Deepfake! A Liar's Dividend for Audiovisual Material

Lara Grohmann, Franziska A. Halle, and Markus Appel

Psychology of Communication and New Media, Human-Computer-Media Institute,

University of Würzburg, Germany

This manuscript was accepted for publication in *Psychology of Popular Media*. This version is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. The final article is available, upon publication, at: 10.1037/ppm0000665

Author Note

All authors of this manuscript declare that we have no conflict of interest.

This research received no funding.

The study's preregistration, materials, data, and code are freely available at https://osf.io/skzu4.

Corresponding author: Lara Grohmann, Psychology of Communication and New Media, University of Würzburg, Oswald-Külpe-Weg 82, 97074 Würzburg; +49-931-3183203; lara.grohmann@uni-wuerzburg.de

Abstract

Whereas a substantial amount of empirical research on misinformation exists, one phenomenon with increasing societal awareness has largely been disregarded: the liar's dividend. This concept describes the leveraging of uncertainty due to the existence of misinformation to one's own advantage. The aim of the current study was to clarify inconclusive prior evidence on the liar's dividend in the context of audiovisual material. In an online experiment, participants were presented with either false deepfake claims or an apology as ostensible responses by a politician to video evidence of one of two potentially scandalous incidents in the political realm. Participants judged the politician's leadership ability and the extent of potential technical manipulation of the video and indicated their general trust in media. Results show that false deepfake claims made by politicians do pay a liar's dividend in the form of higher perceived leadership abilities. This effect was mediated by higher misidentification of the video evidence as a deepfake. In addition, whereas false deepfake claims did not affect trust in media directly, deepfake perception was negatively correlated with general trust in media. False claims of misinformation appear to be effective in discrediting genuine audiovisual material, thereby posing a potential threat to democracy.
Keywords: liar's dividend; trust in media; political communication; misinformation; deepfakes

Public Policy Relevance Statement

Deep learning algorithms enable the creation of audiovisual materials that resemble actual audiovisual recordings, but the depicted incidents never actually happened (deepfakes). This development could benefit political actors who may falsely claim that unfavorable, genuine audiovisual content was created by AI (the liar's dividend). The results of our experiment show that discrediting genuine audiovisual evidence as fake news may indeed be a successful strategy for political actors, a strategy that healthy democracies need to be aware of.

Deepfake! A Liar's Dividend for Audiovisual Material

From a short video showing UK right wing politician Nigel Farage playing Minecraft to a clip of Ukraine's president Selenskyj calling for capitulation to Russia – deep learning algorithms enable the creation of content that resembles actual audiovisual recordings, but the depicted incidents never actually happened (Chesney & Citron, 2019; Farid, 2025). Whereas the use of so-called deepfakes generates new opportunities in areas such as entertainment or education, e.g., creating film scenes with deceased actors or presenting multilingual information in engaging ways, the deepfake technology at the same time entails the potential to harm individuals as well as societies (Chesney & Citron, 2019; Diakopoulos & Johnson, 2021; Farid, 2025; Gambín et al., 2024; Hancock & Bailenson, 2021). AI-generated audiovisual content may exacerbate the general problem of mis- and disinformation in modern societies (Lewandowsky et al., 2017): Citizens may acquire false information, and exposure to potentially deepfaked content might induce an increasing uncertainty in the truthfulness of information encountered and thereby reduce individuals' trust in news, media, and institutions (Vaccari & Chadwick, 2020). Importantly, Chesney and Citron (2019) pointed out another risk generated by the simple existence of deepfakes in the informational environment: the possibility of discrediting genuine audiovisual evidence as deepfake. In this way, individuals might take advantage of extant uncertainty to escape accountability of something they have said or done–a benefit which Chesney and Citron (2019) termed the *liar's dividend*.

Despite the broad discussion of the–hypothetical–liar's dividend in major news outlets (e.g., Carpenter, 2024; Chadwick, 2018; Edwards, 2024; Verma & De Vynck, 2024) empirical research on the effects of lying and attributing incriminating audiovisual evidence to deepfake fabrications is very rare. In fact, only one series of studies examined whether lying in this way actually comes with a dividend. Schiff and colleagues (2025) presented written texts or audiovisual material featuring one of four former politicians in which the politician made insensitive or embarrassing statements. Next, participants read an ostensible response by the politician or no response. Belief in the scandal, support of the politician, and trust in media served as the main dependent variables. The results differed whether text-based or

audiovisual evidence was provided for the scandalous incident: For textual evidence, politicians who claimed that a personal scandal was due to disinformation gained higher support compared to giving no response or apologizing. The authors further showed that the liar's dividend can be achieved both through induction of informational uncertainty by pointing to the widespread distribution of disinformation in general as well as through accusation of political opponents of deliberately defaming them. Concerning audiovisual evidence, however, the results were inconclusive: whereas politicians who claimed that the supposed scandal was due to disinformation by spreading informational uncertainty did not gain higher support compared to giving no response (in the experiments with audiovisual material, there was no apology condition), there were at least some indications for the effectiveness of the oppositional rallying strategy with video evidence.

Schiff and colleagues (2025) concluded that false claims of audiovisual deepfakes might be largely ineffective and not pay a liar's dividend as it might be harder to convince people that authentic video evidence is faked. This reasoning relies on a seeing-is-believing assumption in that individuals tend to readily believe audio-visual information as it resembles reality (Köbis et al., 2021; Sundar et al., 2021).

We assume that such a conclusion is premature. First, theory suggests that truth judgments may be influenced by a variety of heuristics (Brashier & Marsh, 2020; Levine, 2022). According to truth-default theory (Levine, 2014), individuals tend to accept incoming information as true by default unless they have reasons to suspect otherwise and to actively disbelief the information. This suspicion of deception can be triggered to varying degrees by different aspects of, for example, the information, the environment, and/or the source. Regarding the exposure to audiovisual material of political misconduct, nowadays, there might be several reasons for individuals not to believe what they see but to doubt its veracity. Over the last years, deepfake technology has been developing rapidly, rendering deepfakes more and more realistic and increasingly difficult to distinguish from genuine content (Appel & Prietzel, 2022). In addition, the public discussion regarding deepfakes around conflicts in the Middle East and the Russo-Ukrainian War and as part of the 2024 US presidential election (e.g., Edwards, 2024; Taylor, 2024) might have increased

individuals' accessibility of deepfake attributions. Indeed, recent research has shown that the awareness of the threat of misinformation may generalize distrust and reduce the perceived credibility of factually accurate information (see Luo et al., 2022; Van Der Meer et al., 2023). This trend towards 'truth skepticism' (Chesney & Citron, 2019) may thus lead individuals to increasingly doubt the veracity of audiovisual material. In turn, they may rely on a politician's false claims of misinformation as a heuristic cue for judgments of the accuracy of what used to be considered indisputable evidence (see Farid, 2025; Swire et al., 2017). This very possibility of political actors to leverage individuals' informational uncertainty in digital content is what constitutes the liar's dividend (Chesney & Citron, 2019). That is, under certain circumstances, political actors may increasingly benefit from the fact that the existence of audiovisual evidence no longer represents a substantial reason for individuals to move from their state of informational uncertainty to active disbelief of politicians' false deepfake claims (Levine, 2022).

Second, on an empirical level, half of the videos in Schiff et al.'s (2025) studies were rated as somewhat familiar to respondents. A vast amount of research has shown that repeated exposure to information may increase its perceived accuracy–especially compared with new information (Dechêne et al., 2010). This may have decreased the influence of politicians' rebuttals on individuals' veracity judgments in Schiff et al.'s (2025) studies, as compared to the familiar audiovisual material.

The goal of the current study was to re-examine the liar's dividend for audiovisual material. We presented genuine video evidence associated with a political scandal concerning drunkenness at work that may challenge the perceived leadership abilities of the respective politicians, and we manipulated the supposed responses of the politicians. We tested the liar's dividend hypothesis (Chesney & Citron, 2019) by contrasting a denial of the audiovisual evidence through a deepfake lie to an apology as acknowledgment of the video evidence. We assumed that a deepfake attribution response does pay a liar's dividend in terms of perceived leadership ability as compared to an apology (Hypothesis 1). Furthermore, to shed some light on the underlying process through which deepfake claims might result in a liar's dividend, we investigated whether deepfake claims increase the likelihood that video evidence is actually misidentified as a deepfake

(versus the apology condition, Hypothesis 2). In addition to the preregistered hypotheses, we investigated further whether misidentification of the video evidence as a deepfake mediates the effect of deepfake claims on perceived leadership abilities (Hypothesis 3). Moreover, as extant findings on impacts of disinformation on trust in the media are inconclusive, but suggest that increased uncertainty about the authenticity of information may have an impact on trust in media (see Schiff et al., 2025; Vaccari & Chadwick, 2020), we examined the hypothesis that false deepfake attributions relative to apologizing lead to decreased trust in media (Hypothesis 4) and we hypothesized that individuals' trust in media is negatively associated with misidentifying video evidence as deepfake (Hypothesis 5).

Materials and Methods

The hypotheses, methodology, exclusion criteria, sample size, and analyses were preregistered. The preregistration, data, codes, and supplementary material are freely available on the OSF (https://osf.io/skzu4; Grohmann et al., 2025).

Experimental Design and Procedure

As our focal stimuli we chose videos of two politicians with a similar predicament to increase insight into the generalizability of our findings. The experiment was implemented online via Qualtrics at the beginning of 2024 and was based on a 2x2 factorial design with the between-subject factors *response* (deepfake claim vs. apology) and *politician* (Politician A vs. Politician B). Participants were randomly assigned to one of the four experimental conditions. Following electronic informed consent, participants were presented with a short video clip in English showing actual footage of either a speech of Politician A, US politician Dade Phelan, republican speaker of the Texas House of Representatives, or of Politician B, US politician Robin Comey, democratic state representative in Connecticut. Both politicians spoke with a slur and made pauses in speech. Participants were instructed to imagine they would encounter the video on social media together with the claim that the politician was allegedly drunk during the speech. Using US political incidences with a German sample conveyed the advantage of reduced prior familiarity with the events and politicians.

After the video was shown, participants read an ostensible statement by the portrayed politician that included either the claim that the video was a deepfake using the informational uncertainty strategy in the experimental condition (Schiff et al., 2025) or an apology in the control condition (see online supplement S3 for the exact wordings). Subsequently, participants indicated their perceptions of the depicted politician's leadership ability as well as of potential technical manipulation of the video, and their trust in media. After indicating their demographics, potential technical problems and familiarity with the stimulus material, participants were thoroughly debriefed.

Dependent Measures

Perceived leadership ability of the portrayed politician was assessed with the help of five bipolar trait items (incompetent – competent, weak-minded – decisive, unreliable – reliable, not capable of strong leadership – capable of strong leadership, and ineffective – effective) on a seven-point response scale (1 to 7) (Appel & Prietzel, 2022), Cronbach's α = .94. *Trust in media* was measured as agreement to two statements ("I trust the media" and "I believe that the media reports the news fairly") taken from Schiff et al. (2025). Responses were given on a five-point Likert scale ranging from 1 = *strongly disagree* to 5 = *strongly agree* (Spearman-Brown coefficient = .74). *Deepfake perception* was assessed with a single item ("How do you judge this video?") adopted from prior research (Appel & Prietzel, 2022) on a seven-point scale from 1 = *not technically manipulated* over 4 = *undecided* to 7 = *technically manipulated*. To ensure participants' uniform comprehension of what constitutes deepfakes, a general definition preceded the measurement. Additional measures addressed the familiarity with the presented video, potential technical problems, as well as sociodemographic data such as gender, age, educational level, and current occupation.

Participants

A minimum sample size was estimated using G*Power (Faul et al., 2009). Pursuing a test power of 80% and a type I error probability of α = .05, a minimum of 200 participants were required to detect a medium effect of *d* = .40 (two groups, between-subjects design). Participants were recruited via German-

language social media and the platform survey circle and were rewarded with the opportunity to participate in a monetary raffle. In total, 226 participants completed the study. After excluding careless responders (see online supplement S2), the final sample consisted of 182 participants[1] (66.5% female) with a mean age of 29.79 years (*SD* = 11.30, range from 18 to 72). The majority of participants (62.6%) were students at the time of study completion.

Ethics Statement

In Germany, it is not required to obtain institutional ethics approval for psychological research as long as it does not concern issues regulated by law. All reported research was carried out in full accordance with the Declaration of Helsinki. Participants were adults and provided electronic informed consent.

Results

Descriptive statistics of the main study variables are presented in Figure 1. We conducted analyses of variance (ANOVAs) in order to investigate whether reading a statement including a deepfake claim versus an apology after video evidence leads to higher perceptions of the politician's leadership abilities (H1), to a higher tendency to misidentify the authentic video as a deepfake (H2), and to decreased trust in media (H4)[2]. In support of H1, there was a significant main effect of the politician's response on perceived leadership ability, $F(1,178) = 50.01$, $p < .001$, $\eta_p^2 = .22$, in that reading a deepfake claim resulted in significantly higher scores of the portrayed politician's leadership abilities (Politician A: *M* = 4.39, *SD* = 1.56 ; Politician B: *M* = 4.07, *SD* = 1.27) compared to reading an apology (Politician A: *M* = 2.97, *SD* = 0.90 ; Politician B: *M* = 3.01, *SD* = 0.93). There was neither a significant main effect of the politician,

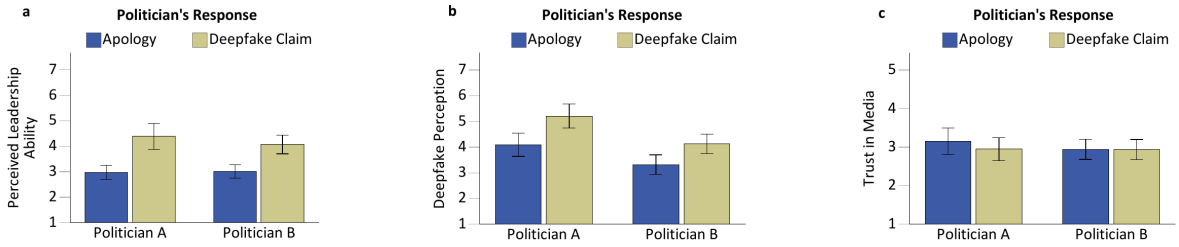
[1] Due to careless responding, the final sample size was lower than anticipated. A sensitivity analysis indicated that a sample size of *N* = 182 provided sufficient power to detect a minimum effect size of *d* = .42. The a priori sample size analysis deviates from the pre-registration (in which 171 participants were pre-registered, see supplement S1).

[2] Highly similar results are obtained when the preregistered t-tests are conducted (see online supplement S1 for details).

$F(1,178) = 0.64$, $p = .426$, nor a significant interaction, $F(1,178) = 1.04$, $p = .308$. Thus, we found a liar's dividend for audiovisual footage.

Figure 1

Graphical Depiction of the Main Results



Note. $N = 182$. a) Perceived leadership ability (scale range 1-7). b) Deepfake perception (scale range 1-7). c) Trust in media (scale range 1-5).

Supporting H2, a significant main effect of the response on recipients' deepfake perceptions emerged, $F(1,178) = 21.36$, $p < .001$, $\eta_p^2 = .11$, in that participants had a significantly higher tendency to misidentify the authentic video as a deepfake when they had read the deepfake claim (Politician A: $M = 5.21$, $SD = 1.42$; Politician B: $M = 4.12$, $SD = 1.32$) than when they had read the apology (Politician A: $M = 4.09$, $SD = 1.46$; Politician B: $M = 3.32$, $SD = 1.38$). No significant interaction was observed, $F(1,178) = 0.57$, $p = .452$, but there was a significant main effect of the politician on deepfake perception, $F(1,178) = 19.96$, $p < .001$, $\eta_p^2 = .10$, in that participants rather tended to misidentify the video of Politician A as a deepfake compared to the video of Politician B.

Contrary to H4, however, there was no main effect of the politician's response on trust in media, $F(1,178) = 0.52$, $p = .470$. Moreover, we observed neither a significant main effect of the politician, $F(1,178) = 0.56$, $p = .453$, nor a significant interaction, $F(1,178) = 0.47$, $p = .495$. That is, there was no significant difference in trust in media between participants who had read a deepfake claim (Politician A:

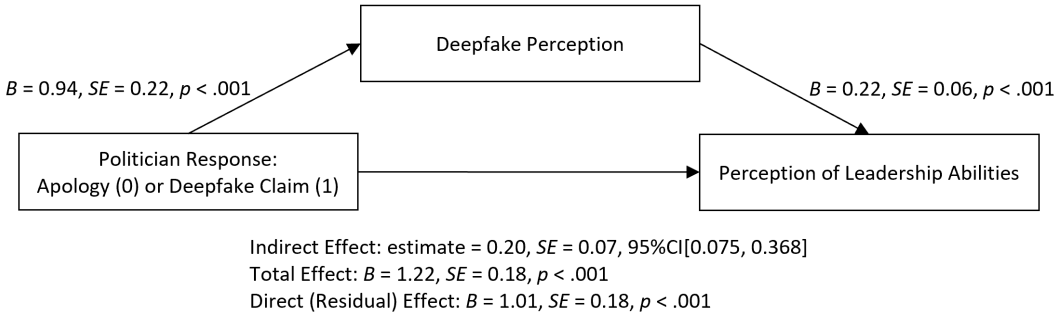
M = 2.95, *SD* = 0.91; Politician B: *M* = 2.94, *SD* = 0.90) versus an apology (Politician A: *M* = 3.15, *SD* = 1.11; Politician B: *M* = 2.94, *SD* = 0.95).

Yet, corroborating H5, a significant negative relationship between misidentification of the video evidence as a deepfake and trust in media was observed, $r = -.20$, $p = .006$. That is, the more participants misidentified the evidence as fake, the less they trusted the media in general.

Moreover, an exploratory mediation analysis was run using PROCESS v.4.3.1, model 4 (Hayes, 2022) to examine the mechanism through which deepfake claims might result in a liar's dividend (H3). Assumptions of normally distributed residuals and homoscedasticity were accounted for by employing percentile bootstrapping with 5000 samples to compute 95% confidence intervals for inferential statistics and regression coefficients. Results revealed a significant mediation through deepfake perceptions (see Figure 2). Reading deepfake claims resulted in higher deepfake perceptions than reading an apology, $B = 0.94$, $SE = 0.22$, $p < .001$. Higher deepfake perceptions, in turn, were associated with positive evaluations of the politician's leadership abilities, $B = 0.22$, $SE = 0.06$, $p < .001$. The indirect effect was significant, effect estimate = 0.20, $SE = 0.07$, 95%CI[0.075, 0.368], with a partially standardized indirect effect of $ab_{ps} = 0.16$, 95%-CI[0.059, 0.275]. The total effect amounted to $B = 1.22$, $SE = 0.18$, $p < .001$. The direct (residual) effect remained significant, $B = 1.01$, $SE = 0.18$, $p < .001$. That is, those who read a deepfake claim instead of an apology were, on average, 0.16 standard deviations higher in their perception of the politician's leadership abilities as a result of the indirect effect through misidentifying the video evidence as a deepfake.

Figure 2

Deepfake Perception Mediates the Effect of Politician Response on Perception of Leadership Abilities



Discussion

The proliferation of mis- and disinformation is considered one of the most severe short-term global risks by undermining truth and amplifying societal polarization (e.g., World Economic Forum, 2024). While there has been extensive research on the occurrence and impact of and counteraction against misinformation (e.g., Aïmeur et al., 2023; Ecker et al., 2022; Kozyreva et al., 2024; Pennycook & Rand, 2021) and deepfakes, specifically (e.g., Appel & Prietzel, 2022; Dobber et al., 2021; Hameleers et al., 2024; Lee & Shin, 2022; Somoray & Miller, 2023), the other side of the coin–falsely attributing true incidents to be misinformation–has largely been overlooked. However, as the weaponization of the 'fake news' claim already represents a global problem and risks to be leveraged for legitimization of censorship by authoritarian governments (World Economic Forum, 2024), there is an urgent need for more empirical evidence on the impacts of false claims of misinformation.

The current findings provide clear evidence that false claims of misinformation compared to apologetic acknowledgment pay a liar's dividend with regard to audiovisual material of a political scandal. The liar's dividend was demonstrated in our experiment in the form of higher perceived leadership abilities. Moreover, compared to apologizing, falsely claiming video evidence to be a deepfake led participants to follow this explanation and to increasingly perceive the genuine video content as a deepfake. Additional

analyses show that the latter variable served as a mediator and explained the effect on perceived leadership abilities.

Our findings suggest that individuals' judgments about the truth status of audiovisual material is more malleable than prior research suggests (Schiff et al., 2025; Sundar et al., 2021). Instead of believing the audiovisual evidence exclusively, individuals use politicians' misinformation attributions for their judgments of veracity of audiovisual material. From the perspective of truth-default theory (Levine, 2014; 2022), we conclude that the default to truth applies to the politicians' statements, the audiovisual material was no reason for individuals to suspect deception and render the politicians' deepfake claims futile. This ultimately yielded a liar's dividend for the politicians (Chesney & Citron, 2019). Whether and how varying degrees of truth skepticism on the side of participants moderate these mechanisms underlying the liar's dividend remains a valuable future research avenue.

Regarding the concerns around the impact of misinformation on trust in media, the current findings are in line with previous research (Schiff et al., 2025) in that the politician's statement did not significantly influence individuals' general trust in media. That is, whereas politicians' false claims of misinformation compared to apologies may affect individuals' perception of video evidence at hand, such deepfake claims may not directly erode individuals' general trust in media news' veracity. However, the more individuals perceived the video evidence to be technically manipulated, the less trust in media news they reported. Importantly, on average, individuals were rather undecisive in their judgment whether the video is a deepfake. In this way, the current findings are in line with previous research (Vaccari & Chadwick, 2020) in that it is perhaps not exposure to misinformation directly influencing media trust, but that misinformation such as deepfake claims may induce in individuals an uncertainty about the veracity of a story or about the authenticity of audiovisual material which in turn may negatively affect individuals' general trust in media news. Future research is warranted to disentangle the specific mechanisms through which misinformation as well as claims of misinformation may impact trust in media and news in general, but also in institutions and individuals disseminating that misinformation.

Limitations of the current study need to be noted that point at promising paths for future research. First, our experiment consisted of two conditions, a deepfake attribution and an apology. The latter condition was chosen because it reflected a response that–prior to the advent of deepfakes–was customary whenever audiovisual evidence for a misdeed was present. When audiovisual proof was presented, it was no feasible option to deny that an incident occurred (Chesney & Citron, 2019). Alternatively, or in addition to the apology condition, researchers may wish to compare the deepfake attribution to a condition in which the politician (or any other culprit) who is faced with incriminating audiovisual evidence does not respond to the accusation at all.

Second, a global measure of perceived leadership abilities served as our main dependent variable (see also Appel & Prietzel, 2022). Related measures on the attitude toward a targeted politician are commonly used in the literature on deepfakes effects (e.g., Dan, 2025; Dobber et al., 2021) and the liar's dividend (Schiff et al., 2025). It is intriguing to consider a more nuanced approach, as false deepfake attributions may have diverging effects on different dimensions, such as the dimensions of competency, benevolence, and integrity as components of trustworthiness (e.g., Mayer et al., 1995). Thus, we encourage future research to assess different evaluative dimensions when examining the liar's dividend.

Third, our results are limited to a single exposure to deepfake claims concerning one specific type of political scandal, i.e., politicians' alcohol abuse in parliament. Future research is encouraged to investigate claims of misinformation and their payout in the context of other types of political scandals. Importantly, an open question remains whether repeated claims of misinformation result in an increasing liar's dividend or whether there exists a turning point at which credibility of such claims wears off and the liar's outcome turns negative. Additionally, the study focused on scandals of active US politicians with a German sample that neither was familiar with the politicians or the scandal nor were the participants impacted by the politicians' work or able to vote for them. While this approach ensured individuals' impartiality towards the politicians' leadership abilities, questions of how claims of misinformation affect

the politician's actual potential electorate and whether partisanship plays a moderating role in this remain subject to future research.

Taken together, the current findings represent a starting point in understanding the potential consequences of a *deep doubt era* (Edwards, 2024). Whereas our results show strong indications that discrediting not just textual but also genuine audiovisual evidence as fake news may be a successful strategy for political actors, future research has yet to investigate in detail under which conditions the liar's dividend is likely to emerge and through which underlying mechanisms this strategy is effective (in addition to inducing deepfake perceptions). Understanding the ways in which individuals' uncertainty about the veracity of the wealth of information online may be exploited in times in which highly realistic deepfakes are possible, may also help to design effective countermeasures that dismantle the mechanisms behind false claims of misinformation.

References

Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: A review. *Social Network Analysis and Mining*, *13*(1), 30. https://doi.org/10.1007/s13278-023-01028-5

Appel, M., & Prietzel, F. (2022). The detection of political deepfakes. *Journal of Computer-Mediated Communication*, *27*(4), zmac008. https://doi.org/10.1093/jcmc/zmac008

Brashier, N. M., & Marsh, E. J. (2020). Judging truth. *Annual Review of Psychology*, *71*(1), 499–515. https://doi.org/10.1146/annurev-psych-010419-050807

Carpenter, P. (2024, February 10). The liar's dividend: How AI is reshaping truth in business communications. *Forbes*. https://www.forbes.com/councils/forbesbusinesscouncil/2024/10/02/the-liars-dividend-how-ai-is-reshaping-truth-in-business-communications/

Chadwick, P. (2018, July 22). The liar's dividend, and other challenges of deep-fake news. *The Guardian*. https://www.theguardian.com/commentisfree/2018/jul/22/deep-fake-news-donald-trump-vladimir-putin

Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy. *California Law Review*, *107*(6), 1753–1819. https://doi.org/10.15779/Z38RV0D15J

Dan, V. (2025). Deepfakes as a democratic threat: Experimental evidence shows noxious effects that are reducible through journalistic fact checks. *The International Journal of Press/Politics*, 19401612251317766. https://doi.org/10.1177/19401612251317766

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, *14*(2), 238–257. https://doi.org/10.1177/1088868309352251

Diakopoulos, N., & Johnson, D. (2021). Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New Media & Society*, *23*(7), 2072–2098. https://doi.org/10.1177/1461444820925811

Dobber, T., Metoui, N., Trilling, D., Helberger, N., & De Vreese, C. (2021). Do (microtargeted) deepfakes have real effects on political attitudes? *The International Journal of Press/Politics*, *26*(1), 69–91. https://doi.org/10.1177/1940161220944364

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, *1*(1), 13–29. https://doi.org/10.1038/s44159-021-00006-y

Edwards, B. (2024, December 28). Welcome to the era of "deep doubt." *Wired*. https://www.wired.com/story/deepfakes-deep-doubt-era-artificial-intelligence/

Farid, H. (2025). Mitigating the harms of manipulated media: Confronting deepfakes and digital deception. *PNAS Nexus*, *4*(7), pgaf194. https://doi.org/10.1093/pnasnexus/pgaf194

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Gambín, Á. F., Yazidi, A., Vasilakos, A., Haugerud, H., & Djenouri, Y. (2024). Deepfakes: Current and future trends. *Artificial Intelligence Review*, *57*(3), 64. https://doi.org/10.1007/s10462-023-10679-x

Grohmann, L., Halle, F. A., & Appel, M. (2025, March 27). *Deepfake! A liar's dividend for audiovisual material*. https://osf.io/skzu4/

Hameleers, M., Van Der Meer, T. G. L. A., & Dobber, T. (2024). Distorting the truth versus blatant lies: The effects of different degrees of deception in domestic and foreign political deepfakes. *Computers in Human Behavior*, *152*, 108096. https://doi.org/10.1016/j.chb.2023.108096

Hancock, J. T., & Bailenson, J. N. (2021). The social impact of deepfakes. *Cyberpsychology, Behavior, and Social Networking*, *24*(3), 149–152. https://doi.org/10.1089/cyber.2021.29208.jth

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (Third edition). The Guilford Press.

Köbis, N. C., Doležalová, B., & Soraperra, I. (2021). Fooled twice: People cannot detect deepfakes but think they can. *iScience*, *24*(11), 103364. https://doi.org/10.1016/j.isci.2021.103364

Kozyreva, A., Lorenz-Spreen, P., Herzog, S. M., Ecker, U. K. H., Lewandowsky, S., Hertwig, R., Ali, A., Bak-Coleman, J., Barzilai, S., Basol, M., Berinsky, A. J., Betsch, C., Cook, J., Fazio, L. K., Geers, M., Guess, A. M., Huang, H., Larreguy, H., Maertens, R., … Wineburg, S. (2024). Toolbox of individual-level interventions against online misinformation. *Nature Human Behaviour*, *8*(6), 1044–1052. https://doi.org/10.1038/s41562-024-01881-0

Lee, J., & Shin, S. Y. (2022). Something that they never said: Multimodal disinformation and source vividness in understanding the power of AI-enabled deepfake news. *Media Psychology*, *25*(4), 531–546. https://doi.org/10.1080/15213269.2021.2007489

Levine, T. R. (2014). Truth-default theory (TDT): A theory of human deception and deception detection. *Journal of Language and Social Psychology*, *33*(4), 378–392. https://doi.org/10.1177/0261927X14535916

Levine, T. R. (2022). Truth-default theory and the psychology of lying and deception detection. *Current Opinion in Psychology*, *47*, 101380. https://doi.org/10.1016/j.copsyc.2022.101380

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, *6*(4), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008

Luo, M., Hancock, J. T., & Markowitz, D. M. (2022). Credibility perceptions and detection accuracy of fake news headlines on social media: Effects of truth-bias and endorsement cues. *Communication Research*, *49*(2), 171–195. https://doi.org/10.1177/0093650220921321

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709. https://doi.org/10.2307/258792

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, *25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Schiff, K. J., Schiff, D. S., & Bueno, N. S. (2025). The liar's dividend: Can politicians claim misinformation to evade accountability? *American Political Science Review*, *119*(1), 71–90. https://doi.org/10.1017/S0003055423001454

Somoray, K., & Miller, D. J. (2023). Providing detection strategies to improve human detection of deepfakes: An experimental study. *Computers in Human Behavior*, *149*, 107917. https://doi.org/10.1016/j.chb.2023.107917

Sundar, S. S., Molina, M. D., & Cho, E. (2021). Seeing is believing: Is video modality more powerful in spreading fake news via online messaging apps? *Journal of Computer-Mediated Communication*, *26*(6), 301–319. https://doi.org/10.1093/jcmc/zmab010

Swire, B., Berinsky, A. J., Lewandowsky, S., & Ecker, U. K. H. (2017). Processing political misinformation: Comprehending the Trump phenomenon. *Royal Society Open Science*, *4*(3), 160802. https://doi.org/10.1098/rsos.160802

Taylor, M. (2024, March 10). An AI deepfake could be this election's November surprise. *Time*. https://time.com/7033256/ai-deepfakes-us-election-essay/

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, *6*(1), 1–13. https://doi.org/10.1177/2056305120903408

Van Der Meer, T. G. L. A., Hameleers, M., & Ohme, J. (2023). Can fighting misinformation have a negative spillover effect? How warnings for the threat of misinformation can decrease general news credibility. *Journalism Studies*, *24*(6), 803–823. https://doi.org/10.1080/1461670X.2023.2187652

Verma, P., & De Vynck, G. (2024, January 22). AI is destabilizing "the concept of truth itself" in 2024 election. *The Washington Post*. https://www.washingtonpost.com/technology/2024/01/22/ai-deepfake-elections-politicians/

World Economic Forum. (2024). *The global risks report*. https://www.weforum.org/publications/global-risks-report-2024/