UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Unmasking political deception: Investigating the Discernment and Emotional Impact of Deepfake Political Speeches Featuring American Presidential Candidates

Permalink
https://escholarship.org/uc/item/3zc6c210

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 47(0)

Authors
Solomon, Eliza
Madsen, Jens Koed
Zulato, Edoardo

Publication Date
2025

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Unmasking political deception: Investigating the Discernment and Emotional Impact of Deepfake Political Speeches Featuring American Presidential Candidates

Eliza Solomon[1] (e.solomon1@lse.ac.uk) **Jens Koed Madsen**[1] (j.madsen2@lse.ac.uk) **Edoardo Zulato**[1] (e.zulato@lse.ac.uk)

[1]Department of Psychological and Behavioural Science, London School of Economics and Political Science, Houghton Street, WC2A 2AE, London, UK.

Abstract

Deepfake videos challenge the quality of information in deliberative democracies. In a mixed-methods study, we examine the role of emotions in the detection of political deepfakes by focusing on trust, empathy, and inspiration to assess how deepfakes influence public perception and engagement with political communication. The research unfolds in two phases: an initial qualitative investigation through 3 focus groups (N = 13), followed by a quantitative survey (N = 261) where focus group insights inform the design and interpretation of the quantitative study. Participants were exposed to real, ChatGPT-generated, and historical speeches presented in modern contexts to gauge perceived authenticity and emotional responses, including trust, empathy, and inspiration. Results indicate no significant difference in perceived authenticity between real and deepfake content, with both eliciting comparable emotional reactions. The quantitative analysis reveals a marginal negative correlation between exposure to deepfakes and trust in political communication. Qualitative findings emphasize the influence of contextual cues and pre-existing biases, showing participants often prioritized emotional resonance over technical accuracy when evaluating content. The study highlights the intricate relationship between AI-generated media and public perception, underscoring the necessity for nuanced regulatory policies and improved media literacy to mitigate the impact of Deepfakes on public trust.

Keywords: Deepfakes; Political Communication; Emotional Impact; Trust; Discerning Authenticity; Mixed-Methods

Introduction

Deepfakes present unprecedented challenges to democratic discourse, as AI can create highly realistic synthetic content that is difficult to distinguish from authentic one (Chesney & Citron, 2019). This may include fabricated videos, audio, or images, portraying political figures making statements aimed to sway voter behavior (Hameleers & Marquart, 2023). The rise of Deepfakes denote authenticity and trust challenges to political discourse. Prior research examined technical aspects of deepfake detection and its role in misinformation (Vaccari & Chadwick, 2020, Ahmed & Chua, 2023). A critical gap exists in exploring how deepfake political messages are interpreted and processed emotionally. This is important as emotional engagement influences voter behavior and decisions (Brader, 2006).

Recent elections shifted the political setting, as candidates increasingly used digital media to engage voters. (Larsson, 2023). This exploratory study focuses on the role of emotions in the detection of political deepfakes. The qualitative Study 1 motivates the design of the main quantitative study 2.

Theoretical framework

Chesney and Citron (2019) highlight that deepfakes can convincingly depict individuals making statements they never uttered, while eliciting emotions such as trust, empathy, or anger. The emotional impact may be amplified by their capacity to manipulate contextual cues that traditionally signal authenticity, such as the fuzziness of the video or audio sync up (López-Gil et al., 2022; Onisha et al., 2024). In turn, emotional impact may outweigh factual accuracy in shaping public opinion (Brader, 2006; Nabi et al., 2008). Thus, deepfakes can elicit some emotional responses, reinforcing pre-existing beliefs and biases (Hameleers et al., 2022). Yet, it is not known if the effect is found across different emotions related to political persuasion. Three emotional responses are central to political messages: trust, empathy, and inspiration.

Trust is essential for political legitimacy and shapes how people use information from political figures (Tschannen-Moran & Hoy, 2000). It is essential in persuasion, as shown by the Elaboration-Likelihood Model (Petty & Cacioppo, 1984), Bayesian argumentation (Hahn et al., 2009), and discriminatory reasoning (Madsen, 2019). Trust in political communication involves both assessments of credibility and emotional connections with the messenger (Marcus et al., 2000). In this study, we operationalize trust as the belief that political actors act in the public's best interests, thus shaping individuals' willingness to engage with and accept political information (Hahn et al., 2009; Madsen, 2019).

Empathy influences political engagement, decision-making, and voting behavior (Renstrom & Ottati, 2020). In this study, we operationalize empathy as the ability to recognize and share others' emotions and experiences, thus enabling connections between political leaders and constituents (Muradova & Arceneaux, 2022).

Inspiration is a central, motivational drive to mobilize political action (Thrash & Elliot, 2003). In this study, we operationalize inspiration as the process through which messages evoke a sense of upliftment and motivation in individuals, prompting them to engage in civic actions or support political causes (Thrash & Elliot, 2003).

In all, studies show that perceived trust, empathy toward the speaker, and inspiration influence political beliefs and decision-making. This approach aligns with Marcus et al.'s (2000), emphasizing the role of emotions in political information processing and decision-making. Likewise, Vaccari and Chadwick (2020) argue that the emotional resonance of deepfakes may be their most significant feature

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

in terms of political influence. However, no research explores if AI-generated content evoke same emotional responses as authentic speeches. Thus, this study investigates if deepfakes elicit emotional responses like genuine speeches and how awareness of content authenticity moderates these responses.

Current literature on deepfakes primarily addresses three areas: detection (Kumari et al., 2021), misinformation effects (Hameleers et al., 2024), and policy implications (de Ruiter, 2021). Detection research focuses on developing algorithms to identify deepfake content. Kumari et al. (2021) describe machine-learning approaches that analyze visual and auditory discrepancies to discern manipulated media. These methods are crucial for maintaining information integrity. However, as detection technologies improve, so do the techniques for creating more convincing deepfakes, resulting in an ongoing arms race between creators and detectors.

On misinformation, Hameleers et al. (2024) examine how deepfake videos shape public perceptions of political figures, often leading to misinformed opinions based on fabricated content. Their findings suggest that deepfakes erode trust in media and political institutions, increasing polarization and misinformation spread. This underscores deepfakes' potential to effect electoral outcomes by distorting public perception through deceptive emotional appeals. It is pivotal to develop new regulatory frameworks and comprehensive strategies to prevent misinformation via deepfakes use (de Ruiter, 2021). Despite research on in technical detection, misinformation effects, and policy discussions, a significant gap remains concerning the emotional impact of political deepfakes.

Only few studies study psychological aspects of deepfakes (e.g., Hameleers et al., 2022). Thus, we take an exploratory approach with a mixed-methods approach, combining focus groups in Study 1 (N = 13) and quantitative surveys in Study 2 (N = 261) to examine responses to real and deepfake speeches attributed to the 2024 U.S. Presidential candidates. We investigate if deepfake political content elicits emotional responses comparable to real speeches and how awareness of content authenticity moderates these responses. From the literature review, we posit the following hypotheses:

H1: Deepfake political speeches yield similar levels of perceived trust compared to real speeches.

H2: Deepfake political speeches elicit comparable levels of empathy compared with real speeches.

H3: Deepfake political speeches elicit comparable levels of inspiration compared with real speeches.

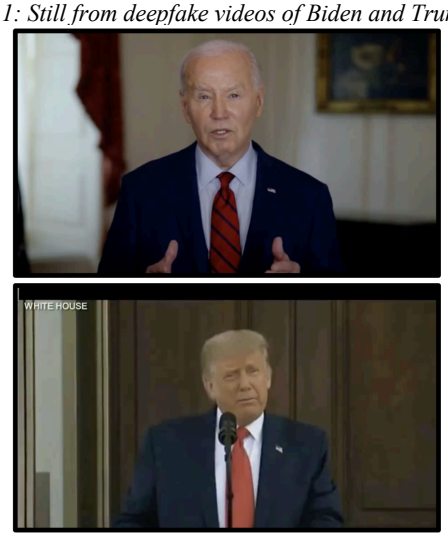
Study design

A mix-method line combines qualitative focus groups and a quantitative survey to investigate how deepfake, historical, and authentic speeches impact emotional responses. Focus groups allow for in-depth explorations of participants' emotional reactions to deepfake content and their authenticity interpretations. Focus groups enable participants to articulate their feelings, thoughts, and contextual understandings, which quantitative measures alone may overlook (Caillaud & Flick, 201; Creswell & Plano Clark, 2017). The qualitative responses shape the quantitative study. Thus, our quantitative survey allowed us to measure deepfake perceptions across a larger population (N=261) and to test the relation between deepfakes trust, empathy, and inspiration.

Generating the deepfake stimuli

Six political videos were divided into three categories: real campaign speeches, deepfakes of historical speeches, and deepfakes AI-generated speeches. Fabricated content was created with Eleven Labs for audio synthesis and synclabs.so for video manipulation. 'Real' stimuli was publicly available campaign speeches from Joe Biden and Donald Trump from the 2024 presidential campaigns, serving as control. These were edited to one-minute segments. In 'historical' deepfakes, Biden delivers Obama's "Yes We Can" speech and Trump gives Lincoln's "Gettysburg Address.", with synthetic versions of Biden and Trump delivering the speeches. This tests participants' ability to detect deepfakes of historically significant, emotionally resonant political rhetoric. Finally, we used ChatGPT to write original political speeches delivered by deepfake versions of Biden and Trump. The speeches were crafted to evoke inspiration, empathy, and trust, allowing us to examine how participants respond to AI-generated political rhetoric engineered for emotional impact. The deceptive nature of the study design was essential for ecological validity, as awareness of potential deepfakes could have biased participants' natural discernment processes. In all, this leaves six videos: real, historical, and AI-generated speeches delivered by Biden or Trump (Fig. 1 visualizes two of the deepfake videos).

Fig. 1: Still from deepfake videos of Biden and Trump



Study 1: Qualitative approach

Three focus groups (N=13) examined emotional reactions to deepfake content and their interpretations in terms of authenticity. The first group met in person (n=6), while the second (n=2) and third (n=5) convened virtually. Participants were US and non-US citizens with English as their native language. Snowball recruitment was utilized for this study, and students or acquittances were recruited via social media and WhatsApp messaging.

Participants were told they would discuss their perception of political speeches, enabling natural deepfake discernment observation. The focus group involved a structured viewing of five videos—four deepfakes and one authentic political speech—across three separate sessions. Participants of the three different groups were shown a unique order of videos to control for potential order effects. To facilitate discussions, a semi-structured topic guide was used. Key areas of inquiry focused on trust, empathy, and inspiration elicited by the content. For example, participants were asked to reflect on if and why they trusted the messages conveyed in each video, whether they felt a sense of empathy towards the speakers, and how inspired they felt by the content. To analyze transcripts, we used a bottom-up thematic analysis, following Braun & Clarke (2006) six-step process. Transcripts were coded using Descript.AI and NVivo. Ethics approval addressed deception use, with comprehensive debriefing provided post-study, no participants withdrew from the study after the deceptive nature of videos was revealed.

Results

Our analysis revealed four themes highlighting the complex interplay of logic reasoning and emotions in evaluating political deepfakes: *Rationalization of fabricated content, Political familiarity and unity message, Cultural relevance and delivery***,** *Bias and partisanship*

Theme 1 – Rationalization of fabricate content. When shown with deepfakes, participant often recurred to technical explanations, such as visual or auditory anomalies. This is well exemplified by extract 1, where Tom (FG1)[1] suspected a deep fake, while Anne (FG1) recured to rationalization.

> Tom: It almost looked to me as if it was a deepfake
> Anne: It was edited so there are two camera angles, so it felt like they could have been taking different takes.

Here, audience presence served as a critical contextual cue. Linda (FG3): "Hearing the crowd… makes it more authentic and trustworthy". This reliance on peripheral cues shows how deepfakes exploit contextual plausibility to influence perceptions (Westerlund, 2019). Moreover, also visual and auditory anomalies served as primary cues for deepfake detection. Such as, Tom (FG1) moments after viewing the Biden ChatGPT speech, observed, "It almost looked to me as if it was a deepfake. Because like the way his mouth was moving wasn't quite lining up with what he was saying".

Theme 2 – Political familiarity and unity message. Two main factors played a role in determining trust in the videos, regardless their real authenticity. First, familiarity with political figure. Second, unity-themed messaging, often overriding skepticism. For instance, Paul (FG3) remarked, "I generally trust Biden more than many other politicians," while Becky (FG1) emphasized the appeal of unity, stating, "The point he's trying to make was more togetherness."

*Theme 3 – Cultural relevance and delivery***.** Empathetic political messages generated strong emotional responses, affecting perceived trustworthiness. Empathy was elicited through culturally resonant language, such as "God bless America," which Josh (FG3) noted, "definitely knows how to reach an emotion." Inspiration, however, was more contingent on delivery, with Chris (FG2) reflecting, "Although I guess it is a bit... dry, I did feel inspired." These findings highlight the nuanced ways deepfakes mimic real speeches to evoke emotional responses; however, their effectiveness often hinges on authenticity and presentation.

Theme 4 – Bias and partisanship. Participants agreed how their pre-existing beliefs can possibly shape their content interpretation. For instance, an American participant (Paul FG3) explicitly acknowledged his bias, explaining, "I wanted to go last on purpose, because I think I'm the most biased person, because I am an American citizen." Critical evaluation often accompanied this bias recognition: "I think again there's maybe a little bit of bias. I, I generally trust, Biden more than many other politicians" (Paul FG3).

Together, these results suggest that our participants interpreted political communication and its veracity through a complex interplay of emotions and logical reasoning. For instance, they highlight the role of their political positions in interpreting political speeches, mentioning the potential role of their own political biases. Second, our results show how participants looked for contextual or technical cues in gauging whether videos were real or fake. This evidence resonates with previous quantitative research (López-Gil et al., 2022; Onisha et al., 2024). This suggests that deepfakes do not operate in a vacuum; rather, they interact with the audience's cognitive frameworks and pre-existing knowledge. On the other hand, our findings inform on the element that contribute to fostering trust, inspiration, and empathy when watching a political speech, regardless their authenticity. This brings novel insights into the literature, showing how political familiarity, unit message, and cultural reference are pivotal elements in fostering emotional reactions. These findings underscore thus the potential efficacy of deepfake in fostering central emotions in political behaviour (Brader, 2006). The qualitative data then suggest that efforts to mitigate the impact of deepfakes should not only focus on detection technologies but also on enhancing media literacy and critical thinking skills among voters.

Study 2: Quantitative approach

We conducted a between-subjects experiment (N=261) examining how political deepfakes affect perceived trust, empathy, and inspiration. We chose a between-subject design to minimize potential carryover effects from repeated exposure to emotionally charged content, ensuring independent reactions across groups. Participants were randomly assigned to view either real (control, n=90), AI-generated (n=81), or historical speech (n=90) of both Biden and Trump, where the latter two were deepfakes. That is, participants saw two videos. Age distribution was within the 25–50 range, with a slight overrepresentation of younger

[1] All participant names were changed to preserve participant anonymity.

participants in the historical condition. Gender distribution was balanced, while political affiliation was coded numerically to represent the political spectrum, with a slight skew toward liberal affiliations in the ChatGPT condition. Presentation order was randomized to control for order effects. The study tested three hypotheses: if deepfake political speeches yield similar levels of perceived trust compared to real speeches (H1), if deepfake political speeches elicit comparable levels of empathy compared with real speeches (H2), and if deepfake political speeches elicit comparable levels of inspiration compared with real speeches (H3). Based on Study 1's findings, we integrated two new hypothesizes addressing the mediation and moderation themes we saw from political bias and emotional resonance. (H4) Deception detection abilities will mediate the relationship between video condition and perceived political authenticity. Similarly, moderation hypotheses will include: (H5) Political affiliation will moderate the relationship between video condition and perceived authenticity.

The Perceived Political Authenticity (PPA) scale (Louden & McCauliff, 2004) was used to measure the main variables, including trust (e.g., "Did you feel you could trust what Trump or Biden was sharing?"), inspiration ("Did you feel inspired by Trump or Biden's speech?"), and empathy ("Do you feel the speaker understands the concerns of ordinary people?") Responses were rated on 5-point Likert scales to quantify participants' emotional reactions. We compare these across conditions to test if participants differ on their views on these emotional states given content type (real, ChatGPT, or historical speeches).

The Deception Detection Scale (Blandon-Gitlin et al., 2014), the Trust in Media Scale (Kohring & Matthes, 2007), and the General Attitudes towards AI Scale (Schepman & Rodway, 2020) were used to measure political authenticity, ability to detect deception, trust in media, and attitudes towards artificial intelligence, respectively. While they do not directly measure trust, empathy, or inspiration, they assess related aspects. The quantitative results clarify that the measures are indirect but captures emotional dimensions relevant to the research, aligning with the study's objectives and ensuring transparency in the findings.

By combining these targeted emotional questions with validated scales, the survey provided a comprehensive evaluation of how participants perceived and emotionally responded to political speeches across different video conditions. This approach allowed for nuanced insights into the impact of deepfake content on trust, empathy, and inspiration in political communication.

Results

Demographic variables were well balanced across the three experimental conditions, as shown in Table Y. No significant differences were observed in age, gender, political affiliation, or eligibility to vote in the U.S., supporting the effectiveness of random assignment.

Table 1. *Participant demographic characteristics by condition (N = 261)*

Variable	ChatGPT (n = 87)	Historical (n = 87)	Control (n = 87)
Mean Age (years)	29.1	28.3	28.7
Female (%)	55%	54%	57%
U.S. Citizen (%)	62%	63%	61%
Political Affiliation (%)			
- **Left-leaning**	49%	44%	46%
- **Centrist**	39%	39%	38%
- **Right leaning**	17%	17%	16%
Education (% Bachelor+)	68%	72%	69%

To test if the type of video impacts perception of trust, empathy, and inspiration, we run ANOVAs for each element across the three conditions (real, ChatGPT, and historical speeches) for both Biden and Trump. We find no significant differences in any condition for either candidate (p-values range between 0.0768 for trust in Biden to 0.872 in empathy for Trump), suggesting that participants responded to all three types of speeches similarly across these emotional states. To further explore the emotional response drivers, we conducted a post hoc comparison of emotional responses based on participants' beliefs about video authenticity (regardless of condition). Participants who believed a video was real reported slightly higher trust ($M = 3.52$) compared to those who believed it was fake ($M = 3.28$), though this difference was not statistically significant ($p = 0.093$). These results suggest emotional responses may be shaped more by perceived than actual authenticity.

ANCOVA (Analysis of Covariance) was chosen for this study to assess the impact of experimental conditions on perceived political authenticity, deception detection, and trust in media while controlling for demographic factors such as age, education, gender, and political affiliation. Across video conditions, the ANCOVA showed no significant difference in perceived authenticity for either politician (Biden: $F(2,239) = 0.279$, $p = 0.757$, $\eta^2 = 0.002$; Trump: $F(2,239) = 2.101$, $p = 0.125$, $\eta^2 = 0.017$). Political affiliation emerged as a powerful predictor of perceived authenticity, especially for Trump ($\eta^2=0.411$) compared to Biden ($\eta^2=0.245$). These large effect sizes indicate that pre-existing political beliefs played a significant role in shaping perceptions of authenticity, regardless of the actual content presented. Our interpretation is, people who were in favor of Biden believed his videos were more authentic, regardless of the type (real, historical, or AI), and vice versa for Trump supporters.

Mediation Analysis: While video condition showed no direct effect on perceived authenticity ($\beta=0.07$, $SE=0.11$, $p=0.52$), it had a significant indirect effect through deception detection abilities ($\beta=0.06$, $SE=0.03$, 95% CI [0.01, 0.12]).

Political Moderation: Political affiliation significantly moderated the relationship between condition and perceived authenticity (Biden: $\beta=0.18$, SE=0.08, $p<0.05$; Trump: $\beta=-0.16$, SE=0.07, $p<0.05$).

Video condition: Video type (authentic, AI-generated, or historical speeches) did not significantly influence perceived authenticity of speeches. This suggests that viewers were unable to reliably distinguish between real and AI-generated content, challenging assumptions about the immediate impact of deepfakes on public perception. Deception detection slightly varied across the three video conditions—control (authentic rally speeches), ChatGPT-generated speeches, and historical speeches—but did not show a statistically significant overall effect ($F(2,239) = 1.643$, $p = 0.196$, $\eta^2 = 0.014$). Control speeches elicited higher confidence in authenticity detection, likely due to familiar contextual markers such as audience reactions and natural delivery styles, which aligned with participants' expectations of political speeches. These elements reduced skepticism make participants less critical of potential deception.

Mediation and Moderation analyses

Mediation and moderation analyses revealed the complex relationships between exposure to political deepfakes, perceived authenticity, and deception detection abilities. The mediation analysis showed a significant indirect effect, indicating that while exposure to deepfakes did not directly alter perceptions of authenticity, individuals' deception detection abilities influenced how they processed and evaluated the content. This show the importance of cognitive mechanisms, such as deception detection, in shaping perceptions of media credibility and underscores the need for improved media literacy to enhance critical evaluation skills.

Moderation analyses demonstrated that demographic factors—particularly political affiliation—significantly affected how participants evaluated deepfake content. Pre-existing beliefs and biases strongly shaped perceptions, with individuals often rationalizing or dismissing content based on their political leanings. Together, these findings emphasize the interplay between cognitive processes and individual differences in evaluating political deepfakes, pointing to the need for targeted interventions to help individuals navigate increasingly sophisticated media environments.

Political affiliation and authenticity The quantitative analysis revealed striking effect sizes for political affiliation's impact on perceived authenticity. The ANCOVA analysis shows that for Biden, political affiliation explained 24.5% of the variance in perceived authenticity ($\eta^2 = 0.245$), while for Trump, it explained 41.1% ($\eta^2 = 0.411$). These substantial effect sizes validate and extend the qualitative findings about pre-existing political beliefs shaping content perception. The strong negative correlation between perceived authenticity ratings for Biden and Trump ($r = -0.42$, $p < 0.001$) quantitatively confirms a pattern first observed in the focus groups. This "opposition effect" suggests that participants who view one politician as authentic tend to view the other as inauthentic, regardless of content authenticity.

Our findings challenge assumptions about deepfakes' impact on public perception, as our results align with research on motivated reasoning in political communication (Taber & Lodge, 2006). Notably, the majority of participants could not consistently detect deepfakes from real videos and used their prior beliefs to guide their perceptions of authenticity. The stability of trust and authenticity judgments across conditions suggests that pre-existing beliefs and demographic factors, particularly political affiliation, play a more crucial role than content authenticity. This has obvious worrying implications for the potential impact of deepfakes – if people are more motivated to rate favorable deepfake content as authentic, it is an epistemic challenge. The findings support the concept of confirmation bias, where individuals tend to seek out and interpret information that confirms their existing beliefs (Nickerson, 1998). The significant indirect effect through deception detection abilities reveals a subtle mechanism by which deepfakes might influence perception, highlighting the importance of media literacy in political communication. This aligns with Marcus et al.'s (2000) Affective Intelligence Theory, which emphasizes the role of emotions in political decision-making. The ability of deepfakes to generate similar emotional responses challenges our understanding of human-AI emotional interaction and supports Vaccari and Chadwick's (2020) argument that the affective potential of deepfakes may be their most significant feature in terms of political influence. This emotional resonance, regardless of content authenticity, suggests that the persuasive power of deepfakes may lie more in their ability to convey emotion convincingly than in their technical perfection.

The quantitative analysis reveals no significant differences in perceived authenticity between real and deepfake political speeches, indicating that both types of content can elicit similar emotional responses. This result challenges existing theories that posit a clear distinction between authentic and manipulated media (Chesney & Citron, 2019). Notably, this contrasts with recent findings. For instance, Groh et al. (2024) reported over 70% accuracy in identifying deepfakes when audio was included. One plausible explanation for the discrepancy is that our real speech clips were edited for length, potentially diminishing their perceived authenticity due to reduced context and visual coherence. Instead, it suggests that audiences may become desensitized to authenticity cues in an era where deepfake technology is prevalent, leading to a normalization of skepticism towards all political messages. Additionally, the marginal negative correlation observed between deepfake exposure and trust in political communication indicates that while deepfakes may not directly undermine perceptions of authenticity, they contribute to an overall climate of distrust. This aligns with research by Guess et al. (2020), which found that exposure to misinformation can erode public trust in media sources even when individuals are aware of the manipulative nature of specific content. Therefore, the quantitative results reinforce the idea that enhancing media literacy is crucial for fostering resilience against misinformation, such as inoculation theory

(Roozenbeek et al., 2022; Harrop et al., 2023) and inoculation hesitancy (Johnson & Madsen, 2024).

Discussion and concluding remarks

The study suggests that the impact of political deepfakes operates through subtle mechanisms. Rather than directly undermining trust or authenticity perceptions, deepfakes appear to interact with pre-existing political beliefs and media literacy skills to shape political perception. The scales used in the quantitative study provide valuable insights into emotional responses, the findings should be interpreted as indicative rather than definitive for trust, empathy, and inspiration as discrete constructs. The study reveals the complex interplay between truth discernment and bias reconciliation in political deepfake perception. The finding of audience presence as an authenticity cue suggests deepfake creators may need to simulate communicative environments for convincing deception. This aligns with Hameleers et al. (2022), who stress that emotional engagement with political content is often mediated by individuals' prior beliefs and experiences. Their research indicates that voters are not merely passive recipients of information; they actively rationalize content based on their political affiliations, which can mitigate the perceived impact of deepfake technology.

The strong effect sizes for political affiliation, combined with qualitative evidence of content rationalization, suggest that current theories of deepfake impact may overestimate content effects while underestimating the role of pre-existing beliefs. This perspective resonates with the work of Fridman et al. (2020), who argue that individuals' cognitive biases significantly influence their interpretation of manipulated media. The significant indirect effects through deception detection abilities indicate that media literacy might serve as a crucial moderating factor in this relationship. Research by Guess et al. (2020) supports this notion, showing that individuals with higher media literacy skills are better equipped to identify misinformation and critically evaluate the authenticity of media content.

The findings have implications for understanding political communication in an era of artificial intelligence, suggesting efforts to combat deepfake influence could focus on building media literacy rather than sole reliance on detection. On public policy, our findings suggest a need for broader systemic solutions beyond technical detection. These could include mandatory AI-content labelling for political videos, public media literacy education initiatives, and stronger platform accountability for deceptive synthetic content amplification. This approach is supported by studies indicating that fostering critical thinking skills can empower voters to navigate complex information environments more effectively (Lewandowsky et al., 2017). By equipping individuals with the tools to discern authenticity in political messages, we can mitigate the risks posed by synthetic media and promote a more informed electorate. The mixed-methods investigation into political deepfakes reveals nuanced mechanisms of influence that challenge prevailing theories. Rather than directly undermining trust or authenticity perceptions, deepfakes interact with pre-existing political beliefs and media literacy skills to shape political perception. This aligns with recent research emphasizing the role of emotional engagement and prior beliefs in mediating responses to political content.

Bias Confirmation: Participants rationalized inconsistencies in deepfakes content, supported by political affiliation ($\eta^2 = 0.245$ for Biden, $\eta^2 = 0.411$ for Trump).

Detection Mechanisms: While focus group participants struggled to explicitly identify deepfakes, the survey's mediation analysis revealed that deception detection abilities implicitly influence content evaluation.

Trust Dynamics: Audience presence served as an authenticity cue, explaining the correlation between trust in media and deception detection abilities ($r = 0.25$, $p < 0.001$).

The study illustrates challenges for detecting deepfakes, supporting arguments that traditional media literacy may be insufficient for increasingly persuasive AI-generated content. The influence of demographic factors and political affiliation on perceptions aligns with literature on motivated reasoning and confirmation bias in political communication. Practical implications include the need for enhanced media literacy education, robust verification processes in journalism, and regulatory approaches such as mandatory labeling of AI-generated content. Future research should explore long-term effects, cross-cultural comparisons, and the development of more effective deepfake detection technologies. This research contributes to a more nuanced understanding of how individuals perceive and respond to political deepfakes, providing a foundation for maintaining the integrity of political discourse in an increasingly digitized world

The findings may be limited by the sample demographics, as participants did not fully represent the general population, particularly older adults and individuals with lower levels of education. The effect of education ($\eta^2 = 0.152$) indicates a potential bias, potentially overestimating the general population's ability to discern deepfakes. Furthermore, the significant effect of age ($\eta^2 = 0.054$) suggests that different age groups may perceive political authenticity differently, and the predominantly younger sample may not accurately represent older age groups. While the study captured the effects of education and age, it may have overlooked other important socio-cultural factors that influence perceptions of political content.

To address the limitations, future research should prioritize broader sampling to include participants from diverse education levels, age groups, and socio-cultural backgrounds. A longitudinal design would help understand how age and education level influence perceptions of deepfakes over time. Additionally, incorporating more socio-cultural variables, such as media consumption habits and technological literacy, and conducting cross-cultural comparisons would provide a more comprehensive understanding of how demographic and cultural factors influence the perception of political deepfakes. These improvements would enhance the study's external validity and provide a more nuanced understanding of this critical issue.

References

Ahmed, S., & Chua, H. W. (2023). Perception and deception: Exploring individual responses to deepfakes across different modalities. *Heliyon, 9*(10), e20383. https://doi.org/10.1016/j.heliyon.2023.e20383

Blandon-Gitlin, I., O'Connor, M., & Duran, M. (2014). Cognitive-load approaches to detect deception. *Applied Cognitive Psychology*, 28(4), 588-597. https://doi.org/10.1002/acp.3035

Brader, T. (2006). *Campaigning for hearts and minds: How emotional appeals in political ads work*. University of Chicago Press.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Caillaud, S., & Flick, U. (2017). Focus groups in triangulation contexts. *A new era in focus group research: Challenges, innovation and practice*, 155-177.

Citron, D., & Chesney, R. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, *107*(6), 1753.

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

de Ruiter, A. (2021). The Distinct Wrong of Deepfakes. *Philosophy & Technology*, *34*(4), 1311–1332. https://doi.org/10.1007/s13347-021-00459-2

Fridman, N., Hwang, T., & Pomerantz, J. (2020). Deepfakes and the new disinformation war: The coming age of synthetic media. *Journal of International Affairs*, 73(1), 65-82.

Guess, A. M., Lockett, D., & Nyhan, B. (2020). Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. *Misinformation Review*. https://misinforeview.hks.harvard.edu/article/misinformation-in-action-fake-news-exposure-is-linked-to-lower-trust-in-media-higher-trust-in-government-when-your-side-is-in-power/

Hameleers, M., Brosius, A., & de Vreese, C. H. (2022). Whom to trust? Media exposure patterns of citizens with perceptions of misinformation and disinformation related to the news media. European Journal of Communication, 37(3),237-268.https://doi.org/10.1177/02673231211072667

Hameleers, M., & Marquart, F. (2023). It's nothing but a deepfake! The effects of misinformation and deepfake labels delegitimizing an authentic political speech. *International Journal of Communication*, 17, 6295 6318. https://ijoc.org/index.php/ijoc/article/view/20777

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

Hahn, U., Harris, A., & Corner, A. (2009). Argument content and argument source: An exploration. *Informal Logic, 29*, 337–367.

Harrop, I., Roozenbeek, J., Madsen, J. K., & van der Linden, S. (2023). The Role of Media in Political Polarization: Inoculation Can Reduce the Perceived Reliability of Polarizing Social Media Content. *International Journal of Communication, 17*, 25.

Johnson, A., & Madsen, J. K. (2024). Inoculation hesitancy: An exploration of challenges in scaling inoculation theory. *Royal Society Open Science, 11*(6), 231711.

Kohring, M., & Matthes, J. (2007). Trust in news media: Development and validation of a multidimensional scale. *Communication Research, 34*(2), 231–252. https://doi.org/10.1177/0093650206298071

Kumari, R., Ashok, N., Ghosal, T., & Ekbal, A. (2021). Misinformation detection using multitask learning with mutual learning for novelty detection and emotion recognition. *Information Processing & Management*, *58*(5), 102631. https://doi.org/10.1016/j.ipm.2021.102631

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

López-Gil, J.-M., Gil, R., & García, R. (2022). Do deepfakes adequately display emotions? A study on deepfake facial emotion expression. *Computational Intelligence and Neuroscience, 2022*, Article 1332122. https://doi.org/10.1155/2022/1332122

Louden, A., & McCauliff, K. (2004). The "Authentic Candidate": Extending candidate image assessment. In K. L. Hacker (Ed.), *Presidential campaign communication: Principles and practice* (pp. 85-103) Routledge.

Madsen, J. K. (2019). Voter reasoning bias when evaluating statements from female and male political candidates. *Politics & Gender, 15*(2), 310–335.

Marcus, G. E., Neuman, W. R., & MacKuen, M. (2000). *Affective intelligence and political judgment*. University of Chicago Press.

Muradova, L., & Arceneaux, K. (2022). Reflective political reasoning: Political disagreement and empathy. *European Journal of Political Research*, 61(4), 740-761. https://doi.org/10.1111/1475-6765.12490

Nabi, R. L., Moyer-Gusé, E., & Byrne, S. (2008). All audiences are not created equal: A taxonomy of audience responses to emotional appeals in entertainment media. In J. Bryant & P. Vorderer (Eds.), *Media effects: Advances in theory and research* (pp. 75-93). Routledge.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175-220. https://doi.org/10.1037/1089-2680.2.2.175

Olof Larsson, A. (2023). The rise of Instagram as a tool for political communication: A longitudinal study of European political parties and their followers. *New Media & Society, 25*(10), 2744–2762. https://doi.org/10.1177/14614448211034158

Onisha, K., Chen, L., & Marchant, G. (2024). *Detecting deception in deepfake political media: The role of audio-visual cues*. *Nature Communications*, *15*(112), 1234–1245. https://doi.org/10.1038/s41467-024-51998-z

Onisha, T. A., Wimmer, H., & Rebman Jr, C. M. (2024). Facial expressions analysis for deepfake and genuine video recognition. *Issues in Information Systems, 25*(1), 159–174.

Petty, R. E., & Cacioppo, J. T. (1984). *Communication and persuasion: Central and peripheral routes to attitude change*. Springer-Verlag.

Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards Artificial Intelligence Scale. *Computers in Human Behavior Reports*, 1, 100014. https://doi.org/10.1016/j.chbr.2020.100014

Renstrom, R. A., & Ottati, V. C. (2020). "I Feel Your Pain": The Effect of Displaying Empathy on Political Candidate Evaluation. *Journal of Social and Political Psychology, 8*(2), 767–787. https://doi.org/10.5964/jspp.v8i2.1292

Roozenbeek, J., Van Der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances, 8*(34), eabo6254.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755-769. https://doi.org/10.1111/j.1540-5907.2006.00214.x

Thrash, T., & Elliot, A. (2003). Inspiration as a Psychological Construct. *Journal of Personality and Social Psychology, 84*, 871–889. https://doi.org/10.1037/0022-3514.84.4.871

Tschannen-Moran, M., & Hoy, W. K. (2000). A multidisciplinary analysis of the nature, meaning, and measurement of trust. *Review of Educational Research*, 70(4), 547-593. https://doi.org/10.3102/00346543070004547

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), 2056305120903408. https://doi.org/10.1177/2056305120903408

Valgarðsson, V. O., Clarke, N., Jennings, W., & Stoker, G. (2021). The good politician and political trust: An authenticity gap in British politics? *Political Studies, 69*(4), 858–880. https://doi.org/10.1177/0032321720928257

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, 9(11), 21-34. https://doi.org/10.22215/timreview/1248