

# Can democratic institutions survive the transition to an era of deepfake-saturated election cycles?

April 16, 2026 | SnugLab Research | [readme.snuglab.com](https://readme.snuglab.com)

---

## Executive Summary

---

Democratic institutions are unlikely to survive the transition to an era of deepfake-saturated election cycles without significant, rapid, and widespread societal adaptation beyond current technical verification standards [21]. While emerging technical solutions offer some mechanisms for managing content authenticity, their slow adoption, limited reach, and the overwhelming scale and psychological impact of deepfakes are projected to erode public trust and democratic stability before these solutions can become universally effective [12, 21]. The "Liar's Dividend," which allows for the dismissal of authentic evidence as AI-generated, poses a more fundamental threat than deepfake circulation itself, by exploiting partisan reasoning to undermine accountability and a shared reality [2, 8, 11, 15].

## Key Findings

---

### The Dual Threat: Deepfakes and the Liar's Dividend

Both deepfakes and the "Liar's Dividend" contribute to eroding trust in democratic institutions, the media, and electoral integrity [1, 3, 5, 10, 11]. Deepfakes involve direct manipulation, such as the 2023 use of voice cloning to falsely suggest a Chicago mayoral candidate condoned police violence [6]. This technology exploits the "realism heuristic," a human tendency to place higher trust in visual and audio information than in text, making it difficult to distinguish authentic from fabricated media [19, 24].

The "Liar's Dividend" is considered a more existential threat to democratic stability because it allows individuals to dismiss authentic, incriminating evidence as AI-generated, exploiting partisan reasoning to undermine accountability [2, 3, 8, 11, 15]. This is achieved through "informational uncertainty" and "oppositional rallying" [7, 8]. Research involving 15,000 American adults found that politicians can increase support across various partisan subgroups by claiming a scandal is "fake news" or a "deepfake"

more effectively than by apologizing or remaining silent [6, 8]. While these claims are effective against text-based reports, their effectiveness against video evidence is debated; some research suggests they are largely ineffective against video, though "oppositional rallying" can, in certain instances, discredit video-based scandals [7, 8, 26]. However, other evidence suggests that advanced digital forensic techniques and human cognitive responses largely mitigate the effectiveness of the Liar's Dividend in discrediting authentic video evidence [33, 34, 35, 36, 37, 38].

The increasing ease, speed, and low cost of deepfake production make the technology accessible to non-experts, creating "synthetic realities" that are hard to audit and can alter perceptions of truth [1, 10, 11, 12, 15]. Both deepfakes and the Liar's Dividend increase the cost of verifying truth for individuals and institutions and exacerbate social polarization [1, 2, 5, 7, 9, 13].

## **Technological Asymmetry and Regulatory Challenges**

The capacity to create deceptive synthetic media is advancing faster than the capacity to detect or regulate it, creating a "technical arms race" [4, 6, 8, 15]. This technological asymmetry renders current regulatory approaches, such as disclosure mandates, largely insufficient to mitigate democratic harm [3, 4, 6, 8, 12, 15]. While some argue for the necessity of new regulations [17], others suggest that "regulation alone cannot change the current trajectory" due to the technological gap [4, 6, 8]. Legislative measures, such as California's AB 2839 and statutes in Texas and Minnesota, have been introduced to require or prohibit deepfakes within certain windows before an election [3, 11, 12]. However, neither technical attribution nor disclosure mandates are considered effective in isolation; a multi-layered defense integrating technology, regulation, and education is required [8, 15, 16, 18, 19].

The impact of rapid, large-scale synthetic media amplification is debated, with some researchers describing an "information apocalypse" and others suggesting "information pollution" more accurately describes the degradation of the information environment [25]. Evidence suggests the situation is better characterized as manageable information pollution that democratic systems can adapt to, given ongoing technological and standardization efforts [12, 28, 29, 30, 31, 32]. However, the scale of the threat is significant: false news diffuses "significantly farther, faster, deeper, and more broadly" than true news [7, 6]. Deepfake-driven fraud resulted in over \$200 million in financial

losses during the first quarter of 2025 [3]. U.S. fraud losses driven by generative AI are projected to grow from \$12.3 billion in 2023 to \$40 billion by 2027 [15, 17].

## **Limitations of Technical Solutions**

Technical protocols like C2PA and watermarking, while intended to verify authenticity, possess inherent vulnerabilities that can be exploited by bad actors, creating a "false sense of security" [1, 2, 15]. C2PA is susceptible to metadata stripping through common processes like re-encoding files or uploading them to social networks [13, 14]. It is also vulnerable to omission attacks, meaning it cannot prove that related files from the same capture session were not removed to manipulate context [27]. Furthermore, C2PA validates signature validity without verifying the honesty of the signer [22].

Watermarks are susceptible to being stripped or weakened by compression, resizing, or adversarial attacks [6, 12, 18]. While watermarking can be a vital and reliable component within a defense-in-depth strategy, despite these vulnerabilities [39, 40, 41, 42, 43, 44], these technical failures create "unverifiable" states that bad actors can exploit to dismiss authentic media as "unverifiable" or fake [1, 4, 13, 14, 18].

The implementation of technical protocols like C2PA relies on the voluntary adoption of hardware-level standards by manufacturers such as Nikon, Sony, Leica, Canon, and Google [9, 12, 14, 17]. This voluntary adoption could create an inequitable information environment where "trustworthiness" is tied to cryptographically signed credentials, potentially marginalizing grassroots or low-resource political communication from non-compliant devices [10, 12, 23].

## **Regional Vulnerabilities and Adaptation Gaps**

The specific vulnerability of democratic regions is determined by the prevalence of different technological threats. In the Global South, "cheapfakes" (low-tech manipulations) and coordinated inauthentic behavior are more common threats [24]. In contrast, significant research and the technological "arms race" primarily focus on high-tech deepfakes utilizing advanced generative models [4, 6, 8, 15, 21].

Despite the urgent need for societal adaptation, the research does not identify any successfully implemented or piloted large-scale media literacy programs or redesigned

verification workflows for journalists. Existing adaptation strategies primarily focus on platform-led disclosure requirements for synthetic content in political ads (Meta, Google) [6] and legislative mandates [3, 11, 12]. The National Institute of Standards and Technology (NIST) is researching risks and mitigation strategies for synthetic content [20].

## Implications

---

The evidence suggests that democratic institutions face a profound challenge from deepfake-saturated election cycles, primarily due to the "Liar's Dividend" and the rapid advancement of deceptive AI technologies outpacing detection and regulation. The ability of political actors to dismiss authentic evidence as fake, coupled with the increasing ease and scale of synthetic media production, threatens the very foundation of a shared reality and public trust. While technical solutions like C2PA and watermarking are emerging, their vulnerabilities, voluntary adoption, and potential to create an inequitable information environment mean they are insufficient as standalone defenses. Without a rapid, widespread, and multi-layered societal adaptation that includes robust regulatory frameworks, enhanced media literacy, and universally adopted, resilient technical standards, democratic processes risk becoming increasingly susceptible to manipulation and delegitimization. The current trajectory indicates that the erosion of public trust and democratic stability may outpace the effectiveness of emerging solutions.

## Limitations and Caveats

---

The research findings present a nuanced and often contested view of the future impact of deepfakes on democratic institutions. There is disagreement among researchers regarding whether the threat constitutes an "information apocalypse" or "manageable information pollution" [25]. While some evidence suggests the "Liar's Dividend" is largely ineffective against video evidence due to forensic techniques [33, 34, 35, 36, 37, 38], other findings indicate that "oppositional rallying" can, in certain instances, discredit video-based scandals [7, 8, 26]. Furthermore, the research does not provide specific quantitative metrics for the performance gap between deepfake generation and detection capabilities [4, 6, 8, 15], nor does it identify successfully implemented large-scale societal adaptation strategies like media literacy programs. The long-term efficacy and scalability of proposed solutions against evolving threats remain subject to ongoing development

and real-world deployment.

## Sources

---

- [1] Deepfakes American Elections - americanbar.org - [https://www.americanbar.org/groups/public\\_interest/election\\_law/american-democracy/our-work/deepfakes-american-elections/](https://www.americanbar.org/groups/public_interest/election_law/american-democracy/our-work/deepfakes-american-elections/)
- [2] journals.sagepub.com - <https://journals.sagepub.com/doi/10.1177/19401612251317766>
- [3] [edu] Deepfakes Elections And Shrinking The Liars Dividend - cset.georgetown.edu - <https://cset.georgetown.edu/article/deepfakes-elections-and-shrinking-the-liars-dividend/>
- [4] [edu] Viewcontent.Cgi - scholarship.claremont.edu - [https://scholarship.claremont.edu/cgi/viewcontent.cgi?article=4903&context=cmc\\_theses](https://scholarship.claremont.edu/cgi/viewcontent.cgi?article=4903&context=cmc_theses)
- [5] Crossing Deepfake Rubicon - csis.org - <https://www.csis.org/analysis/crossing-deepfake-rubicon>
- [6] [peer-reviewed] Articles - pmc.ncbi.nlm.nih.gov - <https://pmc.ncbi.nlm.nih.gov/articles/PMC12351547/>
- [7] [edu] Deepfakes Surge During Election Cycles - news.gatech.edu - <https://news.gatech.edu/news/2024/10/23/deepfakes-surge-during-election-cycles>
- [8] [peer-reviewed] Liars Dividend Can Politicians Claim Misinformation To Evade Accountability - cambridge.org - <https://www.cambridge.org/core/journals/american-political-science-review/article/liars-dividend-can-politicians-claim-misinformation-to-evade-accountability/687FEE54DBD7ED0C96D72B26606AA073>
- [9] [gov] Frontier Research Mitigating Risks Synthetic Content Call Ac - nist.gov - <https://www.nist.gov/aisi/frontier-research-mitigating-risks-synthetic-content-call-action>
- [10] [edu] Liars Dividend Research - cla.purdue.edu - <https://www.cla.purdue.edu/news/college/2024/liars-dividend-research.html>
- [11] Countering Disinformation Effectively An Evidence Based Poli - carnegieendowment.org - <https://carnegieendowment.org/research/2024/01/countering-disinformation-effectively-an-evidence-based-policy-guide>
- [12] [preprint] Html - arxiv.org - <https://arxiv.org/html/2603.02378v1>
- [13] [news] Ai Deepfakes Blur Reality 2026 Us Midterm Campaigns 2026 03 - reuters.com - <https://www.reuters.com/business/media-telecom/ai-deepfakes-blur-reality-2026-us-midterm-campaigns-2026-03-28/>
- [14] Deepfakes And Crisis Knowing - unesco.org - <https://www.unesco.org/en/articles/deepfakes-and-crisis-knowing>
- [15] Safeguarding Trust And Dignity In The Age Of Ai Generated Me - partnershipnai.org - <https://partnershipnai.org/resource/safeguarding-trust-and-dignity-in-the-age-of-ai-generated-media/>
- [16] Ai Powered Disinformation And The Liars Dividend - theinnovator.news - <https://theinnovator.news/ai-powered-disinformation-and-the-liars-dividend/>
- [17] [peer-reviewed] Stable - jstor.org - <https://www.jstor.org/stable/26891938>
- [18] [edu] Watch Out For False Claims Of Deepfakes And Actual Deepfakes - brookings.edu - <https://www.brookings.edu/articles/watch-out-for-false-claims-of-deepfakes-and-actual-deepfakes-this-election-year/>
- [19] Final Deepfakes Elections Report Civicus Ddi Research 2025 - civicus.org - <https://www.civicus.org/documents/ddi/final-deepfakes-elections-report-civicus-ddi-research-2025.pdf>
- [20] [edu] Information Integrity - bridgingbarriers.utexas.edu - <https://bridgingbarriers.utexas.edu/good-systems/projects/information-integrity>
- [21] Deepfakes Elections And Shrinking Liars Dividend - brennancenter.org - <https://www.brennancenter.org/our-work/research-reports/deepfakes-elections-and-shrinking-liars-dividend>
- [22] C2pa Ai Supply Chain Verifying Authenticity - aicompetence.org - <https://aicompetence.org/c2pa-ai-supply-chain-verifying-authenticity/>

- [23] Tackling The Threat Of Deepfakes In Our Elections - nextgenpolicy.org - <https://www.nextgenpolicy.org/lets-talk/tackling-the-threat-of-deepfakes-in-our-elections/>
- [24] Dont Panic Yet Assessing The Evidence And Discourse Around G - knightcolumbia.org - <https://knightcolumbia.org/content/dont-panic-yet-assessing-the-evidence-and-discourse-around-generative-ai-and-elections>
- [25] [peer-reviewed] Articles - pmc.ncbi.nlm.nih.gov - <https://pmc.ncbi.nlm.nih.gov/articles/PMC9453721/>
- [26] [edu] Deepfakes Surge During Election Cycles - gatech.edu - <https://www.gatech.edu/news/2024/10/23/deepfakes-surge-during-election-cycles>
- [27] [edu] Isps24 07 - isps.yale.edu - <https://isps.yale.edu/research/publications/isps24-07>
- [28] [peer-reviewed] Improving authenticity and provenance in digital biomarkers - PMC - <https://pmc.ncbi.nlm.nih.gov/articles/PMC11736138/>
- [29] [edu] Can provenance save us from a barrage of synthetic media? - <https://www.cip.uw.edu/2024/02/21/provenance-synthetic-media-trust-perceptions/>
- [30] [preprint] AI Washing and the Erosion of Digital Legitimacy: A Socio ... - <https://arxiv.org/html/2601.06611v1>
- [31] [blog] Digital watermarking for interoperable and durable Content ... - <https://contentauthenticity.org/blog/digital-watermarking-interoperable-durable-content-credentials>
- [32] Combating Deepfake Threats Using X-FACTS Explainable CNN ... - <https://www.scirp.org/journal/paperinformation?paperid=145087>
- [33] [peer-reviewed] The psychological drivers of misinformation belief and its resistance ... - <https://www.nature.com/articles/s44159-021-00006-y>
- [34] [peer-reviewed] The continued influence of AI-generated deepfake videos despite ... - <https://pmc.ncbi.nlm.nih.gov/articles/PMC12848074/>
- [35] [edu] Trust in voting: How misinformation threatens democracy - USC Today - <https://today.usc.edu/trust-in-voting-how-misinformation-threatens-democracy/>
- [36] [edu] Misinformation is eroding the public's confidence in democracy - <https://www.brookings.edu/articles/misinformation-is-eroding-the-publics-confidence-in-democracy/>
- [37] [edu] Visual Framing by Media Affects Public Perception of Candidates - <https://arkansasresearch.uark.edu/visual-framing-by-media-affects-public-perception-of-candidates/>
- [38] [edu] Hiding secret codes in light protects against fake videos - <https://news.cornell.edu/stories/2025/07/hiding-secret-codes-light-protects-against-fake-videos>
- [39] [peer-reviewed] Deep Learning for Image Watermarking: A Comprehensive Review ... - <https://pmc.ncbi.nlm.nih.gov/articles/PMC12845643/>
- [40] [preprint] SoK: How Robust is Audio Watermarking in Generative AI models? - <https://arxiv.org/html/2503.19176v1>
- [41] C2PA Implementation Guidance - <https://spec.c2pa.org/specifications/specifications/2.3/guidance/Guidance.html>
- [42] [peer-reviewed] Robust Digital Watermarking for Remote Sensing Images Using ... - <https://dl.acm.org/doi/10.1145/3714334.3714350>
- [43] [preprint] A Baseline Method for Removing Invisible Image Watermarks using ... - <https://arxiv.org/html/2502.13998v1>
- [44] [peer-reviewed] A Systematization of Watermarking, Fingerprinting, Model Access ... - <https://ieeexplore.ieee.org/iel7/5962385/10707065/10143370.pdf>