

How does the proliferation of AI-generated deepfakes erode institutional trust in electoral outcomes and reshape accountability structures within democratic governance?

September 7, 2026 | SnugLab Research

Executive Summary

The proliferation of AI-generated deepfakes erodes institutional trust in electoral outcomes by fostering "reality apathy" among voters, who become disengaged due to the overwhelming difficulty of discerning truth from fabrication. Concurrently, deepfakes reshape accountability structures within democratic governance by providing a "liar's dividend," enabling political actors to plausibly deny genuine evidence of misconduct as synthetic, thereby evading responsibility. Evidence suggests both mechanisms contribute to undermining democratic processes, though the extent of widespread deepfake impact on election results to date remains a point of nuanced debate.

Key Findings

Erosion of Institutional Trust

AI-generated deepfakes significantly distort the informational environment, making it increasingly challenging for citizens to distinguish authentic content from fabrications [1, 2, 3, 4, 6, 12]. This corruption of visual content, once a reliable signal of reality, creates cognitive dissonance and fosters pervasive public cynicism [3]. Frequent exposure to synthetic media cultivates "reality apathy," leading individuals to cease attempts at verifying information and adopt an indifferent attitude toward its authenticity [1]. This psychological disengagement impedes informed participation, exacerbates polarization, and diminishes confidence in governance structures [12].

Human detection rates for deepfake images are often low; one study found that "only 60% of deepfake images could be reliably detected by human observers" [1]. This challenge is compounded by the rapid spread of false information, which reaches a cascade depth of 10 about 20 times faster than factual information on platforms like Twitter. Real-world instances include a May 2023 deepfake image of a Pentagon

explosion that caused market turbulence and a July 2024 deepfake video of then-Democratic presidential nominee Kamala Harris [6]. Public concern is high, with 87.4% of UK citizens expressing worry about deepfakes affecting election results [7].

Reshaping Accountability Structures

Deepfakes fundamentally reshape accountability by creating a "liar's dividend," which allows individuals and officials to dismiss genuine evidence of wrongdoing as AI-generated, thereby evading responsibility [4, 6]. This mechanism shifts the burden of proof from the accused to the verifier, effectively decoupling accountability from truth [6]. For example, Elon Musk's legal team has suggested his past statements on Tesla could have been deepfakes [6]. The capacity to generate deepfakes is outpacing detection capabilities, making it difficult to legally track and hold creators responsible [4]. Even when detected, false narratives often dominate the media environment for days or weeks before corrections can take hold [4].

Beyond individual deepfakes, algorithmic governance introduces "black box" opacity and fragmented responsibility among developers, agencies, and vendors, making it difficult to understand decision-making processes and assign blame [8, 9, 10]. As one source notes, "In ever more areas of life, algorithms are coming to substitute for judgment exercised by identifiable human beings who can be held to account" [10]. This insulates power from popular scrutiny and erodes procedural liberties, as "decisions made by algorithm are often not explainable, even by those who wrote the algorithm, and for that reason cannot win rational assent" [10].

Conflicting Evidence on Electoral Impact

While many sources emphasize the significant potential and specific instances of deepfake misuse in elections [3, 4, 6, 11], evidence for widespread political misuse to date remains limited, with only a small number of verified cases [12]. For instance, close examinations of elections in Argentina and Indonesia found no clear impact on election results from deepfakes [5]. This contrasts with the widespread public concern that these technologies will affect future outcomes [7]. Earlier projections indicated that up to 8 million deepfake videos could be shared by 2025 [2].

Implications

The proliferation of AI-generated deepfakes has profound implications for democratic governance by undermining the foundational requirement of an informed citizenry and altering the mechanisms of political responsibility. The erosion of trust through "reality apathy" can lead to voter disengagement and a decline in the perceived legitimacy of electoral outcomes, as the public loses faith in the veracity of information.

Simultaneously, the "liar's dividend" empowers political actors to evade accountability for genuine misconduct, potentially leading to a system where truth is decoupled from consequences. This dual impact challenges the integrity of democratic processes and necessitates robust responses to safeguard informational environments and uphold accountability.

Limitations and Caveats

The research indicates that while the mechanisms by which deepfakes erode trust and reshape accountability are logically sound and supported by evidence, the empirical data on their widespread impact on *electoral outcomes* specifically remains nuanced. There is a distinction between public concern about deepfakes affecting elections and verified instances of deepfakes definitively altering election results. The debate centers on the primacy of "reality apathy" versus the "liar's dividend," suggesting both are co-existing mechanisms rather than mutually exclusive. Direct quantitative data for the global impact of deepfakes on electoral outcomes is still developing, and conclusions should be considered provisional.

Sourcing audit. A check of this report's claims against the source material it was written from could not trace 16 of 17 claims back to a passage in that material. Treat the specific figures and findings above as unverified pending a re-run with a wider source pool.

Sources

[1] [peer-reviewed] When seeing is not believing: self-efficacy and cynicism in the era of intelligent media - Authors: Liu, Qiang; Wang, Lin; Luo, Mengyu - Journal: Humanities and Social Sciences Communications - <https://www.nature.com/articles/s41599-025-04594-5>

[2] [peer-reviewed] AI-driven disinformation: policy recommendations for democratic resilience - Authors: Alexander Romanishyn; Olena Malytska; Vitaliy Goncharuk - Journal: Frontiers in Artificial Intelligence - <https://pmc.ncbi.nlm.nih.gov/articles/PMC12351547/>

[3] [edu] Ai Deepfakes Bias Threaten Elections Healthcare Trust - magazine.tcu.edu - <https://magazine.tcu.edu/endeavors-2026/ai-deepfakes-bias-threaten-elections-healthcare-trust/>

[4] [edu] Is Seeing Still Believing The Deepfake Challenge To Truth In - brookings.edu - <https://www.brookings.edu/articles/is-seeing-still-believing-the-deepfake-challenge-to-truth-in-politics/>

[5] [edu] Ai Enabled Influence Operations Threat Uk General Election - cetas.turing.ac.uk -

<https://cetas.turing.ac.uk/publications/ai-enabled-influence-operations-threat-uk-general-election>

[6] [edu] View - journals.library.columbia.edu -

<https://journals.library.columbia.edu/index.php/stlr/blog/view/669>

[7] [edu] 9 10 Concerned About Deepfakes Affecting Election Results - turing.ac.uk -

<https://www.turing.ac.uk/news/9-10-concerned-about-deepfakes-affecting-election-results>

[8] [edu] Riyanka Roy Choudhury - sloanreview.mit.edu -

<https://sloanreview.mit.edu/panelist/riyanka-roy-choudhury/>

[9] [peer-reviewed] From algorithmic efficiency to democratic legitimacy: rethinking AI governance in the Global South - Authors: Duche-PÃ©rez, Aleixandre Brian; FalconÃ© Picardo, Marco Tulio; ChÃ¡vez

Urquiza, Emmanuel NeptalÃ© Augusto; Vera-Revilla, Cintya Yadira - Journal: Frontiers in Political Science -

<https://www.frontiersin.org/journals/political-science/articles/10.3389/fpos.2026.1875291/full>

[10] Algorithmic Governance And Political Legitimacy - americanaffairsjournal.org -

<https://americanaffairsjournal.org/2019/05/algorithmic-governance-and-political-legitimacy/>

[11] [peer-reviewed] Information Disorder and Global Politics - Authors: Julia C. Morse; Tyler Pratt -

Journal: International Organization -

<https://www.cambridge.org/core/journals/international-organization/article/information-disorder-and-global-politics/76117CC85927164C2BB37099D5BD6E58>

[12] NIJCRHSS 61 P9 2026 - nijournals.org -

<https://nijournals.org/wp-content/uploads/2026/04/NIJCRHSS-61-P9-2026.pdf>