The Design of AI Fact-checking UI to Combat Spreading of Fake News

Duri Lee [*1] **Eunki Joung** [*1] **Hye Soo Park** [*1]

Abstract

To resolve social issues caused by fake news, both quickly and automatically identifying potential fake news and effectively communicating it with users are vital. However, in the existing research, various AI-based fake news detection models are separated from the research on the user interface(UI) and its effects on reducing the user behavior of spreading fake news. Therefore, we replicated a model to detect fake news of the 2016 US presidential election and designed UI to deliver the AI-based rating based on the accuracy of replication. By the result, We found that the FakeBERT model is overfitted in a specific dataset. With the UI design, we conducted a user study, which confirmed that the AI rating-based interface reduces the spreading of fake news. Also, we can find the insights that UI were still valuable despite the low performance of the AI model.

1. Introduction

Fake news has been a significant issue recently as the spread of social media(Zhang & Ghorbani, 2020). Although issues about the trustfulness of news content existed before the internet era, people faced the severity of it through the 2016 presidential election in the United States(Kim et al., 2019). We observed that it became easy to produce and hallucinate readers by mimicking the format of reliable actual news article with the development of IT. At the same time, technological change cause changes in human behavior to accept and share news contents easily(Kim et al., 2019).

Therefore, both of detecting fake news and building/making interface to effectively induce the human behavior change to resolve the fake news related issues. Researchers have conducted natural language processing research to identify fake news and created various datasets for fast and automated detection. Alongside this, there has been research into the design and performance of educational programs to prevent the spread of fake news and the design of interfaces that effectively communicate warnings about fake news. However, research has yet to be done to design and validate the effectiveness of interfaces for communicating fake news determined by natural language processing (NLP) to people, leaving a research gap.

Researching effective interfaces for human-AI interaction, such as how the AI's detection results actually change people's behavior when communicated, and what accuracy is required, is essential to increasing the utility of the technology. In particular, AI models for detecting fake news may be structurally limited in their accuracy due to the difficulty of collecting good quality training datasets that are uncontroversial about labeling and the continuous generation of new information. Nevertheless, it may be meaningful to convey information about fake news judgments through AI models, even if imperfectly, as prior research has shown that information from heuristic judgments can change the behavior of good users.

Hence, we explore two research questions. 1) *Can we detect fake news with the legitimacy predicted by our AI model?* 2) *When users read fake news, will they be less likely to spread it if the news legitimacy predicted by the AI model is provided?*

To answer the 1st question, we replicate the FakeBert, which insists on the high accuracy of classification, and apply it to the health related news and prove the possibility of fake news detection in general issues. To answer the 2nd question, we design the UIs with the AI-rated information and test the user behavioral change while using it.

2. Related work

2.1. What is fake news?

What does fake news mean exactly? Fake news means "news articles that are intentionally and verifiable false and could mislead readers" (Allcott & Gentzkow, 2017). Weiss defines fake news comprehensively as combining action to exchange information, action to invalidate the truth, and intention to achieve power structure(Weiss et al., 2020). Although we can notice what fake news means intuitively, it is controversial how to define the terms(Weiss et al., 2020) when it comes to research because the appearance of fake news has variation in reality(Zhang & Ghorbani, 2020). Therefore, we apply Weiss' broad definition of fake news with the specific scenario for the research.

* Equal contribution

2.2. Approach to detect fake news using AI

Among the diverse effort to build AI models to detect fake news, there are two main approaches (Kaliyar et al., 2021). One approach is news content-based, which extracts the information and unique writing style to detect fake news. The other approach is social context-based, which uses social engagements and instance-based methodologies(Propagation based methodology). Diverse deep learning models were used to detect fake news up to the purpose of research. For example, Jwa use the BERT model to detect fake news by analyzing the relationship between the headline and the body(Jwa et al., 2019). Giachanou used CNN not to classify the fake content, but to classify a user as a potential fact checker or a potential fake news spreader (Giachanou et al., 2020). Fakebert which combine Bert and CNN model to judge fake news shows significantly high accuracy which were over 99% (Kaliyar et al., 2021).

2.3. UI to stop spreading of fake news

Several types of research on the UI show information about fake news detection and the credibility of the news. Mustafa explored the impact of UI on major SNS such as Google Search and Facebook and how the giant tech company changed the UI to avoid spreading false information(Mustafaraj & Metaxas, 2017). Kim studied changes in users' trust in news content according to rating sources(Kim et al., 2019). It generates a realistic fake news article and shows it to the participants with four different UI(now information, expert rating, user article rating, and user source rating). It suggests that the reader's trust significantly differs when the rating is low. Bhuiyan primarily uses the user's collective opinion to design the nudge warning system to readers(Bhuiyan et al., 2021). Recently, in addition to human source-based information, some have tried to apply AI evaluation result for UI to replace a human. Choras suggest UI for presenting overall scoring using multiple evaluations of different AI models(Choraś et al., 2022).

3. Fake News Detection Model

We replicated an existing fake news detection model, FakeBERT, and applied it to health news.

3.1. Model

We replicated FakeBERT, a BERT-based fake news detection model. It generates a embedding of news text and passes it to five convolution layers including three parallel layers and following dense layers. The detailed implementation followed (Kaliyar et al., 2021), but we make changes including using a miniature BERT model(Turc et al., 2019) to reduce computational cost. Our change of implementation is explained in Table 1.

Table 1. Changes in FakeBERT replication

Parameter	Value
BERT Embedding	Miniature bert(12 layers; d=128)
Input length	304 tokens
Fine-tuning	Freeze all BERT layers
Epoch	4
Optimizer	Adadelta (lr=1.0)

For experiments with health domain news, we also implemented a simple CNN model inspired by the experiments in (Dai et al., 2020). This model consists of one convolution layer (kernel size=3) followed by max pooling(pool size=2) and two dense layers (dim=250, dim=1). It uses ReLU as an activation function except the last layer which used sigmoid. Adam optimizer and binary crossentropy loss is used. Each of first 300 words of preprocessed news text is turn into 14-dim embedding vector by tensorflow's Embedding class and fed into the model.

3.2. Dataset

Following (Kaliyar et al., 2021), we trained FakeBERT model to news about 2016 US election. ISOT (Ahmed et al., 2018)(Ahmed et al., 2017) and Kaggle(Lifferth, 2018) fake news datasets were selected as the most similar datasets. ISOT dataset consists of 39k political and world news, mostly from 2016 to 2017. Truth news are from Reuters and fake news are from unreliable sources flagged by Politifact(a fact checking website) and Wikipedia. The Kaggle dataset consists of 20k fake and real news collected up to 2018. Though the sources were not clear, 43.9% of news texts contain 'Trump' and/or 'Hillary', so we assumed it is related to 2016 US election.

After replication of the model, we applied it to health domain which more fits to study AI fact check UI design for three reasons: it is interesting for the participants(university students); credible fact checks are available; timing of the news less matter than political news, so old fact checks can be used. Thus, we selected the most appropriate fake health news dataset.

FakeHealth(Dai et al., 2020) dataset is chosen because it includes a full news text, and the rating process is clear. Experts evaluated the articles regarding ten questions(e.g. 'Does it commit disease-mongering?'), and rated them. We labeled the articles with rating < 2.5 as fake and the others as truth, following (Dai et al., 2020) with a slight difference in threshold.

3.3. Experiment

We replicated the models from scratch and evaluated their performances(Table 1). Accuracy, False Negative Rate(FNR), and False Positive Rate(FPR) were used as

Table 2. Evaluation result of FakeBERT. H1-H5 used FakeHealth dataset and P1-P3 used ISOT or Kaggle dataset. FB, Acc, CNN mean FakeBERT, accuracy, and the simple CNN model, respectively.

Trial	Explanation	Acc	FPR	FNR
H5	CNN + oversampling	0.680	0.213	0.542
H4	CNN + class weights	0.702	0.143	0.618
H3	CNN	0.710	0.071	0.739
H2	FB + undersampling	0.656	0.056	0.931
H1	FB	0.671	0	1
P3	FB trained with ISOT and tested with Kaggle	0.629	0.055	0.677
P2	FB with Kaggle	0.978	0.004	0.041
P1	FB with ISOT	0.999	0.002	0.000

metrics. All experiments are done in a server with single NVIDIA GeForce RTX 3090.

We first tested the FakeBERT model with ISOT and Kaggle dataset and then evaluated different methods to train models with FakeHealth dataset. For FakeHealth, we trained and tested the FakeBERT without change, but all news were predicted as 'true'. Reflecting this result, we identified two reasons that might caused the failure: limitation of (1) model generalizability and (2) the dataset. To address (1), FakeBERT is trained with ISOT dataset and tested with Kaggle dataset. Also, we evaluated the simple CNN model. For (2), we tested undersampling, giving class weights, and oversampling to overcome the dataset limitation. Oversampling was done by summarizing existing fake news by T5 model(Raffel et al., 2020) to generate a new fake news. The results are presented in the next section.

3.4. Result

FakeBERT showed 99.9% accuracy with ISOT dataset and 97.8% accuracy with Kaggle dataset, which is comparable to 98.90% accuracy of the original paper. FakeBERT trained on ISOT dataset marked 62.9% accuracy, 0.055 FPR, and 0.677 FNR when tested with Kaggle dataset.

FakeBERT predicted all news 'true' for FakeHealth dataset (FPR=0, FNR=1). Undersampling leads to make a little amount of 'fake' prediction(FPR=0.06, FNR=0.93). Simple CNN model without consideration of class weights or sampling technique showed higher accuracy, higher FPR, and lower FNR than trials with FakeBERT. Given class weigths or oversampling, the CNN model reported less accuracy and more balanced prediction as shown in Table 2.

4. User Study

The user test is conducted using mid-fi prototype to investigate whether the user is less likely to spread the fake news if the news legitimacy predicted by the AI model is provided. By breaking down the main research question into three sub-questions, we can effectively address the research question in a focused manner. 1) *Could the user interface(UI) that warns about the potential risk of an article as fake news influence users' behavior in sharing it?* 2) *Does it better to provide the AI's judgment in a warning message than just simple warning message?* 3) *How do users think about warnings of potential fake news based on AI model?*

4.1. Settings

4.1.1. Prototypes

The prototype is designed in order to stop the user behavior of sharing the news article in a mobile web browser, which is main target scenario of our research. Therefore, it appears when user tap the Share button in a mobile web browser. Two different UIs (Figure 1) are used for the user test. To mitigate the over-reliance on the AI model in the false negative error, the UI was appeared when the AI rated the article as fake. UI1 warns risk of fake news, however it does not offer the AI prediction for the article user wants to share. UI2 contains the AI prediction, and it provide more information about the AI model they used. Fake news detection accuracy (75%) of the AI model is based on the FakeBert research (68%). For the baseline, no warning message is used.

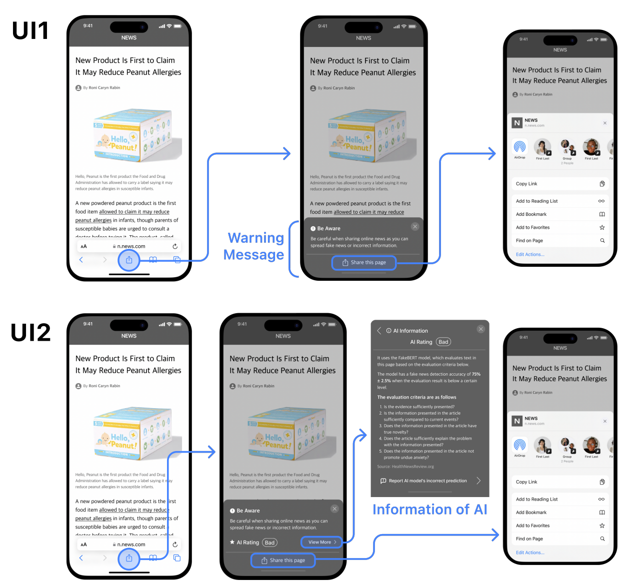

Figure 1. **Design of UI1:** The design components are three parts. The title is *"Be Aware"* and the main text is *"Be careful when sharing online news as you can spread fake news or incorrect information."*. At the bottom, the button for *"Share this page"* is placed. **Design of UI2:** It follows the design of UI1 with he rating by AI(*"AI rating: Bad"* and button for *"View more"*). The AI information includes accuracy of AI model and the evaluation criteria.

The news article and true/fake label for user test are collected from the source of SNU fact check IT & Science and the FakeHealth dataset. The participants are all Korean, therefore the translated version of FakeHealth news is used.

4.1.2. Experimental Environment

For the user test, a mobile phone mirroring the prototype and a laptop to reflects the participants behavior after the test are used. To observe the voluntary sharing behavior of users, the environment for experiment should be semi-natural. We opted articles for each participant using their interests to minimize variations in individuals' "willingness to share" attributed to news content rather than the UI.

After delivering a total of 110 headlines to the participants, 42 news articles that captured the interest of at least one participant was selected. In each of the three cases (baseline with no warning message, UI1, and UI2) as mentioned earlier, every participant read 8 news stories. To ensure even distribution, we organized each participant's set of 24 news articles, ensuring that the news articles they found intriguing were evenly distributed across the different UI sets.

4.2. Participants

The participants for this user study are recruited in KAIST with the snowball sampling. Six KAIST students whose knowledge level for NLP is varied and who are majoring computer science(3), security(1), biology(1) and chemistry(1) are selected.

4.3. Procedure

A brief 5-minute pre-survey, encompassing questions pertaining to the accuracy of our AI model in detecting fake news was commenced before the main user test. Subsequently, we ran three sets lasting 5-8 minutes each, utilizing our prototype: a baseline set, a set of UI1, and a set of UI2. Within each set, participants were presented with eight news articles and given the option to share. In the event of non-sharing, participants pressed a designated button to proceed to the subsequent article. To alleviate cognitive load, a break of 2-3 minutes was provided between sets.

After the user test, individual interviews for 30 minutes with a range of inquiries regarding their behavioral patterns were conducted. This included insights into their motivations for sharing or abstaining from sharing, as well as their reflections upon reviewing video recordings capturing their behavior. A comprehensive list of the questions utilized during the pre-survey and user interview are in Table 3.

Table 3. Questions used in the pre-survey and the user interview.

Number	Questions
S1	What do you think minimum required accuracy of fake news detection by AI model? Please answer in percentage.
S2	What is your NLP background? 1) Haven't heard about it; 2) Know a little; 3) Know how it works; 4) Have run models; 5) Have experienced NLP-based research.
I1	Why didn't you share or try to share the article?
I2	If you saw a message labeled as fake news and still shared it, why?
I3	If you saw a message labeled as fake news and didn't want to share it, why?
I4	How did you think when you saw that the AI rating was "Bad"?
I5	Which interface do you prefer? (UI1, UI2)
I6	What do you think minimum required accuracy of fake news detection by AI model? Please answer in percentage.

4.4. Data Collection

During the testing phase, the participants' interactions with our prototype were recorded through screen capture. The number of times each participants interacted with the sharing button is counted by the researchers meticulous review on the recorded videos. The raw data obtained from this analysis can be found in Table 4.

Additionally, the participants' responses during the interviews were recorded for further analysis. After the interviews concluded, the audio recordings were transcribed verbatim. To manage the extensive verbal data, an affinity diagram was employed to systematically organize and categorize the information obtained from the transcriptions.

4.5. Data Analysis

4.5.1. UI and the user behavior

By the data analysis, it is found that the UI that warns risk of fake news influences users behavior. 3 out of 6 participants paused sharing after seeing UI1 and then went back to reading the article. Also, 5 out of 6 participants withdrew their intent to share after seeing UI2. Several insights about the reason of user behavior were figured out using the qualitative analysis.

Raises suspicion about the article or themselves: P3 who stopped the sharing process and went back to read the article again after seeing the UI1 said that *"I had doubts about whether I had read this news correctly or whether this news was the right information"* P2 claimed that using UI2 made him think about things he hadn't thought about before, said that *"When I read news from a portal site, I never thought*

Table 4. The number of participants' interaction for ①: Trying to share, ②: Watching the UI pop up, ③: Reading the article again (If they saw the UI), ④: Sharing successfully (If they saw the UI) ⑤: Stopping sharing because of UI (②-④). Also the answer for the questions of pre-survey (S1, S2) and interview (I5, I6) described in the table. I6* is answer for minimum accuracy especially for this UI.

	B	UI1									UI2								QUESTIONS				
	④	①	②	③	TN	TP	FN	FP	④	⑤	①	②	TN	TP	FN	FP	④	⑤	S1	S2	I5	I6	I6*
P1	6	3	1	0	1	2	0	0	1	0	4	2	1	1	1	1	1	1	80	3	UI2	80	-
P2	5	5	2	1	1	2	1	1	2	0	3	3	2	0	0	1	2	1	95	5	UI2	95	-
P3	2	3	1	1	1	1	1	0	1	0	1	0	0	0	1	0	0	0	90 - 95	1	UI2	90	70
P4	4	4	2	1	1	1	1	1	2	0	3	3	0	2	0	1	1	2	80 - 90	3	UI2	90	-
P5	4	4	0	0	3	0	1	0	0	0	4	3	1	3	0	0	0	3	90	2	UI2	90	50
P6	4	2	2	0	0	1	0	1	2	0	3	1	1	1	1	1	0	1	90	5	UI2	90	80 - 85

that it might be fake news, but if the UI appears, it makes me think once more that it might be." The propagated effects on the next article they see was observed together in the case of P6, who said *"After seeing this UI1, I was more cautious about sharing other articles, which might be fake news."*

Loss of willingness to share: Furthermore, we found that participants were less willing to share our UI after seeing it, which is what we wanted to prove in our main RQ. This was especially pronounced for participants who saw UI2. *"I thought, why should I share information that might not be true?"* (UI2, P5) *"When they said it was fake, I didn't want to share it anymore."* (UI2, P6)

4.5.2. UI WITH AI BASED INFORMATION

5 out of 6 participants stopped the sharing behavior after they saw the UI2, while no one quit sharing after the UI1. The warning message with AI prediction and information was more influential on users' behavior than simple warning message caused by several reasons mentioned below.

Reduced impact due to easily desensitized to UI1: The participants reported that UI1 *"adds just another touch."* (UI1, P6) or it is *"just as pop-up message that come up usually."* (UI1, P2). Since they easily got desensitized to the simple warning, they didn't think it has significant meaning or they should think about it rather than UI2.

Better understanding of the cautionary wording of UI2: By using UI1, no participant mentioned their concern about that their sharing behavior will spread fake news. In contrast, participants expressed concerns about sharing information, stating that they believed it was *"right not to share it because other people might be exposed to fake information"* (UI2, P5) in the UI2 phase. They refrained from sharing news they suspected to be fake news, as they recognized the potential problems of fake news as P1 mentioned that *"If it was fake news, I stopped sharing it because I thought it will be problematic."* (UI2, P1).

These observations suggest that UI2 can give a better understanding of the cautionary wording and messages presented to users. It appears that participants read the content more carefully in UI2 than those in UI1 and considered it important to exercise caution in sharing information.

4.5.3. USER PREFERENCE

All six participants unanimously preferred UI2 over UI1 with comparison to the alternative UI1.

Increased trustworthy: The participants trusted UI2 more than UI1 because of the offered information details of AI such as evaluation criteria and accuracy as P1 stated that *"I think I would trust something with information much more than a simple warning."*. This resulted in participants placing more importance on UI2 and tending to favor it over UI1 in addition to being more influenced by it as P6 said that *"UI2 made me more alert to the fact that I might spread the fake news, however I feel the same with or without UI1"* and P3 mentioned *"The text in UI1 is actually something that everyone can say, even seems like a strange message or pop-up nagging, but UI2 seems like offering advice to me."*.

5. Discussion

5.1. Fake News Detection Model

5.1.1. LIMITED GENERALIZABILITY OF MODEL

FakeBERT reported high performance with political news, but it did not work well with health news. FakeBERT might be overfitted to each political news dataset. When the FakeBERT trained with ISOT dataset was tested with Kaggle dataset, it it showed decreased accuracy(0.999 $\rightarrow$ 0.629) and increased false negative rate(0.000 $\rightarrow$ 0.677), compared to the performance of same model tested with the same ISOT dataset. This result implies the FakeBERT could not detect fake news from unseen political news dataset, even if it has a similar topic with the train set. It imposes a question that whether fake news detection model showing a good performance with one dataset can work well in a field.

5.1.2. Limitation of Fake Health News Dataset

The characteristics of FakeHealth dataset affected the low performance of the models. First, the dataset consists of of about 2k news articles, which can limit the model performance. The size is more than 10 times smaller than the ISOT dataset(39k) and the Kaggle dataset(20k). Second, the classes are not equally distributed, with true news accounting for 67%. This imbalance led our models predicted the labels in a way that was pretty biased to 'true'(H1, H3), resulting in a lot of false negatives('true' is a negative prediction). Applying oversampling (H5) and weight multiplication (H4) to the simple CNN model, and undersampling (H2) to FakeBERT, reduced prediction imbalances. However, this did not increase overall accuracy. Third, the variety of subtopics in health domain may require a larger training set. The news in FakeHealth dataset covers various topics, such as cancer, diet, and sleep problems. With the size limitation, the model would not be able to see enough amount of ariticles in each topic, leading to low performance. Overall, it would be beneficial to investigate a bigger and more balanced dataset for fake health news detection in future.

5.2. User Interface Design

5.2.1. Needs of more accurate AI models

The participants said that they will be able to trust this system when the base AI model performs at least 90% (on average) in the end of the experiment. P2 expressed the opinion that it is hard to trust AI when it was inaccurate and commented *"If the AI model's performance is not enough like 75%, I don't need to use it."*. Also, P4, who thought that the current accuracy was not suitable even for the assistant and argued that *"The accuracy of 75% seems to be low to adopt and utilize as a second opinion."*

For the case that 3 participants who experienced error situation (FP; The prototype detected the article as a fake news and show the UI, however it was not actually.) during the experiment, they all required higher AI accuracy after the experiment. We argue that the reason is the confusion that users expressed during our experiments. The participants were confused when they figured out the AI detected the article as fake news, contrary to their background knowledge.

However, this intervention was still valuable despite of the low quality of AI. 3 out of 6 participants thought that it will affect users' behavior even if the accuracy is low ($\geq$ 50% , $\geq$ 70%) if the purpose is to warn people not to spread fake news. P5 said that *"I don't like the idea that there is a probability that what I share is fake. So I didn't share if I see the message"* Also P3 mentioned that *"70% or more is enough to make me reconsider my judgment on sharing"*.

5.2.2. Limited ability to completely prevent the spread of all fake news

The participants provided reasons for their decision to continue the sharing process after their experience with UI2. A common theme among participants was that they were unable to stop themselves because the accuracy of 75% described in the AI information was too low. P2 checked the warning message but kept his intention to share, and mentioned that *"I didn't trust the accuracy of the interface because it was too low. Nevertheless, I think it is meaningful to interrupted in the middle by this design"* (P2, UI2)

They don't care if it's fake news in some cases such as P4 said that *"I shared it to my friends in the form of a question 'Isn't this article weird?', so I don't care about whether it is fake or not"* (P4, UI2). Depending on the purpose of sharing, participants shared the article even though they knew it was fake news. When the news was "questionable, open to discussion," "not affecting anyone's profit/loss," or "humorous because it's obviously fake." then they have shared the news regardless of fake or not.

However, by the result mentioned earlier, the proposed UI is sufficient to alert a benign person-person without malicious intent- who reads a fake news article and reproduces it and make them reconsider their behavior.

5.2.3. Needs of Personalized UI

User needs for type of information and effects from the model performance are varied by users' background. For example, the accuracy of AI is offered in this UI, however P6 who had knowledge on the NLP said that F1-score of AI is needed in fake news detection system. Therefore, we claimed that it is necessary to study the effectiveness of providing customized interfaces for the future works.

6. Conclusion

Through this project, we attempted to build a fake news detection model using an AI model to mitigate the problems caused by the rapid spread of fake news and verified the effectiveness of AI-rated informed UI on changing user behavior. Although the model accuracy of replicating Fake Bert on a new dataset was below the expectation in study 1, it was a meaningful finding to understand the difficulty of generalization of the model and the importance of the dataset. In study 2, we design the semi-natural experimental environment, and we can observe the user behavior change up to the interface with meaningful insights on UI design to deliver the detected result from AI. In summary, This study is a worthwhile project to make AI technologies more practical in the real world.

References

Ahmed, H., Traore, I., and Saad, S. Detection of online fake news using n-gram analysis and machine learning techniques. In Traore, I., Woungang, I., and Awad, A. (eds.), *Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, pp. 127–138, Cham, 2017. Springer International Publishing. ISBN 978-3-319-69155-8.

Ahmed, H., Traore, I., and Saad, S. Detecting opinion spams and fake news using text classification. *SECURITY AND PRIVACY*, 1(1):e9, 2018. doi: https://doi.org/10.1002/spy2.9. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/spy2.9`.

Allcott, H. and Gentzkow, M. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–236, 2017.

Bhuiyan, M. M., Horning, M., Lee, S. W., and Mitra, T. Nudgecred: supporting news credibility assessment on social media through nudges. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–30, 2021.

Choraś, M., Pawlicka, A., Kozik, R., and Woźniak, M. How machine learning may prevent the breakdown of democracy by contributing to fake news detection. *IT Professional*, 24(2):25–31, 2022.

Dai, E., Sun, Y., and Wang, S. Ginger cannot cure cancer: Battling fake health news with a comprehensive data repository. *Proceedings of the International AAAI Conference on Web and Social Media*, 14 (1):853–862, May 2020. doi: 10.1609/icwsm.v14i1.7350. URL `https://ojs.aaai.org/index.php/ICWSM/article/view/7350`.

Giachanou, A., Ríssola, E. A., Ghanem, B., Crestani, F., and Rosso, P. The role of personality and linguistic patterns in discriminating between fake news spreaders and fact checkers. In *Natural Language Processing and Information Systems: 25th International Conference on Applications of Natural Language to Information Systems, NLDB 2020, Saarbrücken, Germany, June 24–26, 2020, Proceedings 25*, pp. 181–192. Springer, 2020.

Jwa, H., Oh, D., Park, K., Kang, J. M., and Lim, H. exbake: Automatic fake news detection model based on bidirectional encoder representations from transformers (bert). *Applied Sciences*, 9(19):4062, 2019.

Kaliyar, R. K., Goswami, A., and Narang, P. Fakebert: Fake news detection in social media with a bert-based deep learning approach. *Multimedia tools and applications*, 80 (8):11765–11788, 2021.

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

Lifferth, W. Fake news, 2018. URL `https://kaggle.com/competitions/fake-news`.

Mustafaraj, E. and Metaxas, P. T. The fake news spreading plague: was it preventable? In *Proceedings of the 2017 ACM on web science conference*, pp. 235–239, 2017.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(1), jan 2020. ISSN 1532-4435.

Turc, I., Chang, M., Lee, K., and Toutanova, K. Well-read students learn better: The impact of student initialization on knowledge distillation. *CoRR*, abs/1908.08962, 2019. URL `http://arxiv.org/abs/1908.08962`.

Weiss, A. P., Alwan, A., Garcia, E. P., and Garcia, J. Surveying fake news: Assessing university faculty's fragmented definition of fake news and its impact on teaching critical thinking. *International Journal for Educational Integrity*, 16:1–30, 2020.

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