotty, 2017), (Kowalski & Taylor, 2009), (Tippett, 2010),(Banks, 2017), (Walton & Hepworth, 2011), (Batchelor, 2017), (Delellis & Rubin, 2018),(Jones-Jang et al., 2021), (Lewandowsky et al., 2017), (Ecker et al., 2017), (Lefkowitz, 2017), (Schuenemann & Cook, 2015)
	Block Malicious Accounts on Social Media		(Batchelor, 2017; Delellis & Rubin, 2018; Jones-Jang et al., 2021; Timberg & Dwoskin, 2018), (Coleman, 2021), (Amoruso et al., 2020), (Zhang et al., 2016), (Ng et al., 2021), (Chakraborty et al., 2016)
Detection	Fact-checking (Manual, Crowdsourcing, Computational)		(Wang, 2017), (Shu et al., 2020), (Hassan et al., 2017), (Babakar, 2018), (Gencheva et al., 2017), (Kim et al., 2018), (Ratkiewicz et al., 2011),(Shao et al., 2016), (Konstantinovskiy et al., 2021), (Shi & Weninger, 2016), (Ciampaglia et al., 2015), (Atanasova et al., 2019)
	Automated Algorithmic Solutions	Machine Learning (ML) & Deep Learning (DL)	ML: (Castillo et al., 2011), (Ma et al., 2015), (Kwon et al., 2017),(Hamidian & Diab, 2019), (Wu et al., 2015), (Shu, Wang, et al., 2019), (Yang et al., 2012), (Vosoughi et al., 2017), (Kumar et al., 2016),(Jin et al., 2016), (Ahmad et al., 2020) DL: (Ma et al., 2016, 2018), (Qian et al., 2018), (Bian et al., 2020),(Wang et al., 2018), (Kaliyar et al., 2020), (Sahoo & Gupta, 2021),(Nasir et al., 2021), (Nguyen et al., 2020), (Yuan et al., 2021)

		Other Methods (spread pattern, statistics, etc.)	(Kim et al., 2018), (Papanastasiou, 2020), (Wang & Terano, 2015),(Wang et al., 2017), (Kumar & Geethakumari, 2014), (Chen et al., 2016)
Mitigation (Remedy)	Minimizing the Influence of Fake News		(Fan et al., 2013), (Kotnis & Kuri, 2014), (Wang et al., 2014), (Wang et al., 2017), (Kimura et al., 2009), (Tambuscio et al., 2015), (He et al., 2015)
	Spreading Truth to discredit fake news		(Budak et al., 2011), (Nguyen et al., 2012), (Tripathy et al., 2010),(Tong et al., 2017), (Yang et al., 2020), (He et al., 2015), (King et al., 2021)
	Platform Interventions (Content-level)		(Moravec et al., 2020; Moravec et al., 2019, 2022), (Kim et al., 2019a), (Kim & Dennis, 2019),(Figl et al., 2019), (Garrett & Poulsen, 2019),(Pennycook, Bear, et al., 2020; Pennycook, McPhetres, et al., 2020), (Ng et al., 2021), (Gimpel et al., 2021)

6 Discussion

Several strategies are proposed to combat fake news on social media, mostly focused on detection approaches. However, fake news detection, although necessary, is not enough to stop fake news on social media. First, manual detection of fake news is time consuming and labor intensive. Automated fake news detection addresses this issue but is less accurate. Also, automated fake news detection often suffers from limited explainability. Second, fake news detection happens only after it is disseminated and consumed by people. Since fake news can have severe harmful impacts in a matter of seconds, it is necessary to devise strategies to stop fake news from happening in the first place. In this paper, we used a framework to address the problem of fake news not only after its propagation, but even before it is created. The framework, which is inspired by the Straub Model of Security Action Cycle includes four stages: deterrence, prevention, detection, and remedy/mitigation. A summary of the approaches to combat fake news on social media and example references for each stage is provided in Table 8.

We next pointed to similarities between fake news and security threats (section 3), but also acknowledged key differences between information security threats and fake news on social media. This makes the implementation of some of the countermeasures more challenging in the context of fake news. As mentioned earlier, one difference between fake news and information security is that in case of fake news, people want to believe false news that fit their ideology, while from a behavioral standpoint, information security threats are primarily due to people's sloppiness in detecting threats. Thus, **one interesting direction for future research is to investigate the ways we can combat people's ideology biases** in relation to fake news. Some of the countermeasure approaches in our framework can be helpful in reducing belief in false information. For example, **accuracy-promoting interventions** such as warnings or nudging users to think about information veracity before sharing it can impact judgements about fake news credibility (Bryanov & Vziatysheva, 2021). One approach is **inoculation interventions** which aims at pre-emptively warn users to the threat of fake news and equipping them with the tools to combat it. For example, **media and information literacy** approaches to educate users about deception strategies (Cook et al., 2017) or **guidelines** to help people detect fake news can be helpful. For example, recent research finds that exposing users to simple guidelines to detect misinformation (e.g., "Be skeptical of headlines," "Watch for unusual formatting") improves fake news discernment rate among both nationally representative samples in the U.S. (by 26.5%) and in India (by 17.5%), regardless of whether the headlines are politically concordant or not (Guess et al., 2020). Another approach is **using labels or flags** to trigger critical thinking. To understand whether some type of flagging is more effective than others, three flagging strategies were examined: fact-checker flags, peer-generated flags, and publishers' self-identified humor flags (Garrett & Poulsen, 2019). They found that publishers' self-identified flags were the most effective strategy in reducing people's beliefs and sharing intentions of fake news.

Another challenge in combating fake news on social media is that there are several motivations for fake news creation and spread, while there are not enough demotivation strategies. Deterrents such as "establishing laws and regulations" can be used to demotivate people from creating or spreading fake news on social media. However, there are **several limitations in effectively implementing deterrents and preventive measures to combat fake news on social media**. ***First*****,** fake news has not been legally treated as a crime and there is no agreement on which criteria to consider when recognizing a fake news as a crime. Also, the term "fake news" has been highly politicized and for example, what is considered as fake news by republicans may be considered true by democrats and vice versa. Therefore, as long as fake news is not recognized as a serious threat and there is no overall agreement on what content to consider as fake news, it will be difficult to devise effective regulations and penalties against it. ***Second*****,** there are different types of fake news with different characteristics. Thus, a single strategy cannot be enough to address the variety of behaviors in the fake news context. Fake news can be created and propagated with intention to deceive (disinformation) or without malicious intention (misinformation). There should be a distinction between users who purposefully create and share fake news and those who erroneously share false content with good intention. For example, deterrent strategies can be helpful to deter malicious users who share disinformation but may be less effective against those who may not know that the content they are sharing is false. However, laws and regulations against fake news can still be effective to some extent (even for users with no bad intention) because they make users to think more carefully before sharing any content on social media. Also, legislations and penalties can further focus on the fake news with more harmful impacts. For example, Burkina Faso's parliament adopted a law to punish the publication of "fake news information compromising security operations, false information about rights abuses or destruction of property, or images and audio from a "terrorist" attack."[15] ***Third*****,** the legal punishment of users who unintentionally share fake news is a violation of free speech. Therefore, in case of sharing false content without intention (misinformation), other strategies such as users' inoculation or education may be more effective. Educating users to increase their awareness about fake news characteristics, its' destructive impacts, and ways to spot and counter it on social media is proved to be very helpful in combating fake news. ***Finally*****,** there is a major concern about the compatibility of fake news prevention and the right for free speech. Fake news deterrents and preventive measures can be interpreted as censorship or violations of the right for free speech. On the other hand, authorities may misuse the preventives to filter the opposing views or even filtering the truth. There are many interesting research questions to explore here, such as: how to balance the prevention of fake news and freedom of speech? How to combat fake news while protecting free speech? How to prevent the true information from being mistakenly blocked?

An additional challenge in combating fake news on social media is how to effectively reduce the harmful consequences of fake news. To address this, we described several remedies such as providing true information. However, it is difficult to disbelieve fake news. Due to the Continued Influence Effect (CIE) of fake news (Johnson & Seifert, 1994), information correction often fails to disbelieve fake news and fake news continues to influence people's thinking even after correction. One explanation is that information correction often requires to repeat fake news. The repetition of fake news increases familiarity, which in turn increases believability in fake news (Lewandowsky et al., 2012). An important question in this stage is how to help people disbelieve fake news once they consumed it?

Ultimately, our review has led us to believe that fake news is a multidisciplinary problem that requires various expertise and should be addressed through collective efforts from different fields. The IS discipline can contribute significantly to the research on fake news. IS scholars can draw on theories and empirical findings on the design, use, and impacts of IT artifacts at different levels of analysis (Dennis et al., 2021; Gimpel et al., 2021; Kim et al., 2019; Kim & Dennis, 2019; Moravec et al., 2019, 2022). The focus on human-technology interactions, and design elements and managerial practices that can influence it is a key feature of IS research and is also a cornerstone feature of research on fighting fake news. Thus, the IS scholars can contribute to all stages of the processes.

In addition, IS researchers can learn from the findings in related areas such as fake reviews (Cheng Nie et al., 2022; Xiao & Benbasat, 2011), social behaviors in online social networks (Kuem et al., 2017) and security (Bulgurcu et al., 2010; D'Arcy et al., 2009). For example, since fake news share some similar characteristics with the context of security, IS research can benefit from applying various approaches used in security to the context of fake news. Moreover, the research in psychology and social science can

[15] https://www.poynter.org/ifcn/anti-misinformation-actions/

shed light on psychological and behavioral factors contributing to creation and spread of fake news. They can help in better understanding why people believe fake news, understanding different types of fake news, and how to break echo-chambers and filter bubbles among like-minded users on social media. Other examples for future research include, but not limited to: 1) How to combat people's ideology biases in relation to fake news? 2) Why some people still continue to believe in fake news, even after it is flagged as false? (continued influence effect), 3) Why sometimes some anti-fake news actions such as information correction backfire and increase the spread of fake news? 4) How to balance the policies and regulations against fake news with the need to freedom of expression? 5) How the various Information Communication Technologies (ICT) impact fake news detection? Such questions and beyond can be addressed by the IS research community.

While we review extant works on fake news, classify them by process stages, propose a framework to understand the vast literature on the topic, propose new research directions, and pave the way for future research, our study has limitations that should be acknowledged. First, we analyze each stage of combating fake news separately. Considering that the paper is already lengthy and complex, we include this point as a promising direction for future research. In essence, we present an important starting point for examining combinations of approaches.

Nevertheless, there were some articles in our review which focused on more than one countermeasure. For example, Ng et al., (2021) used two types of platform interventions: 1) an account-level intervention (forwarding restriction) which is a *"preventive"* approach, and 2) a content-level intervention (flagging fake news) as a *"mitigation/remedy".* Also, several papers (mostly in Computer Science) studied both *"detection" and "mitigation"* to counter fake news (Kim et al., 2018; Papanastasiou, 2020; Sharma et al., 2019; Shu et al., 2019). In addition, Helm & Nasu, (2021) discussed three different countermeasures to combat fake news: 1) information correction (mitigation), 2) blocking or removing contents/accounts (prevention), 3) criminal sanctions (deterrence). However, this was a conceptual study, and they didn't examine different approaches. Based on our review, there is dearth of studies on combining more than two stages (e.g., three or all four stages). Future research can look at these combinations to provide a more comprehensive picture. Instead of micro-level look and seeing only each stage at once, future research can take a more holistic view to combat fake news.

Second, some of the countermeasures to combat fake news may be classified under more than one category. For example, *"Fake news influence minimization: limiting the scope of fake news spread by blocking certain nodes (users)"* is referred to a mitigation strategy in all the highly cited papers (Sharma et al., 2019; Shu, Bernard, et al., 2019). Examples of articles using this approach are (Amoruso et al., 2020; Lin et al., 2019; Shrivastava et al., 2020). In this paper, we classified the articles using this approach under "mitigation" strategy, to be consistent with the literature. However, one may also consider it as a preventive approach because it prevents further spreading of fake news by blocking some nodes/users. Also, we classified "Flagging fake news" as a mitigation/remedy strategy because **first**, it happens after fake news is detected (detection can be done manually by e.g., fact-checkers or automatically by e.g., ML & DL approaches). **Second**, mitigation/remedy is defined as "reducing the harmful impacts of abuse (in the context of this study, reducing negative impacts of fake news". Research shows that flagging fake news reduce the impact of fake news by triggering users critical thinking. **Third**, our focus was on combating fake news on social media platforms. However, fake news spread through different channels: social networking sites such as Twitter or Facebook, fake news websites, and peer-to-peer sharing via e.g., messaging apps such as WhatsApp, Telegram, etc. For fake news on social media platforms such as Facebook, given the right incentive, the platform can more easily implement certain control methods. For peer-to-peer sharing via messaging apps, neither the platform nor the government can easily insert itself in the process". In such cases, some of the countermeasures in our framework such as increasing users' awareness, inoculation, educational campaigns, and media literacy initiatives (mentioned in the deterrence and prevention stages) may be helpful in countering fake news. Future research can further investigate the approaches to address the peer-to-peer sharing of fake news.

Last, we acknowledge that we appropriated the countermeasures in the Straub Model as a set of containers for the individual articles. We are not commenting on the truth value of the framework or whether it is better or worse than any other framework, just that it provides sufficient value for decomposing the articles into logical categories for further analysis. We are not testing the framework as though it were a prediction or theory, we are just using it to provide a basis for analysis of the literature.

7 Conclusion

Combating fake news on social media in an extremely complex and challenging problem which requires a multidisciplinary effort. Scholars across various disciplines from computer science and information systems to social science should work collaboratively to address this serious issue. Our findings suggest that most of the fake news research have focused on *detection methods* and was mostly published in computer science outlets. However, there is an opportunity and need to know more about deterring and blocking the creation and dissemination of fake news before the detection phase, and about reducing their harms and further limiting their spread after they are detected. We believe that the IS community take a more active role in addressing the fake news challenge and propose that efforts can be guided by the provided framework. Taking this holistic view can help IS scholars examine important research areas, and ultimately develop more comprehensive, synergetic multi-stage plans for combating fake news on social media. Fake news is an ongoing phenomenon. It is like a virus that will never disappear, but we need to keep fighting it.

Acknowledgement

This research is sponsored by the Discovery Research Grant from Natural Sciences and Engineering Research Council of Canada. The authors are grateful for the valuable and constructive comments from the editor and anonymous reviewers that helped the improvement of this paper's early manuscript.

References

Adams, R. J., Smart, P., & Huff, A. S. (2017). Shades of grey: Guidelines for working with the grey literature in systematic reviews for management and organizational studies. *International Journal of Management Reviews*, *19*(4), 432–454.

Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020). Fake news detection using machine learning ensemble methods. *Complexity*, *2020*.

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Amoruso, M., Anello, D., Auletta, V., Cerulli, R., Ferraioli, D., & Raiconi, A. (2020). Contrasting the spread of misinformation in online social networks. *Journal of Artificial Intelligence Research*, *69*, 847–879.

Atanasova, P., Nakov, P., Màrquez, L., Barrón-Cedeño, A., Karadzhov, G., Mihaylova, T., Mohtarami, M., & Glass, J. (2019). Automatic fact-checking using context and discourse information. *Journal of Data and Information Quality (JDIQ)*, *11*(3), 1–27.

Babakar, M. (2018, May 29). *Crowdsourced Factchecking*. Full Fact. https://fullfact.org/blog/2018/may/crowdsourced-factchecking/

Bakir, V., & McStay, A. (2018). Fake news and the economy of emotions: Problems, causes, solutions. *Digital Journalism*, *6*(2), 154–175.

Banas, J. A., & Miller, G. (2013). Inducing resistance to conspiracy theory propaganda: Testing inoculation and metainoculation strategies. *Human Communication Research*, *39*(2), 184–207.

Banks, M. (2017). Fighting fake news. *American Libraries*, *48*(3–4), 18–21.

Basol, M., Roozenbeek, J., & van der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition*, *3*(1).

Batchelor, O. (2017). Getting out the truth: The role of libraries in the fight against fake news. *Reference Services Review*.

Bian, T., Xiao, X., Xu, T., Zhao, P., Huang, W., Rong, Y., & Huang, J. (2020). Rumor detection on social media with bi-directional graph convolutional networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(01), 549–556.

Biyani, P., Tsioutsiouliklis, K., & Blackmer, J. (2016). " 8 amazing secrets for getting more clicks": Detecting clickbaits in news streams using article informality. *Thirtieth AAAI Conference on Artificial Intelligence*.

Bolsen, T., & Druckman, J. N. (2015). Counteracting the politicization of science. *Journal of Communication*, *65*(5), 745–769.

Bondielli, A., & Marcelloni, F. (2019). A survey on fake news and rumour detection techniques. *Information Sciences*, *497*, 38–55.

Botha, J., & Pieterse, H. (2020). Fake news and deepfakes: A dangerous threat for 21st century information security. *ICCWS 2020 15th International Conference on Cyber Warfare and Security. Academic Conferences and Publishing Limited*, 57.

Bryanov, K., & Vziatysheva, V. (2021). Determinants of individuals' belief in fake news: A scoping review determinants of belief in fake news. *PLoS One*, *16*(6), e0253717.

Budak, C., Agrawal, D., & El Abbadi, A. (2011). Limiting the spread of misinformation in social networks. *Proceedings of the 20th International Conference on World Wide Web*, 665–674.

Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2010). Information security policy compliance: An empirical study of rationality-based beliefs and information security awareness. *MIS Quarterly*, 523–548.

Carrieri, V., Madio, L., & Principe, F. (2019). Vaccine hesitancy and (fake) news: Quasi-experimental evidence from Italy. *Health Economics*, *28*(11), 1377–1382.

Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on twitter. *Proceedings of the 20th International Conference on World Wide Web*, 675–684.

Chakraborty, A., Paranjape, B., Kakarla, S., & Ganguly, N. (2016). Stop clickbait: Detecting and preventing clickbaits in online news media. *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 9–16.

Chen, W., Yeo, C. K., Lau, C. T., & Lee, B. S. (2016). Behavior deviation: An anomaly detection view of rumor preemption. *2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 1–7.

Chen, Y., Conroy, N. J., & Rubin, V. L. (2015). Misleading online content: Recognizing clickbait as" false news". *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*, 15–19.

Cheng Nie, Zhiqiang (Eric) Zheng, & Sarkar, S. (2022). Competing with the Sharing Economy: Incumbents' Reaction on Review Manipulation. *MIS Quarterly*, *46*(3), 1573–1602. https://doi.org/10.25300/MISQ/2022/15666

Ciampaglia, G. L., Shiralkar, P., Rocha, L. M., Bollen, J., Menczer, F., & Flammini, A. (2015). Computational fact checking from knowledge networks. *PloS One*, *10*(6), e0128193.

Coleman, K. (2021). *Introducing Birdwatch, a community-based approach to misinformation*. https://blog.twitter.com/en_us/topics/product/2021/introducing-birdwatch-a-community-based-approach-to-misinformation

Collins, B., Hoang, D. T., Nguyen, N. T., & Hwang, D. (2021). Trends in combating fake news on social media–a survey. *Journal of Information and Telecommunication*, *5*(2), 247–266.

Compton, J. (2013). Inoculation theory. *The SAGE Handbook of Persuasion: Developments in Theory and Practice*, *2*, 220–237.

Cook, J. (2016). Countering climate science denial and communicating scientific consensus. In *Oxford Research Encyclopedia of Climate Science*.

Cook, J. (2019). *Understanding and countering misinformation about climate change.*

Cook, J. (2022). Understanding and countering misinformation about climate change. *Research Anthology on Environmental and Societal Impacts of Climate Change*, 1633–1658.

Cook, J., Bedford, D., & Mandia, S. (2014). Raising climate literacy through addressing misinformation: Case studies in agnotology-based learning. *Journal of Geoscience Education*, *62*(3), 296–306.

Cook, J., Lewandowsky, S., & Ecker, U. K. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PloS One*, *12*(5), e0175799.

Dale, R. (2017). NLP in a post-truth world. *Natural Language Engineering*, *23*(2), 319–324.

D'Arcy, J., Hovav, A., & Galletta, D. (2009). User awareness of security countermeasures and its impact on information systems misuse: A deterrence approach. *Information Systems Research*, *20*(1), 79–98.

De Paor, S., & Heravi, B. (2020). Information literacy and fake news: How the field of librarianship can help combat the epidemic of fake news. *The Journal of Academic Librarianship*, *46*(5), 102218.

Delellis, N. S., & Rubin, V. L. (2018). Educators' perceptions of information literacy and skills required to spot 'fake news.' *Proceedings of the Association for Information Science and Technology*, *55*(1), 785–787.

Dennis, A. R., Galletta, D. F., & Webster, J. (2021). Special Issue: Fake News on the Internet. *Journal of Management Information Systems*, *38*(4), 893–897. https://doi.org/10.1080/07421222.2021.1990609

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*(3), 629.

Di Domenico, G., Sit, J., Ishizaka, A., & Nunan, D. (2021). Fake news, social media and marketing: A systematic review. *Journal of Business Research*, *124*, 329–341.

Douglas, K. M., & Sutton, R. M. (2008). The hidden impact of conspiracy theories: Perceived and actual influence of theories surrounding the death of Princess Diana. *The Journal of Social Psychology*, *148*(2), 210–222.

Ecker, U. K., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition*, *6*(2), 185–192.

Fan, L., Lu, Z., Wu, W., Thuraisingham, B., Ma, H., & Bi, Y. (2013). Least cost rumor blocking in social networks. *2013 IEEE 33rd International Conference on Distributed Computing Systems*, 540–549.

Farajtabar, M., Yang, J., Ye, X., Xu, H., Trivedi, R., Khalil, E., Li, S., Song, L., & Zha, H. (2017). Fake news mitigation via point process based intervention. *International Conference on Machine Learning*, 1097–1106.

Figl, K., Kießling, S., Rank, C., & Vakulenko, S. (2019). *Fake News Flags, Cognitive Dissonance, and the Believability of Social Media Posts*.

Fisher, M., Cox, J. W., & Hermann, P. (2016). Pizzagate: From rumor, to hashtag, to gunfire in DC. *Washington Post*, *6*.

Flew, T., Martin, F., & Suzor, N. (2019). Internet regulation as media policy: Rethinking the question of digital communication platform governance. *Journal of Digital Media & Policy*, *10*(1), 33–50.

Flynn, D. J., Nyhan, B., & Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology*, *38*, 127–150.

Funke, D., & Flamini, D. (2022, January 21). A guide to anti-misinformation actions around the world. *Poynter*. https://www.poynter.org/ifcn/anti-misinformation-actions/

Garrett, R. K., & Poulsen, S. (2019). Flagging Facebook falsehoods: Self-identified humor warnings outperform fact checker and peer warnings. *Journal of Computer-Mediated Communication*, *24*(5), 240–258.

Gencheva, P., Nakov, P., Màrquez, L., Barrón-Cedeño, A., & Koychev, I. (2017). A context-aware approach for detecting worth-checking claims in political debates. *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*, 267–276.

George, J. F., Gupta, M., Giordano, G., Mills, A. M., Tennant, V. M., & Lewis, C. C. (2018). The effects of communication media and culture on deception detection accuracy. *MIS Quarterly*, *42*(2), 551–575.

Gimpel, H., Heger, S., Olenberger, C., & Utz, L. (2021). The effectiveness of social norms in fighting fake news on social media. *Journal of Management Information Systems*, *38*(1), 196–221.

Golbeck, J., Mauriello, M., Auxier, B., Bhanushali, K. H., Bonk, C., Bouzaghrane, M. A., Buntain, C., Chanduka, R., Cheakalos, P., & Everett, J. B. (2018). Fake news vs satire: A dataset and analysis. *Proceedings of the 10th ACM Conference on Web Science*, 17–21.

Gopal, R. D., & Sanders, G. L. (1997). Preventive and deterrent controls for software piracy. *Journal of Management Information Systems*, *13*(4), 29–47.

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences*, *117*(27), 15536–15545.

Gupta, A., Lamba, H., Kumaraguru, P., & Joshi, A. (2013). Faking sandy: Characterizing and identifying fake images on twitter during hurricane sandy. *Proceedings of the 22nd International Conference on World Wide Web*, 729–736.

Haciyakupoglu, G., Hui, J. Y., Suguna, V. S., Leong, D., & Rahman, M. F. B. A. (2018). *Countering fake news: A survey of recent global initiatives*.

Hamidian, S., & Diab, M. T. (2019). Rumor detection and classification for twitter data. *ArXiv Preprint ArXiv:1912.08926*.

Hartley, K., & Vu, M. K. (2020). Fighting fake news in the COVID-19 era: Policy insights from an equilibrium model. *Policy Sciences*, *53*(4), 735–758.

Hassan, N., Zhang, G., Arslan, F., Caraballo, J., Jimenez, D., Gawsane, S., Hasan, S., Joseph, M., Kulkarni, A., & Nayak, A. K. (2017). Claimbuster: The first-ever end-to-end fact-checking system. *Proceedings of the VLDB Endowment*, *10*(12), 1945–1948.

He, Z., Cai, Z., & Wang, X. (2015). Modeling propagation dynamics and developing optimized countermeasures for rumor spreading in online social networks. *2015 IEEE 35Th International Conference on Distributed Computing Systems*, 205–214.

Helm, R. K., & Nasu, H. (2021). Regulatory responses to 'fake news' and freedom of expression: Normative and empirical evaluation. *Human Rights Law Review*, *21*(2), 302–328.

Hemphill, T. A. (2019). 'Techlash', responsible innovation, and the self-regulatory organization. *Journal of Responsible Innovation*, *6*(2), 240–247.

Hensel, P. G., & Kacprzak, A. (2021). Curbing cyberloafing: Studying general and specific deterrence effects with field evidence. *European Journal of Information Systems*, *30*(2), 219–235.

Ireton, C., & Posetti, J. (2018). *Journalism, fake news & disinformation: Handbook for journalism education and training*. Unesco Publishing.

Jang, S. M., & Kim, J. K. (2018). Third person effects of fake news: Fake news regulation and media literacy interventions. *Computers in Human Behavior*, *80*, 295–302.

Jin, Z., Cao, J., Zhang, Y., & Luo, J. (2016). News verification by exploiting conflicting social viewpoints in microblogs. *Proceedings of the AAAI Conference on Artificial Intelligence*, *30*(1).

Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1420.

Jolley, D., & Douglas, K. M. (2017). Prevention is better than cure: Addressing anti-vaccine conspiracy theories. *Journal of Applied Social Psychology*, *47*(8), 459–469.

Jones-Jang, S. M., Mortensen, T., & Liu, J. (2021). Does media literacy help identification of fake news? Information literacy helps, but other literacies don't. *American Behavioral Scientist*, *65*(2), 371–388.

Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2020). FNDNet–a deep convolutional neural network for fake news detection. *Cognitive Systems Research*, *61*, 32–44.

Kapantai, E., Christopoulou, A., Berberidis, C., & Peristeras, V. (2021). A systematic literature review on disinformation: Toward a unified taxonomical framework. *New Media & Society*, *23*(5), 1301–1326.

Keeley, B. L. (1999). Of conspiracy theories. *The Journal of Philosophy*, *96*(3), 109–126.

Khan, T., Michalas, A., & Akhunzada, A. (2021). Fake news outbreak 2021: Can we stop the viral spread? *Journal of Network and Computer Applications*, *190*, 103112.

Kim, A., & Dennis, A. (2019). Says who? The effects of presentation format and source rating on fake news in social media. *MIS Quarterly*, *43*(3).

Kim, A., Moravec, P., & Dennis, A. (2019a). Combating fake news on social media with source ratings: The effects of user and expert reputation ratings. *Journal of Management Information Systems*, *36*(3), 931–968.

Kim, A., Moravec, P., & Dennis, A. (2019b). When Do Details Matter? Source Rating Summaries and Details in the Fight against Fake News on Social Media. *Source Rating Summaries and Details in the Fight against Fake News on Social Media (September 6, 2019). Kelley School of Business Research Paper*, *19–52*.

Kim, J., Tabibian, B., Oh, A., Schölkopf, B., & Gomez-Rodriguez, M. (2018). Leveraging the crowd to detect and reduce the spread of fake news and misinformation. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 324–332.

Kimura, M., Saito, K., & Motoda, H. (2009). Blocking links to minimize contamination spread in a social network. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *3*(2), 1–23.

King, K. K., Wang, B., Escobari, D., & Oraby, T. (2021). Dynamic Effects of Falsehoods and Corrections on Social Media: A Theoretical Modeling and Empirical Evidence. *Journal of Management Information Systems*, *38*(4), 989–1010.

Kogan, S., Moskowitz, T. J., & Niessner, M. (2019). Fake news: Evidence from financial markets. *Available at SSRN 3237763*.

Konstantinovskiy, L., Price, O., Babakar, M., & Zubiaga, A. (2021). Toward Automated Factchecking: Developing an Annotation Schema and Benchmark for Consistent Automated Claim Detection. *Digital Threats: Research and Practice*, *2*(2), 1–16.

Kotnis, B., & Kuri, J. (2014). Cost effective rumor containment in social networks. *ArXiv Preprint ArXiv:1403.6315*.

Kowalski, P., & Taylor, A. K. (2009). The effect of refuting misconceptions in the introductory psychology class. *Teaching of Psychology*, *36*(3), 153–159.

Kreiss, D., & McGregor, S. C. (2019). The "arbiters of what our voters see": Facebook and Google's struggle with policy, process, and enforcement around political advertising. *Political Communication*, *36*(4), 499–522.

Kuem, J., Ray, S., Siponen, M., & Kim, S. S. (2017). What Leads to Prosocial Behaviors on Social Networking Services: A Tripartite Model. *Journal of Management Information Systems*, *34*(1), 40–70. https://doi.org/10.1080/07421222.2017.1296744

Kumar, K. K., & Geethakumari, G. (2014). Detecting misinformation in online social networks using cognitive psychology. *Human-Centric Computing and Information Sciences*, *4*(1), 1–22.

Kumar, S., & Shah, N. (2018). False information on web and social media: A survey. *ArXiv Preprint ArXiv:1804.08559*.

Kumar, S., West, R., & Leskovec, J. (2016). Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. *Proceedings of the 25th International Conference on World Wide Web*, 591–602.

Kwon, S., Cha, M., & Jung, K. (2017). Rumor detection over varying time windows. *PloS One*, *12*(1), e0168344.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., & Rothschild, D. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096.

Lefkowitz, M. (2017, March). *Library tackles fake news with workshops, resources, advice*. Cornell Chronicle. https://news.cornell.edu/stories/2017/03/library-tackles-fake-news-workshops-resources-advice

Lewandowsky, S., Ecker, U. K., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, *6*(4), 353–369.

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131.

Lewandowsky, S., & Van Der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, *32*(2), 348–384.

Li, L., He, W., Xu, L., Ash, I., Anwar, M., & Yuan, X. (2019). Investigating the impact of cybersecurity policy awareness on employees' cybersecurity behavior. *International Journal of Information Management*, *45*, 13–24.

Lim, S. (2020). Academic library guides for tackling fake news: A content analysis. *The Journal of Academic Librarianship*, *46*(5), 102195.

Lin, Y., Cai, Z., Wang, X., & Hao, F. (2019). Incentive mechanisms for crowdblocking rumors in mobile social networks. *IEEE Transactions on Vehicular Technology*, *68*(9), 9220–9232.

Lozano, M. G., Brynielsson, J., Franke, U., Rosell, M., Tjörnhammar, E., Varga, S., & Vlassov, V. (2020). Veracity assessment of online data. *Decision Support Systems*, *129*, 113132.

Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B. J., Wong, K.-F., & Cha, M. (2016). *Detecting rumors from microblogs with recurrent neural networks*.

Ma, J., Gao, W., Wei, Z., Lu, Y., & Wong, K.-F. (2015). Detect rumors using time series of social context information on microblogging websites. *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 1751–1754.

Ma, J., Gao, W., & Wong, K.-F. (2018). *Rumor detection on twitter with tree-structured recursive neural networks*.

Marabelli, M., Vaast, E., & Li, J. L. (2021). Preventing the digital scars of COVID-19. *European Journal of Information Systems*, *30*(2), 176–192.

McCuin, J. L., Hayhoe, K., & Hayhoe, D. (2014). Comparing the effects of traditional vs. Misconceptions-based instruction on student understanding of the greenhouse effect. *Journal of Geoscience Education*, *62*(3), 445–459.

Meel, P., & Vishwakarma, D. K. (2020). Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Systems with Applications*, *153*, 112986.

Mihailidis, P., & Viotty, S. (2017). Spreadable spectacle in digital culture: Civic expression, fake news, and the role of media literacies in "post-fact" society. *American Behavioral Scientist*, *61*(4), 441–454.

Monsees, L. (2020). 'A war against truth'-understanding the fake news controversy. *Critical Studies on Security*, *8*(2), 116–129.

Moody, G. D., Siponen, M., & Pahnila, S. (2018). Toward a unified model of information security policy compliance. *MIS Quarterly*, *42*(1).

Moravec, P., Kim, A., & Dennis, A. (2020). Appealing to Sense and Sensibility: System 1 and System 2 Interventions for Fake News on Social Media. *Information Systems Research*, *31*(3), 987–1006.

Moravec, P., Kim, A., Dennis, A., & Minas, R. (2018). Do you really know if it's true? How asking users to rate stories affects belief in fake news on social media. *How Asking Users to Rate Stories Affects Belief in Fake News on Social Media (October 22, 2018). Kelley School of Business Research Paper*, *18–89*.

Moravec, P. L., Kim, A., Dennis, A. R., & Minas, R. K. (2022). Do you really know if it's true? How asking users to rate stories affects belief in fake news on social media. *Information Systems Research*.

Moravec, P. L., Minas, R. K., & Dennis, A. (2019). Fake News on Social Media: People Believe What They Want to Believe When it Makes No Sense At All. *MIS Quarterly*, *43*(4), 1343–1360.

Moravec, P., Minas, R., & Dennis, A. (2018). Fake news on social media: People believe what they want to believe when it makes no sense at all. *Kelley School of Business Research Paper*, *18–87*.

Morgan, S. (2018). Fake news, disinformation, manipulation and online tactics to undermine democracy. *Journal of Cyber Policy*, *3*(1), 39–43.

Nance, W. D., & Straub, D. W. (1988). *AN INVESTIGATION INTO THE USE AND USEFULNESS OF SECURITY SO TWARE IN DETECTING COMPUTER ABUSE*.

Nasir, J. A., Khan, O. S., & Varlamis, I. (2021). Fake news detection: A hybrid CNN-RNN based deep learning approach. *International Journal of Information Management Data Insights*, *1*(1), 100007.

National Institute of Standards and Technology, F. (2006). *Minimum Security Requirements for Federal Information and Information Systems*.

Ng, K. C., Tang, J., & Lee, D. (2021). The Effect of Platform Intervention Policies on Fake News Dissemination and Survival: An Empirical Examination. *Journal of Management Information Systems*, *38*(4), 898–930.

Nguyen, N. P., Yan, G., Thai, M. T., & Eidenbenz, S. (2012). Containment of misinformation spread in online social networks. *Proceedings of the 4th Annual ACM Web Science Conference*, 213–222.

Nguyen, V.-H., Sugiyama, K., Nakov, P., & Kan, M.-Y. (2020). Fang: Leveraging social context for fake news detection using graph representation. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 1165–1174.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220.

Nugent, C. (2018). *How France's Potential Law Banning Fake News Could Work*. Time. https://time.com/5304611/france-fake-news-law-macron/

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303–330.

Ognyanova, K., Lazer, D., Robertson, R. E., & Wilson, C. (2020). Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. *Harvard Kennedy School Misinformation Review*.

Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. *ArXiv Preprint ArXiv:1107.4557*.

Papageorgis, D., & McGuire, W. J. (1961). The generality of immunity to persuasion produced by pre-exposure to weakened counterarguments. *The Journal of Abnormal and Social Psychology*, *62*(3), 475.

Papanastasiou, Y. (2020). Fake news propagation and detection: A sequential model. *Management Science*, *66*(5), 1826–1846.

Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science*, *66*(11), 4944–4957.

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, *147*(12), 1865.

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, *31*(7), 770–780.

Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences*, *116*(7), 2521–2526.

Petratos, P. N. (2021). Misinformation, disinformation, and fake news: Cyber risks to business. *Business Horizons*, *64*(6), 763–774.

Qian, F., Gong, C., Sharma, K., & Liu, Y. (2018). Neural User Response Generator: Fake News Detection with Collective User Intelligence. *IJCAI*, *18*, 3834–3840.

Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., & Menczer, F. (2011). Truthy: Mapping the spread of astroturf in microblog streams. *Proceedings of the 20th International Conference Companion on World Wide Web*, 249–252.

Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, *5*(1), 1–10.

Roozenbeek, J., & Van Der Linden, S. (2019). The fake news game: Actively inoculating against the risk of misinformation. *Journal of Risk Research*, *22*(5), 570–580.

Ross, L., & Ward, A. (1996). Naive realism in everyday life: Implications for social conflict and misunderstanding. *Values and Knowledge*, *103*, 135.

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), 1–10.

Rubin, V. L. (2019). Disinformation and misinformation triangle: A conceptual model for "fake news" epidemic, causal factors and interventions. *Journal of Documentation*.

Rubin, V. L., Chen, Y., & Conroy, N. K. (2015). Deception detection for news: Three types of fakes. *Proceedings of the Association for Information Science and Technology*, *52*(1), 1–4.

Sahoo, S. R., & Gupta, B. B. (2021). Multiple features based approach for automatic fake news detection on social networks using deep learning. *Applied Soft Computing*, *100*, 106983.

Schuenemann, K. C., & Cook, J. (2015). Using" Making Sense of Climate Science Denial" MOOC videos in a college course. *AGU Fall Meeting Abstracts*, *2015*, ED12A-01.

Shao, C., Ciampaglia, G. L., Flammini, A., & Menczer, F. (2016). Hoaxy: A platform for tracking online misinformation. *Proceedings of the 25th International Conference Companion on World Wide Web*, 745–750.

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *10*(3), 1–42.

Shi, B., & Weninger, T. (2016). Fact checking in heterogeneous information networks. *Proceedings of the 25th International Conference Companion on World Wide Web*, 101–102.

Shin, J., Jian, L., Driscoll, K., & Bar, F. (2018). The diffusion of misinformation on social media: Temporal pattern, message, and source. *Computers in Human Behavior*, *83*, 278–287.

Shirish, A., Srivastava, S. C., & Chandra, S. (2021). Impact of mobile connectivity and freedom on fake news propensity during the COVID-19 pandemic: A cross-country empirical examination. *European Journal of Information Systems*, *30*(3), 322–341.

Shore, J., Baek, J., & Dellarocas, C. (2018). Network structure and patterns of information diversity on twitter. *MIS Quarterly*, *42*(3), 849–872. https://doi.org/10.25300/MISQ/2018/14558

Shrivastava, G., Kumar, P., Ojha, R. P., Srivastava, P. K., Mohan, S., & Srivastava, G. (2020). Defensive modeling of fake news through online social networks. *IEEE Transactions on Computational Social Systems*, *7*(5), 1159–1167.

Shu, K., Bernard, H. R., & 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* (pp. 43–65). Springer.

Shu, K., Dumais, S., Awadallah, A. H., & Liu, H. (2020). Detecting fake news with weak social supervision. *IEEE Intelligent Systems*, *36*(4), 96–103.

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & 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), 171–188.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, *19*(1), 22–36.

Shu, K., Wang, S., & Liu, H. (2019). Beyond news contents: The role of social context for fake news detection. *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, 312–320.

Silverman, C. (2017). What exactly is fake news. *The Fake Newsletter*, *26*.

Smyth, S. M. (2019). The Facebook Conundrum: Is it Time to Usher in a New Era of Regulation for Big Tech? *International Journal of Cyber Criminology*, *13*(2), 578–595.

Straub, D. W., & Welke, R. J. (1998). Coping with systems risk: Security planning models for management decision making. *MIS Quarterly*, 441–469.

Sunstein, C. R. (1999). The law of group polarization. *University of Chicago Law School, John M. Olin Law & Economics Working Paper*, *91*.

Tambuscio, M., Ruffo, G., Flammini, A., & Menczer, F. (2015). Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks. *Proceedings of the 24th International Conference on World Wide Web*, 977–982.

Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2018). Defining “fake news” A typology of scholarly definitions. *Digital Journalism*, *6*(2), 137–153.

Timberg, C., & Dwoskin, E. (2018). Twitter is sweeping out fake accounts like never before, putting user growth at risk. *Washington Post*. https://www.washingtonpost.com/technology/2018/07/06/twitter-is-sweeping-out-fake-accounts-like-never-before-putting-user-growth-risk/

Tippett, C. D. (2010). Refutation text in science education: A review of two decades of research. *International Journal of Science and Mathematics Education*, *8*(6), 951–970.

Tong, G., Wu, W., Guo, L., Li, D., Liu, C., Liu, B., & Du, D.-Z. (2017). An efficient randomized algorithm for rumor blocking in online social networks. *IEEE Transactions on Network Science and Engineering*, *7*(2), 845–854.

Tripathy, R. M., Bagchi, A., & Mehta, S. (2010). A study of rumor control strategies on social networks. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 1817–1820.

Turel, O., & Osatuyi, B. (2021). Biased Credibility and Sharing of Fake News on Social Media: Considering Peer Context and Self-Objectivity State. *Journal of Management Information Systems*, *38*(4), 931–958.

Van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, *1*(2), 1600008.

van Der Linden, S., Roozenbeek, J., & Compton, J. (2020). Inoculating against fake news about COVID-19. *Frontiers in Psychology*, *11*, 2928.

Vosoughi, S., Mohsenvand, M. ‘Neo,’ & Roy, D. (2017). Rumor gauge: Predicting the veracity of rumors on Twitter. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *11*(4), 1–36.

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

Walton, G., & Hepworth, M. (2011). A longitudinal study of changes in learners’ cognitive states during and following an information literacy teaching intervention. *Journal of Documentation*.

Wang, B., Chen, G., Fu, L., Song, L., & Wang, X. (2017). Drimux: Dynamic rumor influence minimization with user experience in social networks. *IEEE Transactions on Knowledge and Data Engineering*, *29*(10), 2168–2181.

Wang, N., Yu, L., Ding, N., & Yang, D. (2014). *Containment of misinformation propagation in online social networks with given deadline*.

Wang, S., Moise, I., Helbing, D., & Terano, T. (2017). Early signals of trending rumor event in streaming social media. *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, *2*, 654–659.

Wang, S., & Terano, T. (2015). Detecting rumor patterns in streaming social media. *2015 IEEE International Conference on Big Data (Big Data)*, 2709–2715.

Wang S.A., Min-Seok Pang, & Pavlou, P. A. (2022). Seeing Is Believing? How Including a Video in Fake News Influences Users’ Reporting of Fake News to Social Media Platforms. *MIS Quarterly*, *46*(3), 1323–1353. https://doi.org/10.25300/MISQ/2022/16296

Wang, W. Y. (2017). “ liar, liar pants on fire”: A new benchmark dataset for fake news detection. *ArXiv Preprint ArXiv:1705.00648*.

Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Su, L., & Gao, J. (2018). Eann: Event adversarial neural networks for multi-modal fake news detection. *Proceedings of the 24th Acm Sigkdd International Conference on Knowledge Discovery & Data Mining*, 849–857.

Wardle, C. (2017). Fake news. It’s complicated. *First Draft*, *16*, 1–11.

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, *9*(11).

Whitman, M. E., Townsend, A. M., & Aalberts, R. J. (2001). Information systems security and the need for policy. In *Information security management: Global challenges in the new millennium* (pp. 9–18). IGI Global.

Wu, K., Yang, S., & Zhu, K. Q. (2015). False rumors detection on sina weibo by propagation structures. *2015 IEEE 31st International Conference on Data Engineering*, 651–662.

Xiao, B., & Benbasat, I. (2011). Product-related deception in e-commerce: A theoretical perspective. *Mis Quarterly*, 169–195.

Yang, F., Liu, Y., Yu, X., & Yang, M. (2012). Automatic detection of rumor on sina weibo. *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*, 1–7.

Yang, L., Li, Z., & Giua, A. (2020). Containment of rumor spread in complex social networks. *Information Sciences*, *506*, 113–130.

Yu, P., Xia, Z., Fei, J., & Lu, Y. (2021). A survey on deepfake video detection. *IET Biometrics*, *10*(6), 607–624.

Yuan, H., Zheng, J., Ye, Q., Qian, Y., & Zhang, Y. (2021). Improving fake news detection with domain-adversarial and graph-attention neural network. *Decision Support Systems*, *151*, 113633.

Zannettou, S., Sirivianos, M., Blackburn, J., & Kourtellis, N. (2019). The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans. *Journal of Data and Information Quality (JDIQ)*, *11*(3), 1–37.

Zhang, H., Alim, M. A., Li, X., Thai, M. T., & Nguyen, H. T. (2016). Misinformation in online social networks: Detect them all with a limited budget. *ACM Transactions on Information Systems (TOIS)*, *34*(3), 1–24.

Zhang, H., Zhang, H., Li, X., & Thai, M. T. (2015). Limiting the spread of misinformation while effectively raising awareness in social networks. *International Conference on Computational Social Networks*, 35–47.

Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management*, *57*(2), 102025.

Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, *53*(5), 1–40.

Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. (2018). Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys (CSUR)*, *51*(2), 1–36.

Appendix A:

Table 9. List of Reviewed Articles Classified by Fake News Combat Stage (this table only contains papers relevant to combating fake news, excluding review papers, conceptual papers, etc.)

	Article	Database(s)	Source	Source Type	Publisher	Citation	Year
Combat (Deterrence)	(Helm & Nasu, 2021)	Google Scholar	Human Rights Law Review	Journal	Oxford University Press	11	2021
	(Li et al., 2019)	Google Scholar/ Scopus/ ScienceDirect	International Journal of Information Management	Journal (IS/ CS/ SS)	Elsevier	172	2019
	(Rubin, 2019)	Google Scholar	Journal of documentation	Journal (CS/IS/SS)	Emerald	83	2019
	(Lewandowsky et al., 2017)	Google Scholar/ Scopus	Journal of applied research in memory and cognition	Journal (Psychology)	Elsevier	1099	2017
	(D'Arcy et al., 2009)	Google Scholar/ EBSCOhost/ ProQuest/ JSTOR	Information systems research	IS Journal	INFORMS	1557	2009
	(Moody et al., 2018)	Google Scholar/ EBSCOhost	MIS quarterly	IS Journal	ACM Digital Library	367	2018
	(Hensel & Kacprzak, 2021)	Google Scholar	EJIS (European Journal of Information Systems)	IS Journal	Taylor & Francis	11	2021
	(Hartley & Vu, 2020)	Google Scholar/ EBSCOhost/ ProQuest	Policy Sciences	Journal (SS)	Springer	89	2020
	(Flew et al., 2019)	Google Scholar	Journal of Digital Media and Policy	Journal	Intellect	102	2019
	(Hemphill, 2019)	Google Scholar	Journal of Responsible Innovation	Journal (Business/IS)	Taylor & Francis	8	2019
	(Morgan, 2018)	Google Scholar	Journal of Cyber Policy	Journal	Taylor & Francis	133	2018
	(Smyth, 2019)	Google Scholar/ ProQuest	International Journal of Cyber Criminology	Journal (SS)	Not provided	10	2019
	(Haciyakupoglu et al., 2018)	Google Scholar	Rajaratnam School of International Studies (RSiS)	Report	RSiS	61	2018
	(Jang & Kim, 2018)	Google Scholar/ Scopus/ ScienceDirect	Computers in Human Behavior	Journal (SS/CS)	Elsevier	321	2018
	(Whitman et al., 2001)	Google Scholar/ Scopus/	Information Security Management: Global challenges in the new millennium	Book	IGI Global	112	2001
	(Kreiss & McGregor, 2019)	Google Scholar/ EBSCOhost	Political Communication	Journal (SS)	Taylor & Francis	71	2019
Combat (Prevention)	(Ng et al., 2021)	Google Scholar/ Scopus/ EBSCOhost	Journal of Management Information Systems (JMIS)	Journal (IS)	Taylor & Francis	3	2021
	(Zhang et al., 2016)	Google Scholar/ Scopus/ACM	ACM Transactions on Information Systems	Journal	ACM	62	2016
	(Gopal & Sanders, 1997)	Google Scholar/ EBSCOhost/ ProQuest/JSTOR	JMIS (Journal of Management Information Systems)	Journal (IS)	Taylor & Francis	405	1997
	(Marabelli et al., 2021)	Google Scholar	EJIS (European Journal of Information Systems)	Journal (IS)	Taylor & Francis	28	2021

	(Banas & Miller, 2013)	Google Scholar/ EBSCOhost/Scopus/ Web of Science	Human Communication Research	Journal (Psychology/Social Science)	Oxford Univ. Press	150	2013
	(Batchelor, 2017)	Google Scholar/ Scopus/Emerald	Reference Services Review	Journal (SS)	Emerald	118	2017
	(Jolley & Douglas, 2017)	Google Scholar/ Scopus/	Journal of Applied Social Psychology	Journal (Psychology)	Wiley	265	2017
	(Chakraborty et al., 2016)	Google Scholar/ Scopus	IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)	Conference	IEEE	372	2016
	(Cook et al., 2017)	Google Scholar/ Scopus	PloS one	Journal (Multidisciplinary)	Public Library of Science	541	2017
	(Cook, 2016)	Google Scholar	Oxford Research Encyclopedia of Climate Science	Book		86	2016
	(Bolsen & Druckman, 2015)	Scholar/EBSCOhost/ Scopus	Journal of Communication	Journal (SS)	Oxford Univ. Press	188	2015
	(Roozenbeek & Van Der Linden, 2019)	Scholar/EBSCOhost/ Scopus	Journal of Risk Research	Journal (Business/ Eng/SS)	Taylor & Francis	277	2019
	(Roozenbeek & van der Linden, 2019)	Google Scholar/ Scopus	Palgrave Communications	Journal (Economy, SS, Psychology)	Palgrave	296	2019
	(Basol et al., 2020)	Google Scholar/ Scopus	Journal of Cognition	Journal	Ubiquity Press	146	2020
	(Van der Linden et al., 2017)	Google Scholar	Global Challenges	Journal	Wiley Online Library	614	2017
	(Papageorgis & McGuire, 1961)	Google Scholar/ Scopus	The Journal of Abnormal and Social Psychology	Journal (Psychology)	American Psychological Association	233	1961
	(Lewandowsky & Van Der Linden, 2021)	Google Scholar/ Scopus	European Review of Social Psychology	Journal (Psychology/SS)	Taylor & Francis	113	2021
	(Ecker et al., 2017)	Google Scholar/ ScienceDirect	Journal of Applied Research in Memory and Cognition}	Journal (Psychology)	Elsevier	222	2017
	(Cook et al., 2014)	Google Scholar/ Scopus	Journal of Geoscience Education	Journal (SS/Geo)	Taylor & Francis	60	2014
	(Cook, 2019)	Google Scholar	Handbook of research on deception, fake news, and misinformation online	Book	IGI global	59	2019
	(De Paor & Heravi, 2020)	Google Scholar/ Scopus	The Journal of Academic Librarianship	Journal (SS)	Elsevier	58	2020
	(Delellis & Rubin, 2018)	Google Scholar/ Scopus	Proceedings of the Association for Information Science and Technology	Journal (SS)	Wiley	10	2018
	(Kowalski & Taylor, 2009)	Google Scholar/ Scopus	Teaching of Psychology	Journal (Psych/SS)	SAGE Publications	234	2009
	(Walton & Hepworth, 2011)	Google Scholar/ Scopus/ProQuest	Journal of Documentation	Journal (SS/ IS)	Emerald	124	2011

	(Jones-Jang et al., 2021)	Google Scholar/ Scopus	American Behavioral Scientist	Journal (Psych/SS)	SAGE Publications	272	2021
	(McCuin et al., 2014)	Google Scholar/ Scopus/ ScienceDirect	Journal of Geoscience Education	Journal (SS/ Education)	Taylor & Francis	48	2014
	(Mihailidis & Viotty, 2017)	Google Scholar/ Scopus	American Behavioral Scientist	Journal (Psych/SS)	SAGE Publications	439	2017
Combat (Detection)	(Nasir et al., 2021)	Google Scholar/ Scopus/ ScienceDirect	International Journal of Information Management Data Insights	Journal (IS, CS, SS)	Elsevier	126	2021
	(Sahoo & Gupta, 2021)	Google Scholar/ Scopus/ ScienceDirect	Applied Soft Computing	Journal (CS)	Elsevier	112	2021
	(Kwon et al., 2017)	Google Scholar/ Scopus	PLoS one	Journal	PLOS	322	2017
	(Kaliyar et al., 2020)	Google Scholar/ Scopus/ ScienceDirect	Cognitive Systems Research	Journal (CS, Psych., NeuroSci.)	Elsevier	140	2020
	(Ahmad et al., 2020)	Google Scholar/ Scopus	Complexity	Journal	Hindawi	124	2020
	(Atanasova et al., 2019)	Google Scholar/ Scopus/ACM	Journal of Data and Information Quality	Journal	ACM	38	2019
	(W. Chen et al., 2016)	Google Scholar/ Scopus/IEEE	IEEE Annual Information Technology, Electronics and Mobile Communication Conference	Conference	IEEE	27	2016
	(Shu, Wang, et al., 2019)	Google Scholar/ Scopus/ACM	ACM international conference on web search and data mining	Conference	ACM	373	2019
	(Wang & Terano, 2015)	Google Scholar/ Scopus/IEEE	IEEE Big Data	Conference	IEEE	67	2015
	(Castillo et al., 2011)	Google Scholar/ Scopus/ACM	WWW	Conference	ACM	2493	2011
	(Ma et al., 2015)	Google Scholar/ Scopus/ACM	ACM	Conference	ACM	496	2015
	(Ma et al., 2016)	Google Scholar/ Scopus	IJCAI International Joint Conference on Artificial Intelligence	Conference	AAAI Press	819	2016
	(Qian et al., 2018)	Google Scholar/ Scopus	IJCAI International JointConference on Artificial Intelligence	Conference	IJCAI	141	2018
	(Wu et al., 2015)	Google Scholar/ IEEE	IEEE 31st international conference on data engineering	Conference	IEEE	498	2015
	(Bian et al., 2020)	Google Scholar/ Scopus	AAAI Conference on Artificial Intelligence	Conference	PKP/OJS	175	2020
	(Ma et al., 2018)	Google Scholar/ Scopus	Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018)	Conference	ACL (Association for Computational Linguistics)	337	2018
	(Y. Wang et al., 2018)	Google Scholar/ Scopus/ ACM	ACM SIGKDD International Conference on Knowledge Discovery & Data Mining}	Conference	ACL (Association for Computational Linguistics)	480	2018

	(Shao et al., 2016)	Google Scholar/ Scopus/ ACM	Proceedings of the 25th International Conference on World Wide Web	Conference	ACM	384	2016
	(F. Yang et al., 2012)	Google Scholar/ Scopus/ ACM	ACM SIGKDD	Workshop	ACM	557	2012
	(Vosoughi et al., 2017)	Google Scholar/ Scopus/ ACM	ACM transactions on knowledge discovery from data (TKDD)	Journal (CS)	ACM	185	2017
	(S. Kumar et al., 2016)	Google Scholar/ Scopus/ ACM	WWW	Conference	ACM	304	2016
	(Jin et al., 2016)	Google Scholar/ Scopus	AAAI Conference on Artificial Intelligence	Conference	PKP/OJS	356	2016
	(J. Kim et al., 2018)	Google Scholar/ Scopus/ ACM	ACM conference on web search and data mining	Conference	ACM	197	2018
	(Papanastasiou, 2020)	Google Scholar/ Scopus	Management Science	IS Journal	INFORMS	81	2020
	(Shu, Dumais, et al., 2020)	Google Scholar/ Scopus/IEEE	IEEE Intelligent Systems	Journal	IEEE	14	2020
	(George et al., 2018)	Google Scholar/ EBSCOhost	MIS Quarterly	IS Journal	MISQ	43	2018
	(Konstantinovskiy et al., 2021)	Google Scholar/ Scopus/ ACM	Digital Threats: Research and Practice	Journal	ACM New York	76	2021
	(Hassan et al., 2017)	Google Scholar/ ACM	Proceedings of the VLDB Endowment	Conference	VLDB Endowment	203	2017
	(Gencheva et al., 2017)	Google Scholar/ Scopus	RANLP 2017	Conference	ACL (Association for Computational Linguistics)	77	2017
	(Ratkiewicz et al., 2011)	Google Scholar/ Scopus/ACM	*Proceedings of the 20th International Conference Companion on World Wide Web* (WWW)	Conference	ACM	519	2011
	(Ciampaglia et al., 2015)	Google Scholar/ Scopus	PLOS one	Journal	PLOS	494	2015
	(Shi & Weninger, 2016)	Google Scholar/ Scopus/ACM	Proceedings of the 25th International Conference Companion on World Wide Web (WWW)	Conference	ACM	80	2016
	(Yuan et al., 2021)	Google Scholar/ Scopus /ScienceDirect	DSS (Decision Support Systems)	IS Journal	Elsevier	14	2021
	(K. K. Kumar & Geethakumari, 2014)	Google Scholar/ Scopus	Human-centric Computing and Information Sciences	Journal (CS)	SpringerOpen	240	2014
	(Gupta et al., 2013)	Google Scholar/ Scopus/ACM	Proceedings of the 22nd International Conference on World Wide Web (WWW)	Conference	ACM	683	2013
	(Lim, 2020)	Google Scholar/ Scopus /ScienceDirect	Journal of Academic Librarianship	Journal (SS/ Education)	Elsevier	22	2020
	(Biyani et al., 2016)	Google Scholar	AAAI conference on artificial intelligence	Conference		167	2016

Combat (Mitigation/Remedy)	(King et al., 2021)	Google Scholar/ Scopus /EBSCOhost	Journal of Management Information Systems (JMIS)	Journal (IS)	Taylor & Francis	1	2021
	(Tripathy et al., 2010)	Google Scholar/ Scopus/ACM	ACM international conference on Information and knowledge management	Conference	ACM	157	2010
	(N. P. Nguyen et al., 2012)	Google Scholar/ Scopus/ACM	ACM Web Science Conference	Conference	N/A	254	2012
	(Budak et al., 2011)	Google Scholar/ Scopus/ACM	Proceedings of the 20th International Conference on World Wide Web	Conference	N/A	895	2011
	(L. Yang et al., 2020)	Google Scholar/ Scopus /ScienceDirect	Information Sciences	Journal (IS)	Elsevier	90	2020
	(Tong et al., 2017)	Google Scholar/ Scopus	IEEE Transactions on Network Science and Engineering	Journal (CS)	IEEE	112	2017
	(H. Zhang et al., 2015)	Google Scholar/ Scopus	International Conference on Computational Social Networks	Conference	Springer	55	2015
	(He et al., 2015)	Google Scholar/ Scopus/IEEE	IEEE 35Th international conference on distributed computing systems	Conference	IEEE	115	2015
	(Tambuscio et al., 2015)	Google Scholar/ Scopus/ACM	Proceedings of the 20th International Conference on World Wide Web	Conference	ACM	164	2015
	(Figl et al., 2019)	Google Scholar/ Scopus	International Conference on Information Systems (ICIS)	Conference (IS)	Association for Information Systems (AIS)	9	2019
	(Fan et al., 2013)	Google Scholar/ Scopus/IEEE	IEEE 33Th international conference on distributed computing systems	Conference	IEEE	129	2013
	(Wang et al., 2017)	Google Scholar/ Scopus/IEEE	IEEE Transactions on Knowledge and Data Engineering	Journal (CS, IS)	IEEE	136	2017
	(Kimura et al., 2009)	Google Scholar/ Scopus/ACM	ACM Transactions on Knowledge Discovery from Data (TKDD)	Journal (CS)	ACM	235	2009
	(Lin et al., 2019)	Google Scholar/ Scopus/IEEE	IEEE Transactions on Vehicular Technology	Journal	IEEE	45	2019
	(Shrivastava et al., 2020)	Google Scholar/ Scopus/IEEE	IEEE Transactions on Computational Social Systems	Journal	IEEE	51	2020
	(Amoruso et al., 2020)	Google Scholar/ Scopus/ACM	Journal of Artificial Intelligence Research	CS Journal	AI Access Foundation	47	2020
	(Nyhan & Reifler, 2010)	Google Scholar/ Scopus/JSTOR	Political Behavior	Journal (Politics)	Springer	2755	2010
	(Farajtabar et al., 2017)	Google Scholar/ Scopus	International Conference on Machine Learning	Conference	N/A	163	2017
	(Gimpel et al., 2021)	Google Scholar/ Scopus	JMIS (Journal of Managament Information Systems)	IS Journal	Taylor & Francis	23	2021
	(Garrett & Poulsen, 2019)	Google Scholar/ Scopus /EBSCOhost	Journal of Computer-Mediated Communication	Journal	Oxford University Press	37	2019

	(A. Kim & Dennis, 2019)	Google Scholar/ Scopus /EBSCOhost	MIS Quarterly	IS Journal	MISQ	186	2019
	(A. Kim et al., 2019a)	Google Scholar/ Scopus	JMIS (Journal of Management Information Systems)	IS Journal	Taylor & Francis	164	2019
	(P. L. Moravec et al., 2019)	Google Scholar/ Scopus	MIS Quarterly	IS Journal	MISQ	173	2019
	(P. L. Moravec et al., 2022)	Google Scholar/ Scopus/INFORMS	Information Systems Research	IS Journal	INFORMS	6	2022
	(P. Moravec et al., 2020)	Google Scholar/ Scopus/INFORMS	Information Systems Research	IS Journal	INFORMS	50	2020
	(Pennycook, Bear, et al., 2020)	Google Scholar/ Scopus/INFORMS	Management Science	IS Journal	INFORMS	338	2020
	(Pennycook & Rand, 2019)	Google Scholar/ Scopus/JSTOR	Proceedings of the National Academy of Sciences	Journal (Multidisciplinary)	National Acad Sciences	448	2019
	(Pennycook, McPhetres, et al., 2020)	Google Scholar/ Scopus	Psychological science	Journal	Sage Publications	1143	2020

Appendix B

Table 10. List of Several Literature Review Papers about Combating Fake News on Social Media

Review Article	Combat Stage	Summary/Highlights	Type of False Information
(Shu et al., 2017)	Detection	• They provided a review of fake news detection on social media from a data mining perspective. • They classified detection models into: Content Models: knowledge-based, style-based, and Context Models: stance-based, propagation-based. T • hey also provided a characterization of fake news based on psychology and social theories.	Fake News
(S. Kumar & Shah, 2018)	Detection	• Provided a comprehensive review of literature on the spread of false information from diverse aspects: actors (spreaders), rationale (why), impacts, characteristics, and algorithms. • Classified detection algorithms into Feature-based, Graph-based, Model-based (Temporal, Propagation models). • They also categorized existing research based on the platform they studied.	Fake News, Fake Reviews, Hoaxes
(Zubiaga et al., 2018)	Detection	• Provided an overview of research about rumors on social media, with the goal of developing a rumor classification system to detect and resolve the veracity of rumors. • Proposed 4 components in the architecture of rumor classification system: detection, tracking, stance classification, and veracity classification	Rumors
(Shu, Bernard, et al., 2019)	Detection, Mitigation	• Reviewed recent methods to study fake news using network properties and how to use these networks for fake news detection and mitigation on social media. • Classified detection methods based on Network properties into: Interaction network embedding, Temporal diffusion, Friendship network embedding, and Knowledge network matching. • Presented news spread ecosystem in 3 dimensions: content, social, and temporal (but they didn't review the literature based on these dimensions)	Fake News
(Zannettou et al., 2019)	Detection, Mitigation	• They proposed four lines of works to study false information on OSNs (user perception, propagation, detection, and politics). • Classified detection methods into Machine learning, Systems, and Other Models/Algorithms. • Also, for each article, they also provided the platform, methodology, and type of false information studied.	Rumors, Hoaxes, Conspiracy Theories, Satire, Clickbait, Fabricated
(Sharma et al., 2019)	Detection, Mitigation	• They reviewed existing methods to detect and mitigate fake news. • They identified 3 characteristics for fake news detection: source, content, and user responses. They classified the existing works into three categories: content-based identification, feedback-based methods (based on user responses), and intervention-based (early identification and containment of fake news)	Fake News, Rumor
(Bondielli & Marcelloni, 2019)	Detection	• They provided a review of different approaches to detect fake news and rumors. They focused on detection techniques: classification approaches (machine learning and deep learning methods), and other techniques (e.g., diffusion patterns, crowdsourcing, etc.)	Fake News, Rumor
(X. Zhang & Ghorbani, 2020)	Detection	• They provided a comprehensive review of online fake news and analyzed it based on four components of fake news: Content, Creator/Spreader, Target/User analysis, and Social context.	Fake News, Fake Review, Rumor or Satire

		• They provided 3 different perspectives to classify the fake news detection approaches:
		  ○ Component-based (Creator analysis, Content analysis, Context analysis)
		  ○ Data mining-based (Supervised learning, Unsupervised learning)
		  ○ Implementation-based (Online/Real-time, Offline detection)
		• They compared different fake news datasets. They proposed a comprehensive fake news detection ecosystem.	

(Zhou & Zafarani, 2020)	Detection	• Provided a comprehensive and systematic overview of fake news research. • Presented fake news lifecycle: creation, publication, propagation. However, they didn't review the literature based on the fake news life cycle. • Proposed 4 perspectives to study fake news: Knowledge, Style, Propagation, and Credibility. They also provided the comparison of the four perspectives and their connection to each stage of fake news lifecycle.	Fake News
(Kapantai et al., 2021)	N/A	• A systematic review on disinformation • Spread and impact of fake news • Collect the various implicit and explicit disinformation typologies proposed by scholars. • Propose three independent dimensions with controlled values per dimension as categorization criteria for all types of disinformation • Excluded studies that addressed fake news problem from computational perspective (e.g., technical approaches for fake news detection)	Disinformation
(Di Domenico et al., 2021)	N/A	• Interdisciplinary and systematic review of the literature on fake news, from a marketing perspective • Implications of social media fake news for consumers, and for companies • Spreading patterns of fake news and its consequences on consumers and firms • Propose a theoretical framework that highlights themes' relationships and research propositions • They excluded studies on fake news detection	Fake news
(Collins et al., 2021)	Detection	• Different types of fake news and trends in combating them • Various methods of combating fake news on social media (all methods are detection): Expert Fact-check, Machine Leaning (ML), Deep Learning (DL), Natural Language Processing (NLP), Crowdsource, Graph-based, Hybrid (expert-crowd source, human-machine approach), Recommender Systems (RS)	Fake news & misinformation
(Lozano et al., 2020)	Detection	• Systematic literature review • Veracity assessment of online (open source) data • Approaches, methods, algorithms, and tools that are used or proposed for automatic veracity assessment of open-source data • Detection and propagation methods • Only included studies from 2013 to 2017	Fake news & misinformation
(Meel & Vishwakarma, 2020)	Detection, Mitigation	• False information ecosystem, from creation to disposition • Propagation models • Approaches for detecting false information (e.g., ML and DL techniques) • Fake information containment and interventions. • Briefly pointed to policies and regulations and blocking malicious accounts under "containment" (but didn't classify these methods as deterrents or preventives)	False information (fake news, rumors, misinformation)

About the Authors

Mona Nasery is a PhD candidate in Information Systems at DeGroote School of Business, McMaster University. Prior to joining McMaster, she studied and worked as a researcher at Polytechnic University of Milan, Italy. She received her BSc. in Software Engineering and MSc. in Computer Science. Her research interest broadly falls within the intersection of Information Systems and Computer Science, such as big data and social media analytics, user behavior, and e-commerce applications. She has several years of research and work experience in the areas of Recommender Systems at Polytechnic University of Milan, and developing applications for smart spaces in EIT Digital, Helsinki, Finland. She also served several volunteer roles at McMaster University including as Social Director at DeGroote Doctoral Students Association (DDSA), Business Faculty Representative at Graduate Students Associations (GSA), and currently as Information Technology Officer at International Graduate Students Association (IGSA).

Ofir Turel is a professor of Information Systems Management at the University of Melbourne, and a Scholar in Residence at the Brain and Creativity Institute, Department of Psychology at the University of Southern California (USC). He has published over 190 papers in leading journals. They include such information systems journals as *MIS Quarterly, Journal of MIS, MIT Sloan Management Review, Communications of the ACM, JAIS, EJIS, and ISJ.* Example psychology and neuroscience outlets include Journal of Psychiatric Research, Addiction Biology, Cognitive, Affective & Behavioral Neuroscience, Appetite, Behavioral Brain Research, and Social Neuroscience. He has been recognized in the top 2% of researchers worldwide in a study conducted by Stanford University. His research has also been featured in numerous media outlets, including for example, the Wall Street Journal, The Washington Post, The Daily Mail, CBC, Clnet, Times Higher Education, The Rolling Stone, PBS, and TV and radio stations, globally. He is currently a Senior Editor for MIS Quarterly, and an editor for Nature's Scientific Reports.

Yufei Yuan is a professor of information systems in DeGroote School of Business at McMaster University, Canada. He received his Ph.D. in computer information systems from the University of Michigan, U.S., and his B.S. in mathematics from Fudan University, China. His research interests are in the areas of artificial intelligence, big data analytics, information security, privacy and trust, mobile commerce, emergency response systems, web-based negotiation support systems, human-computer interaction, fuzzy logic and expert systems, matching problems, and information systems in health care. He has more than 100 papers published in journals such as *MIS Quarterly, Management Science, Journal of Management Information Systems, European Journal of Information Systems, Information & Management, Internet Research, Communications of the ACM, IEEE Security and Privacy, International Journal of Mobile Communications, Group Decision and Negotiation, Decision Support Systems, Fuzzy Sets and Systems, International Journal of Human–Computer Studies, European Journal of Operational Research,* and *Decision Sciences*.