# Toward Informed News Media Consumption: Avoiding Fake News Via Labelling of Online Content

By

Jay Strong

Presented to the Computer Science, Engineering, and Physics Faculty
at the University of Michigan-Flint
in partial fulfillment of the requirements for the degree of
Master of Science

Date: 4/22/2020

First Reader _____
Dr. Matthew Spradling

Second Reader _____
Dr. Mark Allison

# Toward Informed News Media Consumption: Avoiding Fake News Via Labelling of Online Content

Jay Strong
*Computer Science, Engineering, and Physics Department*
*University of Michigan – Flint*
Flint, MI, USA
jasens@umich.edu

***Abstract* – Social media has become a primary source of online news content for the vast majority of consumers in recent years. News content on social networking platforms is inexpensive, easily accessible, readily available, and rapidly disseminated, thus making it a natural preference for news content providers and consumers alike. However, these same social media platforms have also provided an opportunity for malicious users to undermine credible news sources and widely propagate misinformation to the public. This contemporary fake news phenomenon has been associated with a number of adverse social implications, including public confusion and manipulation, and election interference, among others. Fake news has been used as a means to incite hysteria and diversion, which continues to be a growing problem in the United States and across the globe. Much of the efforts to abate the spread of fake news on social media have been focused on detecting and removing the sources directly. However, much less has been done to empower individuals so that they may avoid the threat of potential harm caused by widespread misinformation on social media.**

***Index Terms – Content Labeling, Fake News, Misinformation, Public Relations, Social Networks***

***Goals:* The purpose of this paper is to provide information that would be useful for implementing a potential intervention to end users on social networking platforms that could assist them in mitigating the exposure and influence of fake news. Specifically, we would like to know the following: (i) what consumers are currently using to determine news content credibility; (ii) which information they *should* use to determine news credibility; (iii) determine the overlap between (i) and (ii); (iv) how to present relevant information that consumers will actually use and understand; and (v) determine the limitations of the information we give consumers that could potentially be addressed through education.**

***Methods:* We propose distributing an experimental guided survey to a targeted demographic, which will be aimed at gathering specific information that can be used to aid in designing effective warning labels for news content on social networking platforms.**

***Results:* We have compiled a set of survey questions to be approved for distribution to the public along with a statement of informed consent for the survey participants. We have also outlined the specifics for the target demographic and method of administering the survey, as well as how we will use the information obtained from the survey question responses.**

***Conclusions:* Many other research efforts have been aimed at offensive strategies for mitigating widespread misinformation. We have concluded that more work needs to be done in order to give users a defensive tool for avoiding the common traps of fake news on social networking platforms.**

**This paper is an extension of previous work in the area of labeling online content and other forms of consumer information disclosures.**

## I. INTRODUCTION

It has been observed that more and more consumers tend to seek out and consume news on social media as opposed to more traditional news sources [1]. The reasons for this change in news consumption behaviors are inherent in the nature of the social media platforms themselves: (i) it is often more timely and cost-effective to consume news on social media compared with traditional news media, such as newspapers or television; and (ii) it is easier to further share, comment on, and discuss the news with friends or other readers on social media. [1]. In fact, it was recently reported that 55% of U.S. adults indicate that they consume news content from social media "often" or "sometimes" [2]. Most of the teenagers that Marchi interviewed in [3] reported that they get most, if not all of their information from social networking sites.

Another point to consider is the rise of social botnets. A social botnet is a network of automated accounts that can be programmed to perform coordinated activities, both malicious and benign. Evidence was found that social bots are playing a key role in the spread of fake news and tend to target influential users on social media. Viral sources of false news are heavily supported by social bots, as described in [4] and [5].

These current trends are, without a doubt, inextricably connected to the recent proliferation of fake news on social networking platforms, which has drawn a great deal of attention from both the public and academic communities [6].

So, what is fake news? We discuss this in great detail in the background section of this paper, but for now, we can turn to a definition from Lazer et al., which states "We define "fake news" to be fabricated information that mimics news media

content in form but not in organizational process or intent." [7]. This very concisely sums up the meaning of the term fake news.

It is important that social media users in particular pay attention to the fake news phenomenon and remain vigilant about vetting their sources of information by fact checking articles on current events. Allcott and Gentzkow stated in [8] that many people who see fake news stories believe them. This is partly because like-minded users tend to form "echo chambers" or "filter bubbles" where they are insulated from contrary perspectives [8]. This is especially concerning when you consider the fact that content on some of the most popular social networking platforms, such as Twitter and Facebook, can be shared among users with no significant third-party filtering, fact-checking, or editorial discretion, and an individual user with no reputation of credibility can in some cases reach as many readers as the mainstream media organizations like Fox News, CNN, or the New York Times [8].

Fake news does have an historical precedent in the United States dating all the way back to the widespread use of propaganda during World War I, as well as with the rise of corporate public relations in the 1920s [7]. More recently, it has emerged in the last election cycle for the 45th President of the United States, where the world witnessed the growing fake news epidemic [9].

In section 2, we will first look at the background information on consumer product labelling in the U.S. We will then have an extensive discussion on fake news. This will include how we define fake news, along with a comprehensive topology of the various forms of fake news, and other related works on fake news detection. Then we take a look at some of the work that is being done currently in regard to labelling online sources of news media content.

Section 3 will talk about our own research and methodology, which considers an experimental guided survey that can be distributed to a target demographic for gathering information on users' online behaviours, how they perceive the concept of fake news, and their opinion on warning labels for news content on social media. We will also talk about the specific ways in which we can administer the survey, as well as how we can use the information we obtain from it to aid future work in the area of designing informative warning labels for potentially harmful or misleading fake news content on social networking platforms.

In section 4 we explore some of the potential risks and challenges that may arise as a consequence of further developments in this line of research and what can possibly be done to help mitigate those risks.

Finally, in section 5 we will conclude with a discussion on the benefits of this research to society in regard to online safety, as well as identify topics for future work in the area of consumer protection from, and empowerment against malicious fake news content on social media.

## II. Background & Related Works

| Symbol | Meaning | Precautions |
|---|---|---|
|  | Dangerous container | Symbols inside the inverted triangle refer to the container |
|  | Dangerous contents | Symbols inside the stop sign refer to the contents of the container |
|  | Poison<br>Could be fatal if swallowed or inhaled | Wear gloves<br>Wear a face mask –<br>Do not breathe fumes<br>Wash hands after using |
|  | Corrosive<br>Can burn skin, eyes, respiratory tract<br>Can corrode metal | Wear gloves<br>Wear goggles<br>Wear a face mask -<br>Do not breathe fumes<br>Keep the container sealed |
|  | Flammable<br>Can catch on fire | Use with ventilation<br>Store away from heat |
|  | Explosive<br>Can explode | Handle carefully<br>Don't heat or drop<br>Don't mix with anything else |

Fig. 1. Hazardous consumer product labels [18]

In this section, we draw on previous works in the area of consumer protection and product labelling. We will also take a deeper dive into the ways in which we define and classify fake news for the purposes of this study, along with other related works in the area of fake news detection. Then we will take a closer look at some current work in the specific area of consumer labelling for online content, upon which this paper extends.

### *A. Consumer Product Labelling in the U.S.*

Many consumer products and services are required by the federal government to include disclosures in the form of standardized informational labels or warnings. These disclosures are designed to present important information to consumers in a consistent format and are made visibly obvious. Information labels and warnings are mandatory for many products and services that consumers find in their everyday life [10].

One of the most recognizable consumer information disclosures is the ubiquitous nutrition facts label, which appears on all packaged edible goods in the United States. The first mandatory nutrition facts label appeared after the Nutrition Labelling and Education Act of 1990 was passed [10]. Nutrition

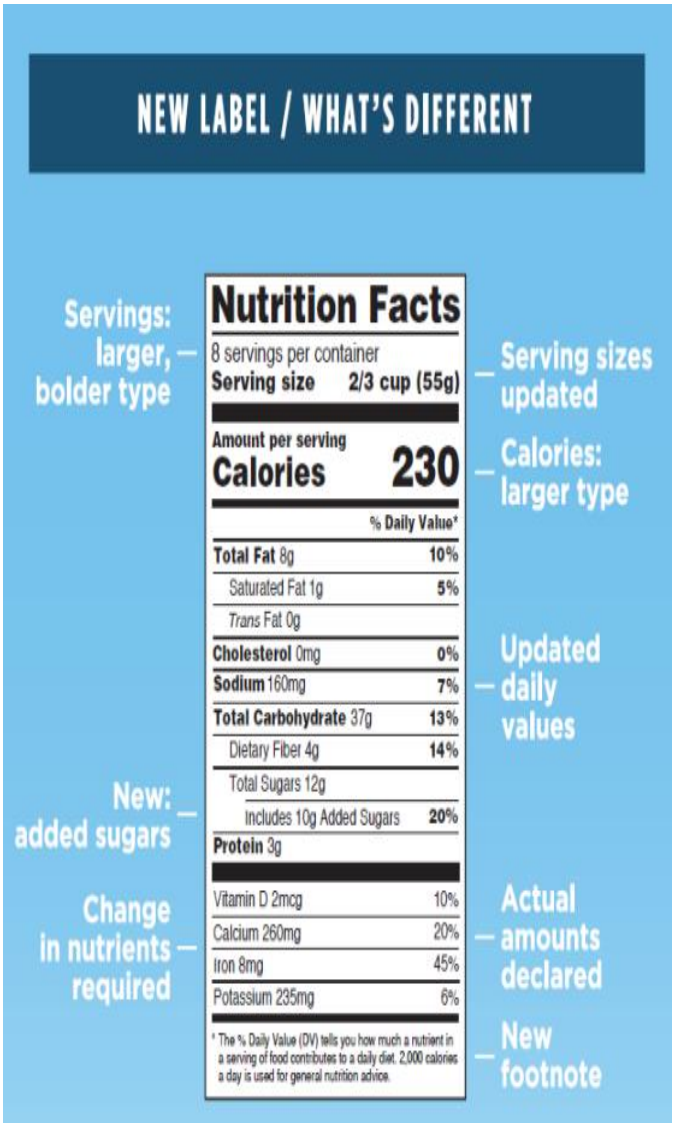


Fig. 2. New nutrition facts label [11]

facts labelling has evolved a number of times since then. The FDA published final rules on the most recent iteration of the nutrition facts label on May 27, 2016 and will include new scientific information, including the link between diet and chronic disease such as obesity and heart disease. The new label makes it easier for consumers to make better food choices (as shown in Figure 2). All packaged food product manufacturers will have until January 1, 2021 to comply with the new regulations of nutrition facts labelling [11].

Another very common form of consumer product labelling are the hazardous materials labels (see Figure 1 for examples). Hazmat labels and placards are required by federal law on all hazardous materials containers, including flammable, combustible, corrosive, explosive, and poisonous substances, some of which can be found in common household cleaning products and consumer goods like gasoline and other combustible fuels.

Other forms of informational labelling and recommendations for consumer products and services include, but are not limited to, film ratings (MPAA); television programming via V-Chip (FCC); video games and entertainment software ratings (ESRB); clothing, wool, and textiles (FTC); energy guides for home appliances and other consumer electronics; standard power consumption metrics for lightbulbs; funeral services, vehicles, dietary supplements, alcohol, tobacco products, firearms, loans, credit cards, and many others [10].

Spradling, Straub, and Strong identified some common themes that emerge from these various federal regulations. (i) to provide consumers with consistent information in a standardized format, such as the case with nutrition facts labelling, (ii) to notify consumers on particular concerns such as warning labels on tobacco products, and (iii) requiring advertising and marketing materials to provide consumers with accurate and complete information, or in other words, requiring honesty in the marketing of goods and services to the public [10].

This paper, along with [10], both recognize an immediate need for similar themes to be integrated into the ethos of publishing news media content online, either through federal regulation, industry self-regulation, or some cooperative form of both, similar to some of the ways in which it has been done in the past with other forms of consumer protection.

### *B. Contemporary Definitions and Topologies of Fake News*

The task of classifying fake news starts with our ability to first define the term *fake news*. Currently, there is no widely accepted precise definition of fake news due to the fact that many individuals, both in the general public and in the academic community, rely on their intuitive understanding of the meaning of fake news. Each research paper adopts its own definition of this term and other related concepts which conflicts with or overlaps terminology and information in other papers [6]. However, some recent studies have done a decent job of providing a broad definition of the term. As we mentioned in the introduction, Lazer et. al. defines fake news to be fabricated information that mimics news media content in form, but not in organizational process or intent [7]. Another definition presented by Shu et. al. states that, "Fake news is a news article that is intentionally and verifiably false." [1]. These definitions are based on two key features, namely, *authenticity* of the information, and *intent* of the author [1].

We must also carefully consider the distinction between fake news and real news. For the purposes of this study, we can point to a description of the nature of authentic news content in a recent paper, where Tandoc, Lim, and Ling present the following statements: "an account of a recent, interesting, and significant event", "an account of events that significantly affect people", "a dramatic account of something novel or deviant" and finally, "A central element in the professional definition of journalism is adherence to particular standards, such as being objective and accurate." [12].

In the same paper, Tandoc, Lim, and Ling conducted a review of published academic literature that used the term "fake news" and curated a relatively comprehensive topology that identifies six categories into which all fake news articles can be classified: satire, parody, fabrication, manipulation, propaganda, and advertising [12].

Another recent study, which focused on the fake news phenomenon that emerged during the 2016 U.S. presidential election campaign, reviewed contemporary academic studies on fake news and adopted a similarly exhaustive classification structure which deconstructs fake news into seven distinct categories: "false connection (where headlines, visuals or captions do not support the content); false context (genuine content shared with false contextual information); manipulated content (genuine imagery/information manipulated to deceive); misleading content (misleading use of information to frame an issue or individual); imposter content (genuine sources are impersonated); fabricated content (100 per cent false, designed to deceive and harm); and satire/parody (with potential to fool but no intention to cause harm)" [13].

### C. *Related Works on Fake News Detection*

Alcott and Gentzkow also published an extensive paper on the social, economic, and political impacts that fake news had on the US presidential election in 2016, in which the authors outlined the level of overall fake news exposure to users on social media platforms and how it may have affected the election results [8].

One major challenge that many statistical approaches face in automatic fake news detection is a lack of labelled benchmark datasets. Wang compiled a manually annotated dataset over a period of ten years, which is an order of magnitude larger than the previously known largest publicly available fake news dataset of a similar type [9]. The author of this study used this new dataset in a supervised machine learning approach, in which they implemented a hybrid convolutional neural network (CNN) to integrate metadata with text in order to investigate automatic fake news detection based on surface-level linguistic patterns, such as excessive profanity, for example. This study demonstrated that significant improvements can be achieved for fine-grained automatic fake news detection using this strategy [9].

Finally, Lazer et. al. identified two potential interventions that could help mitigate the exposure and influence of fake news: "(i) those aimed at empowering individuals to evaluate the fake news they encounter, and (ii) structural changes aimed at preventing exposure of individuals to fake news in the first instance" [7].

In this study, we will focus heavily on the first intervention proposed by Lazer et al. by introducing an experimental survey that can be used to gather information which can help future works on designing informational and recommendation labels for news content on social media.

### D. *Current Work on Labelling Online Content*


Fig. 3. Twitter's new tag for manipulated media (enclosed in red) [19]

The federal government has established itself as an authority in many ways on providing adequate information disclosures, warnings, and recommendations to consumers for products and services that are associated with a potential risk. However, the federal government has maintained a much more 'laissez-faire' approach to the regulation of media content, particularly *news* media content. This is most-likely due to the influence of the First Amendment to the United States Constitution [10].

Spradling, Straub, and Strong outlined labelling design paradigms for online content in [10], where they considered three forms of labelling. The first type of labelling that they discuss is recommendation labels. These types of labels take the form of warnings and recommendations that would "make normative claims and serve to inform the user whether, in what way, or in what context the media artifact should be consumed." [10]. The second type of labelling propose is informational labels, which would provide descriptions and data to the user without bias of interpretation, leaving the responsibility of making informed decisions based on that data to the users. The third and final type of labelling that was outlined in [10] are hybrid labels. Hybrid labels could be any combination of recommended actions that a user could take along with supporting data for the claims.

This paper extends on these ideas brought forth in [10], and seeks a method for gathering information from the public that could be used to make decisions on which of these types of labelling for online content would be most effective and well-received by news media consumers.

Mena conducted a recent study that closely relates to this line of research, in which he examined the impact of warning labels on Facebook users' intentions to share false news stories, using an experimental warning label design on 501 study participants [14].

Labelling for deceptive or fake news content on social media was recently implemented on Twitter in the form of a "Manipulated media" tag (as shown in figure 3). This tag is part of Twitter's official policy on deepfakes and other doctored media found to be intentionally manipulated. Twitter hopes that it will help to safeguard against misinformation ahead of the 2020 U.S. Presidential election [15]. Twitter's new tag made its debut on a tweet posted by the White House social media director Dan Scavino, which was later retweeted by President Donald Trump, which included a video of Joe Biden that was deceptively edited to give false context to the claim that Biden had inadvertently endorsed Trump for re-election in 2020 [16].

## III. Methodology

The purpose of this paper is to provide a potential method for gathering information that could help future research on the design and implementation of a system for labelling online news content. To that end, an experimental guided survey that can be distributed to specific sectors of the public is proposed.

We will outline the details of the experimental survey, including the information we seek, the target demographics, and how we can provide informed consent to the survey participants.

### *A. Gathering Information*

The survey itself is a series of questions that are designed to provide us specific information in three key areas: (i) user perception of informational warning and approval labels for online content; (ii) user perception on what makes a particular news article seem "fake" ; and (iii) user process for consuming online news content. The survey also includes a fourth section for general demographic information (annual income, level of education, age, political affiliation, etc.). The data from the survey responses could be beneficial for future initiatives in this line of research and in working toward the ultimate goal of implementing a real-world labelling system for online news content.

### *B. Identifying the Target Demographics*

For initial investigations, it would be best to distribute the survey to professionals in the field of public relations. More specifically, we would need to target people who work for PR firms, or in the public relations departments of corporations, law enforcement agencies, and political parties. Professionals in this segment of the population have insightful perspectives on the how information is distributed to the public and how public perception comes into play in regard to news content, as well as informative warnings of potential risks and dangers to the public.

Another group that would be a good candidate for survey distribution would be professional journalists. The people who actually investigate news topics and disseminate information on current events would have some valuable input on how fake news is perceived and how they handle misinformation in their personal and professional lives.

Finally, the survey could be distributed to other segments of the general public, such as people who reside in certain geographic locations, university students, and people who work in different industries (i.e. blue-collar manufacturing workers, healthcare professionals, service and hospitality workers, law enforcement, education, etc).

### *C. Survey Subject Protections and Disclaimers*

Email lists would need to be compiled for distributing the survey to the intended subjects. The survey includes specific instructions for free-form responses and a general disclaimer which explains that the survey is entirely voluntary and that the responses will not be connected to individual identities. It further explains that email addresses will be permanently deleted from any records and they will not be added to any mailing lists as a result of taking the survey, nor will their responses be given to any third party.

## IV. Risks & Challenges

In this section we discuss some of the potential risks and challenges that may arise as a consequence of labelling online news content.

The first potential challenge that comes to mind is the inevitable "gaming of the system" types of activities. Malicious users attempting to undermine a consumer protection initiative for online content labelling could target a specific artifact to propagate using a botnet, and artificially enhance credibility metrics. For example, a botnet using its resources to give false citations, or cyclical citations to bolster the number of total citations for an article, are both conceivable scenarios that could pose a real risk to the effectiveness of an online labelling system, and could cause users to unknowingly trust manipulative content. A content labelling system could be combined with or work cooperatively with an offensive type of automatic fake news or botnet detection system similar to some of the examples we discussed in section 2 as a potential strategy for overcoming this challenge.

Another type of challenge that we might expect from an automated, manual, or hybrid labelling system would be false positives and false negatives, either through human error (if manually tagged) or some other error from an automated system. The potential risk here would be falsely flagging a credible and accurate news artifact as a potential danger, or vice versa. In either case, the result would be that the user would lose faith in the effectiveness of the labelling system and either disregard the recommendations and warnings, or simply opt out of the labelling program altogether if that option is available. For this issue, a system of checks would need to be in place. Either a human intervention, or an automated machine learning feature of the labelling system itself, or some combination of the two could be implemented for correcting falsely labelled artifacts. More work must be done to solve this problem if it

arises and could potentially be a subject for future research and testing.

A third potential challenge this work may need to overcome are the profit motives associated with advertising on deceptive content sites, and working against those who have a financial stake in the algorithms on social media platforms and internet search engines, which were all noted by Bakir and McStay in [13]. These types of challenges pose a direct adversarial threat to any kind of online protection for consumers regarding content labelling. Some form of government oversight would need to be discussed to help mitigate this type of risk.

A fourth type of challenge that could potentially interfere with successfully implementing a labelling system for online news content, is the possibility of the news media sources themselves rejecting the idea of information labels and disclaimers, or perhaps even issuing a statement of their own to discredit the information labels. The reason this could become an issue is because news content creators are highly motivated to appeal to as many readers as possible, and a content labelling system might potentially be perceived as a threat to their bottom line. There may exist a balance between responsible journalism, and market demand for sensational or inaccurate news that content providers might need to consider in order to remain competitive. It would obviously be best for a consumer protection initiative for labelling online content to maintain a cooperative relationship with news content providers. Optimally, we would create a type of labelling system that legitimate news sources would be happy to incorporate into their content so that this type of risk might be avoided altogether.

Yet another potential challenge that this work must consider is identifying precisely what types of information would be relevant, useful, and accessible to the average consumer. We could use a nutrition facts analogy to describe this. Say a food product contained the phrase "made with real cane sugar" in their product description. This information is not necessarily relevant for making dietary assessments about the product but might be used by consumers anyway when they make the choice to purchase and use the product. In contrast, a food label which discloses exact amounts of sugar per serving would be a relevant metric for consumers to use when making health-conscious choices about food products, assuming the consumer possesses a certain level of education on the nutritional quality of food products. We must also carefully consider which information would be debated on its efficacy. For example, a "No GMO ingredients" label on a food product has no scientific consensus in regard to food safety [17]. These are the types of information we hope to obtain through the use of the experimental guided survey in this paper.

Lastly, online content labelling could be perceived by some as being an infringement on first amendment rights, where freedom of speech and freedom of the press are declared in the U.S. Constitution. This is also presumably the reason for the federal government's restraint thus far on regulations for online news content. This is another user perspective that the experimental survey hopes to gather more information on.

## V. Conclusions & Future Works

In this paper, we have outlined the need for informed news media consumption on social networking platforms. As observed by Shu et al. a majority prefer to get their news content from online sources [1]. Yet, the average consumer may not be adequately prepared to recognize intentionally misleading news content in its various forms. Consumers may benefit from an intervention to assist them in making informed decisions about the types of news content that they are exposed to online.

This paper has reviewed some of the ways in which informative recommendation and warning labels are currently being used in the United States to inform and protect the public. These existing labelling systems for consumer products and services can provide a type of construct for potential approaches to labelling online content in similar ways (e. g. a 'nutrition facts' type of label for news articles) [10].

We have discussed the details of our experimental survey, including what types of information we are seeking, the target demographic for survey subjects, and how the survey could be distributed. We also included some information on how the survey responses can be used to aid future initiatives that seek to develop and implement informational labelling conventions for news media content online.

Future work will use information from the experimental survey responses to design preliminary label template designs. These designs will be tested to determine effectiveness for delivering useful information to consumers as well as the impact they have on the user experience.

Survey responses will be used to determine metric values to present to the end user. Examples include a number value for the citations of a given news artifact to determine credibility, a credibility rating that considers the trustworthiness of the author and/or organization, a means by which to measure trustworthiness which can be well understood and agreed upon, age of news article, or edit history. The goal will be to determine metrics which are predictive and well-understood by the consumer while also being difficult to fabricate. For metrics which are predictive but not well-understood, goals for public education may be identified. We anticipate that there may be some aspects of fake news which the average consumer finds difficult to identify. Improved education targeting this modern concern would aid in closing the gap left by technology.

## Acknowledgments

I would like to thank my mentor and thesis advisor, Dr. Matthew Spradling, for encouraging me to achieve things of which I did not believe I was capable. I am very grateful for his guidance throughout the course of this project and my academic career in general. Dr. Spradling has played an instrumental role in my professional and academic development.

I would also like to thank my wife, Amanda, for her unwavering patience and support during these past seven years.

It has been a long road and she has sacrificed much so that I may achieve my academic goals. This would not have been possible without her.

Finally, I would like to thank my son, Oliver, for providing me with playful distractions along the way, which I truly appreciated during times of stress.

## REFERENCES

[1] K. Shu, A. Sliva, S. Wang, J. Tang and H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," *ACM SIGKDD Explorations Newsletter,* vol. 19, no. 1, pp. 22-36, 2017.

[2] P. Suciu, "More Americans are Getting Their News from Social Media," Forbes Media LLC., 11 October 2019. [Online]. Available: https://www.forbes.com/sites/petersuciu/2019/10/11/more-americans-are-getting-their-news-from-social-media/#78bc5cf33e17. [Accessed 28 March 2020].

[3] R. Marchi, "With Facebook, Blogs, Fake News, Teens Reject Journalistic 'Objectivity'," *Journal of Communication Inquiry,* vol. 36, no. 3, pp. 246-262, 3 October 2012.

[4] M. Spradling, M. Allison, T. Tsogbadrakh and J. Strong, "Toward Limiting Social Botnet Effectiveness While Detection is Performed: A Probablistic Approach," in *6th Annual Conference on Computational Science & Computational Intelligence*, Las Vegas, Nevada, 2019, in press.

[5] C. Shao, G. L. Ciampaglia, O. Varol, A. Flammini and F. Menczer, "The Spread of Fake News by Social Bots," *arXiv preprint arXiv:1707.07592,* vol. 96, p. 104, 24 July 2017.

[6] N. Ruchansky, S. Seo and Y. Liu, "CSI: A Hybrid Deep Model for Fake News Detection," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017.

[7] D. M. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, C. R. Sunstein, E. A. Thorson, D. J. Watts and J. L. Zittrain, "The Science of Fake News," *Science,* vol. 359, no. 6380, pp. 1094-1096, 2018.

[8] H. Allcott and M. Gentzkow, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives,* vol. 31, no. 2, pp. 211-236, 2017.

[9] W. Y. Wang, "“Liar, Liar Pants on Fire”: A New Benchmark Dataset for Fake News Detection," *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics,* 2017.

[10] M. Spradling, J. Straub and J. Strong, "Introducing & Evaluating 'Nutrition Facts' for Online Content," *Unpublished.*

[11] "Changes to the Nutrition Facts Label," U.S. Food & Drug Administration, 12 March 2020. [Online]. Available: https://www.fda.gov/food/food-labeling-nutrition/changes-nutrition-facts-label. [Accessed 4 April 2020].

[12] E. C. Tandoc Jr., Z. W. Lim and R. Ling, "Defining "Fake News": A Topology of Scholarly Definitions," *Digital Journalism,* vol. 6, no. 2, pp. 137-153, 2018.

[13] V. Bakir and A. McStay, "Fake News and The Economy of Emotions: Problems, causes, solutions," *Digital Journalism,* vol. 6, no. 2, pp. 154-175, 2018.

[14] P. Mena, "Cleaning Up Social Media: The Effect of Warning Labels on Likelihood of Sharing False News on Facebook," *Policy & Internet,* 28 July 2019.

[15] J. Pero, "Twitter to start labelling deepfakes and other deceptive media ahead of 2020 presidential election," The Daily Mail, 4 February 2020. [Online]. Available: https://www.dailymail.co.uk/sciencetech/article-7966923/Twitter-label-deepfakes-deceptive-media.html. [Accessed 4 April 2020].

[16] K. Zakrzewski, "Twitter flags video retweeted by president Trump as 'manipulated media'," The Washington Post, 9 March 2020. [Online]. Available: https://www.washingtonpost.com/technology/2020/03/08/twitter-flags-video-retweeted-by-president-trump-manipulated-media/. [Accessed 4 April 2020].

[17] A. Hilbeck, R. Binimelis, N. Defarge, R. Steinbrecher, A. Székács, F. Wickson, M. Antoniou, P. L. Bereano, E. A. Clark, M. Hansen, E. Novotny, J. Heinemann, H. Meyer, V. Shiva and B. Wynne, "No Scientific Consensus on GMO Safety," *Environmental Sciences Europe,* vol. 27, no. 1, 24 January 2015.

[18] "WHMIS and Hazardous Materials," [Online]. Available: http://www.retailsafety.ca/training/whmis-hazardous-materials. [Accessed 4 April 2020].

[19] D. Scavino, "Sleepy Joe in St. Louis, Missouri today: "We can only re-elect @realDonaldTrump."," Twitter (Tweet), 7 March 2020. [Online]. Available: https://twitter.com/DanScavino/status/1236461268594294785?s=20. [Accessed 4 April 2020].

[20] H. Team, "Survey Questions 101: Over 70 Survey Question Examples, Types of Surveys & FAQs," Hotjar, 23 January 2020. [Online]. Available: https://www.hotjar.com/blog/survey-questions/#6-main-types-of-survey-questions.

[21] M. Anderson and J. Jiang, "Teens, Social Media & Technology 2018," Pew Research Center, 2018.

## APPENDIX

### *A. Survey Sample: Survey Disclaimer Statement*

Your participation in this survey is voluntary. You are free to omit responses to any questions that you do not wish to answer. If you decide to participate in this survey, and later change your mind, you are free to stop at any point.

If you are given an opportunity to submit free-form responses in the survey, please provide relevant and responsible responses; refrain from disparaging, offensive, harassing, or otherwise inappropriate language; and refrain from using any sensitive information of any kind (e.g. credit card or bank account information), or any other forms of information that can be used to identify you or anyone else (e.g. social security number).

The following questions will be used for general analytical purposes only. Although your email address is sent along with your answers, your specific responses will not be connected to you in any way whatsoever after we have received them. Once your results are analysed for the study, your email address will be deleted from any records we may have. Your individual responses will not be given to any third party whatsoever. Furthermore, you will not be added to any mailing lists by taking this survey. Only anonymous data will be displayed.

Proceeding to the survey implies that you understand and agree to the provisions in this disclaimer.

### *B. Survey Sample: Experimental Survey Questions*

<u>Survey Type 1: User perception of warning and approval labels</u>

1. How likely would you be willing to read a news article on social media that has an informative label which states the content was modified from the original source?
    - o Very likely
    - o Somewhat likely
    - o Undecided
    - o Unlikely
    - o Very unlikely

2. How likely would you be willing to share or repost a news article on social media that has an informative label which states the content was modified from the original source?
    - o Very likely
    - o Somewhat likely
    - o Undecided
    - o Unlikely
    - o Very unlikely

3. How likely are you to trust the credibility of a news article on social media that has an informative label which states that the article has five citations or less?
    - Very likely
    - Somewhat likely
    - Undecided
    - Unlikely
    - Very unlikely

4. How likely are you to share or repost a news article on social media that has an informative label which states the article has five citations or less?
    - Very likely
    - Somewhat likely
    - Undecided
    - Unlikely
    - Very unlikely

5. How likely are you to read an article on social media that has an informative label which states the information used in the article is more than six months old?
    - Very likely
    - Somewhat likely
    - Undecided
    - Unlikely
    - Very unlikely

6. How likely are you to share or repost an article on social media that has an informative label which states the information used in the article is more than six months old?
    - Very likely
    - Somewhat likely
    - Undecided
    - Unlikely
    - Very unlikely

7. How likely are you to share or repost an article on social media that has a warning label which states that the author has no established reputation of credibility?
    - Very likely
    - Somewhat likely
    - Undecided
    - Unlikely
    - Very unlikely

8. What types of information would you find useful in helping you make decisions about consuming news content online?

9. Do you feel that an informational warning/recommendation label for an online news artifact infringes upon the right to free speech or freedom of the press, as stated in the First Amendment of the Constitution of the United States of America? Please explain.

<u>Survey Type 2: User perception of what makes an article "fake"</u>

1. How likely are you to trust the credibility of a news article that is published by a mainstream news source?
    - Very likely
    - Somewhat likely
    - Undecided
    - Unlikely
    - Very unlikely

2. How likely are you to trust the credibility of a news article that is published by an independent news source?
    - Very likely
    - Somewhat likely
    - Undecided
    - Unlikely
    - Very unlikely

3. How likely are you to trust a news article that was written by an author that has no affiliation with any news media organization and has no established reputation of credibility?
    - Very likely
    - Somewhat likely
    - Undecided
    - Unlikely
    - Very unlikely

4. How likely are you to trust the credibility of a news article that has a "sensationalist" headline?
    - Very likely
    - Somewhat likely
    - Undecided
    - Unlikely
    - Very unlikely

5. How likely are you to trust the credibility of a news article that has a point of view that is not your own?
    - Very likely
    - Somewhat likely
    - Undecided
    - Unlikely
    - Very unlikely

6. How likely are you to share or repost a news article that has been reposted many times?
    - Very likely
    - Somewhat likely
    - Undecided
    - Unlikely
    - Very unlikely

7. How likely are you to share or repost a news article that has been reposted by a celebrity?
    - Very likely
    - Somewhat likely
    - Undecided
    - Unlikely
    - Very unlikely

8. Please describe in your own words what you believe fake news is.

9. What specific information do you use to make determinations on news credibility, either of news articles, or of the content providers themselves? Please explain.

<u>Survey Type 3: User process for consuming content online</u>

1. How likely are you to review the comment section of an article to determine if the content is fake?
    - Very likely
    - Somewhat likely
    - Undecided
    - Unlikely
    - Very unlikely

2. Upon finding one piece of inaccurate information, how likely are you to determine creditability of the entire article?
    - Very likely
    - Somewhat likely
    - Undecided
    - Unlikely
    - Very unlikely

3. How likely are you to fact check an article before you determine its credibility?
    - Very likely
    - Somewhat likely
    - Undecided
    - Unlikely

- Very unlikely

4. How likely are you to fact check an article before you share or repost?
    - Very likely
    - Somewhat likely
    - Undecided
    - Unlikely
    - Very unlikely

5. At what point do you determine the credibility of a news article?
    - After reading headline
    - After reading first paragraph
    - After reading whole article
    - After fact checking
    - After reading user comments section

6. Which resources do you use (if any) for fact checking? Why or Why not?

7. What is your preferred social media site?
    - Facebook
    - Snapchat
    - Twitter
    - Tik Tok
    - Instagram
    - You Tube
    - Reddit
    - Other_____
    - None

Survey Type 4: User demographic

1. What is your highest level of education?
    - No degree/diploma/certificate
    - High school diploma/GED
    - Bachelor's
    - Master's
    - Doctoral
    - Vocational School
    - Choose not to answer

2. What is your age group?
    - 18-24
    - 25-39
    - 40-64
    - 65+
    - Choose not to answer

3. What is your annual household income?
    - Less than $20,000
    - $20,000 to $40,000
    - $41,000 to $60,000
    - $61,000 to $100,000
    - Greater than $100,000
    - Choose not to answer

4. What is your preferred news media content provider, and why?
    _____

5. What is your political affiliation?
    - Democratic
    - Republican
    - Independent
    - Progressive
    - Green Party
    - Libertarian
    - Other _____
    - Choose not to answer

*C. Survey Sample: Experimental Content Label Designs*

Prototype 1: Recommendation (Warning/Approval) Label

## Reader Discretion Advised

This article has been posted by a private user with no established reputation of credibility. The content in the article has been modified from its original source. Proceed with caution.

## Authentic Content

This article has been screened for authenticity. The content in the article has not been modified from its original source.

Prototype 2: Informative Label

# Article Facts

| | |
|---|---|
| Author Credibility Rating | 87/100 |
| Age of information | Less than 24 hours |
| Posted | Monday April 20, 2020 – 8:23AM EDT |
| Affiliation | Vice News |
| Number of citations | 9 |

Prototype 3: Hybrid Label

# WARNING

- News artifact has less than 5 citations
- Author has no reputation of credibility
- News artifact exceeds the 6-month maximum age for information

Recommended Action: Fact checking is strongly advised.