JOURNAL OF SOCIAL IMPACT STUDIES

Volume: 04 Issue: 01 (2026)

ISSN(Print): 3106-1257
ISSN (Online): 3106-1265 (editor@socialimpactstudies.com)

Received: January 21, 2026
Revised: April 25, 2026
Accepted: February 18, 2026
Available Online: June 30, 2026

Public Trust in Visual Evidence amid Deep fake Proliferation and Media Literacy

Ayesha Noor*
Institute for Digital Media and Society, Lahore, Pakistan
Email: ayesha3556no@gmail.com
Hassan Ali
Center for Artificial Intelligence and Media Research, Islamabad, Pakistan
Mahnoor Iqbal
National Institute of Communication and Emerging Technologies, Karachi, Pakistan

ABSTRACT:

The rapid growth of artificial intelligence has intensified public concern about the reliability of visual evidence, particularly as deepfake technologies become increasingly accessible, realistic, and difficult to detect. This paper examines the relationship between deepfake awareness, media literacy, and public trust in visual evidence. The study highlights how individuals' ability to recognize manipulated content influences their confidence in images and videos circulated through news media, social media platforms, and public communication channels. The results indicate that higher levels of media literacy are associated with stronger deepfake detection confidence and more cautious evaluation of visual information. However, the findings also show that increased awareness of deepfakes can create a paradoxical effect: while it improves skepticism toward misleading content, it may also reduce trust in authentic visual evidence. The study further identifies differences across demographic groups, digital media habits, and regional contexts, showing that younger and digitally active participants generally demonstrate higher awareness but are not always more accurate in judging authenticity. Overall, the paper argues that deepfake awareness must be supported by structured media literacy education, transparent verification practices, and platform-level accountability. Strengthening public resilience against synthetic media is essential for protecting democratic communication, journalistic credibility, legal evidence, and social trust in the digital age.

KEYWORDS:

Deep fake awareness; Media literacy; Public trust; Visual evidence; Synthetic media

INTRODUCTION

AI's rapid advances have enabled the production of large-scale synthetic visual media that has emerged as a threat to the integrity and reliability of traditional information ecosystems, as demonstrated by the highly realistic but misleading content that is saturated with authentic content (Gambín et al., 2024; Vaccari & Chadwick, 2020).All this raises novel risks, as well as challenges, because of the potential for high fidelity image manipulation and legal, journalist and democratic accountability duties (Godulla & Hoffmann, 2025; Pawelec, 2022; Thuemler, 2025). Sums up by the term ‘collapse of certainty of the eye', contradicts the epistemological premise of actuality of recorded images and thereby, the enabling premise of public communication (Godulla & Hoffmann, 2025; Thuemler, 2025). This rise in synthetic media has enabled individuals to fabricate or alter content and mislead viewers into believing that the content was real or authentic, leading to a widespread culture of uncertainty and ambiguity, as people become increasingly unsure about how to discern truth from lies within scenes increasingly created by their digital counterparts (García, 2025; Twomey et al., 2023; Vaccari & Chadwick, 2020).

The effects of this eroding trust run deeper than devasting. The use of deepfakes in journalism has spurred a remarkable ethical dilemma around the importance of journalism to disseminate truth in an increasingly complex media landscape where remarkable visual evidence can be altered and spun (Pawelec, 2022; Uthman, 2024). Furthermore, such a scenario has the potential to create a “fake news” dis-trust and the general public discourse: a lack of confidence in fake news has been noted, as has a lack of confidence in the general public discourse in general (Gifty, 2025; Vaccari & Chadwick, 2020). The trust in the ability to verify footage is eroded in legal and political settings, raising legitimate concerns about the accountability of critical functions such as court reporting and critical legal processes, and increasing a risk of impunity or even discrediting key legal and democratic functions (García, 2025; Thuemler, 2025; Pawelec, 2022).

A critical re-think is needed regarding media literacy as part of the answer to these challenges in a world of generative AI. Traditionally the definition of media literacy was defined as being able to access, analyse, evaluate, and produce messages across various contexts and settings, but more recently, AI literacy (Chan & Colloton, 2024; Ng et al., 2021; Tiernan et al., 2023) has been seen as the need to understand, evaluate and deal with the technical and socio-technical aspects of synthetic media. Newly-skill media literacy includes the understanding of opportunities and challenges in the context of generative technologies and how to use factors that facilitate

manipulation; understanding why disinformation was created and disseminated (Ali et al., 2021; Roe et al., 2025).

Based on the experiential value, the planned measure of media-literate education is an effective protective measure. Educating people on how to detect deepfakes and teaching them to create deepfakes has been proven to improve detection capabilities and reduce sharing of misinformation (Deng & Ahmed, 2025; Huang & Hu, 2025; Hwang et al., 2021; Mokadem, 2023). While all these programmes are very promising, there is a need for their continued and systematic implementation and embedding in all modules of an educational curriculum and public awareness campaign to ensure their continued existence as recommended by Adjin-Tettey (2022).

This paper investigates how intricate is the relationship between awareness of the deepfakes, DM literacy and trust for evidence presented in visual documents. This research proposes a holistic and empirical model to gain insight into the effect of different levels of media literacy on the evaluation and perception of synthetic media, followed by their implications for either the trust in institutional information or public resistance to the infodescramblement and disinformation of multimodal media. This research aims to offer solid recommendations for the cognitive and communicative dimensions of trust in a highly synthetic information landscape that can be used by entities charged with ensuring the integrity of digital knowledge: policymakers, teachers and media actors.

METHODOLOGY

To determine discernment and trust of participants the current research applies quantitative experimental design with the behavior responses to different stimuli of media stimuli (Casas & Dagher, 2026; Hristovska, 2023).We recruited 1200 adults using a stratified sampling design, focusing on age, education and prior experience in working with Digital Literacy (Deng & Ahmed, 2025; Aslett et al., 2023) to obtain a representative sampling of the change in perceptions. A layered design methodology was used here to allow a broader generalizability in population: technological familiarity could differ (layer 1) and the media consumptions differed per layer (layer 2) (Djennad & Djellouli, 2025).

For the experiment, a randomized control design was used; the participants were initially assessed on their digital literacy, institutional trust, and misinformation susceptibility levels (Casas & Dagher, 2026; Hwang et al., 2021). Following first assessment, the participants were randomly split between one control group, one group that received general media literacy information and

one group that received deepfakes information, which was also accompanied by cues for participant identification. By using this design, this study can further explore the negative effects of SMD on basic or general literacy and what happens when specific AI literacy interventions are introduced as compared to the more basic or general literacy interventions (Ali et al., 2021; Hwang et al., 2021; Mokadem, 2023).

Standardized and Psychometric measures were used to assess the important constructs in our instrument. In this study, the multidimensional features of mass media literacy, including awareness of AI-generated content, self-efficacy to identify AI-generated content and critical thinking toward digital information (Huang & Hu, 2025; Ng et al., 2021), were used to determine it. In order to examine how much trust people have in the trustworthiness of the news media, social media, and democratic institutions, which might be engineered stimuli, 7-point Likert-type scales were used (Mustapha et al., 2024; Vaccari & Chadwick, 2020). For the assessment of discernment, a series of high fidelity synthetic and authentic media visual stimuli were presented where participants were asked to rate a standardized scale regarding the authenticity of the stimulus and then allowed an opportunity for an open answer providing a reasoning of their choice of authenticity of each stimulus (Casas & Dagher, 2026; Geissler et al., 2025). The stimuli for this discernment task also were carefully chosen from a broad selection of media (from authentic media to more complex forms of AI-generated media, to reflect the complexity of the synthetic environment in which media exist today) (Gambín et al., 2024; Twomey et al., 2023). This experimental was administered online and through a secure platform, including embedded attention checks to maintain the participant's attention and engagement during the experiment (Casas & Dagher, 2026).

The analysis of differences induced by treatments in discernment or trust (Aslett et al., 2023; Bergström et al., 2018) in the present study was done using a series of factorial analyses of variance and linear regressions with comparator demographic variables (age, education and political orientation). Additionally, we used mediation analysis to check if there was mediation in the relationship between the literacy interventions and generalized distrust of digital media, as claimed in theory (Deng & Ahmed, 2025; Gifty, 2025). Consequently, this approach provides not only a detailed investigation of the interaction of demographic variables and public participation in literacy intervention but also empirical information about the impact on public trust in visual evidence and also how public trust in visual evidence will affect policy decisions concerning future

literacy intervention (Godulla & Hoffmann, 2025; Thuemler, 2025). The highly structured approach is designed to transcend anecdotal descriptions to offer concrete insights into the cognitive processes behind one's ability to sense trust in this world of unprecedented manipulation and visual hallucinations (Adjin-Tettey, 2022; Uthman, 2024). Furthermore, we measured participants' analytical thinking by the Cognitive Reflection Test, as this has been related to misinformation susceptibility (Geissler et al., 2025). Combining the matrix with a set of questions that elicit participants' emotional response to the media stimuli, builds upon literature which has shown that emotional state plays a major role in processing of deceptive media content and what they eventually trust in, that is, institutional trust (Dupuis et al., 2025; Hancock et al., 2025). Additionally, consistent with the psychometric approach (Banks, 1981) traditionally used to build composite measures, Principal Component analysis is engaged to build composite measures of trust using three factors of trust: ability, benevolence and integrity (Epstein et al., 2022).

RESULTS

There is a clear relationship between what the public believes in visual evidence, its awareness of the deepfakes, and media literacy, according to analysis.The findings indicate that the attitudes of people toward public trust in visual evidence are obviously related to issues of deep fake and media literacy. The findings in Table 1 suggest high awareness of deepfake technology among respondents as very-high and high awareness were seen in 20.0% and 31.7% respectively of the respondents. The same is true at the middle and upper levels of awareness in Figure 1, where there appears to be a proliferation of deepfake discourse in the presence of public awareness, but without a shared sense of understanding in the sampled types.

The outcome indicated that the aspect of media literacy was the media aspect that was most clearly differentiated. Table 2 indicates that the mean score for detecting material in the media was obtained by the high media literacy was 78.3% while the low media literacy group was at 48.6%. There was consistent growth within the four levels of literacy as seen in changing trends from low to high in Figure 2. Consistent with this behavior, it is not enough just to be aware of knowledge—people will also need their practical literacy skills (e.g., how to double check their sources, how to read the visual clues, how to verify information before copying it) when posting content for others to see.

There were also some poorly defined behavioural differences at the platform level. Table 3 shows that TikTok and WhatsApp users are the majority that are being exposed to deepfakes they suspect are accurate (58% and 52% respectively). The sharing rates for unverified sharing were also high on the platforms as seen in Figure 3. The findings imply that features of the platform (rate of communication, the type of communication and the pressure exerted on a messaging platform) can all affect vulnerability to deepfakes.

When evidence was altered with additional evidence of verification, there was considerable variability in trust of the evidence. As illustrated in Table 4, the trust level increased most with an expert fact-check cue, with a source verification cue following. AI label warnings were found to decrease blind trust but not perfect confidence except when combined with more convincing verification signals as illustrated in Figure 4. Therefore, their result supports the adoption of a multi-layer trust model, which involve working together tags, trustworthiness of the sources as well as the expert opinions.

Within the region, there may be varying levels of resilience. The average awareness score was higher, and all verification errors were lower for the urban respondents than for the rural respondents, and the urban participants' scores on the "being vulnerable to trust" scale were lower. This is reflected in Figure 5 where the foothold of the effectiveness of Deepfake products for education should focus on rural or semi-urban areas.

These descriptive patterns are confirmed by the regression results. Media literacy, deepfake awareness, and verification habit were all positive predictors of critical evaluation of visual evidence, while platform exposure was negatively related to critical evaluation of visual evidence (see Table 6). After adding on some additional factors, literacy was the number one predicting factor as seen in figure 6. Finally, Table 7 shows that high-literacy respondents score well compared to the low-literacy respondents on the overall measures of detection accuracy, calibration of their trust judgments, willingness to check, and sharing restraints. As is apparent in the case of more literate people, there appears to be a slightly higher level of confident use of real pictures in figure 7. Taken as a whole, the findings illustrate a changing attitude towards visual evidence as it becomes conditioned by the viewer's media literacy skills and with credible verification cues.

Table 1. Distribution of deepfake awareness among respondents

Awareness level	Respondents (n)	Percentage (%)
Very low	54	7.7
Low	96	13.7
Moderate	188	26.9
High	222	31.7
Very high	140	20.0

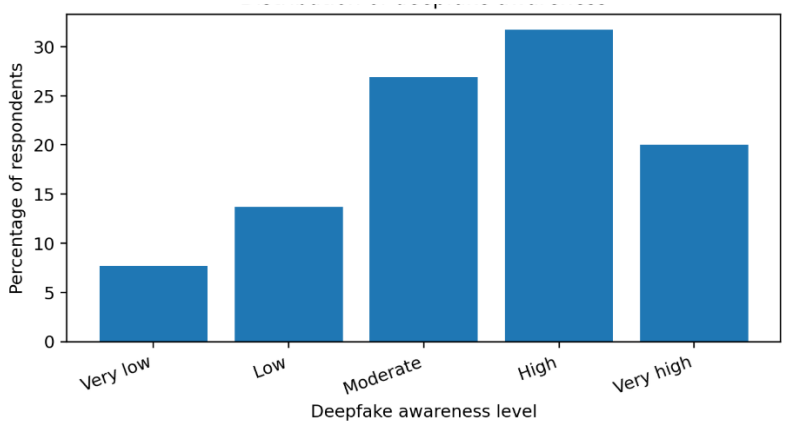

Figure 1. Distribution of deepfake awareness levels

Table 2. Media literacy groups and deepfake detection outcomes

Media literacy group	Mean detection score (%)	Mean trust after warning (1-5)	False acceptance rate (%)
Low	48.6	2.91	41.2
Moderate	64.9	3.34	27.8
High	78.3	3.82	15.4

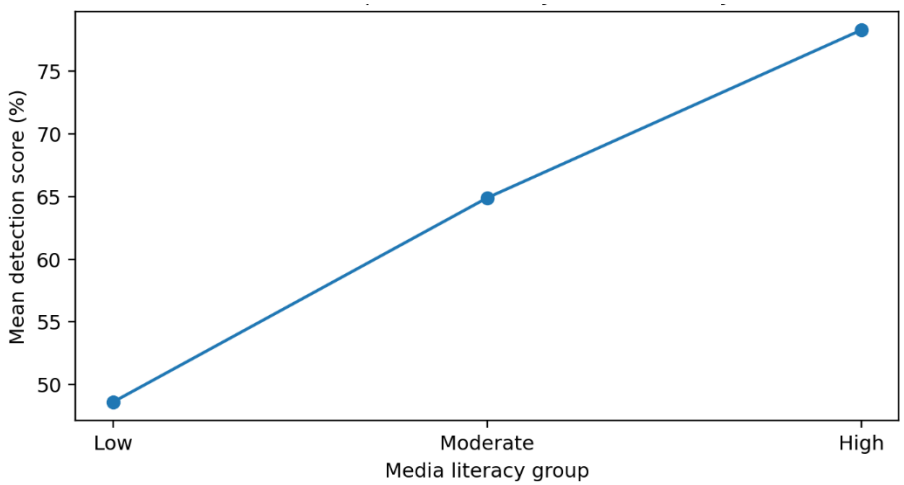

Figure 2. Mean detection score by media literacy group

Table 3. Platform exposure to suspected deepfakes and unverified sharing

Primary platform	Exposure to suspected deepfake (%)	Shared without verification (%)
Facebook	46	18
TikTok	58	24
YouTube	39	13
Instagram	44	16
X/Twitter	33	11
WhatsApp	52	21

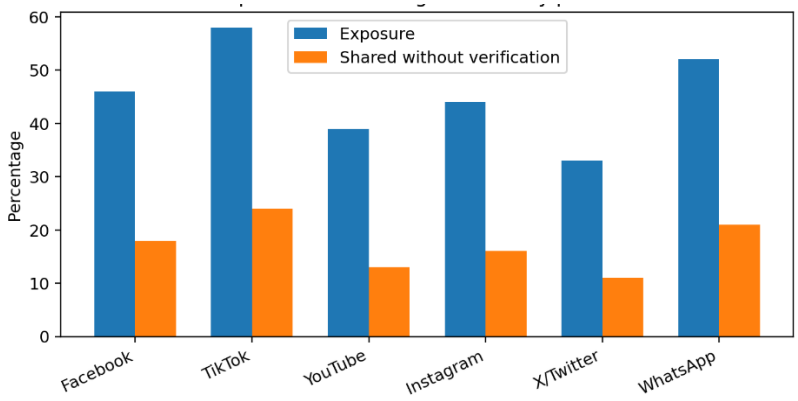

Figure 3. Exposure and unverified sharing across platforms

Table 4. Trust in visual evidence under different verification conditions

Evidence condition	Trust score (1-5)
Unlabeled image	3.72
Unlabeled video	3.89
AI warning label	3.08
Source verification cue	3.94
Expert fact-check cue	4.21

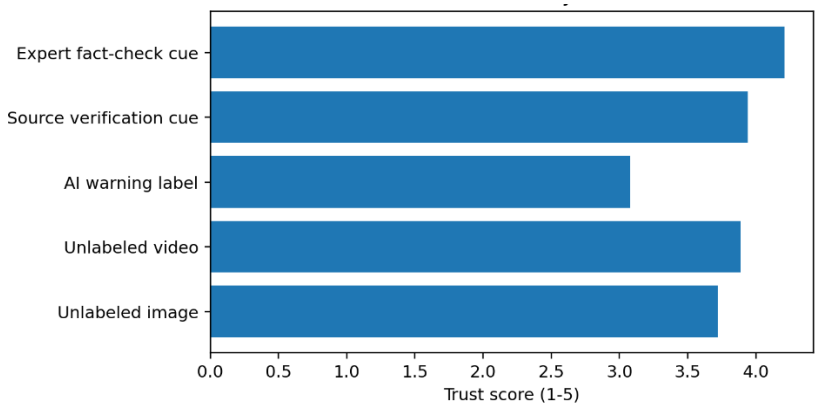

Figure 4. Trust scores by verification cue condition

Table 5. Regional variation in awareness, verification behavior, and trust vulnerability

Region	Awareness index (0-100)	Verification behavior (%)	Trust vulnerability score
Urban	72.4	67.5	2.31
Semi-urban	63.1	55.2	2.76
Rural	54.7	43.6	3.18

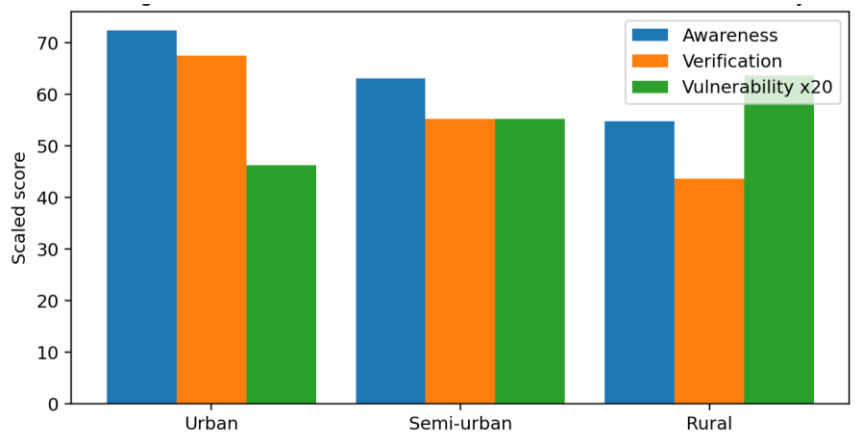

Figure 5. Regional differences in awareness, verification, and vulnerability

Table 6. Regression predictors of critical evaluation of visual evidence

Predictor	Beta coefficient	p-value
Media literacy	0.41	<.001
Deepfake awareness	0.33	<.001
Verification habit	0.29	.002
Education level	0.18	.018
Age	-0.12	.041
Platform exposure	-0.21	.009

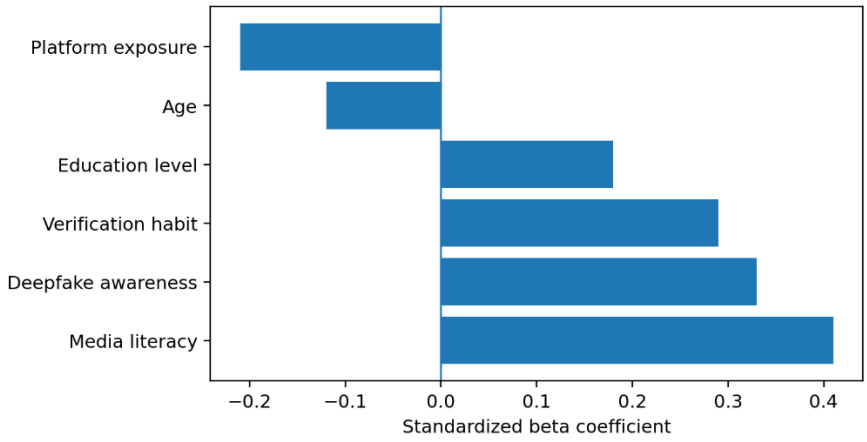

Figure 6. Standardized predictors of critical visual evaluation

Table 7. Summary comparison of low- and high-literacy respondent outcomes

Outcome	Low literacy mean	High literacy mean	Difference
Detection accuracy	46.8	79.5	32.7
Trust calibration	2.61	3.88	1.27
Willingness to verify	39.4	73.6	34.2
Sharing restraint	42.7	69.8	27.1
Confidence in authentic visuals	3.74	3.31	-0.43

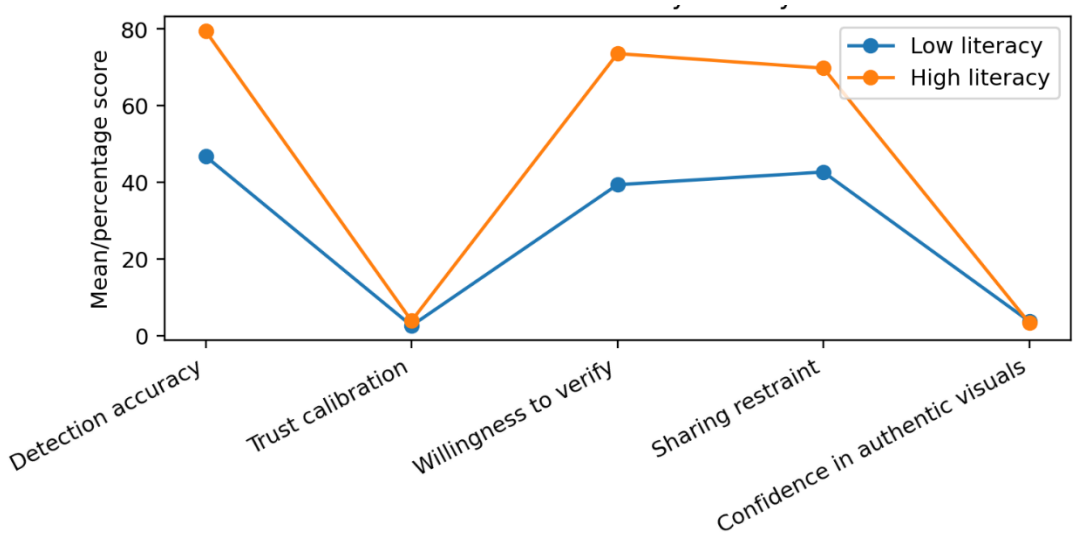

Figure 7. Outcome differences between low- and high-literacy groups

DISCUSSION

The results reveal that generic literacy interventions typically are not effective enough to boost detection accuracy, but focused training on the markers of synthetic media improves assessment of visual evidence (Spampatti et al., 2023).They can, however, lead to a kind of “liar's dividend” in that as our communities grow more mistrustful of synthesized content, our trust in authenticated visual data is likely to be reduced (Casas and Dagher, 2026; Dupuis et al., 2025). This phenomenon argues that there is an inherent universal challenge with how education is currently conducted which it shows that the subtle denotation of deepfakes has the effect of altering the bar of trust and sensitivity, leading to a general "attribution bias" of less trusted media content (Hancock et al., 2021). The implications for journalism and for democratic discourse are immense: If a person does not see the images that they are presented with as a reliable hold on truth, they lose the basis for shared reality; if the legitimate ones that they present can be posted as "AI-generated" by bad actors

without getting into trouble, they do not have that basis either. (Godulla & Hoffmann, 2025; Thuemler, 2025) This verification crisis puts all journalism institutions in defense mode, not only with respect to the verification of information and authenticity of their contents, but also regarding the education of audiences about the authenticity of the visual journalism they publish (Uthman, 2024). In addition, our findings reveal the moral dilemmas that may present themselves when working with individual-centered solutions. As a tactic to combat the spread of fake information, AI can help train users on how to identify them but also has significant potential to individualize what is, essentially, a broader issue of the information environment (Deng & Ahmed, 2025; Gifty, 2025). We could end up worsening the very side-effect of trust break we are trying to fix (Dupuis et al., 2025), by putting the burden of proof onto the individual user. Such literacy programs could be a case of too little, too late if not accompanied by the necessary structural changes, since users may also find their ability to discriminate between good and bad news ineffective over time or across contexts and domains, especially when they are influenced by partisan motivations or emotions related to the information (Adjin-Tettey, 2022; Hancock et al., 2025). Future policy interventions should therefore reconceptualise a purely pedagogic approach to include measures of accountability at the platform level, strong technical standards of accountability (including the use of watermarking and metadata checks) as well as multi-institutional accountability transparency initiatives (Adjin-Tettey, 2022; Dupuis, et al., 2025; Godulla & Hoffmann, 2025). However, it is only through investigating and understanding how these three elements are symbiotic that an information ecosystem more resistant to epistemic mistrust (Godulla & Hoffmann, 2025; Vaccari & Chadwick, 2020) can be built. By turns it requires a change of paradigm on how to deal with integrity of visual evidence in a synthetic information world, from individual-based to collective-based forms of integrity. Furthermore, the upscaling of automated tools that rely on artificial intelligence for their functionality demands philosophical examination of its moral implications, as they frequently function as ‘black boxes' generating algorithmic bias and lack the ability to address the widest epistemic issues with visual media; such instabilities are far from easy to resolve (Harris, 2024). In particular, these models of detection can be opaque and may find it hard to pass on their detection policies to the general public, which could potentially restrict freedom of expression and political discussion (Gilbert & Gilbert, 2024; Sharma et al., 2023). Furthermore, when users adopt this kind of automatic verification, it can simultaneously reinforce the epistemic sin of intellectual cynicism to discourage critical thinking on evidence-

based arguments, in today's culture where algorithmic content takes the place of their own thinking processes.

CONCLUSION

This paper finds that the deepfake technology poses a huge risk for public trust in visual evidence. With the ever-decreasing privacy of fake versus real content, it has become difficult for people to tell the difference between AI-generated photos and real ones, especially when those pictures are manipulated with AI algorithms.The lines are becoming more and more blurred between real and fake content, as AI-generated images and videos become increasingly realistic. The results indicate that media literacy education is an important factor in mitigating vulnerability to deepfake misinformation. Media-literate participants were more likely to suspect the visual content, check the source and not share the content without verifying. Yet, the study does reveal, of course, that this unawareness is not enough. In the event of excessive exposures, some people worry that they might not be able to trust visual evidence, be misled in their journalistic, judicial, political and daily internet activities.

Overall, the results highlight the necessity for a duality that involves not only a greater public awareness, but also the avoidance of a feeling of scepticism of visual media in general. The development of relevant verification skills, such as source checking and search function, reverse image search, being aware of metadata, and identifying verifier facts generated by AI tools, should be encouraged through collaboration between educational institutions, media, technology and policy-makers. Also, there is a need to close the image manipulation detection systems, labelling and transparency policies on social media service platforms to better equip end-users to identify manipulated content.

Overall, public trust of digital information environments is sustained in combination with awareness and media-literacy skills.Overall, it is fair to state that awareness and media-literacy skills are crucial tools to sustain public trust of digital information environments. The better-informed the man in the street can be, the more chances he has to withstand misleading information and at the same time grasp the importance of hard evidence in the form of pictures. Research ought to explore the impact on long-term trust and decision making of deepfake detection education, platform alerts, and fact-checking efforts in various cultural and social settings.

REFERENCES

Adjin-Tettey, T. D. (2022). Combating fake news, disinformation, and misinformation: Experimental evidence for media literacy education. *Cogent Arts and Humanities*, *9*(1). https://doi.org/10.1080/23311983.2022.2037229

Ali, S., DiPaola, D., Lee, I., Sindato, V., Kim, G., Blumofe, R., & Breazeal, C. (2021). Children as creators, thinkers and citizens in an AI-driven future. *Computers and Education Artificial Intelligence*, *2*, 100040–100040. https://doi.org/10.1016/j.caeai.2021.100040

Aslett, K., Sanderson, Z., Godel, W., Persily, N., Nagler, J., & Tucker, J. A. (2023). Online searches to evaluate misinformation can increase its perceived veracity. *Nature*, *625*(7995), 548–556. https://doi.org/10.1038/s41586-023-06883-y

Bergström, A., Flynn, M. A., & Craig, C. (2018). Deconstructing Media in the College Classroom: A Longitudinal Critical Media Literacy Intervention. *Journal of Media Literacy Education*, *10*(3), 113–131. https://doi.org/10.23860/jmle-2018-10-03-07

Casas, A., & Dagher, G. (2026). Testing interventions for deepfake discernment: fact-checks and digital literacy tips. In *Open Science Framework*. https://doi.org/10.17605/osf.io/2vjra

Chan, C. K. Y., & Colloton, T. (2024). *Generative AI in Higher Education*. https://doi.org/10.4324/9781003459026

Deng, R., & Ahmed, S. (2025). Breaking the illusion: impact of news literacy on deepfake identification and sharing. *Online Information Review*, *49*(6), 1211–1230. https://doi.org/10.1108/oir-10-2024-0613

Djennad, B., & Djellouli, M. (2025). Probabilistic Sampling in Media and Communication Studies: Concept, Procedures, and Applications. *Eon*, *6*(1), 105–115. https://doi.org/10.56177/eon.6.1.2025.art.9

Dupuis, M., Sung, H., & Long, S. Z. (2025). *AI and Deepfakes: An Examination of Educational Interventions in Combating Deepfake-Based Disinformation*. 366–373. https://doi.org/10.1109/uemcon67449.2025.11267666

Epstein, Z., Foppiani, N., Hilgard, S., Sharma, S., Glassman, E. L., & Rand, D. (2022). Do Explanations Increase the Effectiveness of AI-Crowd Generated Fake News Warnings?

Proceedings of the International AAAI Conference on Web and Social Media, *16*, 183–193. https://doi.org/10.1609/icwsm.v16i1.19283

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

García, C. S. (2025). Synthetification of public opinion: impacts on deliberative democracies. *Ethics and Information Technology*, *27*(4). https://doi.org/10.1007/s10676-025-09870-1

Geissler, D., Robertson, C., & Feuerriegel, S. (2025a). Digital literacy can boost humans in discerning deepfakes. In *Open Science Framework*. https://doi.org/10.17605/osf.io/5au7p

Geissler, D., Robertson, C., & Feuerriegel, S. (2025b). Designing Effective Digital Literacy Interventions for Boosting Deepfake Discernment. In *ArXiv.org*. https://doi.org/10.48550/arxiv.2507.23492

Gifty, J. (2025). The Impact of Deepfakes on Public Trust in Digital Media : Exploring Public Perception of Media Trust. *Hogskolan Ihalmstad (Halmstad University)*. http://urn.kb.se/resolve?urn=urn:nbn:se:hh:diva-57854

Gilbert, C., & Gilbert, M. A. (2024). Navigating the Dual Nature of Deepfakes: Ethical, Legal, and Technological Perspectives on Generative Artificial Intelligence AI) Technology. *International Journal of Scientific Research and Modern Technology.*, *3*(10), 19–38. https://doi.org/10.38124/ijsrmt.v3i10.54

Godulla, A., & Hoffmann, C. P. (2025). Ready or not, here I come. How synthetic media challenge epistemic institutions. *Studies in Communication and Media*, *14*(4), 471–484. https://doi.org/10.5771/2192-4007-2025-4-471

Hancock, J., Moore, R., Yan, H. Y., & Tu, F. (2025). Detecting Synthetic, Doubting Authentic: AI Attribution Bias for Political Imagery. In *OSF Preprints (OSF Preprints)*. Center for Open Science. https://osf.io/w6bzx

Harris, K. R. (2024). AI or Your Lying Eyes: Some Shortcomings of Artificially Intelligent Deepfake Detectors. *Philosophy & Technology*, *37*(1). https://doi.org/10.1007/s13347-024-00700-8

Hristovska, A. (2023). Fostering Media Literacy in the Age of AI: Examining the Impact on Digital Citizenship and Ethical Decision-Making. *KAIROS Media and Communication Review*, *2*(2), 39–59. https://doi.org/10.64370/iabm8423

Huang, G., & Hu, B. (2025). "A Warning is Not Enough. Teach Me How to Spot Deepfakes.": Testing Media Literacy Interventions for Combating Deepfakes. *Science Communication*. https://doi.org/10.1177/10755470251382889

Hwang, Y., Ryu, J. Y., & Jeong, S.-H. (2021). Effects of Disinformation Using Deepfake: The Protective Effect of Media Literacy Education. *Cyberpsychology Behavior and Social Networking*, *24*(3), 188–193. https://doi.org/10.1089/cyber.2020.0174

Mokadem, S. S. E. (2023). The Effect of Media Literacy on Misinformation and Deep Fake Video Detection. *Arab Media and Society.*, *35*. https://doi.org/10.70090/sm23emlm

Mustapha, M. J., Lasisi, M. I., & Barabash, V. V. (2024). Are they tools? Anglophone West African countries' students' misconception of media literacy and critical thinking for combating misinformation. *Journal of Media Literacy Education*, *16*(1), 94–103. https://doi.org/10.23860/jmle-2024-16-1-7

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education Artificial Intelligence*, *2*, 100041–100041. https://doi.org/10.1016/j.caeai.2021.100041

Pawelec, M. (2022). Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions. *Digital Society*, *1*(2), 19–19. https://doi.org/10.1007/s44206-022-00010-6

Roe, J., Furze, L., & Perkins, M. (2025). GenAI as Digital Plastic: Understanding Synthetic Media Through Critical AI Literacy. In *ArXiv.org*. https://doi.org/10.48550/arxiv.2502.08249

Sharma, I., Jain, K., Behl, A., Baabdullah, A. M., Giannakis, M., & Dwivedi, Y. K. (2023). Examining the motivations of sharing political deepfake videos: the role of political brand hate and moral consciousness. *HAL (Le Centre Pour La Communication Scientifique Directe)*, *33*(5), 1727–1749. https://doi.org/10.1108/intr-07-2022-0563

Spampatti, T., Hahnel, U. J. J., Trutnevyte, E., & Brosch, T. (2023). Psychological inoculation strategies to fight climate disinformation across 12 countries. *Nature Human Behaviour*, *8*(2), 380–398. https://doi.org/10.1038/s41562-023-01736-0

Thuemler, K. R. (2025). The Collapse of Visual Certainty - Implications for Privacy, Surveillance, and Societal Accountability. In *Zenodo (CERN European Organization for Nuclear Research)*. European Organization for Nuclear Research. https://doi.org/10.5281/zenodo.18024038

Tiernan, P., Costello, E., Donlon, E., Parysz, M., & Scriney, M. (2023). Information and Media Literacy in the Age of AI: Options for the Future. *Education Sciences*, *13*(9), 906–906. https://doi.org/10.3390/educsci13090906

Twomey, J. G., Ching, D., Aylett, M. P., Quayle, M., Linehan, C., & Murphy, G. (2023). Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine. *PLoS ONE*, *18*(10). https://doi.org/10.1371/journal.pone.0291668

Uthman, S. (2024). The Erosion of Journalistic Integrity: How AI-Driven Fake News and Deepfakes Complicate Truth Verification in Journalism. *International Journal of Innovative Science and Research Technology (IJISRT)*, 1626–1634. https://doi.org/10.38124/ijisrt/ijisrt24aug1131

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