That's Fake News! Investigating How Readers Identify the Reliability of News When Provided Title, Image, Source Bias, and Full Articles

FRANCESCA SPEZZANO*, Department of Computer Science, Boise State University, USA
ANU SHRESTHA, Department of Computer Science, Boise State University, USA
JERRY ALAN FAILS, Department of Computer Science, Boise State University, USA
BRIAN W. STONE, Department of Psychological Science, Boise State University, USA

As news is increasingly spread through social media platforms, the problem of identifying misleading or false information (colloquially called "fake news") has come into sharp focus. There are many factors which may help users judge the accuracy of news articles, ranging from the text itself to meta-data like the headline, an image, or the bias of the originating source. In this research, participants (n = 175) of various political ideological leaning categorized news articles as real or fake based on either article text or meta-data. We used a mixed methods approach to investigate how various article elements (news title, image, source bias, and excerpt) impact users' accuracy in identifying real and fake news. We also compared human performance to automated detection based on the same article elements and found that automated techniques were more accurate than our human sample while in both cases the best performance came not from the article text itself but when focusing on some elements of meta-data. Adding the source bias does not help humans, but does help computer automated detectors. Open-ended responses suggested that the image in particular may be a salient element for humans detecting fake news.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing**; • **Information systems** → *World Wide Web.*

Additional Key Words and Phrases: news credibility, fake news, meta-data analysis

ACM Reference Format:
Francesca Spezzano, Anu Shrestha, Jerry Alan Fails, and Brian W. Stone. 2021. That's Fake News! Investigating How Readers Identify the Reliability of News When Provided Title, Image, Source Bias, and Full Articles. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 109 (April 2021), 19 pages. https://doi.org/10.1145/3449183

1 INTRODUCTION

The problem of misleading or false information disguised as news content (colloquially called "fake news") has drawn a great deal of renewed attention recently due to the proliferation and popularity of social media as a platform for information diffusion. Fake news has been identified as being more likely to go viral than real news, spreading both faster and wider [47]. Additionally, users are more

*Corresponding author. Email: francescaspezzano@boisestate.edu

Authors' addresses: Francesca Spezzano, Department of Computer Science, Boise State University, Boise, ID, USA, francescaspezzano@boisestate.edu; Anu Shrestha, Department of Computer Science, Boise State University, Boise, ID, USA, anushrestha@u.boisestate.edu; Jerry Alan Fails, Department of Computer Science, Boise State University, Boise, ID, USA, jerryfails@boisestate.edu; Brian W. Stone, Department of Psychological Science, Boise State University, Boise, ID, USA, brianstone984@boisestate.edu.

2573-0142/2021/4-ART109 $15.00
https://doi.org/10.1145/3449183

Fig. 1. Recent examples of false social media posts (as identified by snopes.com) that illustrate use of title and image (T+I) and text excerpts (E).

likely to share headlines they have seen repeatedly than headlines that are novel, even when they know the headline is false [13]. Fake news has become more of an issue, given the world's political climate and the rampant use of social media and a number of studies have sought to identify and mitigate fake news [43].

Within this context, this research seeks to better understand the accuracy of people and automatic detectors when evaluating fake news with varying granularity of information. Specifically, we seek to understand how the article title, image, source bias, and text excerpt affect people's accuracy in identifying reliable or fake news. We selected these meta-data elements because this kind of information is readily available and is often shared on social media (see Figure 1 for recent examples).

Our investigation furthers research in this area by identifying trends among participants with regards to their ability to identify real and fake news, and initiates a discussion on how various meta-data affect people's ability to identify reliable news as compared and contrasted with automatic detection mechanisms using the same news article meta-data and content information. Specifically, we conducted an online survey where we asked participants ($n = 175$) to categorize news as real or fake and provide an explanation for their decision. We varied the way the news was presented to the participants and considered four conditions: news excerpt only and three combinations of various meta-data elements including: (1) news title, image, and source political bias, (2) news title and image, and (3) news title and source bias. We used a combination of quantitative and qualitative research methods to compare participants' accuracy in correctly identifying news veracity and

investigate the reasoning participants gave to justify their choices. Furthermore, we used machine-learning to build an automated fake news detector by using the same type of information provided to the survey participants (news excerpt and meta-data) and compared human versus automated detector accuracy under the same conditions to better understand how humans and computers utilize those elements to determine real and fake news.

Our research shows that humans are less accurate in identifying fake news when they have only the text of the article itself compared to when they rely on meta-data. In particular, participants were more accurate when they had the picture available (accuracy of 0.62 when news title and image are available versus 0.53 when only the news excerpt is available). Analysis of open-ended participant responses revealed that the professionalism of the image was a helpful heuristic that enabled more accurate judgments. Overall, political ideological leaning had little effect, though left-leaning participants were more accurate than right-leaning participants when categorizing an article as real or fake based solely on the headline plus accompanying image. Automatic detection outperformed humans on identifying fake news articles (overall best accuracy of 0.83 versus 0.62).

2 RELATED WORK

The study of how people assess the credibility of online information requires a multidisciplinary effort as it touches information science, psychology, sociology, communication, and education. Metzger and Flanagin have summarized existing research on the factors that influence people when making credibility evaluation decisions under different categories: site or source cues; author cues; message cues; receiver characteristics; and social interactions [29]. Furthermore, the dual processing model of credibility assessment states that people tend to use two general strategies, namely *analytic* and *heuristic*, that reflect a greater and a lesser degree of cognitive rigor, respectively [28]. The *analytic* strategy suggests that people are using what is provided and what they know in order to produce a reasoned assessment. *Heuristic* implies that people use pre-existing rules that they adapt to the case at hand. This also involves a superficial evaluation of the piece of information where the user's gut feelings are often predominant.

People are generally not able to correctly judge news veracity. Specific to the news domain, research by the Pew Research Center [30] showed that only 26% of Americans could accurately classify all five provided factual statements and only 35% could classify all five provided opinion statements. Similarly, Horne et al. [22] showed that when both news title and excerpt are provided, humans' ratings for article reliability were, on average, 6.64 for articles from reliable sources and 5.01 for articles from unreliable sources (on a 10-point scale ranging from "completely unreliable" (1) to "completely reliable" (10)). Several social and psychological theories explain why people are not able to accurately judge news veracity, including the backfire effect [34] and conservatism bias [2] which explain why it is hard for people to revise their beliefs even when presented evidence against them. Also, people are more prone to trust information that confirms an individual's pre-existing beliefs (confirmation bias [33]) and are more likely to believe a claim simply from reading it multiple times (validity effect [4]). Information processing styles can also affect misperceptions related to news. Those who tend toward online processing (forming judgments and updating attitudes immediately when encountering new information) seem to be more open to correction of misinformation than those who tend toward memory-based processing (where information is stored and then retrieved or sampled at the time of judgment, as needed) [6].

Additionally, the dual processing model is relevant to judging the veracity of news. A recent study showed that individuals who tend toward an analytic strategy (scoring higher on the Cognitive Reflection Test) are better at distinguishing real and fake news when presented with a combination of meta-data including headline, image, byline, and source [38]. This pattern held among people on

both sides of the political aisle; a lazier, more heuristic style of thinking rather than ideological leanings seemed to predict difficulty recognizing fake news.

We should not be surprised, then, to learn that users often do not invest the time to fully process information before sharing it, but may instead rely on meta-data such as headline, image, or source. For example, users of a popular social news aggregator, Reddit, can opt to rate content they see through an "upvote" or "downvote", yet research has found that roughly three-quarters of ratings occur without the user actually reading the content; indeed, most users do not read the article they vote on [18]. It is not surprising, then, that user interaction with content, including attention, rating, and engagement (e.g., commenting) are predicted by elements of the title [26]. Likewise, a study of Twitter found that for shortened bitly URLs the majority are shared without being read [15].

Cognitive biases and attentional limitations can come into play not just in judging accuracy but in the decision to share a given news story or not. For example, Pennycook et al. [38] found that accuracy judgments of a story are not strongly influenced by political leanings (though willingness to share a story is). Indeed, they found that most people claim accuracy is extremely important when deciding whether to share a news story. However, they argue that the context of social media can pull users' attention away from that value and toward a stronger weighting of other motives – such as signaling group membership or attracting followers – or a stronger influence by factors like emotional and moral valence [5]. They suggest that the sharing of fake news comes not from a conscious decision to prioritize politics over truth, but from attentional constraints. Bago et al. [1] find results consistent with this account [5]. In one condition, they measured participant accuracy in identifying fake or real news headlines where participants first made a speeded judgment of each headline while under cognitive load (i.e. distracted by a concurrent working memory task) then later deliberated on their earlier judgments under no such constraints. Under these conditions, deliberation did in fact correct previous mistakes made by participants' heuristic system when distracted, and did so regardless of the headline's concordance with political belief.

While some other recent work suggests that asking users to "take their time" and "deliberate" may have little effect on their judgments of fake news headlines [13], Pennycook et al. [38] found that merely focusing users' attention on the concept of accuracy (i.e., making it more attentionally salient) *can* reduce sharing intentions for false content. Knowing this, social media platforms could integrate such information into their presentation layers in order to focus user attention in a way that makes their existing accuracy preference more salient and more likely to intervene in behaviors such as sharing. For example, platforms could label news that an algorithm categorizes as potentially fake, and in so doing call user attention to issues of veracity. Research by communication scholars has shown that labeling, warnings, or corrective information can weaken the effects of misinformation [3, 16, 48, 50]. However, the primary focus of these researchers was on how effective a correction was for users (e.g., a correction from the CDC versus from a friend).

Several methods have been proposed to automatically classify a piece of news as real or fake [24, 54]. These methods use features extracted from the news content and title [10, 21, 39, 40, 43], associated image [53] (but without considering image emotions or quality as we did in this research), social network context [44], and news propagation patterns in social networks [47].

Horne et al. [22] showed that AI assistance with feature-based explanations improves people's accuracy of news perceptions while Yaqub et al. [51] examined the role of different news credibility indicators (fact-checking, news media outlets that dispute news credibility, the public, and AI systems) in decreasing the propensity to share fake news and showed fact-checking to be the most effective of the considered indicators.

Several studies have highlighted that humans are generally poorer at identifying false information in comparison with automated detectors. For instance, people were 66% accurate at detecting Wikipedia hoaxes [25] (while the computer achieved 86% accuracy) and people had an accuracy

ranging between 53% and 62% in identifying fake reviews [35] (whereas the computer achieved approximately 90% accuracy). While Wikipedia hoaxes and fake reviews are related, there is a difference between them and fake news. Specifically, a fake review is a dishonest individual's opinion in a context where there is no absolute ground truth, a Wikipedia hoax is false information pretending to be true (and sometimes intended as a joke), while fake news is false or misleading information that is spread deliberately to deceive (see Molina et al. [31] for a taxonomy distinguishing fake news from other content like satire, commentary, misreporting, and so).

Ringel-Morris et al. [32] systematically manipulated elements of Twitter posts to see how those elements affected credibility assessment by users. They found that users perform poorly in accurately identifying truthfulness by content alone (where the length of a tweet is comparable in length to a news headline), and instead are influenced by shortcut information (e.g., name and image/avatar of content poster). The heuristic shortcut information – or meta-data that condenses, distills, and represents – makes it easier to identify misinformation. Whatever the reason, it is more difficult for people to ascertain whether an article is true or not from longer texts.

To understand which elements of an article drive mistakes in discerning fake news from real news, it would be helpful to compare peoples' accuracy in judging various combinations of meta-data elements and text excerpts. A preliminary attempt in this direction can be found in Zhang et al. [52], who asked six trained readers (three critical thinking instructors and three journalism students) to use specialized tools to annotate news articles based on *content indicators* like 'clickbait' title and emotional tone or *context indicators* like external fact checking results, ads, layout, and impact factor of a journal. Annotators had low inter-rater agreement overall, but in both conditions the researchers found a handful of annotated indicators that correlated with credibility scores given by a few domain experts, more for the *context* condition than the *content* condition (for example, if an article had been fact-checked as false on an external website, that predicted the domain experts finding it less credible). This hints at the possibility that text-based indicators are less reliable at discerning fake news than meta-data and/or external information, but the study used different expert annotators (university instructors vs. journalism students) using different tools in each condition, so it is hard to directly compare the conditions and does not tell us about everyday consumers of news.

The present study builds on previous work by: (1) comparing accuracy while systematically manipulating which information the user has access to (i.e. news excerpt versus various combinations of meta-data), (2) analyzing the specific reasons humans cite for classifying a news item as real or fake under those conditions, and (3) comparing human and computer accuracy at fake news detection under the same conditions (i.e., access to the same type of information).

3 METHODS

This section describes the dataset we used for the evaluation and then presents the methods used to evaluate the accuracy of people and automated detectors (computers) when provided various news article information.

3.1 Dataset: FakeNewsNet

In order to conduct the evaluations, we utilized the FakeNewsNet dataset [44]. While other datasets exist (e.g., [22]), some of those assign the "ground truth" of an article at the source level (whether a source tends to produce true or fake news), whereas the FakeNewsNet dataset contains news articles (title, excerpt, and associated image) individually labeled as real or fake by Politifact and Buzzfeed fact-checking websites. Some articles did not have all of the elements (title, excerpt, and associated image), so those were removed. This resulted in a set of 384 articles (from the full set of

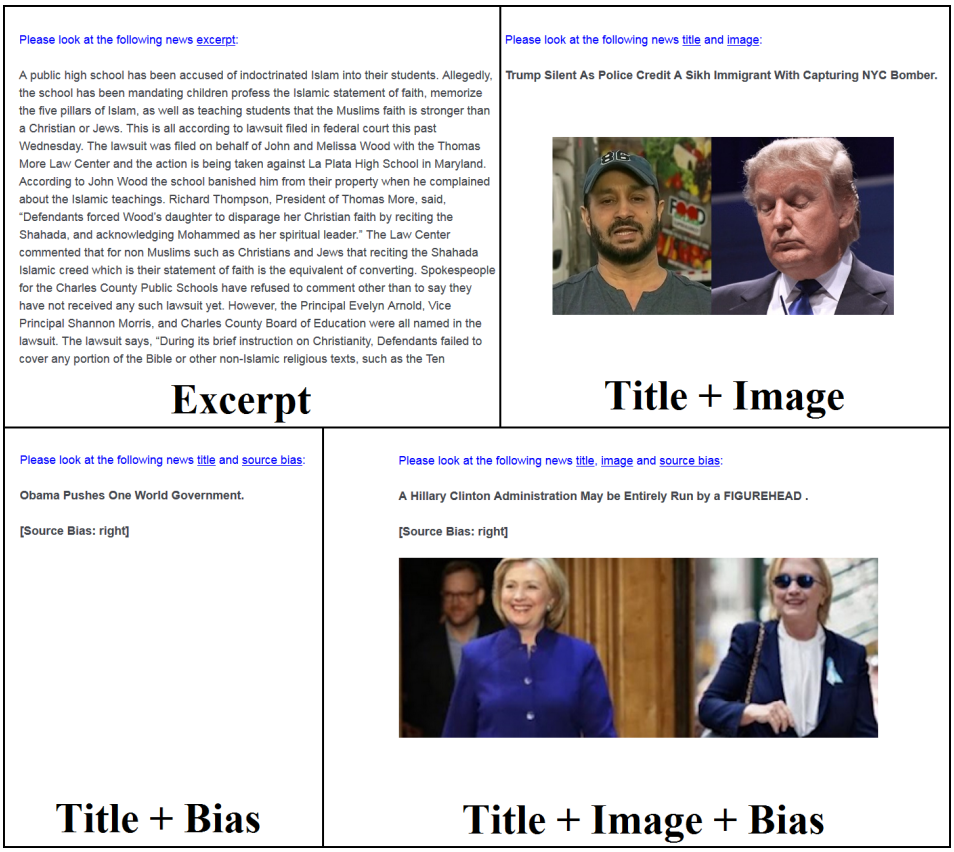


Fig. 2. Example stimuli from each condition: news text excerpt (E); title and image (T+I); title and source bias (T+B); title, image, and source bias (T+I+B). Participants saw 3 examples for the Excerpt condition, and 5 examples for each of the other conditions.

422 articles). We extracted the article source bias from the MediaBias/FactCheck[1] website which assigns seven degrees of bias: extreme-right, right, right-centered, least-biased, left-centered, left, and extreme-left.

3.2 Evaluation by People

This study was conducted using an online "Fake News" survey delivered via Qualtrics. Through this online survey, participants were asked to judge whether news items were real or fake news and then explain the reasoning for their decision. We randomly selected 16 real and 16 fake news articles from the FakeNewsNet dataset described above in Section 3.1. The 16 articles were randomly selected and balanced in terms of the number of left-leaning (extreme, left, and left-center) and right-leaning news (extreme, right, right-center) for each category of news used (real or fake).

The participants responded to four different types of questions where we varied the news information provided: title and image (T+I); title and source bias (T+B); title, image, and source

[1]Other datasets exist that categorize the bias of new sources, for example [42]. We checked the Spearman's rank correlation between the MediaBias/FactCheck scores and the ones provided by [42] and obtained that they are aligned with a correlation of $\rho = 0.90$.

bias (T+I+B); news text excerpt (E). Figure 2 shows sample survey questions for each considered condition. Each participant evaluated five articles for each of the conditions T+I, T+B, and T+I+B, and three articles for condition E[2]. As such, each participant was exposed to a total of 18 articles. For each article in each condition, we asked participants to evaluate the veracity of the article as real or fake and provide an explanation for their judgement. For each condition, the articles were randomly assigned from the set of 32 articles we considered for our study. Different participants frequently evaluated the same articles in the same condition but co-occurrences were randomly distributed across participants due to the random assignment. The order of the presentation of the conditions was randomized and no feedback was given to the participants throughout the experiment.

We recruited undergraduate students (n = 175) from a volunteer pool in general education social science courses (Psychology 101) to participate in our survey (107 F, 68 M; mean age 19.5, SD = 2.4). The research was approved by the university IRB. Participants were compensated with course credit (volunteering for studies being one option for a research experience requirement). Participants received no training. Participants were just asked to evaluate each article given the information presented to them. Since this was conducted online, we cannot be sure that participants did not look at external sources to help them ascertain whether the item was real or fake; however, the median completion time for the survey was about 28 minutes so extensive lateral reading [49] is unlikely.

3.3 Reasoning Given for Evaluations

To investigate the reasoning participants gave for classifying news articles as real or fake, we analyzed open-ended responses that were given by participants as to why they classified each item as real or fake. In doing so we used an inductive approach to generating codes [12]. Specifically, three of the authors first openly coded 160 participant responses, and independently developed codes to describe the observed participant responses on the same sample. The researchers then met and discussed each of the participant responses that were coded. They then compared, discussed, consolidated, and defined the codes that described the data. They then collectively agreed on the codes that should be used for each of the responses in this initial sample. With the codes defined, all three researchers then independently coded another subset of the data, and then met again to confirm the codes identified for each of the responses in the second sample. Codes were used to identify a response only if at least two of the three reviewers utilized that code, and if all reviewers were in agreement on using that code during the review meeting. The purpose of this coding was to identify the prominent themes that illustrated the approaches used by participants to identify whether news items were real or fake. The results of this analysis are found in Section 4.2.

3.4 Evaluation by Computer – Automated Detector

We used the same information evaluated by our participants and utilized an automated machine learning-based fake news detector to compare people vs. machine accuracy and the predictive power of the different conditions (T+I, T+B, T+I+B, E). To build the automatic detector, we considered the whole FakeNewsNet dataset which contains 384 news (half real and half fake) to train and test with a 10-fold cross-validation logistic regression model. We considered the following features in input to the model.

[2]Participants were presented fewer excerpt only (E) articles due to the additional length it took for them to evaluate that condition. Evaluating the title (T), title and image (T+I) and title, image and source bias (T+I+B) did not take participants very long to evaluate.

To encode text data, i.e., news title and excerpt, we considered features that focus on linguistic style, text complexity, and psychological aspects such as Linguistic Inquiry and Word Count (LIWC) features and text readability measures. LIWC features are grouped into: *linguistic features* such as the average number of words per sentence, rate of misspelling, negations, as well as part-of-speech; *punctuation features* that include the kinds and frequency of punctuation; *psychological features* representing emotional, social, and cognitive processes present in the text; and *summary features* defining the frequency of words that reflect the thoughts, perspective, and honesty of the writer. Another approach is the Rhetorical Structure Theory (RST) which captures the writing style of documents [45]. However, since different studies have shown that the performance of LIWC is comparatively better than RST [44, 45], we did not use RST in our analysis.

Readability measures how easily the reader can read and understand a text. We use popular readability measures in our analysis: Flesch Reading Ease, Flesch Kincaid Grade Level, Coleman Liau Index, Gunning Fog Index, Simple Measure of Gobbledygook Index (SMOG), Automatic Readability Index (ARI), Lycee International Xavier Index (LIX), and Dale-chall. We chose to use a topic-agnostic approach for processing the text and did not consider topic-dependent features (from the widely used bag-of-words to the most recent BERT [11] deep learning-based approach) as they are not well-suited for the dynamic environment of news where stories' topics change continuously.

To extract features from the images associated with news articles, we considered several tools including (1) the ImageNet-VGG19[3] state-of-the-art deep-learning based techniques to extract features from the images [46], (2) features describing face emotions, and (3) features referring to image quality such as noise and blur detection. To capture face emotions in images, we used Microsoft Azure Cognitive Services API to detect faces in an image[4] and extract several face attribute features. Among all the features extracted, we consider face emotion (anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise) and smile features. Each of these features ranges in [0,1] and indicates the confidence of observing the feature in the image. To capture news image quality to some extent, we computed the amount of blur in an image by using the OpenCV blur detection tool[5] implementing a method based on the Laplacian Variance [36] along with noise level of face pixels provided by Microsoft Azure Cognitive Service API. We also used the article source political bias as identified by MediaBias/FactCheck as a feature to encode bias.

To avoid overfitting, from this set of features, we selected the top-30 most important title text features, the top-30 most important excerpt text features, and the top-10 most important image features via univariate feature selection.

4 RESULTS

Herein we present the comparisons of the conditions (varying levels of article information) when evaluated by people and a computer.

4.1 Comparing Conditions Evaluated by People

On average across all of the conditions, participants agreed with their assessments of real and fake news 69.8% of the time (min=67.6%, max=71.7%; with a mean standard deviation of 12.8%). Below we present the results of participants' judgments with regards to overall accuracy with respect to the various news elements they were provided, as well as a comparison by participant's ideological leaning.

[3]We removed the classification layer of the VGG19 model, and used the last fully connected layer of the neural network to generate a vector of latent features representing each input image. Moreover, we used PCA to reduce the number of extracted features to 10.

[4]https://docs.microsoft.com/en-us/azure/cognitive-services/face/quickstarts/csharp

[5]https://www.pyimagesearch.com/2015/09/07/blur-detection-with-opencv/

4.1.1 Accuracy. Participants completed multiple examples per condition so we used their average accuracy per condition as the dependent variable. The summary of accuracy for each condition is given in Figure 3. We used a Friedman's test (non-parametric ANOVA for related samples) to compare the overall accuracy between the four conditions (T+I, T+I+B, T+B, E). There was a significant difference between conditions, $X^2(3) = 14.198, p = 0.003$, so we followed up with pairwise comparisons using a Wilcoxon signed-ranks test for matched samples and Bonferroni correction for multiple comparisons. As reported in Table 1, we found that participants were significantly more accurate in the title+image (T+I) (accuracy of 0.621) and the title+image+bias (T+I+B) (accuracy of 0.617) conditions than in the title+bias (T+B) condition (accuracy of .559) or excerpt (E) condition (accuracy of 0.533).

Table 1. Comparison of accuracy between conditions: news text excerpt (E); title and image (T+I); title and source bias (T+B); title, image, and source bias (T+I+B). Note: with Bonferroni correction, only a p-value < 0.0083 would be significant at a family-wise alpha level of 0.05.

	E	T+I	T+B	T+I+B
E		z = -3.270 p = 0.001*	z = -1.325 p = 0.185	z = -3.186 p = 0.001*
T+I			z = -2.881 p = 0.004*	z = -0.095 p = 0.924
T+B				z =-2.751 p = 0.006*

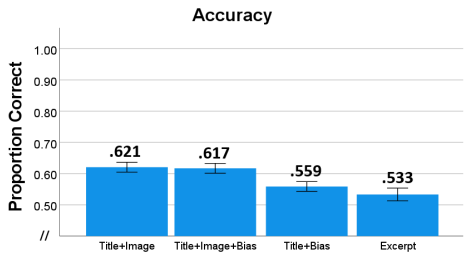
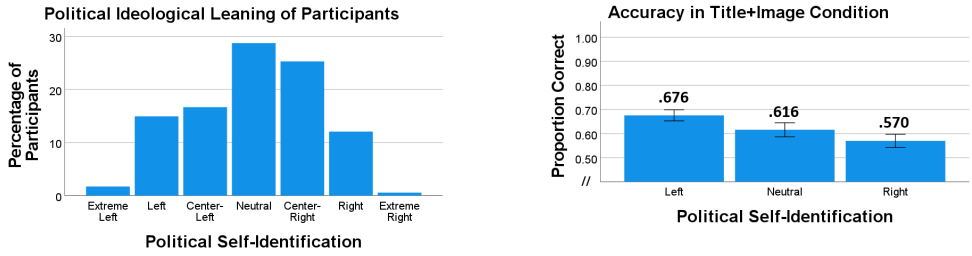


Fig. 3. Mean accuracy for each condition. Error bars are standard error of the means (SEM).

4.1.2 Comparison by Ideological Leaning. Few participants self-identified as extremely left or right with regards to political ideological leaning (see Figure 4a), so we collapsed political ideological leaning into three groups (left-leaning $n = 58$, neutral[6] $n = 50$, and right-leaning $n = 66$) and then compared the accuracy of those groups in each of the conditions using a Kruskal-Wallis test (non-parametric independent ANOVA). Political group did not relate to accuracy in T+B, T+I+B, or E conditions (all p's > 0.5) but there may have been a difference in the T+I condition ($p = 0.036$, not significant after Bonferroni correction). Pairwise comparisons using a Mann Whitney U test showed that left-leaning participants were significantly more accurate than right-leaning participants in the Title+Image condition, as illustrated in Figure 4b (accuracy of 0.676 vs. 0.570, $p = 0.011$, significant after Bonferroni correction; the other two comparisons (left vs. neutral and right vs. neutral) were not significant).

[6]Some established scales use *moderate* instead of *neutral* [23], we acknowledge this as a potential limitation, however this was not central to our analysis as the accuracy of participants was not significantly different from other groups.



(a) Distribution (b) Accuracy

Fig. 4. (a) Distribution of participants' self-identified political ideological leaning; (b) Accuracy in Title+Image condition for each political group. Accuracy in other conditions did not differ significantly between political groups. Error bars are SEM.

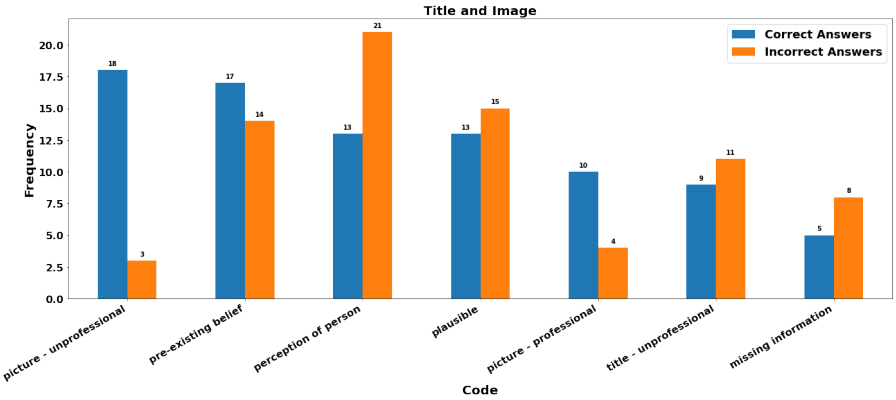
4.2 Reasons for Participant's Judgement of Real or Fake

Given that the highest accuracy was in the title and image (T+I) condition and the lowest accuracy was in the text excerpt (E) condition, we analyzed open-ended responses that were given by participants as to why they classified each item as real or fake from these two conditions. The inductive approach that we used to generate and assign codes was described in Section 3.3. In total, the coded data comprised 320 participant responses as to why they identified the news item as real or fake. Half of these were from the title and image (T+I) condition and half from the text excerpt (E) condition.

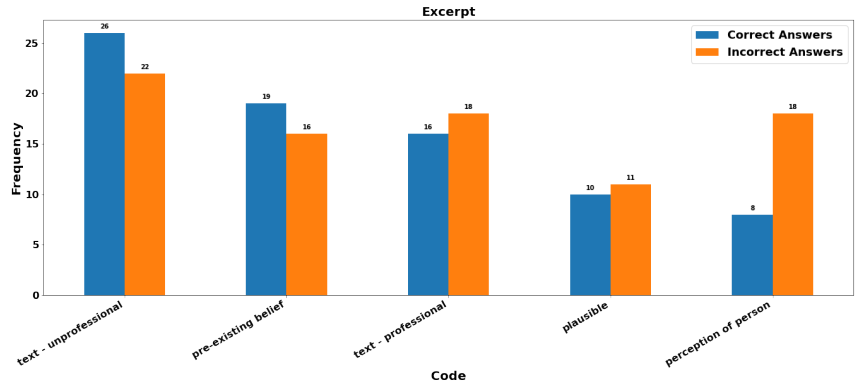
The codes included several dichotomous pairs including plausible/implausible, familiar/unfamiliar, professional/unprofessional (for various news items: title, picture, text), as well as other items including an observation surrounding the emotion the news item intended to evoke, and whether the judgment was based in part on their perception of a particular person, a pre-existing belief, that there was missing information, or whether they simply guessed or were unsure. Table 3 in Appendix A lists the codes, a description of the code, and provides illustrative representative examples. The five most frequently used codes for the T+I and E conditions are indicated in Figure 5.

In the T+I condition, open-ended responses that referred to elements of the image were more often correct than incorrect in accurately labeling the news items as real or fake. For example, correct responses often came with reasoning such as: "Pictures look fake and photoshopped", "the picture seems pretty neutral", and "something seems off about the picture". On the other hand, open-ended responses with more subjective elements (those relating to the respondent, like preconceptions about the person in the news story) were more often incorrect than correct in labeling the news items as real or fake. For example, incorrect responses often came with reasoning such as: "Sounds like something the Clintons would do", "Given that liberals don't tend to be huge fans of Trump this seems like a legitimate even", and "Obama should not say this sort of stuff".

In the text excerpt (E) condition, the pattern in open-ended response themes was less clear, probably reflecting the greater challenge participants faced with this condition (having the lowest accuracy overall). Once again, responses that referred to preconceptions about the person in the news story tended to show up more in incorrect responses than correct ones. For example, "[Hillary] isn't that immature to do something like that", "Donald Trump tends to have lots of fake news", and "I believe this is true because Hillary has been caught in a lot of issues so this doesn't surprise me". However, the other most common themes in open-ended responses (how professional or unprofessional the writing was, plausibility, or references to pre-existing beliefs) did not clearly correlate with accuracy.



(a) Title + Image – Note that for this condition, there are three codes shared within the top five most frequently used codes for correct and incorrect responses. Codes are in decreasing frequency order for correct responses.



(b) Excerpt – Note that for this condition, the top five categories are the same for both the correct and incorrect responses. Codes are in decreasing frequency order for correct responses.

Fig. 5. The five most frequent codes for the Title and Image (5a) and Excerpt (5b) conditions differentiated by when participants responded correctly and incorrectly.

4.3 Comparing Conditions Evaluated by Computer

The computer detector's accuracy was evaluated and compared based on being provided the various news elements (title, image, source bias, and excerpt). This is analogous to the comparison performed on the human identifications in Section 4.1.1.

4.3.1 Accuracy. Table 2 reports the accuracy achieved by the above described automatic detector measured by using a 10-fold cross-validation. As can be seen, the combination of title, image, and bias features achieves the best accuracy of 0.83, while news excerpt features are the least accurate (accuracy of 0.71).[7]

[7]The focus of this work was not to build the best classifier, but rather to see how a machine does in comparison to people evaluating using a good classifier.

Table 2. The accuracy of a computer model in identifying real and fake news when using: excerpt (E); title and image (T+I); title and source bias (T+B); title, image, and source bias (T+I+B).

Condition	Features	Accuracy ± STD
E	LIWC & readability on excerpt	0.71 ± 0.07
T+I	LIWC & readability on title + vgg19 & emotions & quality on image	0.78 ± 0.08
T+B	LIWC & readability on title + bias	0.81 ± 0.07
T+I+B	LIWC & readability on title + bias + vgg19 & emotions & quality on image	0.83 ± 0.05

5 DISCUSSION

In this section, we discuss the results with observations regarding: human accuracy when presented with various combinations of meta-data or an excerpt, the need and value of meta-data, the reasons for participants' judgments of news veracity (real or fake), that source bias seems to not help humans in their determination, and potential reasons for the effect of political leaning in the T+I condition. Also, we discuss the comparison between human and machine evaluations, and examine and explain mis-classifications (for both humans and our automated detector).

As can be seen in Figure 3 (humans) and Table 2 (computer/automatic detector), machine-learning-based techniques are dramatically more accurate than humans. This is not too surprising as automatic detectors have been shown to be more effective in identifying other forms of false information such as fake reviews and Wikipedia hoaxes [25, 35], even though Kumar et al. show that this may not necessarily be true when both humans and computers have access to the same information (66% for humans vs. 47% for the automated detector) [25].

It is of note that accuracy is worse for both humans and the computer when looking at just the excerpt text, which indicates the importance of the other information (title, image, and resource bias) in identifying fake news. This fits with and extends the results of previous work showing that the meta-data of Twitter posts can influence reader's perception of credibility more than the content alone [32]. Indeed, eye tracking data suggests that time spent looking at meta-data such as headline, byline, and timestamp predicts discernment of fake from real news [50]. Other work has demonstrated lasting improvement in fake news discernment when users are trained to look for problematic elements of article titles [20], hinting that automated tools may provide more assistance to users if they can point to specific elements (T+I, T+B, T+I+B, E) rather than a holistic flagging of the entire article.

According to our participants' reasoning, an excerpt with a neutral tone, quotes, and statistics is perceived as professional, while an emotional tone is perceived as unprofessional (see "Text Professional" and "Text Unprofessional" codes in Table 3). However, by analyzing the excerpt of real and fake news from the FakeNewsNet dataset with text-processing techniques, we did not find that, on average, the real news in our dataset had a non-emotional tone and more statistics.[8] In analyzing the text we did find that the real news in our dataset had more quotes than the fake news on average (2.76 vs. 1.61, $p < 0.001$). Thus, we observe that fake news can more readily deceive users if it is written with a neutral tone and contains statistics and quotes. Additionally,

[8]We considered emotional tone and number (as a proxy for statistics) features from LIWC.

we identified that participants' prior perception of the people being reported on more frequently led them to inaccurately identifying whether the news was real or fake. Interfaces could highlight that news about people can lead users to inaccurately identify news and urge them to look at all elements or highlight specific elements that could better help readers discern its veracity (e.g., the emotion found within the headline, the presence of confirmed statistics, and cited quotes).

Previous studies investigating discernment of fake and real news have presented multiple meta-data elements together (e.g., displaying headline, image, and source information for every item [38]) whereas our study begins to tease apart which elements are most relevant. Also we compared human vs. computer accuracy with varying combinations of news elements (T+I, T+B, T+I+B, E) which is also novel. Horne et al. [21] found evidence that the news title is more informative than its excerpt, but, to the best of our knowledge, there is no study comparing people's accuracy in judging combinations of news meta-data elements and excerpt.

While not conclusive from our data, including the source bias did not result in significantly higher accuracy if participants already had the title and the image. There is the possibility that reporting the news source bias does not assist people in their determination of reliability. This could be due to mistrust of the labeled bias and a tendency to over-trust sources that are concordant with their political affiliation (confirmation bias [33]). On the contrary, by adding the source bias in our automatic detector, we increased the accuracy from 0.78 (T+I) to 0.83 (T+I+B). Indeed, several studies in the field of journalism have theorized a correlation between the political bias of a publisher and the trustworthiness of the news content it distributes [14, 17].

Like Pennycook et al. [37], we did not find a large effect of political ideological leaning on accuracy, with the exception of a possible difference in the T+I condition. While limited to only one condition, this is an important case to investigate further as many social media shares include the title and the image associated with the article. Lazer et al. suggest fake news may be more of an issue for those who are right-identified [27], and that conservatives are more suspicious of fact-checking sites. Carraro et al. suggest attentional mechanisms are different in liberals versus conservatives, such that negatively valenced information draws the attention of, and impairs the performance of, conservatives more than liberals [7, 8]. This may be one explanation as to why right-leaning participants appear to show worse accuracy in the T+I condition. If the salient or influential aspects of fake news images come from valence (especially negative valence), that could diminish processing in right-leaning participants when fake news images are present.

Beyond apparently undervaluing the source bias as explained above, we noted in Section 4.2 that people mistakenly judge the veracity of news when they focus on more subjective elements (those relating to the respondent, like preconceptions about the person in the news story). On the other hand, the machine focuses more on objective elements of the news and is not influenced by reader's biases. To better understand where the automatic detector made mistakes, we used LIME[9] – a state-of-the-art technique [41] – to explain the reasons for the false positive and false negative instances. Like our qualitative analysis for humans, we applied this technique to the title and image (T+I) and excerpt (E) conditions. In the T+I condition, the machine mistakenly classified real news as fake (false positive) because the use of punctuation (parentheses and dash), negations, and male related words in the title was more similar to the style of fake news (fewer parentheses, dash, negations, and male related words than other real news in the considered dataset). Also, the machine mistakenly classified fake news as real (false negative) because the use of exclamations, stop words, religion, death, sexual, and tentative related words in the title was similar to the style of

[9]LIME is a technique that explains single instances by creating an interpretable representation that is locally faithful to the classifier. For each instance to explain, LIME computes a local linear classifier and uses the feature weights of the local classifier to assign an importance to the features. The most important features are the ones who explain the label assigned by the global classifier to the given instance.

real news (fewer use of exclamations, religion, death, sexual, and tentative related words and more use of stop words than fake news in the considered dataset). We did not observe a clear pattern in the explanation of false positives and false negatives when we considered the excerpt (E) condition, which is similar to how participants' open-ended responses in this condition did not correlate with their accuracy.

Moreover, we also investigated whether humans and the automatic detector made the same mistakes. We considered a piece of news as human mistake if $\leq 60\%$[10] of the participants who answered that piece of news got its credibility right, while automatic detector's mistakes are given by false positives and false negatives. We found that, in the T+I condition, out of all the mistakes humans made, 37.5% of them were also mistakenly classified by the automatic detector, while, for the E condition, out of all the mistakes humans made, 38.1% of them were also made by the automatic detector. Thus, in both conditions, the majority of humans' mistakes can be corrected by assisting uncertain or confused humans with an automated detector.

6 LIMITATIONS

One potential limitation of our study is that the automated detector was trained on 384 articles whereas the analogous "training" for human participants would be an unknown and heterogeneous amount of past experience with many years of news items. Thus, the comparison between AI and human performance can never be a 'fair' apples-to-apples comparison despite both having access to the same types of information in a given condition (e.g., title+image). However, in no way does it undermine the usefulness of machine learning models to assist humans in discerning real and fake news, even if machine learning models solve the problem differently than a human brain.

One minor limitation is that our political leaning question used "neutral" in the central position of the political leaning scale instead of the more frequently used "moderate". We acknowledge this may have misled participants. It is possible some respondents just assumed it was similar to moderate (as in the question it was shown between "left-centered" and "right-centered"). However, others may have interpreted it to mean not ideological. This was not central to our analysis however, as the accuracy of participants in this category was not significantly different from the one in other groups.

Furthermore, older readers are more likely to share fake news [19], so many studies have focused on older users, whereas the present work focuses on a younger sample. Our results may or may not generalize to older individuals; but young people use social media at a higher rate than older individuals [9] so this subgroup will likely become increasingly important to study. Indeed, our results show that even a relatively young sample is far from accurate at discerning fake from real news.

7 CONCLUSION & FUTURE WORK

In summary, this paper presents findings from a study of people's ability to determine whether news is reliable or fake when given varying combinations of basic information (title, image, source bias, and/or an excerpt from the article). The results show that participants are less accurate when they have only an excerpt and more accurate when provided a title and image, and that for the T+I condition left-leaning participants may be more accurate than right-leaning participants. Our qualitative analysis of responses as to why participants identified news as fake or real revealed that participants' perception of the person being reported on more often mislead them, than helped them in accurately identifying the news as real or fake. While overall people were less accurate than

[10]We considered humans' accuracy in the range 50% to 60% as a condition where humans are confused or undecided, hence we included it as a mistake.

an automated detector in identifying the reliability of news articles, both humans and automated identification was improved when provided meta-data (e.g., title, image, source bias).

Implications of this work could guide designers to focus their attention (and that of their users) to various elements of social media posts. Automated detectors could utilize some of the categories identified in our qualitative analysis that helped humans more correctly identify fake news. To help users better identify fake or real news, interfaces can draw users attention to text professional/unprofessional and image professional/unprofessional. Additionally, platforms may want to prioritize labeling and providing assistance to humans with regards to excerpts since they are the hardest for users to accurately identify as real or fake. This research can inform how computers may be able to train people how to identify the elements of fake news. Beyond just adding a "disputed" label, automated systems could identify or annotate specific credibility indicators [52] (e.g., emotionality of the headline, relevant elements of the image) to help train readers what to look for. Alternatively, platforms could institute a hybrid approach whereby potentially misleading news is signalled to users during their browsing, and the user can choose to expand the labeling to see which elements the algorithm identifies as indicative of real or fake news and its confidence.

Future work will utilize a larger sample size of participants and giving them a larger variety of articles, allowing us to take a closer look at other predictors of accuracy in fake news detection (e.g., media diet and additional demographic information) and measuring additional depending variables (e.g., likelihood to share a news article). Also, as people were able to use the non-professionalism of the image (too emotional, photoshopped, etc.) to make more accurate judgments, we will further investigate using these features to increase the accuracy of automated detection algorithms on larger datasets. Additionally, we will consider the open-ended explanations collected from our participants to train an algorithm which is able to explain to users why they are mistakenly judging a piece of news.

ACKNOWLEDGMENTS

This work has been partially supported by the National Science Foundation under Award #1943370.

REFERENCES

[1] Bence Bago, David G. Rand, and Gordon Pennycook. 2020. Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General* (Jan. 2020). https://doi.org/10.1037/xge0000729

[2] Sudipta Basu. 1997. The conservatism principle and the asymmetric timeliness of earnings. *Journal of Accounting and Economics* 24, 1 (Dec. 1997), 3–37. https://doi.org/10.1016/S0165-4101(97)00014-1

[3] Leticia Bode and Emily K Vraga. 2018. See something, say something: Correction of global health misinformation on social media. *Health communication* 33, 9 (2018), 1131–1140.

[4] Lawrence E. Boehm. 1994. The Validity Effect: A Search for Mediating Variables. *Personality and Social Psychology Bulletin* 20, 3 (1994), 285–293. https://doi.org/10.1177/0146167294203006

[5] William J. Brady, Molly Crockett, and Jay Joseph Van Bavel. 2019. *The MAD Model of Moral Contagion: The role of motivation, attention and design in the spread of moralized content online.* preprint. PsyArXiv. https://doi.org/10.31234/osf.io/pz9g6

[6] Dustin Carnahan and R Kelly Garrett. 2020. Processing style and responsiveness to corrective information. *International Journal of Public Opinion Research* 32, 3 (2020), 530–546.

[7] Luciana Carraro, Luigi Castelli, and Claudia Macchiella. 2011. The Automatic Conservative: Ideology-Based Attentional Asymmetries in the Processing of Valenced Information. *PLOS ONE* 6, 11 (Nov. 2011), e26456. https://doi.org/10.1371/journal.pone.0026456

[8] Luigi Castelli and Luciana Carraro. 2011. Ideology is related to basic cognitive processes involved in attitude formation. *Journal of Experimental Social Psychology* 47, 5 (Sept. 2011), 1013–1016. https://doi.org/10.1016/j.jesp.2011.03.016

[9] Pew Research Center. 2019. Demographics of Social Media Users and Adoption in the United States. https://www.pewresearch.org/internet/fact-sheet/social-media/

[10] Xunru Che, Danaë Metaxa-Kakavouli, and Jeffrey T. Hancock. 2018. Fake News in the News: An Analysis of Partisan Coverage of the Fake News Phenomenon. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '18)*. Association for Computing Machinery, Jersey City, NJ, USA, 289–292. https://doi.org/10.1145/3272973.3274079

[11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[12] Juliet DNSc and Anselm Strauss. 2014. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (fourth edition ed.). SAGE Publications, Inc, Los Angeles.

[13] Daniel A. Effron and Medha Raj. 2019. Misinformation and Morality: Encountering Fake-News Headlines Makes Them Seem Less Unethical to Publish and Share:. *Psychological Science* (Nov. 2019). https://doi.org/10.1177/0956797619887896

[14] Robert M Entman. 2007. Framing bias: Media in the distribution of power. *Journal of communication* 57, 1 (2007), 163–173.

[15] Maksym Gabielkov, Arthi Ramachandran, Augustin Chaintreau, and Arnaud Legout. 2016. Social clicks: What and who gets read on Twitter? *ACM SIGMETRICS Performance Evaluation Review* 44, 1 (2016), 179–192.

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

[17] Matthew Gentzkow, Jesse M Shapiro, and Daniel F Stone. 2015. Media bias in the marketplace: Theory. In *Handbook of media economics*. Vol. 1. Elsevier, 623–645.

[18] Maria Glenski, Corey Pennycuff, and Tim Weninger. 2017. Consumers and Curators: Browsing and Voting Patterns on Reddit. *IEEE Transactions on Computational Social Systems* 4, 4 (Dec. 2017), 196–206. https://doi.org/10.1109/TCSS.2017.2742242 Conference Name: IEEE Transactions on Computational Social Systems.

[19] Andrew Guess, Jonathan Nagler, and Joshua Tucker. 2019. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances* 5, 1 (Jan. 2019), eaau4586. https://doi.org/10.1126/sciadv.aau4586

[20] Andrew M. Guess, Michael Lerner, Benjamin Lyons, Jacob M. Montgomery, Brendan Nyhan, Jason Reifler, and Neelanjan Sircar. 2020. A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences* 117, 27 (July 2020), 15536–15545. https://doi.org/10.1073/pnas.1920498117 ISBN: 9781920498115 Publisher: National Academy of Sciences Section: Social Sciences.

[21] Benjamin D. Horne and Sibel Adali. 2017. This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News. In *The 2nd International Workshop on News and Public Opinion at ICWSM*.

[22] Benjamin D. Horne, Dorit Nevo, John O'Donovan, Jin-Hee Cho, and Sibel Adali. 2019. Rating Reliability and Bias in News Articles: Does AI Assistance Help Everyone?. In *Proceedings of the Thirteenth International Conference on Web and Social Media, ICWSM 2019*. 247–256.

[23] Samara Klar. 2014. A multidimensional study of ideological preferences and priorities among the American public. *Public Opinion Quarterly* 78, S1 (2014), 344–359.

[24] Srijan Kumar and Neil Shah. 2018. False Information on Web and Social Media: A Survey. *CoRR* abs/1804.08559 (2018).

[25] Srijan Kumar, Robert West, and Jure Leskovec. 2016. Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In *Proceedings of the 25th international conference on World Wide Web*. 591–602.

[26] Himabindu Lakkaraju, Julian McAuley, and Jure Leskovec. 2013. What's in a Name? Understanding the Interplay between Titles, Content, and Communities in Social Media. In *Seventh International AAAI Conference on Weblogs and Social Media*. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6085

[27] David Lazer, Matthew Baum, Nir Grinberg, Lisa Friedland, Kenneth Joseph, Will Hobbs, and Carolina Mattsson. 2017. *Combating Fake News: An Agenda for Research and Action*. Technical Report. Harvard & Northeastern University. https://shorensteincenter.org/combating-fake-news-agenda-for-research/

[28] Miriam J Metzger. 2007. Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology* 58, 13 (2007), 2078–2091.

[29] Miriam J Metzger and Andrew J Flanagin. 2015. Psychological approaches to credibility assessment online. *The handbook of the psychology of communication technology* 32 (2015), 445–466.

[30] Amy Mitchell, Jeffrey Gottfriedd, Michael Barthel, and Nami Sumida. 2018. Distinguishing Between Factual and Opinion Statements in the News. *Pew Research Center* (2018).

[31] Maria D Molina, S Shyam Sundar, Thai Le, and Dongwon Lee. 2019. "Fake news" is not simply false information: a concept explication and taxonomy of online content. *American Behavioral Scientist* (2019), 0002764219878224.

[32] Meredith Ringel Morris, Scott Counts, Asta Roseway, Aaron Hoff, and Julia Schwarz. 2012. Tweeting is believing?: understanding microblog credibility perceptions. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. ACM, 441–450.

[33] Raymond S. Nickerson. 1998. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology* 2, 2 (1998), 175–220.
[34] Brendan Nyhan and Jason Reifler. 2010. When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior* 32, 2 (Jun 2010), 303–330. https://doi.org/10.1007/s11109-010-9112-2
[35] Myle Ott, Yejin Choi, Claire Cardie, and Jeffrey T Hancock. 2011. Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1*. 309–319.
[36] José Luis Pech-Pacheco, Gabriel Cristóbal, Jesús Chamorro-Martinez, and Joaquín Fernández-Valdivia. 2000. Diatom autofocusing in brightfield microscopy: a comparative study. In *ICPR*, Vol. 3. 314–317.
[37] Gordon Pennycook, Ziv Epstein, Mohsen Mosleh, Antonio A Arechar, Dean Eckles, and David G Rand. 2019. Understanding and reducing the spread of misinformation online. https://doi.org/10.31234/osf.io/3n9u8
[38] Gordon Pennycook and David G. Rand. 2019. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* 188 (July 2019), 39–50. https://doi.org/10.1016/j.cognition.2018.06.011
[39] Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2018. Automatic Detection of Fake News. In *Proceedings of the 27th International Conference on Computational Linguistics*. 3391–3401.
[40] Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2018. A Stylometric Inquiry into Hyperpartisan and Fake News. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 231–240.
[41] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1144.
[42] Ronald Robertson. 2018. Partisan Bias Scores for Web Domains. https://doi.org/10.7910/DVN/QAN5VX
[43] Karishma Sharma, Feng Qian, He Jiang, Natali Ruchansky, Ming Zhang, and Yan Liu. 2019. Combating Fake News: A Survey on Identification and Mitigation Techniques. *ACM Transactions on Intelligent Systems and Technology* 10, 3 (April 2019), 21:1–21:42. https://doi.org/10.1145/3305260
[44] Kai Shu, Suhang Wang, and Huan Liu. 2019. Beyond News Contents: The Role of Social Context for Fake News Detection. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM 2019, Melbourne, VIC, Australia, February 11-15, 2019*. 312–320.
[45] Kai Shu, Xinyi Zhou, Suhang Wang, Reza Zafarani, and Huan Liu. 2019. The Role of User Profiles for Fake News Detection. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '19)*. 436–439.
[46] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
[47] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.
[48] Emily K Vraga and Leticia Bode. 2017. Using expert sources to correct health misinformation in social media. *Science Communication* 39, 5 (2017), 621–645.
[49] Sam Wineburg and Sarah McGrew. 2017. Lateral reading: Reading less and learning more when evaluating digital information. (2017).
[50] Bartosz W Wojdynski, Matthew T Binford, and Brittany N Jefferson. 2019. Looks Real, or Really Fake? Warnings, Visual Attention and Detection of False News Articles. *Open Information Science* 3, 1 (2019), 166–180.
[51] Waheeb Yaqub, Otari Kakhidze, Morgan L. Brockman, Nasir Memon, and Sameer Patil. 2020. Effects of Credibility Indicators on Social Media News Sharing Intent. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376213
[52] Amy X. Zhang, Aditya Ranganathan, Sarah Emlen Metz, Scott Appling, Connie Moon Sehat, Norman Gilmore, Nick B. Adams, Emmanuel Vincent, Jennifer Lee, Martin Robbins, Ed Bice, Sandro Hawke, David Karger, and An Xiao Mina. 2018. A Structured Response to Misinformation: Defining and Annotating Credibility Indicators in News Articles. In *Companion Proceedings of the The Web Conference 2018 (WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 603–612. https://doi.org/10.1145/3184558.3188731
[53] Xinyi Zhou, Jindi Wu, and Reza Zafarani. 2020. SAFE: Similarity-Aware Multi-modal Fake News Detection. In *Advances in Knowledge Discovery and Data Mining*. 354–367.
[54] Xinyi Zhou and Reza Zafarani. 2018. Fake News: A Survey of Research, Detection Methods, and Opportunities. *CoRR* abs/1812.00315 (2018).

A QUALITATIVE CODES

This appendix consists primarily of Table 3 below that identifies the qualitative codes inductively generated by three reviewers. It also includes the explanation/description of the codes and some examples.

Table 3. Thematic codes (and examples) that were inductively developed by analyzing the question "why did you identify the news item as real or fake". The table identifies the code, provides a description, and example quotes [with the associated accuracy]. The accuracy is indicated as: FN = false negative (identified fake news as real); FP = false positive (identified real news as fake); TN = true negative (identified real news as real); TP = true positive (identified fake news as fake).

Code	Explanation of Code	Example Quotes [and associated accuracy of that case]
Plausible	Believable; possible; reasonable; could happen	• "This is said about many politicians" [FN] • "I haven't heard about this but it seems real enough that it could be believable" [TN] • "It seems like something our former president could have said" [FN] • "I think that article is real because this seems like something that could definitely happen" [TN] • "Nothing pops out at me in a way that would make me think it's fake. All the information seems like it could be legitimate" [FN]
Implausible	Unlikely; unbelievable; unrealistic	• "Seems [...] like something to be seen on Instagram" [FP] • "This seems unrealistic and like one of those click bate stories" [FP] • "I'm not sure Trump is very articulate and able to pull off a speech like this" [FP] • "Seemed like someone made this up and didn't really seem like an actual article" [FP]
Familiar	Heard about it; sense of recognition; seen similar stories	• "I saw something like this on NBC news at some point" [TN] • "The content of the article seems vaguely familiar" [FN] • "I feel like I heard this somewhere but not sure" [FN] • "I've heard similar stories" [TN]
Unfamiliar	Not recognized; haven't heard of it	• "I haven't heard about it before" [TP] • "I am not familiar with the material" [TP] • "I don't remember this much discussion back and forth during the last election on the subject" [FP]
Title Professional	Proper grammar and punctuation; has quotes or statistics; lacks obvious bias	• "Because it uses quotes and statistics in their title" [FN] • "I think it is real because this title has a lot of contexts no bias and it's asking a question" [TN] • "I said this was real because it doesn't seem to have a political title" [TN] • "Used a quote" [FN]
Title Unprofessional	Poor grammar; all caps or exclamations; clickbait; title emotional (see below)	• "Unprofessional language" [FP] • "The title contains one bolded word which you would not normally see" [TP] • "The title seemed like click bate" [FP] • "The word lead being in all caps is sketchy to me as well as the title being so uninformative" [TN]
Title Emotional	Fear-mongering; aggressive; alarmism; evocative	• "The title is a sneak diss" [TP] • "This sounds too dramatic to be a real news story" [FP] • "I think this is fake because the title is trying to evoke people that she is a liar and not being truthful" [TP]
Picture Professional	Neutral image; direct image of the story/event; realistic image (e.g., video still)	• "Picture looks to be taken from a security camera" [TN] • "The picture is video evidence of it actually happening" [TN] • "The photo used matches the topic well and seems to be of high quality" [TN]
Picture Unprofessional	Photoshop; meme; poor image quality or editing; picture emotional (see below)	• "The picture looks like a meme" [FP] • "Pictures look fake and photoshopped" [TP] • "The text has nothing out of the ordinary. But, the picture makes Obama look like he's a dictator or something" [TP] • "The photo uses photoshop which typically is not used within a news article" [TP] • "The image is throwing me off and looks badly edited" [TP]
Picture Emotional	Negative; aggressive; fear-mongering; biased selection of image	• "It is using an aggressive picture to try and get you to feel a certain way about him" [FP] • "The photo associated looks to be manipulative and biased suggesting the article's intentions are emotionally skewed and attempting to insight fear and confusion in the political environment" [TP]

Continued on next page

Continued from previous page

Code	Explanation of Code	Example Quotes [and associated accuracy of that case]
Text Professional	Neutral/calm tone; uses statistics; includes quotes, citations, and/or sources; good grammar or punctuation; detailed; focuses on events rather than opinion and editorial	• "No punctuation errors" [TN] • "This article has facts and quotes that lead me to believe this is real news. It is also professional written and speaks on more facts than opinions" [FN] • "The above news is real because the diction is neutral. In addition, in the end it allows the readers to see the documents as proof that this had really happened" [FN] • "They use facts and statistics that seems to be legit" [FN] • "With the direct quotes and details that this excerpt used, I have the feeling that it is real" [FN] • "It has a calmer tone which makes me think that it is real" [FN] • "Provides clear background information and location and sources" [TN]
Text Unprofessional	Emotional, fear-mongering, or attacking tone; heavy on opinion and editorial; poor grammar and punctuation	• "This article is clearly biased and persuasive, with only using pathos as its rhetoric" [TP] • "The wording seems entirely exaggerated, like an attack" [TP] • "The content is good, but the writing and editing is super unprofessional. It's very opinionated, so it's fake news" [TP] • "This is fake because some of the English in this article is not as professional as one found on a reliable source" [FP] • "It seems like this article is used for propaganda because in the end they advertise for Ted Cruz" [TP]
Source Untrusted or Unfamiliar	Skepticism about where the article itself is found (e.g., Facebook) or sources cited within it (e.g., "as reported by..."); skepticism about the type of media that did the reporting	• "It states that Hillary's website is HillaryClinton.com. Which doesn't sound credible at all" [FP] • "The source comes straight from Facebook so new stories can easily be changed and published by anyone" [TP] • "This is believable, but I do not know the source" [FP] • "Seems to be a radicalized right wing website" [TP] • "If it is on live television don't trust the news to quote someone the right way" [FP]
Perception of Person	Bases reasoning on strong opinions the person who is the subject of the news item; mentions trust or lack of trust in an individual person	• "Unlikely Donald Trump, we can trust what Obama told us is actually real news" [FN] • "I believe this to be true because Hillary has been caught in a lot of issues so this doesn't surprise me that this is something she would do" [FN] • "It just does not sound like the President. He cites unity and I feel that he is not for unity to the public" [FP] • "Don't believe it was the Illuminati or anything. I do believe Obama worked for the world not the USA. Reasoning is he gave USA enemies nuclear rights and a pallet full of cash. Why? Would you like ever provide your enemy" [FN]
Pre-existing Beliefs	Bases reasoning on information/beliefs not in the article itself (and not specifically about a person in the news item)	• "I feel like this isn't too far from the beliefs of the democrats" [TN] • "This article doesn't seem to be real or logical in my eyes because of how important the internet is in the lives of any U.S. citizen" [TP] • "Military news is usually pretty real" [FN]
Missing Information	Not enough evidence to consider it true; not enough evidence to consider it false; lack of sources, quotes, statistics or details	• "I don't have enough information to give a real or fake" [FP] • "Not enough information" [FP] • "It's difficult to believe something when there is zero explanation behind it" [TP] • "Doesn't have any solid evidence that this is true" [FP] • "Probably is true but not backed by solid evidence" [TP]
Guess / Unsure	Explicitly mentions not being sure or making a guess	• "This could be fake or real. There is no way for me to gauge its fakeness" [TN]

Received June 2020; revised October 2020; accepted December 2020