User Perceptions of Article Credibility Warnings: Towards Understanding the Influence of Journalists and AI Agents

Matthew Sumpter, Tempestt Neal
University of South Florida
mjsumpter@usf.edu, tjneal@usf.edu

Abstract

Due to online sharing, false and misleading information spreads at an alarming rate. In response, social media sites have begun to append credibility warnings to articles that human fact-checkers have deemed unreliable. However, information spreads at rates exceeding that at which humans are capable of fact-checking. This has lead to a boom in computational fact-checking with artificial intelligence (AI). While the computational assessment of credibility is becoming more accurate, users' perceptions of computational fact-checking as a trustworthy supplement for humans and the impact of computational fact-checking on users' abilities to correctly decide on the credibility of an article are not well studied. We conducted a cross-sectional survey in which 204 survey respondents rated the credibility of four news articles, each randomly assigned a credibility warning (i.e., an assessment of the article's credibility determined by either an AI agent or human journalist, or no assessment at all). We found that AI warnings were as successful, if not more so, than warnings provided by a journalist, at influencing participants' assessments of a news article's credibility (regardless of the warning's accuracy). Additionally, our results show that an article's magnitude of sentiment, along with the user's understanding of AI, both play a vital role in determining the effectiveness of AI warnings.

1 Introduction

Information on the web is in great abundance. However, due to freedom of speech, anonymity, and lack of standardization, there is no clear indication of whether or not something read on the internet is credible. Although this has been a widely recognized issue, the modern political landscape has brought it to the forefront of our collective conversation, especially in the United States (Berghel 2017; Lazer et al. 2018; Allen et al. 2020). There have been various studies which aim to identify the means by which internet users interpret, clarify, and remember information, as well as how misinformation can be mitigated against (Pingree, Hill, and Mcleod 2013; Pingree, Brossard, and Mcleod 2014; Winneg et al. 2014; Lewandowsky, Ecker, and Cook 2017). Studies have found that preexposure warnings - explicit warnings that precede the article content - are effective at mitigating the impact of misinformation on an individual (Lewandowsky et al. 2012). However, the responsible party for generating these warnings is not well established.

Fact-checking in the form of easily interpretable and widely available warnings requires comprehensive analysis. For human journalists, fact-checking can be time-consuming (and incredibly slow compared to the time by which misinformation spreads (Vosoughi, Roy, and Aral 2018)), leading to inconsistencies, bias, and manipulation (Birdwatch 2021) in human reviews (Nieminen and Rapeli 2019). These issues have spurred the growth of computational fact-checking (Wu et al. 2014; Ciampaglia et al. 2015; Ciampaglia 2018). Numerous approaches and early prototypes for computationally fact-checking online information, including stance detection (Hanselowski et al. 2018; Bhatt et al. 2018), framing-bias detection (Morstatter et al. 2018), online network analysis (Pérez-Rosas et al. 2017), and the application of adversarial neural networks (Wang et al. 2018), have emerged. For instance, Popat et al. (2016) demonstrated the use of artificial intelligence (AI) at classifying the credibility of text samples, achieving over 70% accuracy.

Considering the significant effort by psychologists and computer scientists to understand and prevent the spread of misinformation (Shu et al. 2017; Lazer et al. 2018; Ruths 2019), connecting the psychological understanding of misinformation with the computational methods meant to control its spread remains an open problem. Specifically, an improved understanding of how internet users might respond to computational fact-checking, particularly as a substitute (or supplement) for human fact-checkers, is necessary to truly gauge **1**) if users will trust that an AI is providing them with an accurate assessment of the credibility of online information, and **2**) if so, will users generally agree with an assessment of credibility offered by an AI agent? In addition, because AI is largely a buzzword to much of the general public (Livingston 2018), it is also important to understand user's perception of AI in general, and how their perceptions are influenced by their knowledge of AI. Together, these components will help measure the potential impact of computational fact-checking, especially in comparison to human journalists, before such AI applications reach the broader population. As such, we conducted a cross-sectional survey ($N = 204$) to probe how different internet users assess the credibility of various articles, each labelled with a credibility warning seemingly made by either an AI agent or human

Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

journalist, to answer the following research questions:

RQ1. *Are users more persuaded by credibility assessments offered by an AI agent in comparison to a human journalists?*

RQ2. *What factors, article- or user-dependent, correlate with the influence of AI credibility warnings?*

Our results show that users generally trust an AI agent's assessment of credibility. We also found that this trust may be dependent on the user's general perception of AI as an effective problem-solving tool. In addition, we found evidence suggesting that the article's level of sentiment impacts the influence of credibility warnings on users' likelihood to agree with the warning. To our knowledge, this is the first study to compare the responses of internet users to human and AI credibility warnings.

This paper is outlined as follows: Section 2 summarizes related literature on computational fact-checking and psychological contributors to misinformation assimilation. Section 3 details the experimental setup and overviews the survey design. Section 4 provides our research findings and analyses. Section 5 summarizes the paper, notes study limitations, and details future work. Finally, Section 6 lists the articles presented to the study participants.

2 Background

Credibility is defined as being believable, trustworthy, and having some quality of expertise (Metzger and Flanagin 2013). The question of how a user determines if something they read online is credible is attributed to a great number of variables; source and receiver characteristics, message content and style, as well as the presentation medium have all been identified as potential variables that can influence a user (Wathen and Burkell 2002). This complex determination can therefore be naturally flawed on the receiving end, or worse, intentionally manipulated by the purveyor of information (Volkova and Jang 2018). In the modern age, the barriers to entry that existed in traditional media have effectively vanished leaving a proverbial "wild-west" of information propagation. Consequently, researchers have been searching for effective methods to combat the spread of misinformation (Figueira and Oliveira 2017; Vosoughi, Roy, and Aral 2018).

Journalistic fact-checking is the most obvious and straightforward method for addressing the credibility of a claim; a non-biased journalist or team of journalists researches a claim, addressing specific standards for credibility, and comes to an ultimate conclusion (Nieminen and Rapeli 2019). While this is a noble service, its issues are two-fold. First, it is time consuming to fact-check a claim, as there are thousands being circulated every second. Some research focuses on computational methods to make this work easier for journalists. For instance, Vlachos and Riedel (2014), Adler and Boscaini-Gilroy (2019), and Shaar et al. (2020) have all proposed natural language processing (NLP) solutions for comparing new claims to existing fact-checks to prevent repetitive credibility reviews. Additionally, Shiralkar et al. (2017) designed algorithms to exploit knowledge graphs to help increase the productivity of fact-checkers. These methods may prove useful, however they still result in a fact-check after the original presentation of information. This raises the second issue: studies have shown that retractions and corrections to misinformation are relatively ineffective (Ecker, Lewandowsky, and Apai 2011; Lewandowsky et al. 2012). For example, Johnson and Seifert (1994) found that retractions had no effect in reducing the reliance on misinformation. These outcomes have been attributed to the way a user creates a mental model of a claim (Johnson-Laird, Gawronski, and Strack 2012). Because a user creates a coherent story of an occurrence, the inaccurate claim gets incorporated as a foundational and logical part of the story. Therefore, although the claim may be debunked, the misinformation remains as an artifact in the user's mind.

Thus far, there are three solutions that have been proposed to increase the effectiveness of retractions: **1**) corrections that tell an alternative, coherent story, **2**) consistent repetition of the retraction, and **3**) warnings at the time of initial exposure (Lewandowsky et al. 2012). Unfortunately, the first two options attempt to correct for misinformation that has already been consumed by a user – resulting in the user potentially contributing to the spread of misinformation themselves by propagating it further. This leaves the general approach of the third option: cut off the misinformation at the source. However, this remains difficult, primarily since it is difficult to check a claim before it reaches the user (Budak, Agrawal, and Abbadi 2011).

One possible solution, that of using NLP and AI, has been proposed to recognize the style and patterns of misinformation such that misinformation can be flagged as it is being presented to the user (Hanselowski et al. 2018; Bhatt et al. 2018; Morstatter et al. 2018). This would allow for the spontaneous nature of web information to be tagged, increasing skepticism, which has been noted as a keystone in encouraging readers to tread carefully (Lewandowsky et al. 2012). Long-term trends in public perception of AI have improved, with the majority of user's optimistic about its place in society (Fast and Horvitz 2017). However, there is increasing evidence of human error in the form of bias (Ciampaglia 2018; Yapo and Weiss 2018), which may lead to skepticism of an AI model to accurately assess the credibility of a news source. Consequently, as this technology continues to be developed, it is important to determine how, why, and to what extent a user is willing to trust AI to analyze their news.

To our knowledge, only one study has explored the interplay of automated fact-checking and end users' interpretation of the credibility of news articles. In a qualitative survey, Horne et al. (2019) asked users ($N = 654$) to rank various news sources according to their reliability and bias. The goal of their study was to determine how algorithmic assistance could improve users' perception of these two factors. Their findings were that AI assistance improved human perceptions, but feature-based explanations of the decision-making process are needed to improve "participants' ratings on unreliable and biased articles" (Horne et al. 2019). They also found that those who were more adept news readers tended to be more successful in making determinations of reliability and bias, as opposed to those who relied on social media sharing for their news. Our study builds upon these find-

ings by comparing the effectiveness of AI at improving user perceptions of credibility to what they are meant to replace - journalists. Further, Pennycook et al. (2020) and Clayton et al. (2019) also explored the effectiveness of credibility warnings when applied to news headlines on Facebook and coined the "implied truth" effect. This phenomenon occurs when the absence of a credibility warning implies to the user that it has been fact-checked and is valid, although the truth is that it simply hasn't been checked at all. This reflects a current trend in these AI solutions to only flag information that has been deemed to be untrustworthy that needs to be accounted for. Our study begins to address this behavior by simulating an environment where *all* information is tagged with a credibility warning, indicating whether it was deemed to be credible or not explicitly. Untagged articles were reserved for the control group.

3 Method

3.1 Demographics

We conducted a cross-sectional survey of adult internet users ($N = 204$) recruited through email, Reddit ads, social media posts, and university newsletters to complete an anonymous survey on Qualtrics. Participants ranged from 18 to 74 years old, with an average age of 30 years old (SD = 13.5 years). Most respondents were male (50.8%); 44.2% self-identified as female, while all others declined to answer. The majority of respondents identified as White (79%), with 9.7% identifying as Asian, 6.6% as Black, and the remaining 3.4% as Native American, Hawaiian or Pacific Islander; 7.7% of participants identified as Hispanic. Participants indicated a range of educational backgrounds: about one-third (33.5%) have a 4-year degree, 27.4% reported some college education, 17.3% have a professional degree, 11.2% completed high school only, 7.6% have a Doctorate degree, with the remaining 3% declining to answer. Our Institutional Ethics Board approved the research plan and participants gave informed consent to complete the survey.

3.2 Overview of Survey

General Assessment We first asked participants a series of questions to gain a general assessment of their understanding and trust in AI to solve problems and to use various AI agents in their daily lives, such as in their homes or for automated driving. Similarly, we gauged users' trust in journalists to deliver accurate reporting, in addition to asking users to report how they stay up to date on current events and what factors they deem important when determining the credibility of a news source. The specifics of these questions can be found in Table 1.

Overview of Articles We compiled a total of six articles (175 average words/article, SD=31 words) all having a topic within the scope of politics for the purpose of consistency; half presented true information while half were false (see Section 6 for article content). Each participant was randomly assigned four of the six articles. Each article was labeled with one of three warnings: **1**) an AI model's assessment of credibility (*AI warning*), **2**) a journalist's assessment of credibility (*Journalist warning*), or **3**) no assessment (*Control warning*) (see Figures 1 and 2). These warnings used color-coded tags analogous to traffic light colors to serve as a strong visual marker. With three possible warnings applied to each article, there were 18 (6 articles x 3 warnings) total possibilities of warning/article combinations ($\approx$ 35 participants assigned to each article).

Survey Question [*Possible Responses*]
How knowledgeable are you about artificial intelligence? [*5-point Likert scale of knowledge*]
How comfortable would you be using a self-driving car? [*5-point Likert scale of comfort*]
How comfortable are you using smart home devices (Amazon Alexa, Google Home, etc.) [*5-point Likert scale of comfort*]
How effective do you think artificial intelligence is at solving problems? [*5-point Likert scale of effectiveness*]
How do you stay up to date on current events? (Select all that apply) [*Options: Print, Social Media, Online News Sites, Online Forums, Television, Youtube, Radio, Word of Mouth, Other*]
How engaged are you in politics? [*5-point Likert scale of engagement*]
Please rank from most important to least important the factors you look for when determining if something you read online is credible. [*Options: Domain name, News organization, Quality of written content, Believable argument, Website presentation, Authorship, Number of external links or citations*]
Journalists can be trusted to deliver accurate reporting. [*5-point Likert scale of agreement*]

Table 1: Survey questions on participants' understanding and trust in AI and journalism, news-seeking habits, and fact-checking.

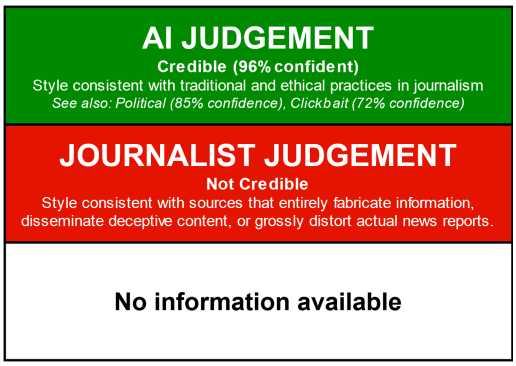

Figure 1: Article Warnings

For each article, participants were provided information as if an AI agent or journalist had already assessed the credibility of the article, and rated it as *credible*, indicated by a green label (with a confidence level for the AI), or *not credible*, indicated by a red label. Further, since Horne et al. (2019) found that explanations of the decision on credibility were important in helping users make accurate decisions, ex-

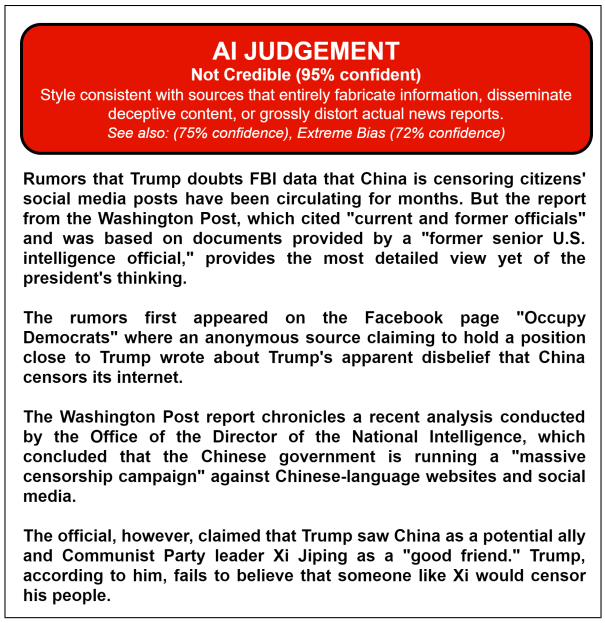

Figure 2: Example Participant View

planations were also included on both the *journalist* and *AI* warnings. Further, warnings were intentionally misleading in two of the six cases (Table 2). Participants were asked to read the article, rate (5-point scale for familiarity) whether they had any prior knowledge of the article's content, and whether they found the article to be credible (7-point scale for credibility). After judging four articles, participants were provided the correct information about the articles they read to lessen the likelihood that misinformation would be spread as a result of this study.

Article #	True?	Pre-exposure Warning
1	Yes	Correct
2	Yes	Incorrect
3	Yes	Correct
4	No	Correct
5	No	Correct
6	No	Incorrect

Table 2: Breakdown of the qualities of the articles presented to participants

Finally, we note that AI and journalist warnings were fake - there was no such determination made by a third-party. Fake articles were intentionally written by human volunteers to mimic credible news publications; by using fabricated articles rather than sourcing unreliable ones, we sought to eliminate any room for error on the credibility of the 'false' articles. Credible articles were extracted from Reuters, The Bipartisan Press, and the Bellingham Herald. Included articles were purposefully selected to cover recent current events, and to reflect a range of viewpoints of both the left- and right-wing political ideologies.

4 Results and Analysis

4.1 Participant Views Toward AI and Journalism

Most participants self-reported some knowledge of AI systems. Specifically, 40.1% of participants reported as slightly knowledgeable, 38.5% reported as moderately knowledgeable, 10.4% reported as very knowledgeable, and 3.13% reported as extremely knowledgeable. Only 7.8% reported no prior knowledge. In addition, nearly half of all participants (46.02%) considered AI to be moderately effective at problem solving, while overall, participants had a positive view of AI as a problem solving tool. Only 2.27% of the sample felt that AI is ineffective for problem solving. We also found that the more knowledge a user has about AI, the more likely they are to believe it to be an effective problem solving tool (Table 3). However, participants' views of AI-powered technologies such as self-driving cars and smart home devices were not as favorable. Specifically, higher percentages of subjects reported some level of discomfort with self-driving cars and smart home devices (45.84% and 46.35%, respectively) in comparison to those comfortable with these technologies (41.67% and 40.1%, respectively).

As shown in Figure 3, over half of participants (59.74%) use online resources to receive their news. About one-fifth of participants receive their news from online news sites, while 17.04% rely on social media. Only 8% of participants reported staying up to date through printed media (newspapers, magazines, etc.).

Over one-third of participants somewhat or strongly agreed that journalists can be trusted to deliver accurate reporting (3.26% and 38.59%, respectively), while one-third somewhat or strongly disagreed (23.91% and 9.78%, respectively). The remaining 24.46% were indifferent.

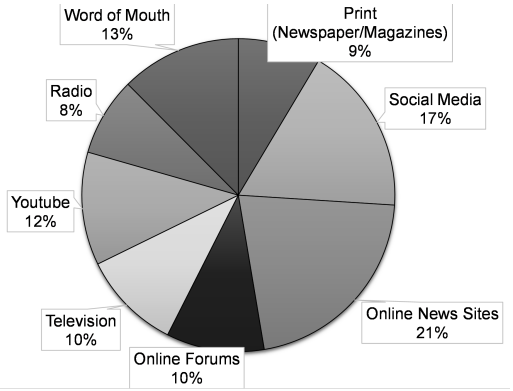

Figure 3: Survey participants mostly receive news from online news sites.

4.2 Credibility and Political Involvement

Recent findings have suggested that those on the extremes of the political spectrum are more likely to share fake news (Guess, Nagler, and Tucker 2019; Hopp, Ferrucci, and Vargo 2020), and for that reason we gathered self-reported political leaning in case any trends emerged. A majority of participants (70%) identified as liberal in their political views, 12% identified as centrist, and 18% identified as conservative.

	How effective is artificial intelligence at solving problems?				
AI Knowledge	**Extremely effective**	**Very Effective**	**Moderately effective**	**Slightly effective**	**Not effective**
Extremely Knowledgeable	16.67%	**66.67%**	16.67%	0.00%	0.00%
Very Knowledgeable	15.00%	**40.00%**	35.00%	10.00%	0.00%
Moderately Knowledgeable	4.05%	36.49%	**45.95%**	10.81%	2.70%
Slightly Knowledgeable	2.63%	27.63%	**51.32%**	15.79%	2.63%

Table 3: Participants knowledgeable of AI were more likely to view AI as an effective problem solving tool.

Liberal-leaning individuals typically believe that the government should be active in supporting social and political change (Conover and Feldman 1981), while conservative-leaning individuals seek to preserve a range of institutions such as religion, parliamentary government, and property rights, with the aim of emphasizing social stability and continuity (Conover and Feldman 1981). Centrists hold a moderate view. Figure 4 shows an almost even split between trust and distrust of journalists' reporting; however, liberal-leaning participants were much more likely to trust journalists (65%) than conservative-leaning participants (38%).

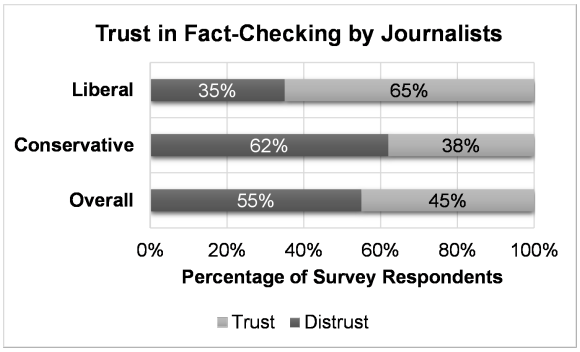

Figure 4: Participants self-declared trust in journalists to provide accurate reporting. Liberal-leaning participants were almost twice as likely to trust journalists when compared to conservative-leaning participants.

4.3 Effects of AI and Journalist Warnings

To assess if AI or journalist warnings were effective at persuading the user to correctly determine the credibility of an article, we conducted six one-way ANOVA tests corresponding with the six different articles presented (See Table 4 for results). For each test, we used the user ratings of credibility as the dependent variable, and the article treatments (i.e., the AI warning, journalist warning, or no warning (control group)) as the independent variables to determine if the differences in the users' responses in comparison to the control group were statistically significant.

ANOVA tests for articles #1, #4, and #6 revealed no statistically significant difference between the users' ratings of credibility for those articles, despite the articles' treatments. However, ANOVA tests on articles #2, #3, and #5 did show a statistically significant difference in the users' ratings of credibility between the three treatments. Specifically, the ANOVA test for articles #2 and #3 showed a statistically significant difference between the control group's ratings of credibility and both the journalist and AI warning-treated articles, although a post hoc test showed no significant difference between the AI and journalist warning-treated articles. The ANOVA test for article #5 showed a statistically significant difference in user ratings of credibility between the AI warning-treated articles and both the control and journalist warning-treated groups, with no significant difference between the latter two.

Art. #	F**-score**	p**-value**	**Explanation**
1	0.5447	0.5818	-
2	6.3330	**0.0026**	AI and journalist warnings resulted in statistically significant differences in user ratings of credibility compared to the control group.
3	4.107	**0.0194**	AI and journalist warnings resulted in statistically significant differences in user ratings of credibility compared to the control group.
4	0.4018	0.6702	-
5	3.8970	**0.0236**	AI warnings resulted in statistically significant differences in user ratings of credibility compared to the journalist warning and control groups.
6	0.4642	0.6300	-

Table 4: One-way ANOVA ($\alpha < 0.05$) results for participant assessments of credibility per article across treatments (i.e., AI warnings, journalist warnings, and no warnings).

To identify the factors that may have resulted in the statistically significant impact on users' ratings of credibility (articles #2, #3, and #5), we analyzed the positive and negative sentiment of each article using the VADER sentiment analyzer (Hutto and Gilbert 2014); sentiment being defined as an opinion regarding a situation or event. As shown in Figure 5, when articles contained language consistent with strong sentiment (i.e., $|\texttt{VADER\_positive\_sentiment\_score} - \texttt{VADER\_negative\_sentiment\_score}| > 0.10$), participants were less likely to be persuaded by the pre-exposure warnings (regardless of whether the sentiment was positive or negative). This insight is strengthened by the observation of the inverse; article #2, which was rated as the most neutral, saw the largest change in participant perception of credibility from the control group. In short, a reader of an emotional or opinionated article is less likely to be influenced by a credibility warning. Additionally, we also found that the more likely a user is to think of AI as an effective problem solving tool, the more likely they were to be in agreement with the AI warning (Table 5).

In summary, Figure 6 shows the impact of AI and journalist warnings across the six articles. AI warnings proved as

	How effective is artificial intelligence at solving problems?				
	Extremely effective	**Very Effective**	**Moderate effective**	**Slightly effective**	**Not effective**
Agreed with AI Assessment	**75.00%**	**60.00%**	**60.98%**	45.45%	0.00%
Disagreed with AI Assessment	25.00%	40.00%	39.02%	**54.55%**	**100%**

Table 5: Participants more likely to view AI as a problem solving tool were also more likely to align their opinion of an article's credibility with that of the AI warning.

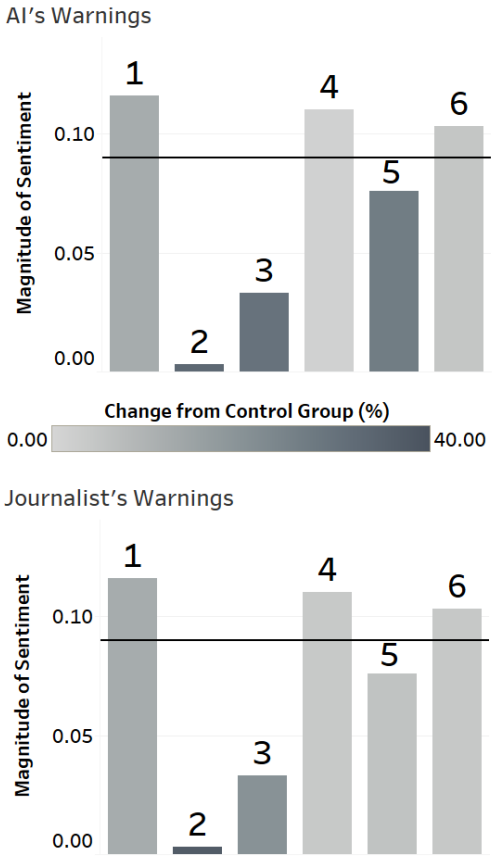

Figure 5: Warnings were less effective at swaying participants when there was strong sentiment (> 0.09). Below this threshold, lower sentiment led to a increasingly significant impact of warnings.

effective as journalist warnings in influencing users to correctly determine the credibility of an article in comparison to the control group which received no warnings. For 4 of 6 articles (articles 2 through 5), AI warnings were more effective than journalist warnings in assisting users in correctly assessing the credibility of the article. These findings indicate that AI-generated ratings of credibility may effectively replace a journalist's ratings. Our results showed that not only did the AI warning influence our participants to shift their perception of an article's credibility, but it did so to a similar degree of the journalist's warning. In all articles excluding article #5, the journalist's warning and AI's warning showed no significant difference in their influence on users. Additionally, in regard to article #5, there is some indication that AI can be more effective at influencing a user's opinion than a journalist, although the contributors to this require further investigation.

5 Conclusion

This study provides insight into the efficacy of using credibility warnings to influence internet users' perception of news credibility. We conducted a cross-sectional survey asking internet users to rate the credibility of news articles when given preexposure warnings of credibility attributed to journalists and an AI algorithm. In doing so, we found evidence that AI may be effective at influencing a user's perception of an article's credibility. More importantly, we found that AI may be at least as effective as a journalist in doing so, and in some cases, more effective. This influence worked to increase the accuracy of a user's rating of credibility. We also found that language sentiment may influence the degree to which a user perceives and believes preexposure warnings. Heavily sentimental language may indicate bias or strong emotional appeal. This kind of language has the tendency to elicit emotion in the reader and solidify their own beliefs in the information presented (Zhang et al. 2011), potentially suggesting that sentimental language is counterproductive in the fight against misinformation. Overall, consistent with Horne et al. (2019), we found evidence that "AI assistance improves human perceptions about reliability...in news articles" and "political leaning has little impact on rating reliability." Finally, our results suggests that AI could potentially deceive a user into thinking an article is credible (articles #2 and #6). If a preexposure warning is incorrect, it could increase the spread of misinformation, leading to severe consequences (Yapo and Weiss 2018). As such, it is important to continue improving the accuracy of such algorithms.

When considering AI as a support for human decision-making, having a perfectly accurate algorithm is only a partial success. It is also necessary to convince your target audience that it actually will work for them. The findings noted in Table 5 show a positive relationship between AI understanding and success of the AI credibility warnings. In other words, the more a user understands about AI, the more effective they believe it to be. This increased belief leads to more trust in AI systems, and allows for an artificially-generated fact-checking to be more successful in moving public opinion.

Finally, we note that this study has the following limitations. First, our study did not have a diverse participant

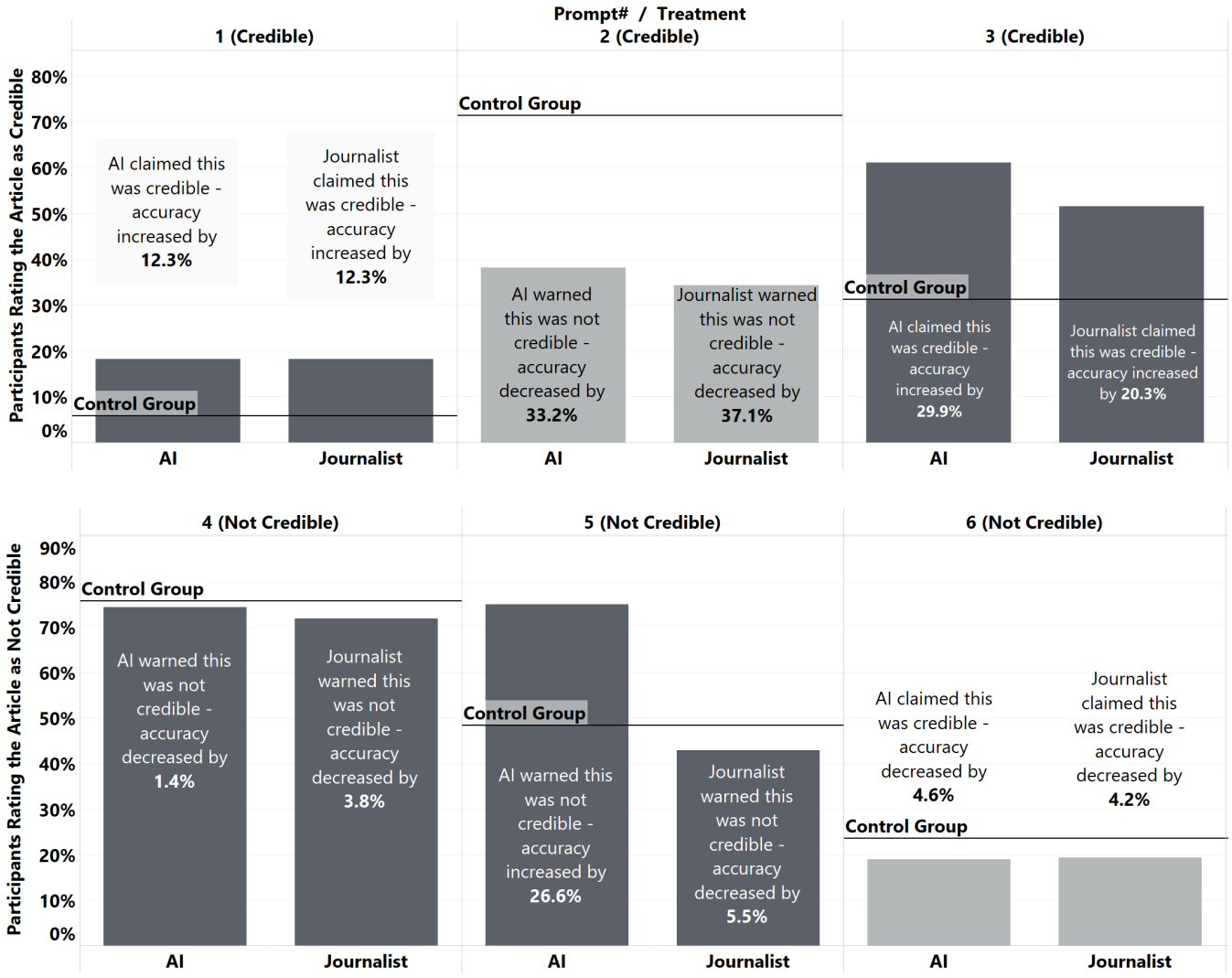

Figure 6: AI warnings prove as effective as journalist warnings in influencing users to correctly determine the credibility of an article in comparison to the control group which received no warnings. For 4 of 6 articles (articles 2 through 5), AI warnings were more effective than journalist warnings in assisting users in correctly assessing the credibility of the article.

pool, and thus may not generalize to the broader population. We also did not define credibility for participants, such that users may have had various understandings and definitions of the term. Additionally, our articles were topically focused on current events and politics; although these are ripe topics for misinformation, different findings may emerge across different topics. In regard to the survey design, we did not evaluate the change of users' trust in AI over time, which could evolve as the relevance of current events in their personal lives change. We also did not monitor users engagement level with each article; much of news consumption takes the form of scrolling headlines and news blurbs, so the effectiveness of AI as a function of different levels of engagement and article lengths is left for future work. In addition to this, future work will also need to determine how different presentation styles of credibility warnings influence users perception of credibility.

References

Adler, B.; and Boscaini-Gilroy, G. 2019. Real-Time Claim Detection from News Articles and Retrieval of Semantically-Similar Factchecks. *CEUR Workshop Proceedings* 2411. ISSN 16130073.

Allen, J.; Howland, B.; Mobius, M.; Rothschild, D.; and Watts, D. J. 2020. Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances* 6(14): eaay3539. ISSN 23752548. doi:10.1126/sciadv.aay3539. URL http://advances.sciencemag.org/.

Berghel, H. 2017. Lies, Damn lies, and fake news. *Computer* 50(2): 80–85. ISSN 00189162. doi:10.1109/MC.2017.56.

Bhatt, G.; Sharma, A.; Sharma, S.; Nagpal, A.; Raman, B.; and Mittal, A. 2018. Combining Neural, Statistical and External Features for Fake News Stance Identification. In *Companion Proceedings of the The Web Conference 2018*, WWW '18, 1353–1357. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee. ISBN 9781450356404. doi:10.1145/3184558.3191577. URL https://doi.org/10.1145/3184558.3191577.

Birdwatch. 2021. Challenges. URL https://twitter.github.io/birdwatch/about/challenges/.

Budak, C.; Agrawal, D.; and Abbadi, A. E. 2011. Limiting the spread of misinformation in social networks. In *Proceedings of the 20th International Conference on World Wide Web, WWW 2011*, 665–674. New York, New York, USA: ACM Press. ISBN 9781450306324. doi:10.1145/1963405.1963499. URL http://portal.acm.org/citation.cfm?doid=1963405.1963499.

Ciampaglia, G. L. 2018. Fighting fake news: a role for computational social science in the fight against digital misinformation. *Journal of Computational Social Science* 1(1): 147–153. ISSN 2432-2717. doi:10.1007/s42001-017-0005-6. URL https://doi.org/10.1007/s42001-017-0005-6.

Ciampaglia, G. L.; Shiralkar, P.; Rocha, L. M.; Bollen, J.; Menczer, F.; and Flammini, A. 2015. Computational fact checking from knowledge networks. *PLoS ONE* 10(6): 1–13. ISSN 19326203. doi:10.1371/journal.pone.0128193.

Clayton, K.; Blair, S.; Busam, J. A.; Forstner, S.; Glance, J.; Green, G.; Kawata, A.; Kovvuri, A.; Martin, J.; Morgan, E.; et al. 2019. Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior* 1–23.

Conover, P. J.; and Feldman, S. 1981. The Origins and Meaning of Liberal/Conservative Self-Identifications. *American Journal of Political Science* 25(4): 617–645. ISSN 00925853, 15405907. URL http://www.jstor.org/stable/2110756.

Ecker, U. K. H.; Lewandowsky, S.; and Apai, J. 2011. Terrorists brought down the plane!–No, actually it was a technical fault: processing corrections of emotive information. *Quarterly journal of experimental psychology (2006)* 64(2): 283–310. ISSN 1747-0226. doi:10.1080/17470218.2010.497927. URL http://www.ncbi.nlm.nih.gov/pubmed/20694936.

Fast, E.; and Horvitz, E. 2017. Long-Term Trends in the Public Perception of Artificial Intelligence. *Proceedings of the AAAI Conference on Artificial Intelligence* 31(1). URL https://ojs.aaai.org/index.php/AAAI/article/view/10635.

Figueira, Á.; and Oliveira, L. 2017. The current state of fake news: Challenges and opportunities. *Procedia Computer Science* 121: 817–825. ISSN 18770509. doi:10.1016/j.procs.2017.11.106. URL https://doi.org/10.1016/j.procs.2017.11.106.

Guess, A.; Nagler, J.; and Tucker, J. 2019. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances* 5(1). doi:10.1126/sciadv.aau4586. URL https://advances.sciencemag.org/content/5/1/eaau4586.

Hanselowski, A.; PVS, A.; Schiller, B.; Caspelherr, F.; Chaudhuri, D.; Meyer, C. M.; and Gurevych, I. 2018. A Retrospective Analysis of the Fake News Challenge Stance-Detection Task. In *Proceedings of the 27th International Conference on Computational Linguistics*, 1859–1874. Santa Fe, New Mexico, USA: Association for Computational Linguistics. URL https://www.aclweb.org/anthology/C18-1158.

Hopp, T.; Ferrucci, P.; and Vargo, C. J. 2020. Why Do People Share Ideologically Extreme, False, and Misleading Content on Social Media? A Self-Report and Trace Data–Based Analysis of Countermedia Content Dissemination on Facebook and Twitter. *Human Communication Research* ISSN 0360-3989. doi:10.1093/hcr/hqz022. URL https://doi.org/10.1093/hcr/hqz022. Hqz022.

Horne, B. D.; Nevo, D.; O'Donovan, J.; Cho, J.-H.; and Adalı, S. 2019. Rating Reliability and Bias in News Articles: Does AI Assistance Help Everyone? *Proceedings of the International AAAI Conference on Web and Social Media* 13(01): 247–256. URL https://ojs.aaai.org/index.php/ICWSM/article/view/3226.

Hutto, C. J.; and Gilbert, E. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.

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

Johnson-Laird, P.; Gawronski, B.; and Strack, F. 2012. Mental models and consistency. *Cognitive consistency: A fundamental principle in social cognition* 225–243.

Lazer, D. M. J.; Baum, M. A.; Benkler, Y.; Berinsky, A. J.; Greenhill, K. M.; Menczer, F.; Metzger, M. J.; Nyhan, B.; Pennycook, G.; Rothschild, D.; Schudson, M.; Sloman, S. A.; Sunstein, C. R.; Thorson, E. A.; Watts, D. J.; and Zittrain, J. L. 2018. The science of fake news. *Science* 359(6380): 1094–1096. ISSN 0036-8075. doi:10.1126/science.aao2998.

Lewandowsky, S.; Ecker, U. K.; and Cook, J. 2017. Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era. *Journal of Applied Research in Memory and Cognition* 6(4): 353–369. ISSN 22113681. doi:10.1016/j.jarmac.2017.07.008. URL http://dx.doi.org/10.1016/j.jarmac.2017.07.008.

Lewandowsky, S.; Ecker, U. K.; Seifert, C. M.; Schwarz, N.; and Cook, J. 2012. Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest, Supplement* 13(3): 106–131. ISSN 15291006. doi:10.1177/1529100612451018.

Livingston, G. 2018. Artificial Intelligence: The Most Unfortunate Buzzword. URL https://medium.

com/datadriveninvestor/artificial-intelligence-the-most-unfortunate-buzzword-a5f0b678a7b1.

Metzger, M. J.; and Flanagin, A. J. 2013. Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics* 59: 210–220. ISSN 03782166. doi:10.1016/j.pragma.2013.07.012. URL http://dx.doi.org/10.1016/j.pragma.2013.07.012.

Morstatter, F.; Wu, L.; Yavanoglu, U.; Corman, S. R.; and Liu, H. 2018. Identifying Framing Bias in Online News. *ACM Transactions on Social Computing* 1(2): 1–18. ISSN 2469-7818. doi:10.1145/3204948.

Nieminen, S.; and Rapeli, L. 2019. Fighting Misperceptions and Doubting Journalists' Objectivity: A Review of Fact-checking Literature. *Political Studies Review* 17(3): 296–309. ISSN 1478-9299. doi:10.1177/1478929918786852. URL http://journals.sagepub.com/doi/10.1177/1478929918786852.

Pennycook, G.; Bear, A.; Collins, E. T.; and Rand, D. G. 2020. The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science* .

Pérez-Rosas, V.; Kleinberg, B.; Lefevre, A.; and Mihalcea, R. 2017. Automatic Detection of Fake News. *CoRR* abs/1708.07104. URL http://arxiv.org/abs/1708.07104.

Pingree, R. J.; Brossard, D.; and Mcleod, D. M. 2014. Mass Communication and Society Effects of Journalistic Adjudication on Factual Beliefs, News Evaluations, Information Seeking, and Epistemic Political Efficacy. *Taylor & Francis* 17(5): 615–638. ISSN 1520-5436. doi:10.1080/15205436.2013.821491. URL https://www.tandfonline.com/action/journalInformation?journalCode=hmcs20.

Pingree, R. J.; Hill, M.; and Mcleod, D. M. 2013. Distinguishing Effects of Game Framing and Journalistic Adjudication on Cynicism and Epistemic Political Efficacy The Influence of Postdebate News Framing and Fact Checking on Epistemic Political Efficacy and Cynicism. *Communication Research* 40(2): 193–214. doi:10.1177/0093650212439205. URL http://www.

Popat, K.; Mukherjee, S.; Strötgen, J.; and Weikum, G. 2016. Credibility assessment of textual claims on the web. In *International Conference on Information and Knowledge Management, Proceedings*, volume 24-28-October-2016, 2173–2178. New York, NY, USA: Association for Computing Machinery. ISBN 9781450340731. doi:10.1145/2983323.2983661. URL https://dl.acm.org/doi/10.1145/2983323.2983661.

Ruths, D. 2019. The misinformation machine. *Science* 363(6425): 348–348. ISSN 0036-8075. doi:10.1126/science.aaw1315. URL https://science.sciencemag.org/content/363/6425/348.

Shaar, S.; Babulkov, N.; Da San Martino, G.; and Nakov, P. 2020. That is a Known Lie: Detecting Previously Fact-Checked Claims. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 3607–3618. Online: Association for Computational Linguistics. doi:10.18653/v1/2020.acl-main.332. URL https://www.aclweb.org/anthology/2020.acl-main.332.

Shiralkar, P.; Flammini, A.; Menczer, F.; and Ciampaglia, G. L. 2017. Finding streams in knowledge graphs to support fact checking. *Proceedings - IEEE International Conference on Data Mining, ICDM* 2017-Novem: 859–864. ISSN 15504786. doi:10.1109/ICDM.2017.105.

Shu, K.; Sliva, A.; Wang, S.; Tang, J.; and Liu, H. 2017. Fake News Detection on Social Media: A Data Mining Perspective. *SIGKDD Explor. Newsl.* 19(1): 22–36. ISSN 1931-0145. doi:10.1145/3137597.3137600. URL https://doi.org/10.1145/3137597.3137600.

Vlachos, A.; and Riedel, S. 2014. Fact Checking: Task definition and dataset construction. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, 18–22. Baltimore, MD, USA: Association for Computational Linguistics. doi:10.3115/v1/W14-2508. URL https://www.aclweb.org/anthology/W14-2508.

Volkova, S.; and Jang, J. Y. 2018. Misleading or Falsification: Inferring Deceptive Strategies and Types in Online News and Social Media. In *Companion Proceedings of the The Web Conference 2018*, WWW '18, 575–583. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee. ISBN 9781450356404. doi:10.1145/3184558.3188728. URL https://doi.org/10.1145/3184558.3188728.

Vosoughi, S.; Roy, D.; and Aral, S. 2018. The spread of true and false news online. *Science* 359(6380): 1146–1151. ISSN 10959203. doi:10.1126/science.aap9559.

Wang, Y.; Ma, F.; Jin, Z.; Yuan, Y.; Xun, G.; Jha, K.; Su, L.; and Gao, J. 2018. EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery Data Mining*, KDD '18, 849–857. New York, NY, USA: Association for Computing Machinery. ISBN 9781450355520. doi:10.1145/3219819.3219903. URL https://doi.org/10.1145/3219819.3219903.

Wathen, C. N.; and Burkell, J. 2002. Believe it or not: Factors influencing credibility on the Web. *Journal of the American Society for Information Science and Technology* 53(2): 134–144. ISSN 15322882. doi:10.1002/asi.10016.

Winneg, K. M.; Hardy, B. W.; Gottfried, J. A.; and Jamieson, K. H. 2014. Deception in Third Party Advertising in the 2012 Presidential Campaign. *American Behavioral Scientist* 58(4): 524–535. doi:10.1177/0002764214524358. URL https://journals.sagepub.com/doi/abs/10.1177/0002764214524358.

Wu, Y.; Agarwal, P. K.; Li, C.; Yang, J.; and Yu, C. 2014. Toward computational fact-checking. *Proceedings of the VLDB Endowment* 7(7): 589–600. ISSN 21508097. doi:10.14778/2732286.2732295. URL http://dl.acm.org/doi/10.14778/2732286.2732295.

Yapo, A.; and Weiss, J. 2018. Ethical Implications of Bias in Machine Learning. doi:10.24251/HICSS.2018.668.

Zhang, J.; Kawai, Y.; Nakajima, S.; Matsumoto, Y.; and Tanaka, K. 2011. Sentiment bias detection in support of news credibility judgment. In *Proceedings of the Annual Hawaii International Conference on System Sciences*. ISBN 9780769542829. ISSN 15301605. doi:10.1109/HICSS.2011.369.

6 Appendix

6.1 Article 1 Content

A man in Bellingham was recently arrested in the murder of his wife and their two pets, both of who he shot using a gun. After he killed them, he went on Twitter to brag about it and thank the NRA, as well as give his address to be arrested. He claims that it would not have been possible without an AR-15, apparently advocating for stricter gun control. The liberal apparently felt the need to shoot and kill his entire family to prove his point on gun control.

An account appearing to be his on Twitter tweeted a few hours before he was arrested and around the estimated time of the killing, "Guns don't kill people, people do. Guns just make it a lot easier. AR-15 makes it super easy. I jst killed my whole family, and i couldnt have done it without a gun! I'm too much of a coward, a knife would have been waaay too hard. So, thanks to everyone at the NRA."

While the link between the account and killer has not yet been confirmed, the account most likely belonged to the killer, using both the same phone number and email as him, as well as name. The killer also replied to tweets by President Trump and the NRA saying similar things, and had a history of anti-Trump and anti-gun sentiment.

6.2 Article 2 Content

New York Governor Andrew Cuomo on Friday threatened to sue the Trump administration over its decision to restrict New Yorkers' access to some programs that allow faster security checks at ports of entry, part of a dispute about the state's limits on cooperation with current U.S. immigration policy.

President Donald Trump, a Republican who was born and raised in New York, has criticized the state and other states and cities his administration deems "sanctuary jurisdictions" because of their policies limiting information sharing between local law enforcement and federal immigration authorities.

"It's an abuse of power. It is extortion. And it's exactly what you did at Ukraine," Cuomo said in a reference to an impeachment charge that Trump pressured Ukraine to investigate a political rival, former Vice President Joe Biden. "You didn't learn the lesson."

The Department of Homeland Security said on Thursday that it would bar New York residents from both new passes and renewals of a program known as Global Entry, as well as three other programs that allow faster travel between the United States, Canada and Mexico.

6.3 Article 3 Content

Trump strikes again, going on Twitter to boldly announce that he believed that vaccines are linked to autism. Autism is a serious developmental disorder that impacts the ability to socialize and communicate.

In a stunning display of misinformation, Trump said that"Healthy young child goes to doctor, gets pumped with massive shot of many vaccines, doesn't feel good...–AUTISM."

The tweet has since been retweeted almost twenty thousand times with many pointing out that the connection is obviously false, and has been debunked by experts in the CDC, World Health Organization, and various peer-reviews publications and journals, yet Trump still denies those facts.

Trump has been an avid believer in conspiracy stories, from calling climate change a hoax to saying that he would "date" Ivanka Trump if she wasn't his daughter.

6.4 Article 4 Content

The left is at it again, this time with a new video that's gotten the attention of the right wing media. On the video, a woman standing in a crowd of people on a college campus is shown saying "we should start a race war" and "we should start a war with Russia."

The video appears to have been uploaded to YouTube in January. The footage was posted on The Daily Wire on Monday by author and political commentator Tanya Rivero, who says she took the video down because "the video was not meant to be shared. I apologize for any offense it has caused."

The Daily Wire also reported that the previous video of the woman's speech was removed from YouTube for violating a rule requiring the removal of content "that advocates for violence or hatred against an identifiable group."

Rivero posted the video at the far-right site, Gateway Pundit, on Monday. The Gateway Pundit posted an article titled "Here is the video that proves liberal campus activists are plotting to start a war with Russia."

6.5 Article 5 Content

Rumors that Trump doubts FBI data that China is censoring citizens' social media posts have been circulating for months. But the report from the Washington Post, which cited "current and former officials" and was based on documents provided by a "former senior U.S. intelligence official," provides the most detailed view yet of the president's thinking.

The rumors first appeared on the Facebook page "Occupy Democrats" where an anonymous source claiming to hold a position close to Trump wrote about Trump's apparent disbelief that China censor's its internet.

The Washington Post report chronicles a recent analysis conducted by the Office of the Director of National Intelligence, which concluded that the Chinese government is running a "massive censorship campaign" against Chinese-language websites and social media.

The official, however, claimed that Trump saw China as a potential ally and Communist Party leader Xi Jiping as a "good friend." Trump, according to him, fails to believe that someone like Xi would censor his people.

6.6 Article 6

On January 23, 2018, the Trump administration announced the opening of the first offshore leasing program in U.S. history, in the Gulf of Mexico.

It was the first major expansion of U.S. oil production in decades. The U.S. had previously leased less than 10 percent of the U.S. Outer Continental Shelf.

The country's offshore oil production had increased rapidly since the 1970s, but production remained low compared to other countries like Russia and Saudi Arabia.

In the oil-rich Gulf, the U.S. is now the largest oil producer. In 2014, the U.S. had about 1.7 billion barrels of oil reserves, but only about 1.1 billion were fully employed or commercially producible.

Total U.S. oil production during 2017 was 9.78 million barrels of oil equivalent per day (BOE/d). This is a significant increase from the 7.84 BBOE/d in 2016. The U.S. is currently the world's largest oil producer. To be clear, the U.S. is also the world's largest settler of oil. The U.S. had about 2.1 million barrels of oil per day in 2016, which is more than any other country in the world. The U.S. is also the world's largest producer of natural gas.

The U.S. produced about 35% of the world's natural gas in 2016, but the country only produced about 10% of its natural gas reserves.