

# Introducing & Evaluating ‘Nutrition Facts’ for Online Content

Matthew Spradling  
Computer Science, Engineering, and  
Physics Department  
University of Michigan – Flint  
Flint, MI, USA  
mjspra@umflint.edu

Jeremy Straub  
Institute for Cyber Security Education  
and Research  
North Dakota State University  
Fargo, ND, USA  
jeremy.straub@ndsu.edu

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

**Abstract**—So-called ‘fake news’ – deceptive online content that attempts to manipulate readers – is a growing problem. It has been blamed for election interference, public confusion and other issues, both in the United States and beyond. A tool of state intelligence agencies, scammers and marketers alike, deceptive online content is poised to have growing consequences. This problem is made particularly pronounced as younger generations choose social media sources over journalistic ones for their information. This paper considers a prospective solution in the form of providing consumers with ‘nutrition facts’ style information for online content. To this end, it reviews prior work in product labeling and disclaimers, considers several possible approaches to the challenge and the tradeoffs between them.

**Keywords**—fake news, nutrition facts, content labeling, online content, deceptive content

## I. INTRODUCTION

Deceptive online content, commonly called ‘fake news,’ has been argued to be responsible for impacting election results [1], [2], confusing the public [3] and even causing an armed standoff [4]. Problematically, individual consumers seem under-prepared to deal with information that is not pre-vetted for them by a conventional news source (though concerns about conventional news sources circulating untrue news have also been raised [5], [6]). Reader laziness [7] and repetition of exposure [8] have been shown to increase consumer belief in this deceptive content.

With 55% of Americans indicating that they are, at least sometimes, getting news from unvetted social media sites [9] and 75% of Americans believing fake news headlines [10], [11], the lack of accuracy of some of this content and the lack of consumers’ perception of this lack of accuracy is problematic. Marchi shows that this is a problem that will only get worse, with most teens preferring to get news from social media (and not typical journalistic sources) and many being unconvinced of journalists’ objectivity [12].

Bakir and McStay [13] discuss several prospective solutions to the problems of online content accuracy and accuracy projection. These include increasing the prevalence of more accurate news articles in feeds, verifying facts in articles manually, automated detection of deceptive content and warning labels [13]. A collection of journalists suggested changing algorithms to give individuals greater exposure to different perspectives, increasing source transparency in reporting, enhancing access to elected officials and collaborative journalism [13]. Bakir and McStay [13] note a clear profit motive for deceptive content sites: online advertising. Given this motive, there seems to be little hope of the issues of deceptive content being resolved on their own or by those involved in its dissemination.

This paper considers one prospective solution: a form of digital ‘nutrition facts’ or ‘consumer facts’ for online content. This type of approach would provide a politically agnostic overview of a media artifact’s source and content based on a set of metrics. These metrics could be used to provide an evaluation to inform the consumer (like the United States Surgeon General’s warning on cigarettes). Alternatively, the component data could be presented directly to the consumer to allow them to form their own decision (like the “nutrition facts” food labeling system in the United States and energy guides for appliances). This paper discusses prior work in related areas of consumer notification. It considers the need for and the prospective efficacy of a content labeling system for online content. It outlines three main forms this labeling system could take, the data that it could be based upon and the pros and cons of each approach. The paper concludes with a discussion of the societal benefits that such a system could prospectively have and identifies topics for future work.

## II. BACKGROUND

The considered content labeling system draws from prior work in several areas. This section reviews this prior work. First, the United States nutrition facts food labeling system is reviewed. Then, prior fake news classification techniques are considered. Finally, consumer facts disclosures and labeling in other industries are discussed.

### A. History of Nutrition Labeling in the USA

The building blocks of federally regulated nutrition labeling in the United States date back to the Pure Food and Drug Act of 1906 and the Federal Meat Inspection Act. These were among several consumer protection laws signed by Theodore Roosevelt during the Progressive Era. The overall purpose of this and other legislation from 1890 – 1920 was to prevent consumer fraud rather than addressing public health concerns. Consumers wished to know what was in their food, but did not have the necessary understanding of health science for evaluating its nutritional value [14], [15].

The supreme court decision in *U.S. v. Johnson* determined that the Food and Drug Act of 1906 did not prohibit false claims regarding the therapeutic nature of a drug, only inaccuracies regarding the factual nature of its ingredients. This led to the Sherley Amendment in 1912, prohibiting “false and fraudulent” labeling. The Sherley Amendment represented the first act of congress toward limiting product labeling based upon the intent of the manufacturer rather than simply addressing the factual nature of the product’s contents [15].

The modern Food and Drug Administration began to take form in the late 1920s and early 1930s. By 1933, it recommended major revisions to the Food and Drug Act of

1906. This recommendation led to the Food, Drug and Cosmetic Act of 1938. This legislation required artificial colors, flavorings, and preservatives to be listed. This represented the first step toward addressing nutritional content in mandatory labeling. This act also removed the Sherley Amendment's requirement to prove intent to defraud in cases of drug misbranding, placing more burden on the manufacturer to ensure the product was safe and accurately labeled [15].

During the following decades, scientific understanding of vitamins and their role in disease prevention slowly increased. As this information became more available to the public, consumer demand grew for food products that could promote health. Industry responded through their own labeling and ingredients, but a lack of regulation on what was required to claim food to be "healthy" generated demand for oversight. The FDA regulated in 1973 that manufacturers whose packaging claimed health benefits would be required to label the product with factual information relevant to the claim. For example, if a product were claimed to be "high in fiber," the actual fiber content of the product needed to be provided. At this point, the requirement did not carry with it any standard for how the label should be presented. It was possible to technically meet the requirement using small hard-to-read print that a consumer might not notice. Furthermore, there was no agreement upon what is "high" when it comes to fiber content. The ambiguous nature of this claim, along with others such as "low in fat," made it difficult for consumers to evaluate, even with the grams of fiber known [15].

The first mandatory 'nutrition facts' label came about through the passage of the Nutrition Labeling and Education Act of 1990. This act standardized several health-benefit terms, giving a common meaning to claims such as the aforementioned "high in fiber." Manufacturers making health benefit claims on their packaging were required to use terms from an approved list and only those for which the product met the FDA regulated definition. In addition, the act granted the FDA the authority to mandate nutrition labeling on most packaged foods. The standardized format of the nutrition facts label was designed to limit the ability of a manufacturer to conceal certain facts with smaller font. The standardized definitions of health-benefit terms were designed to eliminate the ambiguity of such claims, so that "high in fiber" or "low in fat" meant the same thing for each product [15].

The initial attempt to regulate ill-intention of the manufacturer was problematic because it was difficult to prove intent. The revision in 1938 placed the burden on the manufacturer to avoid misrepresenting the product. Philosophically, it was determined to be more important to protect the consumer from the negative impact of misrepresented facts than to protect the manufacturer from blame for unintentional or incidental misrepresentation.

Requiring simply that the manufacturer disclose their own factual basis for claims of benefit proved insufficient, as 1) manufacturers were left to come up with their own definitions for what constitutes a "benefit" and 2) manufacturers were able to conceal inconvenient facts using small print. The Nutrition Labeling and Education Act of 1990 standardized the size and scope of information which should be presented to describe contents of the package, which helped to eliminate concealment. The act additionally defined terminology and standards for "benefits" which could be advertised on the product label, preventing the manufacturer from coming up

with their own definitions which might lead to misrepresentation.

The original nutrition facts label format proved to be difficult to interpret in some respects for some customers. Deficiency in math education seemed to be linked to being unable to compute actual nutrition information when eating more than a recommended serving. Furthermore, what is recommended as a serving may not match what consumers eat in practice. The latest revisions have been made to address these challenges, though the underlying causes (choosing to eat more than the recommended serving size and not being good at math) remain.

### *B. Classification of Fake News*

This section discusses the classification of fake news items. First, approaches to measure media credibility are considered. Then fake news detection and classification techniques are reviewed.

Measures of media credibility have been studied for several decades in the field of journalism [16] and more recently in investigating the credibility of websites [17] and social media blogs [18]. Gaziano and McGrath investigated the dimensions of credibility, the impact of the local nature of news coverage and the impact of the delivery medium on perceived credibility [16]. Meyer argued that certain aspects of credibility can come into conflict, such as balancing believability with respect of the community [19]. Corritore considers several factors of credibility in websites including honesty, reputation, and predictability [17]. Kang observed that while there is nothing preventing "anyone" from making a blog, consumers may find blogs more credible than mainstream news sources due to their independent nature [18]. Unfortunately, many of the factors (honesty, integrity, etc.) used to measure credibility would be difficult to self-report in a meaningful way. It is unlikely that a newspaper would label itself "dishonest" for example.

The task of classifying fake news starts with an 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 [20]. However, some recent studies have done a decent job of providing a broad definition of the term. Lazer, et al. define fake news to be fabricated information that mimics news media content in form, but not in organizational process or intent [21]. A more concise definition presented by Shu, et al. states that, "fake news is a news article that is intentionally and verifiably false" [22]. These definitions are based on two key features: namely, authenticity of the information, and intent of the author [22].

The distinction between fake news and real news must also be carefully considered. For the purposes of this study, several descriptions of the nature of real news are relied upon. Edson, et al. define real news with 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" [10].

In the same paper, Edson, et al. 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 [10]. Another recent study, which focused on the fake news phenomenon that emerged during the 2016 US presidential election campaign, reviewed contemporary academic studies on fake news and adopted a similarly comprehensive 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].

Alcott and Gentzkow published an extensive review of the social, economic, and political impacts that fake news had on the US presidential election in 2016. They discussed the level of overall fake news exposure to users on social media platforms and how it may have affected the election results [1].

One major challenge that many statistical approaches face in automatic fake news detection is a lack of labeled 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 [23]. This new dataset was used in a supervised machine learning approach, which implemented a hybrid convolutional neural network to integrate metadata with text in order to investigate automatic fake news detection based on surface-level linguistic patterns. This study demonstrated that significant improvements can be achieved for fine-grained automatic fake news detection using this strategy [23].

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” [13].

The goal of the current work is to use the second intervention proposed by Lazer, et al. [13] to effectively facilitate the first. The goal is to demonstrate the use of state-of-the-art computer detection methods to produce relevant information. This can then be presented to the end user in a concise format, enabling them to make informed decisions about the news content that they are exposed to on social media and elsewhere.

### C. Other Governmental Labeling in the United States

In addition to the ubiquitous ‘nutrition facts’ that many people will see multiple times per day on numerous types of food packaging, the federal government also regulates consumer disclosures on a wide variety of other products and for a myriad of services. State governments, in some locations, may also have additional non-conflicting labeling

requirements or require the labeling of additional products (states cannot require companies to replace federal labels with state ones, due to federal preemption doctrine [24]; however, in some cases where voluntary standards exist these would be overridden by mandatory state standards). The numerous sources of federal regulation range from the Federal Trade Commission (FTC) Act, which covers most commercial transactions’ representations and advertising to specific regulations for products, such as wool and textiles [25]. Table 1 summarizes products with regulated disclosures in the United States, and the source of the disclosure requirements.

Across all of these various regulations, a number of consistent themes and purposes emerge. The first is to provide consumers with standardized information in a consistent form. This is similar to the nutrition facts label. Heaters, air conditioners, other appliances (as shown in Figure 1) and light bulbs all have disclosures that allow consumers to easily compare products based on standard metrics of power consumption, a common estimated power cost value and a common estimated use per year. Other features, for example, the brightness of lightbulbs, are also described in a standard way for some products. For clothing, material content and cleaning instructions are presented in a common form.

The second theme is that regulations notify consumers of particular concerns. For example, cigarettes and smokeless tobacco have warning labels [26] and clothing that cannot be cleaned using normal methods of washing must provide a notice to that effect and cleaning instructions [27].

The third theme is that the advertising and marketing materials must provide consumers with accurate and complete information. The FTC has issued instructions to social media

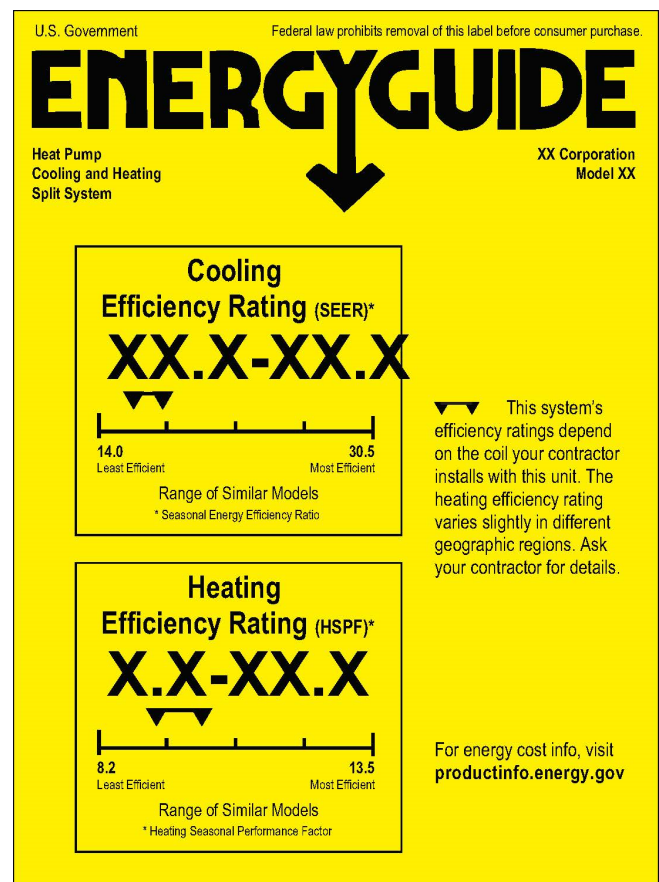


Fig. 1. Energy Guide Template for Heat Pump [48].

influencers [28] regarding how to disclose affiliation and compensation. It also requires funeral homes and related providers to provide price lists and other information to customers in a standardized way [29].

In many cases, the government has worked with relevant industries to develop government standards, cooperative standards or private standards. In content rating labeling, two different standards are prominent: MPAA ratings for films and V-Chip ratings for television programming. The two standards, which are summarized in Table 2, are relatively similar. The MPAA ratings were developed by the Motion Picture Association of America (MPAA) [30] while the V-Chip ratings were developed by the Federal Communications Commission (FCC) [31]. Grossman [31] contends that the V-Chips were never successful in meeting their objectives (though they are still included in modern-day televisions). The V-Chip introduction raised complaints about the ratings assigned and only 30% of families surveyed actually set it up and only 8% regularly used it. Both the MPAA and V-Chip ratings assign an appropriateness level (largely associated with viewer age) to each program. MPAA ratings include a free text field where the reason for the rating (i.e., what caused the level of rating) can be explained, for the PG, PG-14, R and NC-17 ratings [32]. V-Chip ratings also include up to four letters (D, L, S and V) which provide additional information about the reason for the rating [33]. The D, L, S and V value meanings are described in Table 3.

For video games, similarly, the Entertainment Software Rating Board (ESRB) has developed a ratings system quite similar to the MPAA ratings. Each rating has a letter rating (listed in Table 4) and a text description field [34]. There is also a supplemental field to describe interactive parts of the game, such as in-game purchases [34]. ESRB ratings are either assigned by a team of reviewers (for boxed games), based on a survey and a video of game play, or a survey and automatic assignment (for downloadable games) [35].

ESRB content descriptors include “Alcohol Reference,” “Animated Blood,” “Blood,” “Blood and Gore,” “Cartoon Violence,” “Comic Mischief,” “Crude Humor,” “Drug Reference,” “Fantasy Violence,” “Intense Violence,”

“Language,” “Lyrics,” “Mature Humor,” “Nudity,” “Partial Nudity,” “Real Gambling,” “Sexual Content,” “Sexual Themes,” “Sexual Violence,” “Simulated Gambling,” “Strong Language,” “Strong Lyrics,” “Strong Sexual Content,” “Suggestive Themes,” “Tobacco Reference,” “Use of Alcohol,” “Use of Drugs,” “Use of Tobacco,” “Violence,” and “Violent References” [34]. Interactive element descriptors include “In-Game Purchases,” “Users Interact,” “Shares Location,” and “Unrestricted Internet” [34].

The government has established a clear role for itself in ensuring that consumers have accurate information and are warned about products that represent a danger, in most areas. However, in the area of content, the government has been far more restricted in its labeling and warnings (both those it requires and those it participated in developing), focusing on protecting children and ensuring that accurate and standardized details are provided for commercial transactions.

TABLE I. GOVERNMENT LABELING / PRODUCTION INFORMATION REQUIREMENTS AND THEIR SOURCES.

Product	Regulating Law(s)
Weight Loss Products / Services [49]	Voluntary guidelines (Industry Panel & FTC)*
Clothing	Textile Act [25] Wool Act [25] Care Labeling Rule [27]
Fur [50]	Fur Products Labeling Act
Business Opportunities [51]	FTC Business Opportunity Rule
Native Advertising [52]	Enforcement Policy Statement on Deceptively Formatted Advertisements*
Ceiling Fans [53]	Energy Labeling Rule
Showerheads [53]	Energy Labeling Rule
Faucets [53]	
Toilets [53]	
Urinals [53]	
Water Heaters [53]	Energy Labeling Rule
Pool Heaters [53]	Energy Labeling Rule
Furnaces [53]	Energy Labeling Rule
Boilers [53]	
Dishwashers [53]	Energy Labeling Rule
Air Conditioners [53]	Energy Labeling Rule
Heat Pumps [53]	
Refrigerators [53]	Energy Labeling Rule
Freezers [53]	
Washing Machines [53]	Energy Labeling Rule
Televisions [53]	Energy Labeling Rule
Light Bulbs / Lamps [54]	Energy Labeling Rule
Diamonds / Pearls / Gemstones [55]	FTC Jewelry Guides*
Environmental Benefits / Carbon Offsets [56]	FTC Green Guides*
Feather / down / bedding products [57]	FTC Business Guide* (with information from the International Association of Bedding and Furniture Law Officials)
Social media posts [28]	FTC Endorsement Guides*
Funeral and Related Services [29]	Funeral Rule*
Vehicles	Energy Independence and Security Act of 2007 [58] Automobile Information Disclosure Act of 1958 / 15 U.S. Code § 1232 American Automobile Labeling Act [59]
Fuel [60]	FTC Fuel Rating Rule*
Dietary Supplements [61]	Dietary Supplements Health and Education Act FTC Enforcement Policy Statement on Food Advertising
Made in USA claims & country of origin labeling [59]	FTC Enforcement Policy Statement Tariff Act
Layaways [62]	FTC Act Truth in Lending Act
Loans, Credit Cards & Other Lending [63]	Truth in Lending Act
Cigarettes [26]	Federal Cigarette Labeling and Advertising Act of 1965 Public Health Cigarette Smoking Act Comprehensive Smoking Education Act of 1984
Smokeless Tobacco [26]	Comprehensive Smokeless Tobacco Health Education Act of 1986

\* Interpretation of the FTC Act

TABLE IV. MPAA AND V-CHIP RATINGS.

MPAA Rating [32], [64]	V-Chip Rating [33]	Meaning [32], [33], [64]
N/A	TV-Y	For the very young, targeted at ages 2-6
N/A	TV-Y7	For children aged above 7
N/A	TV-Y7-FV	TV-Y7 with “fantasy violence that may be more intense or more combative” [33]
G	TV-G	Suitable for “all ages” [33] and “general audiences” [32]
PG	TV-PG	Programs where “parental guidance” is recommended [32], [33] and which may not be appropriate for young children
PG-13	TV-14	Programs which may have content inappropriate for children under age 13 or 14
R	TV-MA	Programs for older audiences, 17 or above
NC-17	N/A	Children under 17 are not allowed, even with parental supervision

TABLE II. TABLE 3. V-CHIP CONTENT DESCRIPTIONS.

Letter	Meaning [33]
D	“suggestive dialogue”
L	“course or crude language”
S	“sexual situations”
V	“violence”

TABLE III. ESRB RATINGS [34].

Rating	Meaning
E	“Everyone” – content for all age levels
E 10+	“Everyone 10+” – content “generally suitable for ages 10 and up”
T	“Teen” – content “generally suitable for ages 13 and up”
M	“Mature” – content “generally suitable for ages 17 and up”
A	“Adult” – content “only for adults ages 18 and up”
RP	“Rating Pending”

This is an undoubtable influence of the First Amendment to the United States Constitution which instructs that “congress shall make no law ... abridging the freedom of speech, or of the press” [36].

### III. METHODS OF LABELING FOR ONLINE CONTENT

This section discusses methods for labeling online content. First, label design considerations are discussed. Then, potential labeling authority sources are considered.

#### A. Labeling Design Paradigms for Online Content

Three forms of labeling for online content are considered. Recommendation labels make normative claims and serve to inform the user whether, in what way, or in what context the media artifact should be consumed. Informational labels make descriptive claims about the content of the media, allowing the user to interpret this data and make an informed decision about consumption. Hybrid labels provide recommended actions along with data to justify the recommendations.

1) *Recommendation Labels*: Recommendation labels take the form of warnings and approvals. A warning label serves as a recommendation to the reader they should not consume the content, while an approval label claims that the content is safe for consumption. Existing online labeling includes warnings that a public Wi-Fi connection is insecure and indicators suggesting that a site is secure, before the user inputs credit card information. Warning labels on tobacco products inform the user of the dangers of smoking. Approval labels on gluten free products inform the consumer the product is safe to eat.

For media content, warning labels could inform the user that a media artifact is “opinion,” “poorly cited,” “deceptively edited,” or otherwise communicate potential dangers of consuming the content. Approval labels could state that the

media is “well cited,” “from a trusted source,” or otherwise inform the user of positive qualities that may imply the content is safe for consumption.

2) *Informational Labels*: An informational label provides data to the user without bias of interpretation. The goal is to give the consumer enough information to

make an informed decision as to whether the content is safe to consume. Informational labels avoid suggesting a course of action. Existing examples of informational labels include nutrition facts labeling, which provides the consumer with data such as total fat and calories. V-chip labeling informs the consumer that a video contains violence.

Rather than claiming a video is “deceptively edited,” an informational label may state that a video “has been edited from the original source content.” Instead of stating an article is “poorly cited,” the number of citations could be provided without a statement as to how many would be desirable.

3) *Hybrid Informational and Recommendation Labels*: A hybrid approach to labeling provides both data and a recommendation to the user. The goal is to provide the user with a recommended course of action along with supporting data that justifies the recommendation. This design attempts to bridge the knowledge gap when consumers may not know what constitutes “good” or “bad” content. Nutrition facts include both the total fat content and the percent of daily recommended fat. ESRB ratings include both content-specific labels such as “blood” and recommendations for who should consume the content, such as “teen” or “everyone.”

For media content, a hybrid label could include both the number of citations as well as the average number of citations for an article of similar age and topic. In addition to stating that a video has been “edited from the original source,” the label could make the implication that “this video may not represent an accurate representation of the facts.”

#### B. Labeling Central Authority

Historically, as discussed in section 2, labeling regulations have either been self-imposed or come from government-mandated standards. In this section potential central authorities and third party oversight are considered as possible sources of content labeling.

1) *Industry Self-Regulation*: In recent years, online news media organizations have begun to label certain articles “opinion.” The news outlet is the arbiter as to what constitutes an opinion article. Definitions and reasoning for how these determinations are made are not necessarily consistent or made available to the reader.

Most recently, it has been reported that Twitter intends to label video content which it has deemed to have been manipulated to convey misleading information. Twitter may choose to flag “deep fake” videos or videos edited to make the speaker appear to say something they have not or in deceptive context. From Twitter’s statement, “we may label Tweets containing synthetic and manipulated media to help people understand the media’s authenticity and to provide additional



context.” This new labeling is intended to roll out in March 2020 ahead of the 2020 United States Presidential election [37].

Industry self-regulation runs a potential risk that the industry’s labeling mechanisms may not necessarily serve an altruistic purpose for society. For example, labeling certain articles as “opinion” may serve to make other unlabeled articles be taken to be pure fact. An industry-wide regulatory body, such as the ESRB for the video game industry, could provide definitions for labels that are shared by all outlets agreeing to follow these self-imposed regulations.

2) *Government Regulation*: When the ESRB was created, it was due to the threat of action from the United States Congress. Congress planned to pass regulatory legislation unless the video game industry regulated itself [38]. Nutrition facts labeling came into being after such attempts at industry self-regulation failed. As such, there is historical precedent for regulation to come through Congress. However, government regulation of the media industry in the United States has implications related to the First Amendment to the United States Constitution, which protects freedom of speech and freedom of the press. Furthermore, different nations and future legislative authorities may have different interpretations of what constitutes “safe” versus “dangerous” content. While claiming to protect the public from fake news, a government may instead further confound the consumer with targeted misinformation.

3) *Third Party Applications*: Third party applications, such as web browser extensions, may be put into place by end-users and others in order to filter online content at the consumer side. Ad-blocker extensions allow users to browse online content without advertisements or to avoid sites which otherwise might contain undesirable media. A third-party labeling application could be designed to work as an overlay to news articles and videos. Open-source development could allow consumers to collectively define what is meant by “opinion,” for example, and apply this commonly agreed upon definition when labeling content. One significant challenge of third party labeling is the need to determine the safety of content in an automated fashion. Industry self-regulation and government regulation could require that labels be applied before the media is published. Third-party labeling would need to determine this at the time of viewing using an algorithm.

#### IV. ASSESSMENT OF ONLINE CONTENT NUTRITION FACTS FORM AND COMPONENT DATA

Numerous components could prospectively be included in standardized disclosure for online content. The first question is one of form: the disclosure could take the form of a warning, such as the Surgeon General’s warning used for cigarette packs (shown in Figure 2). Alternately, it could be more informative like the energy guide (shown in Figure 1) or lighting facts. It could also be possible that the disclosure would take an energy guide / lighting facts form unless a key indicator showed potential harm and then a cigarette-style disclosure would be used, in these instances.

Another key question is whether a similar format would be used for children and adults. Of course, nutrition facts are the same whether the consumer of a piece of food is a five-year-old or ninety-year-old; however, online there may be more



Fig. 2. FDA’s proposed new cigarette warning labels [65].

ability to be responsive. Content providers and search engines could be required to show a specialized disclosure for children for any children-targeted websites and when it is known that a child is operating the computer. Children’s disclosures could potentially vary by age, starting with a very simple cartoonish disclosure for young children and moving towards a disclosure similar to the adult one as children mature (also preparing them to use the adult disclosure). In addition, mandatory (for example, V-Chip style) filtering could be implemented for children, based on age and other factors.

Irrespective of the exact form that the disclosure takes (which is a topic for future study), it is key to determine what information elements should be included. These components could form the basis for making the warning / no-warning decision, if a warning-style system was used, or present to the consumer if a ‘facts’ / ‘guide’ style system was used.

Key elements that may be effective for the consumer to view or to analyze include the title of the article, author’s name, publication’s name and a labeled or detected political perspective [39]. A rating (based on manual or automatic ratings of prior work) could potentially be provided for both the author and publication, to further inform evaluation. In addition, the classification from one or several analyzers could be included. Analyzers have been developed based on content simplicity [40], satirical cues [41] and language analysis [42]. They have been developed using technologies including deep learning [20], data mining [22], reinforcement learning [43] and Bayes classifiers [44]. Tacchini, et al. [45] have even proposed techniques specific to social media networks.

Additionally, content filtering based on the presence of profanity [46] may be desirable, both to identify prospectively deceptive content and to help determine the appropriateness for children. Other age-level appropriateness classification information could be supplied by the content provider, detected through analysis of words and phrases present, or both. A meta tag exists, already, to allow web page authors to advise the appropriateness of the content of each web page [47].

The content accuracy disclosure, whatever form it could take, would join a sea of other descriptors that users may use to evaluate web pages. These include information about its security (SSL certificate, privacy policy evaluation), business trustworthiness (such as a Better Business Bureau rating) and user content ratings. It will be critical to ensure that users don’t mistake a rating in one area for suitability in another. For example, a content trustworthiness rating would make no

statement about whether a website was well-secured to make an online purchase from.

## V. CONCLUSIONS & FUTURE WORKS

This paper has explained the need for consumers to have access to additional information when determining what content they want to consume and while consuming it. Consumers are not well equipped to recognize intentionally deceptive content and can, thus, benefit from an automated, manual or combined system that can provide them with relevant information regarding page characteristics or, potentially, even a warning about particularly problematic content.

This paper has reviewed a number of prior labeling systems for consumer products, in the United States, which provide examples of prospective approaches that can be taken for an online content labeling system. Potential considerations with different approaches – and with labeling online content in general – have also been discussed.

Future work is necessary to advance this concept to a point where it would be ready for consumer use. Work is needed in three key areas. First, metric values to be included in the consumer disclosure must be identified. These will be needed irrespective of what format the disclosure takes.

Second, determining what format the disclosure will take will be critical. It could prospectively be a nutrition / energy facts style information panel, which provides information without making a recommendation as to what action a consumer should take. Alternately, it could take a form similar to cigarette warnings, where particularly problematic pages are flagged with a warning. Finally, a combined approach could be designed where, in most cases, the facts panel is presented, but a warning is presented in some extreme cases.

The third area where future work is needed is on the development of the browser plug-in (or similar) to actually implement and facilitate the testing of this type of a system. This plug-in could prospectively be configurable to allow it to ingest different content sources, to facilitate user customization and different configurations to comply with consumer desires and applicable regulations in multiple countries.

Once the foregoing have been completed, different system configurations will need to be tested to determine which are most effective at informing consumers about deliberately deceptive content. Their impact on the consumer experience will also need to be assessed.

## REFERENCES

- [1] H. Allcott and M. Gentzkow, "Social Media and Fake News in the 2016 Election," *J. Econ. Perspect.*, vol. 31, no. 2, pp. 211–236, 2017.
- [2] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer, "Fake news on Twitter during the 2016 U.S. presidential election," *Science* (80-. ), vol. 363, no. 6425, pp. 374–378, 2019.
- [3] M. Balmas, "When Fake News Becomes Real: Combined Exposure to Multiple News Sources and Political Attitudes of Inefficacy, Alienation, and Cynicism," *Communic. Res.*, vol. 41, no. 3, pp. 430–454, 2014.
- [4] C. Kang and A. Goldman, "In Washington Pizzeria Attack, Fake News Brought Real Guns," *New York Times*, 05-Dec-2016.
- [5] J. Acosta, "How Trump's 'fake news' rhetoric has gotten out of control," *CNN Website*, 11-Jun-2019.
- [6] G. Brobeck, "Tenn. lawmaker introduces resolution calling CNN, Washington Post 'fake news,'" *WKYT*, 30-Jan-2020.
- [7] G. Pennycook and D. G. Rand, "Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning," *Cognition*, vol. 188, pp. 39–50, Jul. 2019.
- [8] G. Pennycook, T. D. Cannon, and D. G. Rand, "Prior exposure increases perceived accuracy of fake news," *J. Exp. Psychol. Gen.*, vol. 147, no. 12, pp. 1865–1880, Dec. 2018.
- [9] "More Americans Are Getting Their News From Social Media." [Online]. Available: <https://www.forbes.com/sites/petersuciu/2019/10/11/more-americans-are-getting-their-news-from-social-media/#589ec4d43e17>. [Accessed: 01-Feb-2020].
- [10] E. C. Tandoc, W. Lim, and R. Ling, "Defining 'Fake News': A typology of scholarly definitions," *Digit. Journal.*, vol. 6, no. 2, 2017.
- [11] C. Silverman and J. Singer-Vine, "Most Americans Who See Fake News Believe It, New Survey Says," *BuzzFeed News*, 06-Dec-2016.
- [12] R. Marchi, "With Facebook, Blogs, and Fake News, Teens Reject Journalistic 'Objectivity,'" *J. Commun. Inq.*, vol. 36, no. 3, pp. 246–262, 2012.
- [13] V. Bakir and A. McStay, "Fake News and The Economy of Emotions," *Digit. Journal.*, vol. 6, no. 2, pp. 154–175, Feb. 2018.
- [14] I. D. Barkan, "Industry invites regulation: The passage of the Pure Food and Drug Act of 1906," *Am. J. Public Health*, vol. 75, no. 1, pp. 18–26, Oct. 1985.
- [15] A. T. Borchers, F. Hagie, C. L. Keen, and M. E. Gershwin, "The history and contemporary challenges of the US food and drug administration," *Clin. Ther.*, vol. 29, no. 1, pp. 1–16, Jan. 2007.
- [16] C. Gaziano and K. McGrath, "Measuring the Concept of Credibility," *Journal. Q.*, vol. 63, no. 3, pp. 451–462, Sep. 1986.
- [17] C. L. Corritore, R. P. Marble, S. Wiedenbeck, B. Kracher, and A. Chandran, "Measuring Online Trust of Websites: Credibility, Perceived Ease of Use, and Risk," in *AMCIS 2005 Proceedings*, 2005, p. 370.
- [18] M. Kang, "Measuring Social Media Credibility: A Study on a Measure of Blog Credibility Measuring Social Media Credibility: A Study on a Measure of Blog Credibility," 2010.
- [19] P. Meyer, "Defining and Measuring Credibility of Newspapers: Developing an Index," *Journal. Q.*, vol. 65, no. 3, pp. 567–574, Sep. 1988.
- [20] N. Ruchansky, S. Seo, and Y. Liu, "CSI: A hybrid deep model for fake news detection," in *International Conference on Information and Knowledge Management, Proceedings*, 2017, vol. Part F1318, pp. 797–806.
- [21] D. M. J. Lazer et al., "The science of fake news," *Science* (80-. ), vol. 3, no. 9, pp. 1094–1096, 2018.
- [22] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," *ACM SIGKDD Explor. Newsl.*, vol. 19, no. 1, pp. 22–36, Sep. 2017.
- [23] W. Y. Wang, "'Liar, Liar Pants on Fire': A New Benchmark Dataset for Fake News Detection," pp. 422–426, 2017.
- [24] D. L. Gregory, "The Labor Preemption Doctrine: Hamiltonian Renaissance or Last Hurrah," *William Mary Law Rev.*, vol. 27, 1985.
- [25] "Threading Your Way Through the Labeling Requirements Under the Textile and Wool Acts | Federal Trade Commission." [Online]. Available: <https://www.ftc.gov/tips-advice/business-center/guidance/threading-your-way-through-labeling-requirements-under-textile>. [Accessed: 31-Jan-2020].
- [26] "2000 Surgeon General's Report Highlights: Warning Labels | Smoking & Tobacco Use | CDC." [Online]. Available: [https://www.cdc.gov/tobacco/data\\_statistics/sgr/2000/highlights/label/index.htm](https://www.cdc.gov/tobacco/data_statistics/sgr/2000/highlights/label/index.htm). [Accessed: 01-Feb-2020].
- [27] "Clothes Captioning: Complying with the Care Labeling Rule | Federal Trade Commission." [Online]. Available: <https://www.ftc.gov/tips-advice/business-center/guidance/clothes-captioning-complying-care-labeling-rule>. [Accessed: 31-Jan-2020].
- [28] "Disclosures 101 for Social Media Influencers | Federal Trade Commission." [Online]. Available: <https://www.ftc.gov/tips-advice/business-center/guidance/disclosures-101-social-media-influencers>. [Accessed: 31-Jan-2020].
- [29] "Complying with the Funeral Rule | Federal Trade Commission." [Online]. Available: <https://www.ftc.gov/tips-advice/business-center/guidance/complying-funeral-rule>. [Accessed: 31-Jan-2020].

- [30] "History." [Online]. Available: <https://www.filmratings.com/History>. [Accessed: 01-Feb-2020].
- [31] D. Grossman, "Whatever Happened to the V-Chip?," Popular M, 12-Mar-2018.
- [32] "WELCOME TO FilmRatings.com." [Online]. Available: <https://www.filmratings.com/>. [Accessed: 01-Feb-2020].
- [33] "The V-Chip: Options to Restrict What Your Children Watch on TV | Federal Communications Commission." [Online]. Available: <https://www.fcc.gov/consumers/guides/v-chip-putting-restrictions-what-your-children-watch>. [Accessed: 01-Feb-2020].
- [34] "Ratings - ESRB Ratings." [Online]. Available: <https://www.esrb.org/ratings/>. [Accessed: 02-Feb-2020].
- [35] "Ratings Process - ESRB Ratings." [Online]. Available: <https://www.esrb.org/ratings/ratings-process/>. [Accessed: 02-Feb-2020].
- [36] United States Constitution, First Amendment. United States, 1791.
- [37] J. Pero, "Twitter to start labeling deepfakes and other deceptive media ahead of 2020 presidential election," Daily Mail, 04-Feb-2020.
- [38] J. O'Holleran, "Blood Code: The History and Future of Video Game Censorship," J. Telecommun. High Technol. Law, vol. 8, p. 571, 2010.
- [39] W. Fleck, N. Snell, T. Traylor, and J. Straub, "Development of a 'Fake News' machine learning classifier and a dataset for its testing," in Disruptive Technologies in Information Sciences II, 2019, vol. 11013, p. 9.
- [40] B. D. Horne and S. Adali, "This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News," in NECO Workshop, 2017.
- [41] V. L. Rubin, N. J. Conroy, Y. Chen, and S. Cornwell, "Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News," in Proceedings of NAACL-HLT 2016, 2016, pp. 7-17.
- [42] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, Y. Choi, and P. G. Allen, "Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking," in Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 2931-2937.
- [43] Y. Wang et al., "Weak Supervision for Fake News Detection via Reinforcement Learning," in Proceedings of the 34th AAAI Conference on Artificial Intelligence, 2019.
- [44] G. Gykhailo and Volodymyr Mesyura, "Fake News Detection Using Naive Bayes Classifier," 2017.
- [45] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro, "Some Like it Hoax: Automated Fake News Detection in Social Networks," Apr. 2017.
- [46] B. Kalvoda, B. Stoick, N. Snell, and J. Straub, "Evaluation of Algorithms for Fake News Identification," in Proceedings of SPIE, 2019.
- [47] "meta name = rating | How do I use the metatag Rating?" [Online]. Available: <https://www.metatags.org/all-meta-tags-overview/meta-name-rating/>. [Accessed: 02-Feb-2020].
- [48] "EnergyGuide Labels: Templates for Manufacturers | Federal Trade Commission." [Online]. Available: <https://www.ftc.gov/tips-advice/business-center/guidance/energyguide-labels-templates-manufacturers>. [Accessed: 31-Jan-2020].
- [49] "Voluntary Guidelines for Providers of Weight Loss Products or Services | Federal Trade Commission." [Online]. Available: <https://www.ftc.gov/tips-advice/business-center/guidance/voluntary-guidelines-providers-weight-loss-products-or-services>. [Accessed: 31-Jan-2020].
- [50] "How to Comply with the Fur Products Labeling Act | Federal Trade Commission." [Online]. Available: <https://www.ftc.gov/tips-advice/business-center/guidance/how-comply-fur-products-labeling-act>. [Accessed: 31-Jan-2020].
- [51] "Selling a Work-at-Home or Other Business Opportunity? Revised Rule May Apply to You | Federal Trade Commission." [Online]. Available: <https://www.ftc.gov/tips-advice/business-center/guidance/selling-work-home-or-other-business-opportunity-revised-rule>. [Accessed: 31-Jan-2020].
- [52] "Native Advertising: A Guide for Businesses | Federal Trade Commission." [Online]. Available: <https://www.ftc.gov/tips-advice/business-center/guidance/native-advertising-guide-businesses>. [Accessed: 31-Jan-2020].
- [53] "EnergyGuide Labeling: FAQs for Appliance Manufacturers | Federal Trade Commission." [Online]. Available: <https://www.ftc.gov/tips-advice/business-center/guidance/energyguide-labeling-faqs-appliance-manufacturers>. [Accessed: 31-Jan-2020].
- [54] "The FTC 'Lighting Facts' Label: Questions and Answers for Manufacturers | Federal Trade Commission." [Online]. Available: [https://www.ftc.gov/tips-advice/business-center/guidance/ftc-lighting-facts-label-questions-answers-manufacturers#Which general service lamps must have Lighting Facts labels?](https://www.ftc.gov/tips-advice/business-center/guidance/ftc-lighting-facts-label-questions-answers-manufacturers#Which%20general%20service%20lamps%20must%20have%20Lighting%20Facts%20labels?) [Accessed: 31-Jan-2020].
- [55] "In the Loupe: Advertising Diamond, Gemstones and Pearls | Federal Trade Commission." [Online]. Available: <https://www.ftc.gov/tips-advice/business-center/guidance/loupe-advertising-diamond-gemstones-pearls>. [Accessed: 31-Jan-2020].
- [56] "Environmental Claims: Summary of the Green Guides | Federal Trade Commission." [Online]. Available: <https://www.ftc.gov/tips-advice/business-center/guidance/environmental-claims-summary-green-guides>. [Accessed: 31-Jan-2020].
- [57] "Down...But Not Out: Advertising and Labeling of Feather Down | Federal Trade Commission." [Online]. Available: <https://www.ftc.gov/tips-advice/business-center/guidance/downbut-not-out-advertising-labeling-feather-down-0>. [Accessed: 31-Jan-2020].
- [58] Energy Independence and Security Act of 2007. United States, 2007.
- [59] "Complying with the Made in USA Standard | Federal Trade Commission." [Online]. Available: <https://www.ftc.gov/tips-advice/business-center/guidance/complying-made-usa-standard>. [Accessed: 31-Jan-2020].
- [60] "Complying with the FTC Fuel Rating Rule | Federal Trade Commission." [Online]. Available: <https://www.ftc.gov/tips-advice/business-center/guidance/complying-ftc-fuel-rating-rule>. [Accessed: 31-Jan-2020].
- [61] "Dietary Supplements: An Advertising Guide for Industry | Federal Trade Commission." [Online]. Available: <https://www.ftc.gov/tips-advice/business-center/guidance/dietary-supplements-advertising-guide-industry>. [Accessed: 31-Jan-2020].
- [62] "Offering Layaways | Federal Trade Commission." [Online]. Available: <https://www.ftc.gov/tips-advice/business-center/guidance/offering-layaways>. [Accessed: 31-Jan-2020].
- [63] "15 USC CHAPTER 41, SUBCHAPTER I: CONSUMER CREDIT COST DISCLOSURE." [Online]. Available: <https://uscode.house.gov/view.xhtml?req=granuleid%3AUSC-prelim-title15-chapter41-subchapter1&edition=prelim>. [Accessed: 01-Feb-2020].
- [64] "Classification and Rating Rules," Sherman Oaks, CA, 2010.
- [65] "FDA Proposes New Health Warnings for Cigarette Packs and Ads | FDA." [Online]. Available: <https://www.fda.gov/tobacco-products/labeling-and-warning-statements-tobacco-products/fda-proposes-new-health-warnings-cigarette-packs-and-ads>. [Accessed: 02-Feb-2020].