MEDAAD
Vol. **(2026)**, 2026, **pp**. 39–49
ISSN: 3078-3550

Research Article

The Erasure of Truth? Public Trust in News Media Amid the Proliferation of AI-Generated Video and Deepfakes

Nasr Al-Din Abdul Qader Othman[1, *], Abdu Mohamed Dawood Hafiz[1]

[1] *Associate Professor of Public Relations, College of Mass Communication, Ajman University, UAE*

[2] *Faculty Member, College of Arts and Humanities - Mass Communication in Public Relations Program, Fujairah University, UAE*

ARTICLEINFO

Article History

Received 08 Mar 2026

Revised: 04 May 2026
Accepted 02 Jun 2026
Published 20 Jun 2026

Keywords

Deepfakes,

synthetic media,

public trust,

news media credibility,

AI-generated video,

Misinformation,

media literacy,

epistemic security.

ABSTRACT

Background: The fast development of generative artificial intelligence technology allows machines to create large amounts of synthetic media which consists of machine-generated audiovisual content that raises deep concerns about how people determine what information is true in their public communications.

Purpose: This article studies how AI-generated video content spreads throughout society while simultaneously making people doubt news organizations although the main problem emerges from an overall breakdown of social verification systems instead of individual cases of false information.

Methods/Approach: The article uses critical discourse analysis together with comparative case study methods to analyze empirical data about deepfake trust levels and people's ability to detect them and their production of the "liar's dividend" which describes how people use deepfake knowledge to doubt genuine proof. The study analyzes three distinct situations which include the worldwide voting process of 2024 and the political use of deepfake accusations and the way news organizations handle their source information.

Findings: The research shows that people develop increasing levels of skepticism about synthetic media after multiple viewings which continue to affect them even after scientists prove the media as fake. The 2024 electoral disinformation spread through cheapfakes which appeared as the most common type of content instead of deepfakes with complex features. The liar's dividend presents a greater threat to journalistic authority because it affects public trust in media organizations through longer periods than synthetic content does. Media literacy functions as a protective factor which stops people from losing trust in news media.

Implications: The article introduces the "Synthetic Media Trust Deficit" concept which establishes a new element in media credibility theory while it supports the need for combined efforts between journalism and platform governance and media literacy policy to protect epistemic security beyond technological detection systems.

1. INTRODUCTION

The year 2024 brought a widespread prediction that generative artificial intelligence would create a disastrous conflict with democratic political systems. Multiple national elections will take place throughout sixty jurisdictions which together represent about half of the global population during what experts call the "super election cycle" that includes India, Indonesia, the United States, the European Union, Mexico, and South Africa. Security agencies together with technology firms and election commentators predicted that AI-generated deepfake videos would create an election crisis because these videos would fool voters at a large scale which would destroy trust in election results. The prophesied apocalypse did not appear in its expected form throughout the entire duration. The 2024 election cycle research shows that basic low-tech disinformation operations which people call cheap fakes that spread through encrypted messaging apps and social media platforms had a stronger impact on election misinformation than advanced synthetic video content which mainly targeted developing countries that lacked robust media institutions and verification systems [1]. The election process in Namibia faced challenges because WhatsApp users shared old images which they incorrectly described to others. The Ecuador election faced hybrid information warfare threats through attackers who used both actual information leaks and fake content

*Corresponding author email: n.ali@ajman.ac.ae
DOI: https://doi.org/10.70470/MEDAAD/2026/004

to disrupt the process. The public in Germany experienced no major loss of trust when synthetic media incidents emerged in their country [1].

This apparent anticlimax should not be mistaken for reassurance. The central argument of this article is that the deepfake threat was misdiagnosed from the outset. Artificial intelligence generated video content creates the biggest danger because it produces synthetic content which fools viewers during single viewing instances. The current situation leads to an emergency because people experience an epistemic breakdown which destroys all social verification methods that help citizens determine reality [2]. When any video may plausibly be synthetic, the evidentiary status of all video becomes negotiable. People in today's world doubt what they see but their skepticism about truth remains based on random factors which have no connection to actual facts. The public now doubts what they see because they use audiovisual evidence as their main validation tool which used to be journalism's strongest evidence.

The legal and communication fields use the term "liar's dividend" to describe this particular situation. Chesney and Citron describe the strategic advantage which dishonest actors gain from public knowledge about deepfakes because people now doubt genuine wrongdoing recordings by treating them as fake content. The liar's dividend creates a negative impact on synthetic media that supports an opposite effect to what people normally expect from these media types. The fake video itself does not create the most severe damage but the fake claim which pretends to detect fake content when it targets real evidence does. The liar's dividend reduces the trustworthiness of authentic information while traditional disinformation methods work by introducing bogus data into the system. The first problem shows potential for correction but the second issue destroys the entire system which enables correction to happen.

News media function as the central point which drives this major social transformation. The social purpose of journalism depends on a sequence of epistemic trust which starts when people believe journalists have confirmed their news content and journalists base their trust on evidence which includes documents and testimony and most importantly recorded sound and visual materials for their verification process. The two links become the target of synthetic media attacks which attack them together. The situation contaminates the evidence which journalists use for their work while it creates a permanent reason for both audiences and hostile political groups to doubt all journalistic content. Research studies have shown that public confidence in news organizations has steadily decreased throughout democratic nations during past decades. The article investigates how AI video content expansion affects public trust in news media while it transforms the current downward trend into either faster declines or different patterns of media trust.

Three research questions structure the inquiry:

- RQ1: How does repeated exposure to AI-generated video content affect public trust in traditional news media?
- RQ2: What is the role of the "liar's dividend" in undermining journalistic authority and the evidentiary status of authentic audiovisual material?
- RQ3: How, and under what conditions, can media literacy and detection or provenance frameworks mitigate trust erosion?

The questions hold importance which stretches past the boundaries of media studies. A society needs to achieve what I call epistemic security for democratic self-government to work because it must produce dependable information which people can share and accept for making decisions together [3]. The social proof of audiovisual documentation enables people to determine their authenticity which supports elections and public health campaigns and judicial proceedings and historical memory preservation. The argument collapses because the warrant fails or because people start to doubt its effectiveness which turns post-truth into an actual system instead of a rhetorical device. The main civilian organization which protects that warrant exists through journalism even though journalism contains its own set of flaws. The investigation into synthetic media effects on public trust toward journalism requires analysis of democratic epistemology systems which go beyond simple audience reactions to the media sector.

This article presents the problem through a particular perspective which holds significance. Scientists and policymakers continue to struggle with technological determinism because they believe new technology functions as an automatic system which produces specific social outcomes through its operational capabilities. The evidence from 2024 shows that the initial assumption does not stand. Different countries experienced diverse levels of impact from synthetic media because they shared similar technological advancement but their media systems and institutional trust levels and platform control and civic knowledge about media influenced how synthetic media affected them [1], [4]. A complete theory about deepfake technology needs to study how social elements connect with technological systems instead of focusing only on technological aspects.

The article proceeds as follows. The second section presents a literature review which studies four main areas including deepfake technology development and audiovisual deception psychology and the liar's dividend and epistemic distrust and media trust during the algorithmic age. The theoretical framework of Section 3 emerges through the combination of epistemic security theory with media trust theory and cognitive heuristics which positions the study between technological determinism and social construction of technology. The research methodology in Section 4 employs critical discourse analysis together with comparative case study methodology for its analytical framework. The fifth section contains three case studies which include the 2024 election cycle and the liar's dividend in political communication and the methods newsrooms use to identify sources and track information origins.06 The research findings come together in Section 6 while

Section 7 presents the Synthetic Media Trust Deficit concept which explains its effects on journalists and platforms and policymakers and specifies the research area restrictions and potential study paths.

2. LITERATURE REVIEW

2. 1 The Evolution of Deepfakes: From Entertainment to Information Warfare

The term "deepfake" a portmanteau of "deep learning" and "fake" entered public circulation in late 2017, when a Reddit user of that name distributed face-swapped pornographic videos produced with open-source machine learning tools. The core technology stems from computer vision research which began its development during the early phases of this field. Researchers introduced Generative adversarial networks (GANs) in 2014 to create an adversarial system which trains two networks against each other until they generate fake images that both the discriminator network and human viewers find indistinguishable from real photographs. The development of synthetic media generation has evolved through multiple architectural improvements which started with autoencoder-based face-swapping and neural voice cloning and reached its peak with diffusion models that enable text-to-video synthesis. The process of synthetic media generation has evolved through architectural improvements which now let users create content using standard applications and free software libraries without needing expert knowledge or extensive data processing power [5], [6]. The rapid growth of generative AI has spread its capabilities to the public before any rules or systems for tracking or controlling it could be established which Hagendorff describes as an expanding difference between what AI can do and what people can manage [7].

The analytical process requires clear identification between deepfakes and cheapfakes because deepfakes represent AI-generated content which AI systems modify extensively while cheapfakes consist of simple deceptive audiovisual content created through basic editing tools and speed adjustments and content redefinition and incorrect text overlays [4]. The technical differences between these two elements create effects which affect how people view them and how government officials create their official policies. Hameleers' research tested political content through two methods which showed that cheapfakes achieved equal or higher persuasive power than deepfakes because they used original footage which showed its true origin even though it was presented in misleading ways [8]. The research made its point through actual data from the 2024 election cycle because studies which compared different countries showed that cheapfakes became the main source of electoral disinformation in Global South elections including Namibia's election where WhatsApp networks spread miscaptioned content that fact-checkers and platforms could not handle [9]. The predicted surge of advanced synthetic video appeared in few isolated cases yet basic video manipulation techniques generated more widespread impact throughout the world at levels which could be measured by researchers [10].

The strategic register of synthetic media evolved from their original entertainment and harassment functions to their current use in information warfare operations. Alanazi and his team documented the spread of deepfake technology from its original use in creating non-consensual intimate images which represent the majority of deepfake content to its current applications in political disinformation and financial fraud and state-backed influence activities [11], [12]. The researchers from multiple fields studied social effects which showed that different countries maintain separate regulatory systems which react to events instead of predicting them and struggle to balance free speech with protective measures [13]. Gregory who writes about human rights documentation explains that the same synthesis abilities endanger "frontline witnessing" which represents the evidentiary methods that citizens and activists and journalists use to document abuses and that distant viewers accept as credible. The deepfake era threatens two main problems because it affects how voters choose candidates and it jeopardizes the entire system which supports public surveillance.

2. 2 The Psychology of Deception: Credibility, Realism, and the Uncanny Valley

People tend to find audiovisual manipulation more effective because it uses various methods which create distinct emotional reactions. Communication psychology research shows that people follow the "realism heuristic" according to Sundar because they automatically trust information which they receive through channels that duplicate their natural sensory experiences [14]. Video reaches the peak of this heuristic which causes users to respond most strongly. People who lack advanced knowledge understand text as something which someone has created but they view moving images with matching sounds as direct evidence of actual occurrences. The experimental results from Hameleers show that people trust manipulated videos more than text-based disinformation which produces similar false information because videos seem more real to low-information viewers who don't have enough background to doubt what they see [4].

People believe in realism but their belief system does not follow a simple path. The Utah Valley University Center for National Security Studies has documented research which shows that people develop automatic detection responses to AI-generated faces and voices through their physiological and biometric reactions which show signs of discomfort similar to the "uncanny valley" effect although they maintain their belief in the content's authenticity [15]. Two essential effects result from the discrepancy which exists between what people consciously judge and their automatic responses which occur without their awareness. First, it suggests that human perceptual systems retain some residual sensitivity to synthesis artefacts, a sensitivity that training might amplify. Second, and more troublingly, it implies that conscious credibility

judgements are unreliable guides to the actual detection capacities audiences possess, since the two operate on partially independent tracks [16].

Köbis, Doležalová, and Soraperra conducted experiments which proved that people fail to identify deepfakes correctly yet they show an unjustified belief about their ability to spot these fake videos [17]. The authors describe a common pattern of overconfidence through their study which shows people achieved random results in spotting fake videos but they maintained strong belief in their video identification skills [18]. People experience deception through video content which leads them to believe their detection abilities function perfectly when in reality they have no actual skill. Self-assessed detection ability creates a paradox because individuals either become more suspicious or more trusting based on their confidence in spotting deception. People who think they can detect fake information tend to avoid checking with official sources while they ignore fact-checking efforts which makes them more likely to fall for the exact same deceptions they believe they can detect [19].

Synthetic media delivers cognitive effects which extend beyond the initial evaluation period. The MIT Media Lab conducted research which shows AI-generated content leads people to create synthetic details in their memories of events although they should know these details do not exist [20]. Deepfake "resurrections" research shows that viewers perform active interpretation of synthetic media content through multiple stages of understanding which include checking for authenticity and assessing moral considerations and decoding social meanings [14]. Ahmed's study about accidental deepfake distribution reveals that people spread synthetic content without knowing its nature because their political interests and mental skills and social connections determine their behavior which leads to content distribution patterns that depend on viewer traits and message elements [10].

2. 3 The Liar's Dividend and Epistemic Distrust

The concept of the liar's dividend, introduced by Chesney and Citron in their foundational legal analysis, names the perverse benefit that public awareness of deepfakes confers upon dishonest actors [7]. The public now understands that video content can be altered which leads to decreased trust in genuine recordings; politicians who appear on video while committing crimes now have the power to claim their footage was manufactured through technology that did not exist before. The dividend expands when more people become aware of it because every news story about deepfake technology and media literacy campaign which tells people to doubt their eyes provides both honest people with doubt and dishonest people with ways to deny their actions [7]. Public understanding of the initial damage creates a secondary harm known as liar's dividend which forms its own unique educational challenge.

The consequence, this article argues, is a "trust recession": a generalised depreciation in the credibility of all audiovisual evidence, authentic and synthetic alike. Scientists have identified this pattern because they have started to recognize its existence through their research findings. The research by Weikmann, Greber, and Nikolaou shows that people who watch deepfakes develop stronger doubts about what they know which spreads to all political audiovisual content [20]. Murtuza and Oliullah's research shows that political deepfake viewing produces effects which affect how people process information and evaluate following content instead of directly altering their existing beliefs [19]. The research by Barari, Lucas, and Munger shows that political deepfake content achieves similar credibility levels with other fake media content and actual media content under particular circumstances which reveals deepfake content harms the evidence system more than it enhances its persuasive power [9]. Deepfake technology operates independently from traditional disinformation when it executes its destructive impact on public media trust which serves as the foundation for all media content evaluation.

2. 4 Media Trust in the Algorithmic Age

The synthetic media crisis arrives atop an already depleted reservoir of institutional trust. The Edelman Trust Barometer tracks longitudinal indicators which show that public trust in news media has steadily decreased throughout the past two decades while political beliefs now determine trust levels more than media institutions do. Lazer and his team conducted their vital fake news research to prove that information systems underwent a core restructuring which removed traditional media controls and allowed algorithms to choose content distribution and made viewers' attention time the main economic factor which determines financial returns. The system generates misinformation as a natural system characteristic because false information spreads to more people at a slower pace than the original false information which started the spread.

The combination of algorithmic curation systems with the realism heuristic creates circumstances which make users more vulnerable to negative impacts. Sundar presents a psychological framework about human-AI interaction which shows people now experience content through automated systems that use hidden selection methods to choose their selections. The assessment of machine agency identification signals shapes how people determine content reliability according to Sundar's research [11]. The historical heuristics which audiences used for content evaluation have become vulnerable to exploitation because recommendation systems which focus on user engagement promote emotionally intense audiovisual material that includes synthetic content [4], [11], [16]. The interdisciplinary study by Godulla, Hoffmann, and Seibert discovered that communication research has yet to fully adopt these fields because it concentrates on deepfake detection methods instead of studying their influence on media system trust [8]. The current research answers this challenge by

shifting its focus from individual artifacts to entire ecosystems while studying how trust evolves instead of focusing on deceptive practices.

3. THEORETICAL FRAMEWORK

3.1 Epistemic Security, Media Trust, and Cognitive Heuristics

The research study draws from three theoretical resources which combine to form a unified analytical system. The first is epistemic security theory, which conceptualises a society's capacity to produce, circulate, and validate reliable knowledge as a form of critical infrastructure analogous to energy grids or financial systems [2]. The system which validates claims encounters threats that go beyond simple false information because its fundamental structure of norms and institutions and technologies which give claims their trusted status becomes damaged. Johnson uses a systemic approach to study how deepfakes affect digital media trust by focusing on verification system impacts instead of tracking how many people fell for these synthetic media content [2]. Epistemic security theory thus reframes the research questions: the dependent variable of interest is not belief accuracy but trust calibration—whether publics extend and withhold trust in ways that track actual reliability.

The second resource is media trust theory which explains that people develop trust in news media through multiple dimensions which include their beliefs about accuracy and their assessments of intentions and their competence levels which build up from multiple encounters yet decrease when they experience betrayal. The tradition maintains an uneven trust system which requires extended periods to develop but collapses instantly because any single failure in an outlet or artifact damages the entire institutional category's trustworthiness. Organizations use synthetic media to take advantage of these unbalanced situations between various groups. The entire journalistic industry suffers from trust losses when a news brand faces a viral deepfake attack or when an actual news report gets dismissed as a deepfake.

The third resource is the cognitive heuristics tradition, particularly the realism heuristic and metacognitive overconfidence documented in Section 2.2 [4], [11], [12].The micro-level trust dynamics which heuristics describe became the foundation for understanding how trust operates at the macro level: people base their trust on specific signals and their ability to identify these signals because complete verification exceeds human mental limits and synthetic media disrupts these signals at rates which audiences struggle to match.

3.2 Between Technological Determinism and Social Construction

The article bases its main theoretical framework on the rejection of technological determinism which predicts that generative model features will create direct social effects. The evidence from 2024 fails to support determinism because identical synthesis abilities appeared worldwide but different countries experienced different political effects from these abilities [1]. The social construction of technology (SCOT) framework shows how artefact effects develop through their interactions with social institutions which consist of media systems and political traditions and regulatory frameworks and digital literacy access points [8]. The research team led by Godulla created an interdisciplinary framework which reveals how deepfake effects develop through mediating systems while they warn communication scholars to avoid adopting deterministic views which dominate technical and policy discussions [8].

The concepts of determinism and constructivism exist outside of the need to be treated as opposition to each other. The author supports a soft determinist approach which shows that generative AI produces new possibilities though human choices determine when negative outcomes will occur. The Deepfake Risk Matrix stands as the operational basis of this position through research which CIVICUS Digital Democracy Initiative conducted to break down risk elements into five separate but connected aspects which include political-institutional and social and economic and digital ecosystem and actor behaviour dimensions [1]. The matrix operates as an analytical tool in this article which Section 5 uses for case comparison purposes to determine how synthetic media stimuli generate different trust results based on the arrangement of five key dimensions. The upcoming 2024 cases will serve as a natural experiment to test the accuracy of these prediction models.

Figure 1 shows how these theoretical resources connect to form a conceptual model which this article calls the Synthetic Media Trust Deficit. The independent variables consist of exposure frequency together with detection difficulty and media literacy which affect public trust in news media through the mediating mechanism of generalised epistemic distrust which operates as the liar's dividend at the population level. The research demonstrates that political polarization together with platform algorithmic amplification systems affect the process through which generalized epistemic distrust functions as the liar's dividend in public trust towards news media.

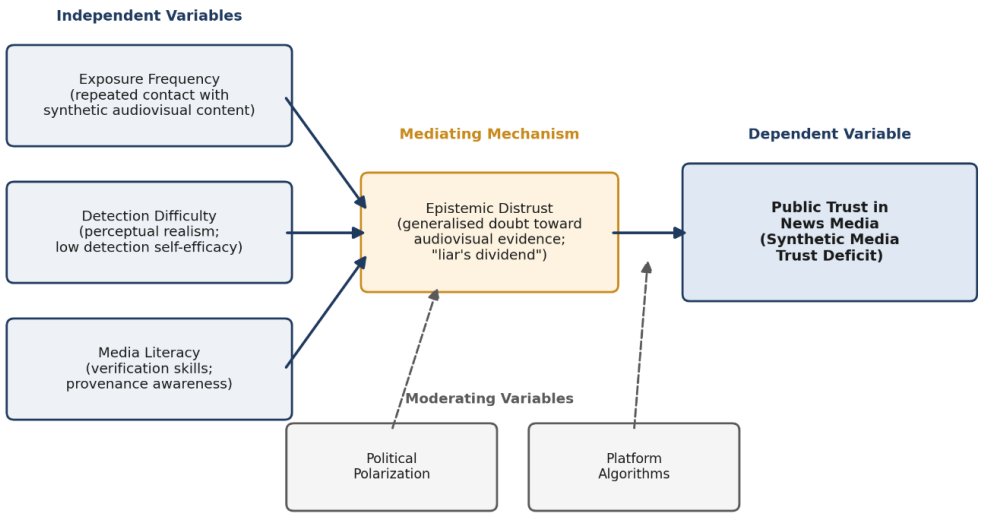

Fig. 1. Conceptual model of the Synthetic Media Trust Deficit. Exposure frequency, detection difficulty, and media literacy influence public trust in news media through the mediating mechanism of epistemic distrust, moderated by political polarization and platform algorithms.

4. ANALYTICAL APPROACH

The research paper uses a theoretical-analytical approach instead of collecting original data through primary empirical methods. The study employs critical discourse analysis (CDA) to analyze news content and policy discussions while using comparative case study methods. The decision rests on three separate foundations.

The research focuses on a system-wide process which studies how people change their belief systems about knowledge throughout various institutions and digital platforms and public spaces during different time periods. Researchers need experimental designs to prove causal relationships between exposure and micro-level effects but these designs remove social systems from their experimental environment which leads to social trust patterns. Case study methodology enables researchers to study aggregation through fieldwork which shows how different social environments create different results from the same technological features [1],[8].

The liar's dividend operates through media claims which include both fake content allegations and real information confirmation as well as discussions about our inability to verify truth in media [7]. CDA functions as the correct tool for this analysis because it studies how actors use specific terms like "deepfake" and "AI-generated" and "manipulated" to change the meaning of evidence while they assign trustworthiness and avoid taking responsibility. The research used news reports and platform policies and political statements about deepfake occurrences and deepfake accusations which spread between 2023 and 2025 from the Section 5 case studies.

Research into deepfake reception practices uses the comparative case design which follows established research methods. Barari, Lucas, and Munger's experimental study shows that synthetic media evaluation becomes more effective when researchers compare these media against existing fake content and authentic content instead of studying them independently [9]. The current design applies the same comparative method to study institutional structures through three separate cases which examine national environments and communication approaches and organizational conduct. The selection of cases follows a most-different-systems approach for Case 1 which includes Namibia and Ecuador and Germany because these countries show maximum variation across Deepfake Risk Matrix dimensions yet they all face election cycle risks for 2024 [1]. The selection process for Cases 2 and 3 depends on specific criteria which identify proven instances of liar's dividend application and major newsroom institutional reactions to these events.

The limitations of this approach are acknowledged at the outset. The analytical case study method fails to calculate specific causal effect sizes but it reveals the underlying mechanisms and their established configurations which demonstrate their feasibility. The research results receive their definitions from theoretical frameworks which connect to the experimental and survey literature that Section 2 presents instead of forming statistical conclusions. Researchers need to perform longitudinal studies across different cultures to develop standardized testing methods which will transform these research findings into scientific hypotheses.

5. DISCUSSION AND CASE STUDIES

5.1 Case 1: The 2024 Global Election Cycle Namibia, Ecuador, Germany

The 2024 election cycle functions as an experimental space which enables researchers to study the effects of media on voter choices according to the mediated-determinism thesis. Three national cases, documented in the CIVICUS Digital

Democracy Initiative's cross-national research, illustrate how similar technological exposure produced sharply divergent epistemic outcomes [1].

- Namibia shows the typical pattern of cheapfake election manipulations which occur throughout the Global South. The main platform for spreading electoral disinformation in the country proved to be WhatsApp which users received through incorrect image captions and repeated unrelated video content and basic audio manipulation without any need for artificial intelligence content generation [1]. The platform operated through encrypted peer-to-peer systems which blocked fact-checkers and platform moderators from seeing the content while the media industry in the country lacked sufficient staff to handle the fast-paced and extensive information spread. The Namibian case suggests that where verification infrastructure is weak, adversaries have no need for technical sophistication: the cheapest manipulation suffices, because the corrective system that sophisticated fakes are designed to defeat barely operates in the first place [1], [4]. People lose trust in these situations because they have encountered so many fake things that they doubt all information in their surroundings.
- Ecuador shows a combination of information warfare strategies which merge artificial content with real information during extensive operations that include both information leaks and targeted attacks and organized support systems [1]. The Ecuadorian case shows a vital analytical feature because organizations actively create confusion between genuine and fake information which they use to establish their operations. The operations of the liar's dividend strategy focus on creating evidence confusion which prevents people from understanding any specific truth [1], [7]. The process of verification for authentic recordings becomes difficult because Ecuadorian journalists face constant allegations of fabricated content which drains their resources that should support their news gathering activities.
- The German system shows strong institutional stability when compared to other systems. The AI-generated audio content which targeted political figures during Synthetic media occurrences did not create any significant impact on how people voted or their overall trust in media sources [1]. The Deepfake Risk Matrix shows that public service media in combination with fact-checking networks and low affective polarisation and elite norms and platform accountability systems create the conditions which help CIVICUS maintain its current state [1]. Resilience developed through different element combinations because no individual element showed enough strength to stand alone.

The comparative lesson is stark: the three cases show different results which cannot be explained by their equal access to technology because their technological resources stayed unchanged yet their media freedom and institutional trust and platform governance and civic verification systems differed [1], [8]. Media systems determine the effects of deepfake content before deepfakes start to affect their systems. The study produces results which disprove determinist views because it supports the SCOT-based theoretical framework from Section 3 which shows how the same artefact class created knowledge disruption in one setting but handled changes effectively in another environment.

5.2 Case 2: The Liar's Dividend in Political Discourse

The second case examines the discursive weaponisation of deepfake awareness by political actors seeking to evade accountability. The pattern, which Chesney and Citron predicted in 2019, now shows up in real cases from different countries because political figures and their teams use the same strategy when dealing with genuine evidence of wrongdoing. The evidence shows politicians and their teams reject AI material but they accept the evidence content while they work to eliminate suspicion about its creation process [7].

Critical discourse analysis of such episodes reveals a recurring rhetorical architecture.

The news industry demands quick answers so the fabrication claim emerges as an immediate and absolute statement which takes advantage of this demand while attackers need only seconds to spread false information but scientists must spend multiple days to demonstrate such claims are false. The claim receives support through technological feasibility because people understand these tools enable voice imitation which helps transform a defensive denial into a plausible scientific doubt according to source [7]. The proof requirement exists as a reversed matter because journalists need to show their information meets forensic standards but the person making the accusation remains free from any need to prove their false claim. Forensic analysis methods reduce the amount of remaining uncertainty but they will never achieve complete elimination of all uncertainties. The organization uses their inability to prove document authenticity as their main evidence to support their case.

The architecture will perform based on how audiences approach information according to the experimental research findings. People tend to think they can identify fake content better than they actually do [12] which leads to increased public mistrust of all audiovisual information [19], [20]. The claim of fabrication finds ready acceptance among viewers who already doubt videos because they find the denial easy to understand through their cognitive processes. People accept denials of fake videos because their mental processes match their political beliefs which drive their decision-making. The evidence acceptance of out-partisans along with their co-partisans who reject the evidence as fake illustrates how the liar's dividend grows through the process of polarisation which serves as the moderating variable in the Figure 1 model.

Journalistic authority faces direct impacts which build up to create long-term consequences for its institution. The liar's dividend creates a negative impact on investigative journalism because it reduces the value of their core product which consists of proven evidence about important misconduct when they deliver their work. The dismissal process creates legal

authority which allows organizations to repeat the same process for different situations. Gregory shows that the most severe damage to authenticity infrastructure exists in the practice of citizens and journalists who record events for accountability purposes because society accepts these recordings as valid evidence. The practice of journalism loses its ability to hold people accountable because people now take away their recognition through unprovable accusations instead of detecting actual deception during fact-checking processes. The article presents its main argument that the liar's dividend presents a greater danger than deepfake technology does.

5.3 Case 3: Newsroom Responses and Detection Frameworks

The third case studies how institutions protect themselves through their established methods which track content origins and identify fake information while these practices create problems for official oversight organizations.

The BBC along with Reuters and other international news organizations now use provenance as their main approach which involves cryptographically secured metadata that attaches to content when it first appears and stays intact throughout the editorial process. The main tool for this purpose exists in the Coalition for Content Provenance and Authenticity (C2PA) standard which stores signed origin and capture device and edit information directly into media files to enable content tracking through subsequent processing stages. The BBC uses provenance initiatives together with Reuters which applies capture-time authentication during news gathering operations to create a new method for solving this problem. The BBC and Reuters now use proven systems to validate real content which changed the way they handle fake news detection from their previous adversarial system that failed to match new fake content [17] and [13] and [15].

The planned provenance strategy shows theoretical value but people have different levels of interest when it comes to using this approach. The security system protects only content which originates from authorized production lines while most audiovisual content that people share remains unprotected because it lacks signatures which threatens to make citizen evidence from underfunded areas disappear into digital oblivion [15]. Gregory expresses concern that the implementation of authenticity infrastructure without proper accessibility measures will create problems for the frontline witnesses who need protection because institutional producers will gain control over evidence distribution [15].

News organizations have started using AI detection systems which operate like provenance systems but these tools show decreasing effectiveness in their current applications. The detection systems encounter their worst performance when they face new generation systems together with adversarial system modifications and social media distribution artifacts which result from compression processes [17], [18]. The research from Center for News Technology and Innovation has unveiled a hidden danger which platform labelling systems present through their selective labelling process. The research shows that unlabelled false content creates an "implied truth" problem because the system treats such content as verified information which leads to the "false information untagged" paradox. The research shows that audiences develop better understanding of content through unlabelled content compared to labelled content [13]. Complete provenance disclosure which organizations should apply uniformly has become the most effective method to replace warning labels according to the emerging best-practice consensus [13].

The final regulatory system creates an opposition which exists between efforts to stop artificial disinformation and the need to defend free speech rights. Singapore maintains strict control over election periods through its ban on digitally altered content which misleads voters about candidates while the government holds power to remove false information and force corrections [1]. The systems enable quick elimination of harmful synthetic content yet they grant governments the power to determine what constitutes truthful political communication which could lead to abuse of power by current rulers who control limited media freedom regions [1], [18]. The comparison between different literary works shows that the solution to this problem remains absent in complete theoretical form because the same legal instrument functions as knowledge protection in one institutional structure yet serves as censorship tool in another [1], [8]. Regulation exists as a social construct which shapes the technological systems it governs.

The table presents a unified view of deepfake and cheapfake characteristics which merge technical data with perceptual effects and trust-related effects from this research and previous Section 2.

TABLE I. COMPARATIVE MATRIX OF DEEPFAKE AND CHEAPFAKE CHARACTERISTICS.

Dimension	Deepfakes (AI-generated)	Cheapfakes (low-tech manipulation)
Production technology	GANs, autoencoders, diffusion models; voice cloning; text-to-video synthesis [3], [17]	Selective editing, speed alteration, miscaptioning, recontextualisation of authentic footage [4]
Skill and cost barrier	Falling rapidly; consumer applications and open-source tools [3], [17]	Minimal; standard editing software or none
Scale of observed use (2024 elections)	Present but episodic	Dominant, especially in Global South contexts via encrypted messaging [1]
Perceptual realism	High and increasing; may trigger non-conscious uncanny valley responses [5]	Variable; exploits authentic footage, so provenance cues remain genuine [4]
Human detectability	Near chance; strong overconfidence in self-assessed detection [12]	Moderate; depends on contextual knowledge rather than perceptual inspection [4]

Machine detectability	Adversarial arms race; degrades against novel generators and compression [17], [18]	Poorly suited to forensic detection; manipulation is contextual, not artefactual
Primary deception mechanism	Fabricated depiction of events that never occurred	Deceptive framing of events that did occur [4]
Trust impact	Generalised epistemic distrust of audiovisual evidence; fuels liar's dividend [7], [19], [20]	Incremental credibility erosion; cumulative unverifiability of information environment [1]
Principal countermeasures	Provenance standards (C2PA), detection tools, disclosure labelling [13], [15]	Contextual fact-checking, media literacy, platform friction [1], [16]

6. FINDINGS AND ANALYSIS

The research evidence from Section 2 literature review combines with Section 5 case studies to produce four essential findings which take the form of theoretically structured propositions.

Finding 1: Deepfake videos create permanent damage to media trust which stays strong even after people discover the videos are fake. Research findings from experiments and surveys show that synthetic media exposure creates trust issues which correction methods fail to resolve. People tend to doubt all political information which appears in audiovisual format after they learn about its existence [19], [20]. People tend to forget which parts of AI-generated content they discovered to be false when they first learned the information [6]. The process of debunking fake news requires people to review deceptive content which leads to increased credibility through the illusory truth effect because repeated exposure makes information easier to process which people mistakenly believe means it is true [10], [16]. The fact-checking system produces an unpleasant outcome because it requires verification yet fails to build public confidence which simultaneously reduces its own monitoring effectiveness. Media trust theory demonstrates that general suspicion does not disappear after people receive correct information about the original item which caused their suspicion. The Ecuadorian case study shows that disputed situations led to permanent evidence stalemates which did not convert into short-term information gaps [1].

Finding 2: Media literacy functions as a moderating variable higher literacy is associated with maintained trust in verified sources rather than indiscriminate scepticism. People will definitely doubt information because of synthetic media but their level of distrust will depend on how they choose to handle their doubts. The evidence indicates that media literacy, understood as procedural knowledge of verification (provenance checking, source triangulation, awareness of manipulation techniques) rather than mere awareness that fakes exist, channels scepticism toward discrimination between verified and unverified sources [11], [16]. The human-AI interaction framework developed by Sundar shows how literate people replace their failed realism heuristic with two new heuristics which help them keep their trust levels separate [11]. The combination of threat-based awareness campaigns which lack verification procedures will create general public distrust which liars use to their advantage [7], [12]. The German population demonstrates proper use of scepticism throughout their entire society but Namibian people do not show this behavior [1]. Media literacy serves as the main factor which enables technical countermeasures to achieve their desired trust results.

Finding 3: Research shows that deceptive activities will harm journalistic work more than deepfake technology will in the future. The main analytical argument of this article emerges from the following collection of supporting evidence. Deepfake videos function as deceptive media content which people find equally believable as other fake content but they do not create a new way to influence people through persuasion [9]. Their observed electoral role in 2024 was secondary to cheap fakes [1]. The deepfake period has developed its own unique type of damage which affects communication because people now use the fabrication claim to doubt genuine evidence that appears without any delay [7]. The claim operates through four rhetorical techniques which Case 2 demonstrates by showing how it denies categories quickly while building technological support and shifting responsibilities and creating doubt which leads to increased political polarization. The fabrication claims functions differently from deepfakes because it does not need production or distribution or security checks yet it gains strength from each new generative technology and every educational program about these technologies [7], [12]. The attack focuses on journalism at its most vital social point which involves confirming actual instances of misconduct. The detection-based threat model for fake video identification causes institutions to direct their resources to the wrong problem because their main challenge is protecting the credibility of authentic video evidence.

Finding 4: People believe they possess deepfake detection skills but they actually lack the ability to identify these fake videos. The research study by Köbis, Doležalová, and Soraperra shows that people fail to detect patterns at random levels while they still maintain high levels of confidence [12]. Biometric data shows that people develop automatic detection responses which their brain does not convert into conscious thoughts thus revealing their hidden ability to perceive things which they cannot recognize or move [5]. The systematic failure of metacognition produces various effects which affect multiple parts of the system. The public shows little interest in institutional verification while they accept viral content blindly and unknowingly spread fake material through their social media activities which depend on their political interests and their ability to think and the number of people they connect with [10], [12]. The liar's dividend grows because citizens who doubt their ability to identify deception will ignore proven facts that challenge their existing beliefs through their individual evaluation process. The evidence shows that audiences fail to use detection knowledge effectively so they need

to learn how to align their confidence with their actual performance instead of receiving detection information which they cannot use in practice [12],[5].

The research results show a progressive development instead of an all-or-nothing system. The social system which verifies truth undergoes continuous decline because different media systems operate differently in their operational environments [1], [8]. The Figure 1 model demonstrates that exposure and detection difficulty levels affect the development of generalised epistemic distrust which media trust results from polarisation and algorithmic amplification through their influence on the mediating state.

7. CONCLUSION

This article investigates how the fast development of AI-generated videos impacts public trust in news media which leads to public belief in truth. The research shows that truth continues to exist because the predicted "deepfake apocalypse" has not become a reality. The use of advanced AI video content remains uncommon in political messaging because different nations with strong media networks have successfully handled individual occurrences which did not harm public trust in their institutions. The research shows that people have started to lose their ability to spot authentic audiovisual evidence but they have not lost their ability to identify fake evidence. People develop mistrust about information because synthetic media creates new ways to deceive them which strengthens the "liar's dividend" that lets fake evidence replace real evidence.

The article explains the Synthetic Media Trust Deficit through its definition which shows how people trust news media when they have proven audiovisual evidence but their trust decreases when synthetic content enters their information space. The research extends traditional media credibility theory through its new approach which shows that audience trust levels can decrease even when people correctly identify fake and real content. The primary issue exists in two ways because individual facts show errors and people across society doubt all news sources.

The research reveals multiple elements which drive public distrust of AI technology because people encounter AI-generated content at different rates and they face rising challenges to identify altered material while their ability to understand media and their political beliefs and social media algorithms also play a role. The way different elements interact determines how people assess news trustworthiness while they decide how information systems maintain their stability.

The research findings led the article to develop multiple operational recommendations which help organizations move forward. Journalists need to establish open verification methods together with dependable content origin tracking systems and they must show viewers every stage of AI-based content production. Digital platforms need to create standardized labeling systems which show content origins to decrease user confusion but they should avoid selective labeling because it might create false approval for unmarked deceptive material. The government needs to establish rules which defend society from AI-based criminal activities including fraud operations and election interference and unauthorized synthetic media production but must avoid creating excessive state interference in political information. The article demonstrates media literacy as an essential educational asset because schools should teach students verification techniques and critical thinking skills instead of basic knowledge about AI-generated threats.

The research paper establishes theoretical conclusions through comparative analysis instead of conducting extensive empirical data collection. The study suggests researchers should conduct longitudinal studies with different cultures and intervention programs to analyze how synthetic media affects public trust throughout extended time periods. The article shows that truth remains intact but people now doubt audiovisual evidence which creates an urgent need for both public institutions and society to maintain their trust.

Funding:

The authors declare that no specific financial aid or sponsorship was received from governmental, private, or commercial entities to support this study. The research was solely financed by the authors' own contributions.

Conflicts of Interest:

The authors declare that there are no conflicts of interest in this study.

Acknowledgment:

The authors express their heartfelt appreciation to their institutions for the essential support and motivation provided throughout the research period.

References

[1] CIVICUS Digital Democracy Initiative, Future-Proofing Elections Against Deepfake Disinformation. CIVICUS Research Report, 2025. [Online]. Available: https://www.civicus.org/documents/ddi/final-deepfakes-elections-report-civicus-ddi-research-2025.pdf

[2] E. P. Boediman, "Exploring the impact of deepfake technology on public trust and media manipulation: A scoping review," Jurnal Komunikasi, vol. 19, no. 2, pp. 131–152, 2025.

[3] T. Hagendorff, "Mapping the ethics of generative AI: A comprehensive scoping review," Minds and Machines, vol. 34, Art. no. 39, 2024, doi: 10.1007/s11023-024-09694-w.

[4] M. Hameleers, "Cheap versus deep manipulation: The effects of cheapfakes versus deepfakes in a political setting," International Journal of Public Opinion Research, vol. 36, no. 1, Art. no. edae004, 2024, doi: 10.1093/ijpor/edae004.
[5] Utah Valley University, Center for National Security Studies, Research Summary: Impact of AI Generated Media, 2024. [Online]. Available: https://www.uvu.edu/news/2024/10/media/research-summary.pdf
[6] P. Pataranutaporn, C. Archiwaranguprok, S. W. Chan, E. Loftus, and P. Maes, "Synthetic human memories: AI-edited images and videos can implant false memories and distort recollection," in Proc. CHI Conf. Human Factors in Computing Systems, Apr. 2025, pp. 1–20.
[7] R. Chesney and D. K. Citron, "Deep fakes: A looming challenge for privacy, democracy, and national security," California Law Review, vol. 107, no. 6, pp. 1753–1819, 2019, doi: 10.2139/ssrn.3213954.
[8] A. Godulla, C. P. Hoffmann, and D. Seibert, "Dealing with deepfakes—An interdisciplinary examination of the state of research and implications for communication studies," Studies in Communication and Media, vol. 10, no. 1, pp. 72–96, 2021, doi: 10.5771/2192-4007-2021-1-72.
[9] S. Barari, C. Lucas, and K. Munger, "Political deepfakes are as credible as other fake media and (sometimes) real media," Journal of Politics, vol. 87, pp. 510–526, 2025, doi: 10.1086/729741.
[10] S. Ahmed, "Who inadvertently shares deepfakes? Analyzing the role of political interest, cognitive ability, and social network size," Telematics and Informatics, vol. 57, Art. no. 101508, 2021, doi: 10.1016/j.tele.2020.101508.
[11] S. Sundar, "Rise of machine agency: A framework for studying the psychology of human-AI interaction (HAII)," Journal of Computer-Mediated Communication, vol. 25, no. 1, pp. 74–88, 2020, doi: 10.1093/jcmc/zmz022.
[12] N. C. Köbis, B. Doležalová, and I. Soraperra, "Fooled twice: People cannot detect deepfakes but think they can," iScience, vol. 24, Art. no. 103364, 2021, doi: 10.1016/j.isci.2021.103364.
[13] Center for News Technology and Innovation (CNTI), "Synthetic Media & Deepfakes: Issue Primer," 2025. [Online]. Available: https://cnti.org/issue-primers/synthetic-media-deepfakes/
[14] M. T. Soto-Sanfiel and Q. Wu, "How audiences make sense of deepfake resurrections: A multilevel analysis of realism, ethics, and cultural meaning," Computers in Human Behavior, vol. 159, Art. no. 108822, 2025, doi: 10.1016/j.chb.2025.108822.
[15] S. Gregory, "Deepfakes, misinformation and disinformation and authenticity infrastructure responses: Impacts on frontline witnessing, distant witnessing, and civic journalism," Journalism, vol. 23, no. 4, pp. 708–729, 2022, doi: 10.1177/14648849211060644.
[16] D. M. J. Lazer et al., "The science of fake news," Science, vol. 359, no. 6380, pp. 1094–1096, 2018, doi: 10.1126/science.aao2998.
[17] S. Alanazi and S. Asif, "Exploring deepfake technology: Creation, consequences and countermeasures," Human-Intelligent Systems Integration, vol. 6, pp. 49–60, 2024, doi: 10.1007/s42454-024-00054-8.
[18] S. Alanazi et al., "Unmasking deepfakes: A multidisciplinary examination of social impacts and regulatory responses," Human-Intelligent Systems Integration, 2025, doi: 10.1007/s42454-025-00060-4.
[19] H. M. Murtuza and M. D. Oliullah, "Exploring the impacts of political deepfakes on cognitive processing," 2025.
[20] S. Weikmann, A. Greber, and D. Nikolaou, "Exposure to deepfakes and epistemic distrust," 2024. [Online]. Available: https://publications.civicus.org/publications/future-proofing-elections-against-deepfake-disinformation/executive-summary/