International Journal of Communication 17(2023), 518–540 1932–8036/20230005

User Perceptions and Trust of Explainable Machine Learning Fake News Detectors

JIEUN SHIN
SYLVIA CHAN-OLMSTED
University of Florida, USA

The goal of the study was to explore the factors that explain users' trust and usage intent of the leading explainable artificial intelligence (AI) fake news detection technology. Toward this end, we examined the relationships between various human factors and software-related factors using a survey. The regression models showed that users' trust levels in the software were influenced by both individuals' inherent characteristics and their perceptions of the AI application. Users' adoption intention was ultimately influenced by trust in the detector, which explained a significant amount of the variance. We also found that trust levels were higher when users perceived the application to be highly competent at detecting fake news, be highly collaborative, and have more power in working autonomously. Our findings indicate that trust is a focal element in determining users' behavioral intentions. We argue that identifying positive heuristics of fake news detection technology is critical for facilitating the diffusion of AI-based detection systems in fact-checking.

Keywords: AI, fake news, media literacy, trust, explainability

Almost half of the U.S. news consumers use social media for news daily, and more than 60% of them are worried about the accuracy of news they encounter on social media (Newman, 2022). As a means to mitigate the news consumers' concerns, Ali and Hassoun (2019) stressed the essential role of AI in combating fake news. It was suggested that the use of smart journalistic algorithms is critical to identifying fake news and inaccurate content online, especially on social media. With the exponential growth of AI in all sectors, scholars have argued that the implications of AI should be foregrounded in the context of news and journalism (Lewis, 2019).

From a technical perspective, there has been significant progress in the last few years in machine learning (ML) methods, a major application of AI, that automatically detect fake news (Ozbay & Alatas, 2020). A key approach of these detection models is to train an algorithm such that it can predict news as fake with high accuracy. This automatic detection of fake news can help prevent false information from going viral on social media and assist journalists and fact-checkers in responding more effectively to misinformation (Graves, 2018). Two significant challenges of today's ML research in this area are that the true accuracy level of these models is

Jieun Shin: jieun.shin@ufl.edu
Sylvia Chan-Olmsted: chanolmsted@jou.ufl.edu
Date submitted: 2022-02-21

Copyright © 2023 (Jieun Shin and Sylvia Chan-Olmsted). Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at http://ijoc.org.

not confirmed and that the explanation as to why the model made a given classification is often invisible to users (Kim, Tabibian, Oh, Schoelkopf, & Gomez-Rodriguez, 2018). Such lack of transparency can contribute to hesitance in using the automated fake news detectors. In fact, we argue that advancing fake news technology alone is insufficient in tackling the problem of fake news and misinformation, because the most advanced solutions offer no utility when they are not used or trusted.

For these reasons, explainable AI, which emphasizes the provision of explanations for the decisions, have gained momentum in recent years (Guo, Ding, Yao, Liang, & Yu, 2020; Zhou & Zafarani, 2020). This "explanation" aspect of the AI technology offers hope in reducing the mystery of AI applications and even solidifies the role of machine from channel to communicator (Guzman & Lewis, 2020). In other words, it is plausible that the progression from a "black box" to a "white box" ML with some interpretability of the results would influence how users perceive the AI application. Compared with the recent advances made in the engineering domains of explainable AI, relatively little is known about how users perceive and trust these automated AI-based detection systems. Because trust is a basis of the adoption decision for an innovative technology (Rogers, 1995), it deserves more attention in research and practice.

Thus, in this study, we examined various factors that might influence the perception and usage intention of fake news detectors. As an exploratory study, we included a comprehensive list of predictors of the adoption of a new technology based on the previous studies (Kim, Kim, & Hwang, 2009; Lyons, Wynne, Mahoney, & Roebke, 2019; Ramos, Thieme, Utne, & Mosleh, 2020; Tussyadiah, Zach, & Wang, 2020). One cluster focused on individual characteristics such as demographics, attitude toward AI, and prior experience and knowledge. The other cluster pertained to the perceived characteristics of a specific tool such as ease of use, performance, and utility. Using a survey, we found that both individual characteristics and users' perceptions of the fake news detector influence their adoption intention. Specifically, individuals with low confidence in detecting fake news, more experience in fact-checking tools, more experience with AI technology, and higher levels of overall trust in AI were more likely to use the detector. Most important, we found that users' trust levels in the detector was the strongest predictor of usage intention, thereby serving as the key factor leading to the adoption. This study contributes to a growing body of literature on the implementation of AI in news by emphasizing the user's perspectives.

Literature Review

The topic of fake news was thrust into the spotlight during the presidential election in 2016. Since then, many fake news campaigns with political or commercial motives have been uncovered on social media platforms. Fake news refers to false news stories that are created with a deliberate intention to mislead people (Tandoc, Lim, & Ling, 2018). Fake news has become an umbrella term to encompass statements that are false. However, fake news is distinguished from other types of misinformation[1] in that it is disguised as legitimate news reports to exploit journalistic styles and formats (Molina, Sundar, Le, & Lee, 2021). The growth of fake news is said to be one of the biggest threats to democracy not only because it hampers

[1] In general, misinformation refers to inadvertent creation and sharing of false information, whereas disinformation assumes that falsehood stems from deliberate intention (Tandoc et al., 2020).

individuals' rational decision making but also because it undermines new media's credibility overall (Tandoc et al., 2021).

In recent years, researchers have aimed at understanding the phenomenon of fake news, including recognizing their common patterns, through the development of ML-based fake news detection models (Zhou & Zafarani, 2020). In the private sector, Facebook, Amazon, Microsoft, and Google have all joined the effort to partner with academic institutions to develop digital forensics techniques in the arms race between digital manipulations and detections (Metz, 2019). Although the technical aspect of such models is beyond the scope of this study, the novelty of the recent approach to enhance the ML models' interpretability offers some interesting implications from a communication perspective. For example, it is reasonable that some explanations on how the fake news classification was made might affect consumer confidence or trust on the result.

Given the context and the specific goals of the study, the next section will review the overriding construct of trust, especially in AI-related entities. Next, factors that could affect human trust in the context of fake news detectors are discussed from the perspectives of users.

Trust in AI

The term "artificial intelligence" covers a broad range of related technologies that aim to simulate human intelligence (Chan-Olmsted, 2019). AI can be classified into subfields such as ML, computer vision, speech recognition, natural language processing, planning, expert systems, and robotics (de-Lima-Santos & Ceron, 2021). ML is one of the subfields in AI that is dedicated to designing algorithms that build statistical models from data without preexisting solutions to a problem (Castro & New, 2016).

In pursuit of user acceptance of AI systems, the literature emphasizes interpersonal trust as an essential component (Gillath et al., 2021). The Computers are Social Actors (CASA) paradigm suggests that people would interact with computers as they would toward humans using similar social norms (Nass & Moon, 2000). Similarly, some researchers suggest that viewing AI application as teammates rather than tools helps our understanding of human trust in AI (Seeber et al., 2020; Wynne & Lyons, 2018). The current study adopts such a notion and incorporates factors related to considering AI applications as human-like collaborators. A key aspect of such collaborations is that the users trust the machine. Gillath et al. (2021) stressed that the lack of trust is one of the main hurdles standing in the way of taking full advantage of AI.

Users of any AI application must have the judgment that they can rely on an AI agent to achieve their goals when there is uncertainty (Okamura & Yamada, 2020). Interpersonal trust literature conceptualized trust as a multidimensional concept with cognitive, affective, and behavioral dimensions (Lewicki, Tomlinson, & Gillespie, 2006; McAllister, 1995). It encompasses one's willingness to be vulnerable to and act on the another's words, actions, and decisions (Lewicki et al., 2006; McAllister, 1995). Accordingly, this study defines trust in AI as the extent to which a person is confident and comfortable in recommendations, actions, or decisions of an AI technology and is willing to act according to the information provided.

Similarly, Hancock et al. (2011) identified three types of factors that might affect human trust in machines and robots. They are mainly human-related, robot-related, and environmental factors.

Specifically, both human-related and robot-related factors might cover capacity-based variables like human competency or machine dependability, as well as characteristic considerations like human propensity to trust or robot anthropomorphism. Some of these factors are independent of human-robot interaction contexts, while others are context specific. The third group of factors, environmental factors, addresses the human-robot team collaboration (e.g., communication) and nature of the task (e.g., task complexity). We adapt this framework and propose to examine the perceptions and trust of the ML-based fake news detection from both the human and machine perspectives while integrating the contexts of fake news detection and AI technology in the analysis.

The Human Perspective

Within the perspective of humans, users' ability to understand and use the AI technology can affect their trust in the technology (Hancock et al., 2011; Siau & Wang, 2018). For instance, expertise, typically resulting from knowledge and experience, helps users develop a more accurate understanding of AI's working rationale and capabilities, which is a fundamental antecedent of trust in AI (Mohseni, Zarei, & Ragan, 2020; Mohseni et al., 2020). Such understanding allows users to establish a proper expectation of the AI's performance, which would positively affect trust formation (Matthews, Lin, Panganiban, & Long, 2020; Nguyen et al., 2018). Also, users with more knowledge of AI often have more positive attitudes toward AI (Araujo, Helberger, Kruikemeier, & de Vreese, 2020). Beyond AI-related expertise, previous studies have repeatedly supported the contribution of trust to the intention to adopt new technologies. Most trust literature have suggested that initial trust must be established before the trustor becomes willing to depend on and develop a relationship with the trustee (Lewicki et al., 2006; Mcknight & Chervany, 2001). Empirical studies in AI-related areas have also concluded that people are more likely to interact with machines that they feel they can trust (Lee & See, 2004) and more willing to adopt AI service agents when trust is established (Komiak & Benbasat, 2006; Tussyadiah et al., 2020). In sum, users' general AI knowledge, experience, and trust are likely to play a role in forming the specific trust on the actual ML technology they proceed to interact with.

Although the aforementioned topics are domain specific, human characteristics such as personality traits and demographics are relatively domain-free and constant. This study also proposes that general characteristics like trust propensity and demographics might play a role in such human-AI interactions. Trust propensity refers to individuals' inherent, dispositional tendency to trust others (Mayer, Davis, & Schoorman, 1995). It is generally independent of trustee and environment characteristics (Mcknight, Carter, Thatcher, & Clay, 2011; Mcknight, Cummings, & Chervany, 1998), but has the strongest predictive power when information about the trustee's ability, integrity, and benevolence is unclear (Gill, Boies, Finegan, & McNally, 2005). The concept is applicable to AI trustees, because people have a general propensity to trust not only other people but also machines (Merritt & Ilgen, 2008). In the context of human-AI trust, research on the association between trust propensity and trust is limited, but a positive relationship between them has been initially supported (Tussyadiah et al., 2020). Demographic variables such as age and gender could also influence trust in AI (Hancock et al., 2011; Hoff & Bashir, 2015). Demographic-based research on human-automation interaction has yielded inconsistent patterns. But overall, it is safe to conclude that different demographic groups have ways to assess the trustworthiness of automation and respond to automation (Hoff & Bashir, 2015).

In addition, literature has often shown a positive relationship between other human factors like self-efficacy and trust in the new technology environment (Kim et al., 2009; Zhou, 2012). Self-efficacy refers to individuals' belief of their capability of performing a particular behavior (Bandura, 1977). Self-efficacy could contribute to trust building because users who have more positive expectations of the outcomes and more positive attitudes toward the trustees may have a greater chance to experiment with novel interventions (Bandura, 1977; Kim et al., 2009). In another relevant context, fact-checking tools/systems, an efficient method to verify news authenticity through human efforts, have been adopted as a main remedy to combat fake news or misinformation in recent years (Zhou & Zafarani, 2020). Users have different, and even polarized attitudes toward fact-checking services, with many expressing distrust (Brandtzaeg & Følstad, 2017). It is plausible that people who frequently use fact-checking services are more likely to be skeptical about news authenticity; they may also be more likely to trust or use news evaluation tools and services. Accordingly, the following research question is posited:

RQ1: How are various user characteristics, such as (a) demographics, (b) trust propensity, (c) fake news self-efficacy, (d) fact-checking service usage, (e) AI expertise, (f) and overall AI trust, associated with their trust in specific explainable ML fake news detectors?

The Machine Perspective

One of the most challenging things in adopting AI technology is to situate AI in humans' beliefs and intentions (Chakraborti & Kambhampati, 2018). At the same time, whether consumers feel positive or negative about algorithms-based technology varies depending on the type of task for which the algorithm is used (Castelo, Bos, & Lehmann, 2019). In fact, Hancock et al. (2011) found that the perceptions of the technology's performance or competence are the most influential factors in the trust building process. Therefore, it is our proposition that the perceived performance of the ML fake news detector would affect users' trust of the technology.

As indicated earlier, this study is interested in exploring the role of AI in the context of fake news from a collaborative perspective (Seeber et al., 2020). Miller (2019) suggests that in the age of AI, humans and machines need to work symbiotically and collaboratively to enhance each other. Collaborative capacity implies several qualities, such as the AI's ability to work autonomously, understand the context, and communicate with the human effectively (Lyons et al., 2019). When AI appears as a collaborator or with more agency, users are more likely to trust it and more willing to accept it (Dietvorst, Simmons, & Massey, 2016; Nguyen et al., 2018). These notions suggest that a certain degree of perceived agency and collaborative capacity is needed for users of a fake news detector to trust it.

Finally, it has been argued that the goals of explainable AI should include improved trustworthiness, informativeness, and confidence (Arrieta et al., 2019). Ultimately, explainable ML aims at increasing the degree to which a human can comprehend why certain decisions or predictions have been made with transparency (Emmert-Streib, Yli-Harja, & Dehmer, 2020). Nevertheless, moving from a black box of fake news diagnostics to offering information that maintains certain technical information, the ML detector inherently presents layers of unavoidable complexity, which might result in diminished confidence or desire to adopt (Ramos et al., 2020). Therefore, it is plausible that the perceived complexity of the ML fake news

detector would play a role in users' trust of the technology. Accordingly, the following research question is posited.

RQ2: How are the perceived machine characteristics such as (a) performance, (b) collaborative capacity, (c) agency, and (d) complexity associated with users' trust in explainable ML fake news detectors?

Trust and Adoption Intention

Theoretical arguments and empirical evidence have frequently supported the contribution of trust to intention to adopt new technologies. Many scholars have advocated the integration of the trust construct and classic theories used to explain adoption behaviors, including the technology acceptance model (TAM; Davis, 1989) and the theory of planned behavior (TPB; Ajzen, 1991), to better learn adoption intention of new technologies (e.g., Wu & Chen, 2005; Xie et al., 2017). These studies conceptualized and empirically validated that trust associated with TAM and TPB constructs including perceived ease of use, perceived usefulness, attitude, and perceived behavioral control contributes to adoption. The interpersonal trust literature also suggested that initial trust must be established before the trustor becomes willing to depend on and develop a relationship with the trustee (Lewicki et al., 2006; Mcknight & Chervany, 2001). Empirical studies in AI-related areas have supported that people are more likely to interact with machines that they feel they can trust (Lee & See, 2004) and more willing to adopt AI service agents when trust is established (Komiak & Benbasat, 2006; Tussyadiah et al., 2020). The last research question is thus posited to explore how users' trust in the ML detector predicts their adoption intention of the application.

RQ3: How does trust in the explainable ML fake news detector predict adoption intention of this AI application?

The Context: Explainable ML Fake News Detectors

The lack of transparency in algorithm processes had led to calls for research on explainability in AI (Castelvecchi, 2016). Many computational studies have aimed at understanding the phenomenon of fake news, including recognizing their common features and patterns and configuring fake news detection models based on ML. Most of these fake news detection methods focus on applying news contents and social contexts in the process and classifying fake news without explanations of how the decisions were derived (Karimi & Tang, 2019; Shu et al., 2019). More recent studies addressing algorithm acceptance have specified the heuristic role of explainability in AI application acceptance and the importance of human judgment and context (Shin, Zhong, & Biocca, 2020).

The recent growth of explainable ML has attempted to enhance the understandability, as well as trust (Lundberg & Lee, 2017), of the detection models by offering additional information such as accuracy levels and visual contribution of certain texts. For example, dEFEND is a leading interpretable fake news detector uses both social media posts and user feedback to identify fake news (Shu, Mahudeswaran, Wang, Lee, & Liu, 2018). In addition, there are a few other examples of explainable fake news detectors such as Propagation2Vec (Silva et al., 2021) and xFake (Yang et al., 2019). Propagation2Vec (Silva et al., 2021) is a network-based model that is capable of explaining the logic of determining fake news diffusion in the early

stage. xFake (Yang et al., 2019) determines fakeness of news articles based on linguistic features and visualizes detection results.

Conceptually, these fake news detectors define fake news as something that is verifiably false, and consider the news consumption ecosystem as part of the detection modeling (Shu et al., 2019). Although these detection methods have produced relatively accurate classifications, Alharbi, Vu, and Thai (2021) evaluated the leading explainable fake news ML models and found those with high accuracy rates do not necessarily deliver a higher level of trust among their users. Because the past study has not approached the subject with established trust measures and has used only a small sample when testing for the trust level of these explainable ML detectors (Alharbi et al., 2021), the current study offers a more comprehensive view of explainable ML in the context of fake news detection.

Method

Data Collection

Two national surveys were administered to collect data in May 2021. The questionnaire, including a dEFEND fake news detection output (Figure 1), was first tested in a pretest with 50 participants from MTurk. The participants' feedback was gathered, the reliability of the scales was assessed, and the questionnaire was revised accordingly. The main test recruited 1,127 participants with at least a 90% approval rating from MTurk. Quality checks were in place to exclude straight-liners and inattentive respondents. A total of 1,052 valid respondents were included in the final analysis. Males accounted for 51.43% of the final sample, and females accounted for 47.72%; the rest (0.85%) were nonbinary. The mean age of the sample is 23.67 years old (SD = 13.00). Of the participants, 0.38% had less than a high school degree, 9.32% completed high school, 16.44% attended college but did not earn a degree, 56.56% earned an associate or bachelor's degree, and 17.30% earned a master's degree or higher. About 10.27% of the participants are of Hispanic origins. As for race, 74.24% were White, 12.45% were Black or African American, 8.94% were Asian, 1.14% were American Indian or Alaska Native, 0.10% were Native Hawaiian or Pacific Islander, and 3.14% were biracial or multiracial. For income, 31.9% of the participants had an annual household income under $39,999, 38.6% were between $40,000 and $79,999, and 29.5% were above $80,000.

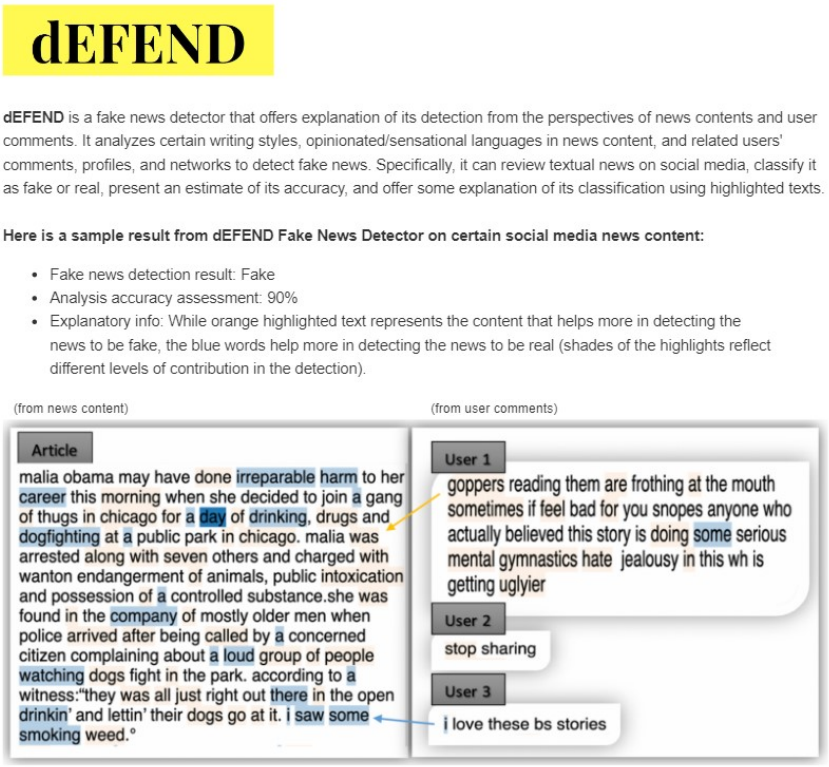

Figure 1. Example of dEFEND fake news detector output.

In the survey, questions about the respondents' dispositional characteristics (i.e., trust propensity) and general perceptions of AI technology were presented first. Then, various questions were given about the respondents' self-assessed fake news related efficacy and expertise. To provide a better context for the ML detector, the respondents were then offered explanations of what ML and explainable ML are. Next, the result of a leading explainable ML fake news detector, dEFEND, was presented in the survey (see Figure 1). Other detector-specific questions such as the detector's performance and agency were then assessed. Finally, trust in the detector and intent to adopt were posed. Demographic variables were collected at the end of the survey.

Measures

Most measures used in the survey were adapted from previous studies whenever possible. Five-point Likert and semantic differential scales were used. The variables of education and income were converted into an 8-point scale and a 12-point scale, respectively. The scales and questions used in the survey and their associated sources are presented in Table 1, including Cronbach α coefficients for multi-item variables. The scales used in the study exhibited satisfying reliability, with Cronbach's α ranging from 0.95 to 0.76.

Table 1. Means and Standard Deviations of All Study Variables.

Variable	Mean (SD)	Scale reliability
Age	23.67 (13.00)	-
Education (8-point ordinal scale)	4.49 (1.32)	-
Income (12-point ordinal scale)	6.54 (3.18)	-
Fact-checking tools use • How often do you use fact-checking services (e.g., Snopes, FactCheck.org) to verify information you get online?	2.39 (1.14)	-
Fake news experience • "Fake news" is deliberately false or misleading content presented as news. How experienced would you say you are in identifying fake news?	3.16(1.02)	
AI knowledge • AI (artificial intelligence)-related tools/technologies are used in many areas nowadays (e.g., smart speakers, chatbots, self-driving cars, virtual assistants like Siri/Alexa, and smart e-mail categorization). How knowledgeable would you say you are about the facts and issues related to AI?	2.92 (0.93)	-
AI experience • How experienced would you say you are in using AI-powered tools/technologies?	2.83 (0.95)	-
Trust propensity (Frazier et al., 2013) • I usually trust people/things until they give me a reason not to trust them. • Trusting another person or thing is not difficult for me. • My typical approach is to trust new acquaintances/things until they prove I should not trust them. • My tendency to trust others is high.	3.56 (1.01)	0.90
Fake news self-efficacy (Chan-Olmsted & Qin, 2022) • I believe that I can identify fake news by myself. • I know how to verify fake news by using appropriate tools for checking. • I believe that I can reduce the likelihood of sharing fake news.	4.07 (0.74)	0.76
Overall AI trust (Beltramini, 1988; Sirdeshmukh et al., 2002) • Undependable-Dependable • Incompetent-Competent • Low Integrity-High Integrity • Unresponsiveness-Responsive • Unsafe-Safe • Bad Intention-Good Intention • Untrustworthy-Trustworthy • Dishonest-Honest	3.80 (0.74)	0.84

Detector performance (Mcknight et al., 2011) • Inaccurate-Accurate • Unreliable-Reliable • Hard to use-Easy to use • Ineffective-Effective	3.69 (0.95)	0.91
Detector collaborative capacity (Calhoun, Bobko, Gallimore, & Lyons, 2019) • It is possible for the detector and I to work together. • I can work with the detector as a team to identify fake news.	4.03 (0.91)	0.88
Detector agency (Wynne & Lyons, 2018) • The detector has the full ability to decide if a content is fake news • The detector can offer the best decision on what is fake news • The detector can operate effectively with little oversight from you to determine what is fake news • It does not matter what the detector indicates. You will decide completely on your own if a content is fake news (R) • The best way to identify fake news is to rely on both this detector *and* what you know about fake news • It would be more effective to identify fake news when you also use this detector	3.36 (1.07)	0.79
Detector complexity (Maynard & Hakel, 1997) • It would be a complex task to use this detector. • It would be mentally demanding to use this detector. • It would require a lot of thoughts and problem-solving to use this detector.	2.68 (1.15)	0.91
Trust in Detector (Calhoun et al., 2019) • I would rely on it without hesitation. • I think using it will lead to positive outcomes. • I would feel comfortable relying on it. • I think I could depend on it if the detection task is hard. • I would trust its results.	3.20 (1.07)	0.93
Intention to adopt Detector (Teo, 2011) • I would use it on a regular basis. • I expect that I would use it in the future. • I would use it without hesitation.	3.25 (1.17)	0.92

Note. Scale reliability was measured with Cronbach's alpha.

Data Analyses

Two hierarchical regression models were used to answer RQ1 and RQ2: How various user characteristics and the perceived machine characteristics explain and predict trust in the explainable ML fake news detector. We first entered the user characteristics into Model 1 to account for the variance in trust level of the detector. These variables include participants' demographics (i.e., age, education level, income), their trust propensity, prior fact-checking tool use, fake news self-efficacy, and their AI knowledge, experience, and trust in general. In Model 2, we added four variables measuring the specific machine characteristics of the fake news detector, including perceived performance, collaborative capacity, agency, and complexity.

In addition, we performed a hierarchical three-stage analysis of regression models (Model 3, Model 4, Model 5) step-by-step to examine RQ3, which asks how trust in the detector is associated with the adoption intention of this application. We entered the variables regarding the user characteristics in Model 3 and the detector characteristics in Model 4. Last, Model 5 included *trust in the detector* along with all the other variables to account for the variance of adoption intention.

Multicollinearity among independent variables was checked using tolerance and variance inflation factor (VIF) values. The tolerance value ranged from 0.30 to 0.91 and the range of VIF was from 1.09 to 3.38, suggesting that multicollinearity was not an issue.

Results

Overall, both individual characteristics and fake news detector characteristics influenced trust levels. Table 2 summarizes the results of the two analyses for predicting trust in the AI detector. RQ1 asked how various user characteristics are associated with their trust in the fake news detector. The analysis showed that individual characteristics accounted for 32% of the variance of trust in Model 1, $F(9, 1042) = 56.74$, $p < .001$. Specifically, we found that younger age ($\beta = -.13$, $p < .001$), trust propensity ($\beta = .06$, $p < .05$), prior use of fact-checking tool ($\beta = .17$, $p < .001$), self-efficacy to detect fake news ($\beta = -.05$, $p < .05$), $p < .01$, prior experience of AI technology ($\beta = .08$, $p < .05$) and trust in AI technology in general ($\beta = .41$, $p < .001$) were significant predictors of trust in the fake news AI application. Education ($\beta = .04$, $p = .10$) and income ($\beta = -.02$, $p = .37$) did not emerge as significant factors for trust in the detector. AI knowledge ($\beta = .08$, $p = .60$) was not a significant predictor either. However, these findings indicate that overall human factors such as individuals' prior experiences with fake news tools and their inherent attitudes about AI are key factors in trust perception of the fake news detector.

Table 2. Predicting Trust Levels in the Application.

	Model 1	**Model 2**
	Trust in Fake News Detector	**Trust in Fake News Detector**
Individual Characteristics		
Age	−0.13 (.00)***	−0.06 (.00)***
Education	0.04 (.02)	0.06 (.01)***
Income	−0.02 (.01)	−0.03 (.01)*
Trust propensity	0.06 (.03)*	0.01 (.02)
Fact-checking tools use	0.17 (.03) ***	0.11 (.02)***
Fake news self-efficacy	−0.05 (.04)*	−0.04 (.02)*
AI expertise—knowledge	0.02 (.02)	0.00 (.03)
AI expertise—experience	0.08 (.09)*	0.06 (.03)*
AI trust	0.41 (.60)***	0.08 (.03)***
Detector Characteristics		
Detector performance		0.38 (.04)***
Detector collaborative capacity		0.12 (.02)***
Detector agency		0.31 (.02)***
Detector complexity		−0.06 (.02)**
N	*1,052*	*1,052*
Adjusted R^2	*.32*	*.70*
R^2 change	*.32*	*.38*
F	*F(9, 1042)=56.74*	*F(13, 1038)=190.1*

Note. All coefficients represent standardized regression coefficients. $*p < .05$, $**p < .001$, $***p < .001$.

RQ asked how the machine characteristics are related to users' trust in the fake news detector. As seen in Model 2, the addition of the fake news detector-related variables explained 38% more variance of trust, $F(13, 1038) = 190.1$, $p < .001$. All four detector-related variables—perceived performance (β = .38, $p < .001$), collaborative capacity (β = .12, $p < .01$), agency (β = .31, $p < .001$), and complexity (β = −.06, $p < .01$)—were significant predictors in the expected direction for trust in the detector. Younger age (β = −.06, $p < .001$), prior use of fact-checking tools (β = .11, $p < .001$), self-efficacy to detect fake news (β = −.04, $p < .05$), prior AI experience (β = .06, $p < .05$), and trust in AI in general (β = .08, $p < .001$) remained significant predictors in Model 2. These findings suggest that the perceived qualities of the software greatly influence the extent to which users are likely to trust the detector.

Furthermore, three analyses (Table 3) were conducted to examine the extent to which *trust in the fake news detector* influences *adoption intention (RQ3).* Model 3 and Model 4 were fitted to the data to serve as baseline models for comparison. Model 3 included the block of user characteristics variables which accounted for 31% of the variance, $F(9, 1042) = 53.97$, $p < .001$. Younger age (β = −.07, $p < .01$), trust propensity (β = .06, $p < .05$), prior use of fact-checking tools (β = .19, $p < .01$), prior AI experience (β = .10, $p < .05$), and trust in AI in general (β = .40, $p < .001$) were significant predictors of intention to adopt

the fake news AI detector. Model 4 with the additional block of detector-specific variables explained 27% more variance, $F(13, 1038) = 111.0$, $p < .001$. Similarly, all four variables measuring perceptions of the detector—performance (β = .31, $p < .001$), collaborative capacity (β = .14, $p < .001$), agency (β = .22, $p < .001$), and complexity (β = −.07, $p < .01$)—were significantly associated with adoption intention. Lastly, when *trust in the fake news detector* was added to Model 5, it explained 13% additional variance, $F(14, 1037) = 185.7$, $p < .001$. In fact, trust was the strongest predictor (β = .67, $p < .001$) for *intention to adopt*. These findings confirm the importance of trust in technology adoption such that people are more willing to use AI systems when trust is established.

Table 3. Predicting Adoption Intention of the Detector.

	Model 3	**Model 4**	**Model 5**
	Intention to Adopt	**Intention to Adopt**	**Intention to Adopt**
Individual Characteristics			
Age	−0.07 (.00)**	−0.01 (.00)	0.03 (.00)
Education	0.02 (.02)	0.04 (.02)	0.00 (.01)
Income	−0.04 (.01)	−0.06 (.01)**	−0.03 (.01)
Trust propensity	0.06 (.03)*	0.03 (.02)	0.02 (.02)
Fact-checking tools use	0.19 (.03) ***	0.15 (.02)***	0.07 (.02) ***
Fake news self-efficacy	−0.03 (.04)	−0.04 (.03)	−0.01 (.02)
AI expertise—knowledge	0.01 (.05)	0.00 (.04)	0.00 (.03)
AI expertise—experience	0.10 (.05)*	0.07 (.04)*	0.04 (.03)
AI trust	0.40 (.05)***	0.12 (.04)***	0.07 (.03)**
Detector Characteristics			
Detector performance		0.31 (.04)***	0.06 (.04)
Detector collaborative capacity		0.14 (.03)***	0.06 (.02) **
Detector agency		0.22 (.03)***	0.01 (.02)
Detector complexity		−0.07 (.02)**	−0.03 (.02)
Trust in the Detector			0.67 (.04)***
N	*1,052*	*1,052*	*1,052*
Adjusted R^2	*.31*	*.58*	*.71*
R^2 change	*.31*	*.27*	*.13*
F	*F(9, 1042) = 53.97*	*F(13, 1038) = 111.0*	*F(14, 1037) = 185.7*

Note. All coefficients represent standardized regression coefficients. $*p < .05$, $**p < .001$, $***p < .001$.

Discussion

The goal of the current study was to explore the factors that might explain users' perceptions and usage intent of the leading explainable AI fake news detection technology. Adopting the notion that a fake news detection system with an interpretable nature developed through the use of AI technology has the potential to establish trust collaboratively. Against this backdrop, we explored user perceptions and trust on the result of a major ML fake news detector with a focus on both user and machine

characteristics. The findings indicated that users' trust levels in the ML detector were influenced by both individuals' inherent characteristics and their perceptions of the AI application. The results also showed that users' adoption intention was ultimately influenced by trust in the detector which explained a significant amount of the variance.

Specifically, we found that age played a role in explaining trust in the fake news detection application. Interestingly, other demographic variables like education and income were not associated with the trust on the detector (i.e., education levels and income only emerged as significant predictors after we controlled for the perceptions of the detector). It seems that the heuristic role of explainability or ML detection technology resonates better with younger users regardless of their socioeconomic backgrounds. In addition, individuals with more experience in fact-checking tools, more experience with AI technology, and higher levels of overall trust in AI tended to trust the fake news detector. It suggests that topical interest and experience (rather than knowledge), with a rooted trust in AI overall, are more relevant user characteristics in predicting trust in the specific AI application. Note that those who reported low levels of self-efficacy to detect fake news were also more likely to trust the application. Because self-efficacy refers to individuals' belief of their capability of performing a particular behavior (Bandura, 1977), it is possible that the participants saw the detector as a useful means to compensate for their perceived lower efficacy.

We also found that the variables measuring different facets of the fake news detector were all significantly associated with trust levels in the AI application. That is, trust levels were higher when users perceived the application to be highly competent at detecting fake news (performance), highly collaborative in terms of working with human users (collaboration), and have more power in working autonomously (agency). Also, users were more likely to trust the application, when the technology was perceived to have lower levels of complexity. The fact that these variables explained a significant portion of the variance in trust points to the importance of exposure to an explainable fake news detector with heuristic cues to facilitate user trust. The significance of perceived performance is consistent with previous studies (Hancock et al., 2011) that user perception of the technology's performance are the most influential factors in the trust building process. The agency of the detector in predicting trust also shows the role of autonomy users placed on the machine. Perhaps in a politically charged environment where partisans have different views of fake news (Calvillo, Garcia, Bertrand, & Mayers, 2021), such machine agency (and assumed neutrality) offers more potential for "trust."

Lastly, our findings revealed that trust significantly contributed to intention to adopt the application. Trust was the strongest predictor in determining adoption intention in the full model which included all previously examined variables. This result is consistent with the view which advocates the integration of the trust construct in explaining adoption behavior (Lee & See, 2004; Wu & Chen, 2005; Xie et al., 2017). In particular, trust has been seen as a central element in news consumption because the level of trust directly determines news consumers' behaviors such as paying for news and installing news apps (American Press Institute, 2016; Fletcher & Park, 2017). In a similar vein, our study shows that trust in the explainable ML fake news detector is critical for its adoption intention (Komiak & Benbasat, 2006; Tussyadiah et al., 2020).

Taken together, the findings point to the important role of the fundamental trust in AI technology and the acceptance of alternative means to detect fake news. Currently, there are mixed feelings about the

use of AI in fact-checking among news consumers as well as journalists. Although some hail such technologies as a breakthrough in combating misinformation, others express concerns over delegating the complicated task of truth validation to computers (Graves, 2018). However, given the speed and volume of misinformation spreading in online space, it is inevitable for fact-checkers, platforms, and consumers to, at least partially, rely on software that helps them identify false information (Abel, 2022). Thus, efforts to understand the resistance to the adoption of AI systems should accompany the development of explainable fake news detectors.

Trust is a focal concept in the human-automation interaction literature, because trust often determines users' willingness to interact with and rely on technologies like automation (Hoff & Bashir, 2015). In the emotion-laden context of fake news/misinformation (Ceci & Williams, 2020), it is especially vital to understand the human factors that play a role in the perception and usage of anti-fake news technology. There have been numerous studies on the development of fake news detection technology (Zhou & Zafarani, 2020), but little is known about how users perceive and trust these automated AI-based detection systems (Auernhammer, 2020). This lack of understanding is further complicated by the novelty of ML in the news sector and negative connotations associated with algorithmic decision making (Araujo et al., 2020). In this sense, this study contributes to the interpersonal trust literature (Lewicki et al., 2006; McAllister, 1995) that conceptualized trust as a multidimensional concept in the context of explainable fake news detectors. Thus, the findings may serve as a reference for future analysis that takes a holistic approach to designing human-centered AI.

This study has a number of limitations. First, the survey participants skewed younger with 84% of respondents under 40. This is perhaps because of the fact that MTurkers tend to be younger than the general U.S. population or professional panels (Chandler, Mueller, & Paolacci, 2014). Another limitation is related to the geographic location of the respondents. Because the sample is U.S. based, the findings cannot be generalized into other parts of the world where their familiarity of AI might be different. In addition, the study relied on one type of fake news detector (i.e., dEFENSE) to measure participants' perception of the application. Although this application is a typical example of fake news detectors that are currently available, future research could use more than one application to examine whether different features of an application influence trust levels and adoption intention. Lastly, this study did not include all possible factors that could relate to users' perception and adoption. For example, political ideology or attitude toward fake news may play a role. According to a Pew Research Center's survey (Mitchell et al., 2019), Republicans expressed far greater concern about fake news and perceived them as a bigger problem than Democrats. Future research could explore attitudinal factors associated with fake news.

Despite its limitations, this study offers insights into the antecedents of trust in fake news detection technology and adoption intention. The purpose of this study was to serve as a starting point for future research to expand the scope of current analysis. Our findings indicate that trust is a focal element in determining users' behavioral intentions. Such trust is largely influenced by individuals' inherent characteristics and their perceptions of the AI application. We argue that identifying positive heuristics of fake news detection technology is critical for facilitating the diffusion of AI-based detection systems.

References

Abel, G. (2022, June 28). *What is the future of automated fact-checking? Fact-checkers discuss*. Retrieved from https://www.poynter.org/fact-checking/2022/how-will-automated-fact-checking-work/

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. doi:10.1016/0749-5978(91)90020-T

Alharbi, R., Vu, M. N., & Thai, M. T. (2021). Evaluating fake news detection models from explainable machine learning perspectives. In *Proceedings of 2021-IEEE International Conference on Big Data* (pp. 705–714). Montreal, Canada: IEEE. doi:10.1109/ICC42927.2021.9500467

Ali, W., & Hassoun, M. (2019). Artificial intelligence and automated journalism: Contemporary challenges and new opportunities. *International Journal of Media, Journalism and Mass Communications, 5*(1), 40–49. Retrieved from https://www.arcjournals.org/pdfs/ijmjmc/v5-i1/4.pdf

American Press Institute. (2016, April). *A new understanding: What makes people trust an rely on news*. Retrieved from https://www.americanpressinstitute.org/publications/reports/survey-research/trust-news/single-page/

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society, 35*(3), 611–623. doi:10.1007/s00146-019-00931-w

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., . . . Herrera, F. (2019). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115.

Auernhammer, J. (2020, August 11–14). Human-centered AI: The role of human-centered design research in the development of AI. In S. Boess, M. Cheung, & R. Cain (Eds.), *Synergy—DRS International Conference 2020* [Virtual]. doi:10.21606/drs.2020.282

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. doi:10.1037/0033-295X.84.2.191

Beltramini, R. F. (1988). Perceived believability of warning label information presented in cigarette advertising. *Journal of Advertising*, *17*(2), 26–32.

Brandtzaeg, P. B., & Følstad, A. (2017). Trust and distrust in online fact-checking services. *Communications of the ACM*, *60*(9), 65–71. doi:10.1145/3122803

Calhoun, C. S., Bobko, P., Gallimore, J. J., & Lyons, J. B. (2019). Linking precursors of interpersonal trust to human-automation trust: An expanded typology and exploratory experiment. *Journal of Trust Research, 9*(1), 28–46. doi:10.1080/21515581.2019.1579730

Calvillo, D. P., Garcia, R. J. B., Bertrand, K., & Mayers, T. A. (2021). Personality factors and self-reported political news consumption predict susceptibility to political fake news. *Personality and Individual Differences, 174*, 110666. doi:10.1016/j.paid.2021.110666

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research, 56*(5), 809–825. doi:10.1177/0022243719851788

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature News*, *538*(7623), 20–23. doi:10.1038/538020a

Castro, D., & New, J. (2016). The promise of artificial intelligence. *Center for Data Innovation*, *115*(10), 32–35.

Chakraborti, T., & Kambhampati, S. (2018). *Algorithms for the greater good! On mental modeling and acceptable symbiosis in human-AI collaboration*. arXiv preprint arXiv:1801.09854. Retrieved from https://arxiv.org/pdf/1801.09854.pdf

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods, 46*(1), 112–130. doi:10.3758/s13428-013-0365-7

Chan-Olmsted, S. M. (2019). A review of artificial intelligence adoptions in the media industry. *International Journal on Media Management, 21*(3–4), 193–215. doi:10.1080/14241277.2019.1695619

Chan-Olmsted, S., & Qin, Y. S. (2022). The effect of news consumption on fake news efficacy. *Journal of Applied Journalism & Media Studies*, *11*(1), 61–79. https://doi.org/10.1386/ajms_00041_1

Ceci, S. J., & Williams, W. M. (2020, October 25). The psychology of fact-checking. *Scientific American*. Retrieved from https://www.scientificamerican.com/article/the-psychology-of-fact-checking1/

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. doi:10.2307/249008

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2016). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science, 64*(3), 1155–1170. doi:10.1287/mnsc.2016.2643

de-Lima-Santos, M. F., & Ceron, W. (2021). Artificial intelligence in news media: Current perceptions and future outlook. *Journalism and Media*, *3*(1), 13–26. https://doi.org/10.3390/journalmedia3010002

Emmert-Streib, F., Yli-Harja, O., & Dehmer, M. (2020). Explainable artificial intelligence and machine learning: A reality rooted perspective. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10*(6), e1368. doi:10.1002/widm.1368

Fletcher, R., & Park, S. (2017). The impact of trust in the news media on online news consumption and participation. *Digital Journalism, 5*(10), 1281–1299. doi:10.1080/21670811.2017.1279979

Frazier, M. L., Johnson, P. D., & Fainshmidt, S. (2013). Development and validation of a propensity to trust scale. *Journal of Trust Research*, *3*(2), 76–97. https://doi.org/10.1080/21515581.2013.820026

Gill, H., Boies, K., Finegan, J. E., & McNally, J. (2005). Antecedents of trust: Establishing a boundary condition for the relation between propensity to trust and intention to trust. *Journal of Business and Psychology*, *19*(3), 287–302. doi:10.1007/s10869-004-2229-8

Gillath, O., Ai, T., Branicky, M. S., Keshmiri, S., Davison, R. B., & Spaulding, R. (2021). Attachment and trust in artificial intelligence. *Computers in Human Behavior, 115*(52), 1066607. doi:10.1016/j.chb.2020.106607

Graves, L. (2018). *Understanding the promise and limits of automated fact-checking. Reuters Institute for the study of journalism factsheets*. Reuters Institute for the Study of Journalism. Retrieved from https://tinyurl.com/yb3f7969

Guo, B., Ding, Y., Yao, L., Liang, Y., & Yu, Z. (2020). The future of false information detection on social media: New perspectives and trends. *ACM Computing Surveys, 53*(4), 1–36. doi:10.1145/3393880

Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A human-machine communication research agenda. *New Media & Society, 22*(1), 70–86. doi:10.1177/1461444819858691

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors, 53*(5), 517–527. doi:10.1177/0018720811417254

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors, 57*(3), 407–434. doi:10.1177/0018720814547570

Karimi, H., & Tang, J. (2019). Learning hierarchical discourse-level structure for fake news detection. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 3432–3442). Minneapolis, MN: Association for Computational Linguistics.

Kim, J., Tabibian, B., Oh, A., Schoelkopf, B., & Gomez-Rodriguez, M. (2018). Leveraging the crowd to detect and reduce the spread of fake news and misinformation. In *Proceedings of the ACM International Conference on Web Search and Data Mining* (pp. 324–332). New York, NY: Association for Computing Machinery. doi:10.1145/3159652.3159734

Kim, Y. H., Kim, D. J., & Hwang, Y. (2009). Exploring online transaction self-efficacy in trust building in B2C e-commerce. *Journal of Organizational and End User Computing, 21*(1), 37–59. Retrieved from https://www.igi-global.com/article/journal-organizational-end-user-computing/3851

Komiak, S. Y. X., & Benbasat, I. (2006). The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS Quarterly, 30*(4), 941–960. doi:10.2307/25148760

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80. doi:10.1518/hfes.46.1.50_30392

Lewicki, R. J., Tomlinson, E. C., & Gillespie, N. (2006). Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management, 32*(6), 991–1022. doi:10.1177/0149206306294405

Lewis, S. C. (2019). Artificial intelligence and journalism. *Journalism & Mass Communication Quarterly, 96*(3), 673–675.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the Advances in Neural Information Processing Systems* (pp. 4765–4774). Long Beach, CA: NIPS. Retrieved from https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

Lyons, J. B., Wynne, K. T., Mahoney, S., & Roebke, M. A. (2019). Trust and human-machine teaming: A qualitative study. In W. Lawless, R. Mittu, D. Sofge, I. S. Moskowitz, & S. Russell (Eds.), *Artificial intelligence for the Internet of everything* (pp. 101–116). San Diego, CA: Academic Press. doi:10.1016/B978-0-12-817636-8.00006-5

Maynard, D. C., & Hakel, M. D. (1997). Effects of objective and subjective task complexity on performance. *Human Performance*, *10*(4), 303–330.

Matthews, G., Lin, J., Panganiban, A. R., & Long, M. D. (2020). Individual differences in trust in autonomous robots: Implications for transparency. *IEEE Transactions on Human-Machine Systems, 50*(3), 234–244. doi:10.1109/THMS.2019.2947592

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709–734. doi:10.5465/AMR.1995.9508080335

McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal, 38*(1), 24–59. doi:10.5465/256727

Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems, 2*(2), 1–25. doi:10.1145/1985347.1985353

Mcknight, D. H., & Chervany, N. L. (2001). What trust means in e-commerce customer relationships: An interdisciplinary conceptual typology. *International Journal of Electronic Commerce, 6*(2), 35–59. https://doi.org/10.1080/10864415.2001.11044235

Mcknight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *Academy of Management Review, 23*(3), 473–490. doi:10.5465/amr.1998.926622

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors, 50*(2), 194–210. doi:10.1518/001872008X288574

Metz, C. (2019, November 24). Internet companies prepare to fight the "deepfake" future. *The New York Times*. Retrieved from https://www.nytimes.com/2019/11/24/technology/tech-companies-deepfakes.html

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence, 267*, 1–38. doi:10.1016/j.artint.2018.07.007

Mitchell, A., Gottfried, J., Stocking, G., Walker, M., & S, Fedeli. (2019, June 5). *Many Americans say made-up news is a critical problem that needs to be fixed*. Retrieved from https://www.journalism.org/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/

Mohseni, S., Yang, F., Pentyala, S., Du, M., Liu, Y., Lupfer, N., . . . Ragan, E. (2020). *Machine learning explanations to prevent overtrust in fake news detection*. ArXiv:2007.12358 [Cs]. Retrieved from http://arxiv.org/abs/2007.12358

Mohseni, S., Zarei, N., & Ragan, E. D. (2020). *A multidisciplinary survey and framework for design and evaluation of explainable AI systems*. ArXiv E-Prints, arXiv:1811.11839v5 [cs.HC]. Retrieved from http://arxiv.org/abs/1811.11839

Molina, M. D., Sundar, S. S., Le, T., & Lee, D. (2021). "Fake news" is not simply false information: A concept explication and taxonomy of online content. *American Behavioral Scientist, 65*(2), 180–212. doi:10.1177/0002764219878224

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103. doi:10.1111/0022-4537.00153

Newman, N. (2022). *Overview and key findings of the 2022 digital news report*. Reuters Institute. Retrieved from https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022/dnr-executive-summary

Nguyen, A. T., Kharosekar, A., Krishnan, S., Krishnan, S., Tate, E., Wallace, B. C., & Lease, M. (2018). Believe it or not: Designing a human-AI partnership for mixed-initiative fact-checking. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (pp. 189–199). New York, NY: Association for Computing Machinery. doi:10.1145/3242587.3242666

Okamura, K., & Yamada, S. (2020). Adaptive trust calibration for human-AI collaboration. *PLoS One, 15*(2), 1–20. doi:10.1371/journal.pone.0229132

Ozbay, F. A., & Alatas, B. (2020). Fake news detection within online social media using supervised artificial intelligence algorithms. *Physica A: Statistical Mechanics and Its Applications, 540*, Article 123174. doi:10.1016/j.physa.2019.123174

Ramos, M. A., Thieme, C. A., Utne, I. B., & Mosleh, A. (2020). A generic approach to analysing failures in human—System interaction in autonomy. *Safety Science, 129*, Article 104808. doi:10.1016/j.ssci.2020.104808

Rogers, E. M. (1995). *Diffusion of innovation*. New York, NY: The Free Press.

Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elkins, A., . . . Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management, 57*(2), Article 103174. doi:10.1016/j.im.2019.103174

Sirdeshmukh, D., Singh, J., & Sabol, B. (2002). Consumer trust, value, and loyalty in relational exchanges. *Journal of Marketing*, *66*(1), 15–37.

Shin, D., Zhong, B., & Biocca, F. (2020). Beyond user experience: What constitutes algorithmic experiences. *International Journal of Information Management, 52*, 1–11. doi:10.1016/j.ijinfomgt.2019.102061

Shu, K., Cui, L., Wang, S., Lee, D., & Liu, H. (2019). dEFEND: Explainable fake news detection. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 395–405). New York, NY: Association for Computing Machinery. doi:10.1145/3292500.3330935

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2018). FakeNewsNet: A data repository with news content, social context and spatialtemporal information for studying fake news on social media. *Big Data, 8*(3), 171–188. doi:10.1089/big.2020.0062

Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. *Cutter Business Technology Journal, 31*(2), 47–53.

Silva, A., Han, Y., Luo, L., Karunasekera, S., & Leckie, C. (2021). Propagation2Vec: Embedding partial propagation networks for explainable fake news early detection. *Information Processing & Management*, *58*(5), 102618. https://doi.org/10.1016/j.ipm.2021.102618

Tandoc Jr., E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news": A typology of scholarly definitions. *Digital Journalism, 6*(2), 137–153. doi:10.1080/21670811.2017.1360143

Tandoc Jr., E. C., Duffy, A., Jones-Jang, S. M., & Pin, W. G. W. (2021). Poisoning the information well?: The impact of fake news on news media credibility. *Journal of Language and Politics*, *20*(5), 783–802.

Teo, T. (2011). Technology acceptance research in education. In T. Teo (Ed.), *Technology acceptance in education: Research and issues* (pp. 1–5). Rotterdam, The Netherlands: Sense.

Tussyadiah, I. P., Zach, F. J., & Wang, J. (2020). Do travelers trust intelligent service robots? *Annals of Tourism Research, 81*, Article 102886. doi:10.1016/j.annals.2020.102886

Wu, I.-L., & Chen, J.-L. (2005). An extension of trust and TAM model with TPB in the initial adoption of on-line tax: An empirical study. *International Journal of Human-Computer Studies, 62*(6), 784–808. doi:10.1016/j.ijhcs.2005.03.003

Wynne, K. T., & Lyons, J. B. (2018). An integrative model of autonomous agent teammate-likeness. *Theoretical Issues in Ergonomics Science, 19*(3), 353–374. doi:10.1080/1463922X.2016.1260181

Xie, Q., Song, W., Peng, X., & Shabbir, M. (2017). Predictors for e-government adoption: Integrating TAM, TPB, trust and perceived risk. *The Electronic Library, 35*(1), 2–20. doi:10.1108/EL-08-2015-0141

Yang, F., Pentyala, S. K., Mohseni, S., Du, M., Yuan, H., Linder, R., . . . Hu, X. (2019, May). Xfake: Explainable fake news detector with visualizations. In *The World Wide Web Conference* (pp. 3600–3604). New York, NY: Association for Computing Machinery. doi:10.1145/3308558.3314119

Zhou, T. (2012). Understanding users' initial trust in mobile banking: An elaboration likelihood perspective. *Computers in Human Behavior, 28*(4), 1518–1525. doi:10.1016/j.chb.2012.03.021

Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys, 53*(5), 1–40. doi:10.1145/3395046