Physical Education, Health and Social Sciences

https://journal-of-social-educatin.org

E-ISSN: 2958-5996

P-ISSN: 2958-5988



# Deepfakes and AI-Generated Disinformation: A Threat to Democracy in the 2026 Elections

**Tony T. Williams**

MHA, Ed.S. PhD(h.c.), Department of Health and Human Services, Ashford University- UAGC.
Email: williamscaddola@gmail.com

***DOI: https://doi.org/10.63163/jpehss.v3i2.1725***

**Abstract**

Since 2024, the dissemination of electoral disinformation has been radically changed by the rapid emergence of affordable, easy-to-use generative artificial intelligence (AI) tools that can create believable-looking images, audio, and video of candidates and public figures saying or doing things that never actually happened. This paper draws upon recent academic research, government reports, and investigative reporting mostly from the past two years to explore the impact deepfakes and AI-generated disinformation pose to democratic processes as they enter the 2026 election cycle. The paper, based on case studies from the United States, Pakistan, India, Indonesia, Bangladesh, Canada, Ireland, Ecuador and the European Union, illustrates the extent, nature and impact of synthetic political content, particularly when used against women in public life. It explores the increasing difficulty of creating deepfakes versus combating their detection, the shifting nature of centralized platform fact checking to crowd-sourced approaches, the quick growth of state and international regulation, including the EU's Artificial Intelligence Act, and the psychological impact of deepfakes, which undermines trust in genuine evidence itself. The paper also looks at how trustworthy AI chatbots are as sources of election information and examines new election risks for the midterm and global elections of 2026. It ends with three levels of technical, regulatory and societal solutions and the view that resilience against AI-generated disinformation rests on a coordinated approach of increasing the capacity of disinformation detection, disclosure and media literacy among the public.

## Introduction

In just a few years, generative artificial intelligence (GAI) has transitioned from a niche research tool to a consumer utility, making its impact on how elections are conducted more apparent than ever.In the last few years, generative artificial intelligence (GAI) has gone from being a specialized research tool to a consumer product, and its impact on the way elections are run is more visible than ever. It is now possible to generate a photorealistic image of a candidate in a compromising position, or a fluent video address that was never recorded, and platforms that can accomplish this, albeit from several seconds of audio and using artificial intelligence, are widely available and cheap, lowering the technical and financial barriers that once made convincing forgeries possible only for well-resourced state actors or film studios (CIVICUS & Digital Democracy Initiative, 2025). It creates an environment in which citizens no longer know what's being said or done on a video, audio clip or even a live-seeming interview, and a fast-moving threat to the integrity of democratic deliberation.

The threat has been unusually dense this year over the past few years (2024-26). In 2024, national elections took place in more than sixty countries—it has been called the biggest simultaneous tests of democratic institutions against AI-generated disinformation, in history (Frontiers in Political Science, 2024). By 2025 and 2026, deepfake examples could be found in the elections in Ireland, Canada, Ecuador, Germany, Namibia and Singapore; the former premier of Pakistan used an AI-generated voice to give an address at a campaign rally while imprisoned (Dawn.com, 2024); and even political consultants in the United States paid for an AI clone of President Biden's voice to suppress turnouts in the New Hampshire primary (Federal Communications Commission, 2024). Meanwhile, the mechanisms and institutions that citizens have traditionally turned to to separate fact from fabrication have also been on the move: Meta scrapped its program of third-party fact checking in the United States early in 2025 for a crowd-sourced "Community Notes" approach (Poynter, 2025), and the European Union has rolled out sweeping new transparency requirements for AI-generated content under Article 50 of its Artificial Intelligence Act (European Commission, 2026).

The race against time to understand whether democracies can intelligently adapt their information ecosystems rapidly enough, to maintain electoral integrity, is a pivotal moment in the years ahead, and especially in the days ahead to the 2026 elections. The stakes are more than just technical. Misinformation and disinformation are among the most dangerous near-term global risks, as the World Economic Forum's Global Risks Report has identified for three years running, building on the increased awareness that the loss of common ground on facts will have negative consequences for social cohesion, economic stability, and the basic legitimacy of elected government (Zurich, 2025; World Economic Forum, 2026a, 2026b).

This paper serves to answer three key questions. 1. What has been observed regarding the extent and nature of AI-driven disinformation in recent elections, and how has it changed since 2024? Second, what technical, regulatory and platform-level responses have been created and how effective have they been? Third, what does the history of these trends tell us about threats to the 2026 election cycle and what mix of measures could likely be effective in reducing those threats? To answer these questions, the paper draws from a wide range of recent academic research, government and intergovernmental reports, investigative journalism and legal analysis, in almost every instance published in the past two years, to offer a comprehensive and up-to-the-minute evaluation of a rapidly evolving policy issue.

**Theoretical Framework and Literature Review**

The democratic dangers of synthetic media have been the subject of scholarly interest before the advent of today's generative AI frenzy, but the theoretical construct of synthetic media, and the potential dangers for democracy, have gained new urgency because synthetic media technologies are now accessible to the public political actors, not just those with substantial resources. The two theoretical frameworks central to this paper include: the "liar's dividend" and psychological inoculation theory as it relates to media literacy.

Once a population learns that video and audio can be readily fabricated, the liar's dividend kicks in: If a political operative is known to be committing real wrongdoing, and is discovered to have been doing so with a video or audio file, that evidence can be called fake, made up by a computer, and a population that has become suspicious of digital evidence will be more ready to take that claim at its word than it would be if the evidence were physical—and more willing to ignore it entirely (Brennan Center for Justice, 2024). The concept has been discussed in great detail over the last few years in legal and political commentary, and it is one of the most significant indirect harms of the deepfake era, not because it's used to trick people with a fake video, but because it's used to make people question whether a video is real.The concept has been talked about

extensively over the past couple years, both in legal and political commentary, and one of the most significant indirect harms of the deepfake era is not because it's used to trick people with a fake video, but because it makes people question whether a video is real. The threat of deepfake may come not just from any one viral fabrication, but from an overall sense of epistemic uncertainty.

A second body of literature investigates, if it is possible, and how, citizens can be trained to resist being deceived by synthetic media. Research on digital literacy and psychological inoculation indicates that warnings of the existence of deepfakes are not particularly effective and that efforts to teach concrete, specific methods for detecting manipulated media yield measurable improvements in detections and self-reported confidence in detection compared to a warning alone (Huang & Hu, 2025). The implications of this finding for policy design are significant, because many of the current promotional messages on platforms about 'AI-generated content' – often just an attestation placed in the corner without further elaboration – may do little to help voters, as compared to interventions that foster longer-lasting evaluative skills.

A third thread of relevant scholarship is built around the empirical magnitude of the deepfake problem, which has been complicated by subsequent measurement and in which early alarmist predictions have been overly complicated. The fewest resources to be found in any disinformation campaign, whether it be election-related or otherwise, is audience attention, not the quantity of fabricated content (Knight First Amendment Institute, 2024); indeed, in a widely-cited review of the 78 election-related deepfakes documented globally during the 2024 election cycle, the Knight First Amendment Institute reported that the problem is "a distribution problem, not a generation problem. A parallel quantitative research project by the International Panel on the Information Environment (IPIE) revealed 215 examples of AI-generated deepfake electoral disinformation in the 50 countries that held competitive national elections in 2024 – significantly less than the number suggested by the press at the time (CIVICUS & Digital Democracy Initiative, 2025). Notwithstanding, a subsequent study on the 2025 Canadian federal election revealed that the proportion of AI-generated images among the total number of election-related images shared on major platforms was 5.86%, with right-leaning accounts sharing images nearly twice as often as left-leaning accounts, and with the most realistic fabrications receiving disproportionately high engagement rates despite their low frequency of distribution (arXiv preprint discussed in CETAS, 2025). Combined, these results point to a more complex matrix than either condemnation or worry: while the absolute numbers of the deepfakes may be relatively small, their targeting, partisanship, and engagement dynamics are a disproportionately powerful force for harm.

Lastly, there is a growing awareness in the literature about the inseparability of the AI-driven disinformation sphere from the wider information landscape. It affects the platform architecture, the platform's moderation policy, and the presence or absence of trusted intermediaries, like fact checks or election officials. The multi-domain approach of the remainder of this paper draws on this systemic framing, in which technology, gender, regulation, platform-governance, and psychological aspects of the deepfake threat are mutually dependent, and therefore not separable.

**Methodology**

The approach adopted in this study is a Desk Based Research with a qualitative research design, whereby secondary sources of information are synthesized, with the majority published from 2024 to 2026. To ensure the analysis was grounded in the most recent and relevant empirical evidence and policy shifts, a focused search of the last two years of publications was conducted, with structured searches of academic literature, government and intergovernmental publications, legal analysis and policy analysis also being used to identify sources. Sources include peer-reviewed journal articles from journals such as International Journal of Press/Politics, Journalism & Mass Communication Quarterly and Frontiers in Political Science, reports from research institutions like

the Centre for Emerging Technology and Security at the Alan Turing Institute, the Knight First Amendment Institute, the Brennan Center for Justice, and the Center for Democracy and Technology, regulatory and legal documents from the European Commission, the U.S. Federal Communications Commission, and Ballotpedia's legislative tracking service, and reporting from established news organizations like Reuters-affiliated Context, Al Jazeera, NBC News and the Associated Press. Key empirical claims were corroborated with triangulation across these source types, focusing on statistics on the prevalence of deepfakes, legislative counts, and survey findings, the latest version of each of which was available at the time. The policy area is moving rapidly and the paper explicitly states that the results reflect the evidence base at mid-to-late 2026, but adds that technology and the policy landscape are likely to continue moving rapidly.

**The Global Landscape of AI-Generated Election Disinformation, 2024–2026**

The 2024 global election cycle — in which over 60 countries conducted national elections — was the first opportunity to assess the impacts of generative AI on politics, and resulted in a publicly documented collection of case studies that grew in 2025 and 2026. A summary of some of the most important incidents is given in Table 1.

***Table 1. Selected AI-Generated Election Disinformation Incidents, 2023–2026***

| Date | Country | Incident | Documented Response |
|---|---|---|---|
| Sept. 2023 | Slovakia | AI audio clip falsely depicted a candidate discussing vote manipulation, released days before the election. | Widely cited as an early case demonstrating timing vulnerability; no formal takedown before polling. |
| Dec. 2023–Feb. 2024 | Pakistan | Jailed former PM Imran Khan used AI voice/image clone to address rallies and claim election victory. | Disclosed as AI-generated by his party; no legal sanction; cited in later legislative debate. |
| Jan. 2024 | United States (NH) | Robocall used AI-cloned Biden voice urging Democrats not to vote in the primary. | $6M FCC fine; felony indictment; jury acquittal on state charges (2025). |
| Feb.–June 2024 | India, Indonesia, Bangladesh | AI-manipulated videos on welfare policy claims; digitally revived deceased leaders for campaign events. | Election Commission of India issued SOPs; no criminal prosecutions documented. |
| Aug. 2024 | United States | Grok chatbot spread false ballot-deadline information in nine states. | Five secretaries of state sent formal letter; X redirected queries to Vote.gov. |
| Dec. 2024 | United States | Study found 35,000+ non-consensual deepfake images of 26 members of Congress (25 women). | DEFIANCE Act and TAKE IT DOWN Act advanced in Congress; signed into law May 2025. |

| Date | Country | Incident | Documented Response |
|---|---|---|---|
| Jan. 2025 | United States (platform policy) | Meta ended third-party fact-checking, adopted Community Notes model. | Rollout completed April 2025; independent research raised concerns about speed and coverage. |
| Feb. 2025 | Ecuador | AI-generated content mimicking genuine news formats circulated during presidential election. | Studied via post-election survey research (n = 201, n = 400). |
| Apr.–Oct. 2025 | Ireland | Library of 120+ fabricated images of Irish politicians circulated before presidential election. | Debunking dilemma noted by researchers; no coordinated takedown. |
| 2025 | Canada | AI-generated images made up 5.86% of election-related social media images analyzed. | Detection framework (OpenFake) used for post-hoc academic analysis. |
| May 2026 | United States (federal law) | TAKE IT DOWN Act enforcement began; FTC issued warning letters to platforms. | 15 major platforms warned; litigation over state deepfake laws continues. |
| Aug. 2026 | European Union | EU AI Act Article 50 transparency/labeling obligations entered into force. | Code of Practice on AI-generated content applied across signatory platforms. |

The series of elections in South Asia in countries including Bangladesh, Pakistan, Indonesia and India between January and June 2024 were a good first experiment in some ways. Former Pakistani prime minister Imran Khan, who is under house arrest and facing a ban on speaking in person, leveraged AI to speak at a virtual rally with over 1.4 million views, while his party issued an AI-generated video in which a synthetic version of Khan announced victory for his party (Dawn.com, 2024; Malay Mail, 2024). Digital rights watchers in Pakistan pointed out that it would be hard for voters to verify that the information they found on social media was correct, given the comparatively low literacy rate in the country and the fact that audio and video content is becoming more important to them than reading (Context by TRF, 2024). In India, AI-generated videos were disseminated to suggest the government was planning to do away with affirmative action in the country if it returns to power, and another regional party used synthetic media to digitally reanimate a deceased former leader for campaign appearances, highlighting how synthetic media is not just about deceiving live figures but also resurrecting political symbols (Al Jazeera, 2024). Reports of elections based on AI-generated content and generating concerns about the integrity of the electoral process have also been reported from Bangladesh's January 2024 parliamentary election and Indonesia's February 2024 presidential election (Frontiers in Political Science, 2024). The most impactful documented event of the 2024 election cycle wasn't a video deepfake, it was an audio deepfake: Two days prior to the upcoming presidential primary in New Hampshire on January 2024, a robocall, using an AI-generated voice profile of President Biden, called on participating Democratic voters to refrain from voting in the presidential primary, implying that those who did would not be eligible to vote in November 2024 (Federal Communications

Commission, 2024). The call was received by approximately 25,000 voters, and the political consultant responsible for the call, Steve Kramer, was indicted on felony voter-suppression charges, fined $6 million by the Federal Communications Commission, and fined $1 million by the telecommunications company that made the call (The Register, 2024; NBC News, 2024a). Kramer was finally acquitted of the state felony charge after a jury accepted his defense that the New Hampshire primary was an "unsanctioned straw poll," which was the case due to the difficulty of proving that existing laws to suppress voter-suppression were being applied to AI-generated content, and helped spur follow-up efforts by state legislatures to target deepfake content specifically.The jury ultimately acquitted Kramer, in part because of the difficulty in applying existing voter-suppression laws to AI-generated content, and in part because it was his defense that the New Hampshire primary was an "unsanctioned straw poll," which helped catalyze follow-on state-level legislative efforts specifically aimed at deepfakes (Yahoo News, cited via CBS News, 2025).

By 2025, the use of deepfakes in elections was expanding into areas that were previously regarded as low-risk. Researchers identified a library of over 120 fabricated images of Irish politicians in circulation in advance of the upcoming Irish presidential election, which presents a challenging dilemma for the people depicted: either to debunk the content publicly, which risks promoting it, or to ignore it, which risks people believing it (CETAS, 2025). In February 2025, Ecuador's general election, AI-generated content was spread that resembled the look and feel of real news broadcasts, but not the realism of a particular image, and took advantage of audiences' trust in these familiar news formats (CETAS, 2025; Frontiers in Political Science, 2025). Earlier research on the 2024 election revealed that foreign influence operations contributed only a fraction to the overall amount of misleading content produced, whereas in the mixed-methods study of that election, domestic political actors were responsible for the lion's share of the misleading AI content (Frontiers in Political Science, 2025). Analysis of over 187,000 social media posts revealed that AI-generated images accounted for 5.86% of all election-related visual content, with right-leaning accounts twice as likely to share AI content as left-leaning accounts, but overall there was only a small share of all platform engagement that was related to harmful fabrications (CETAS, 2025).

The most impactful uses of AI-generated content in elections have often been the work of domestic political actors—from candidates to their own campaigns using synthetic media as a regular campaign tool (CETAS, 2025; Knight First Amendment Institute, 2024)—as opposed to the anonymous foreign influence operations that are most commonly cited as a danger in early public reporting. What's important is that the normalization highlighted by the CDT (2026) also seems to be changing by 2026, as candidates are not just using AI-generated political content but doing so without significant public reactions, which could make its use in more deceptive ways even easier in the future.

The phenomenon is not limited to a specific region, media system or level of economic development: In addition to the cases mentioned above, CETAS (2025) shows how deepfakes were used in the context of elections in three other countries, which were selected for comparative analysis because they differ in their digital contexts, economic and political systems. The Irish and Ecuadorian cases are also striking for demonstrating two types of fabrication that are present throughout the dataset: the production of large collections of discrete fabricated images that cover multiple political figures, thereby creating a “class of candidates” risk to their reputation, and the mimicry of the news broadcast format, which taps into the audience's familiarity with the visual and stylistic cues of news reporting, instead of depending on the technical quality of any given fabrication (CETAS, 2025). The same pattern emerged very early, when an audio segment was sent around in Slovakia appearing to be a conversation between an alleged leading candidate and a reporter day before the nation's parliamentary election, in September 2023, and is widely

believed to be an early, influential deepfake—often cited by researchers as the case that first proved how much deepfakes had come to grip with that early on and for which the clip was not available for verification or rebuttal (Wikipedia, 2026).

**The Gendered Dimension: Non-Consensual Deepfakes and Women in Politics**

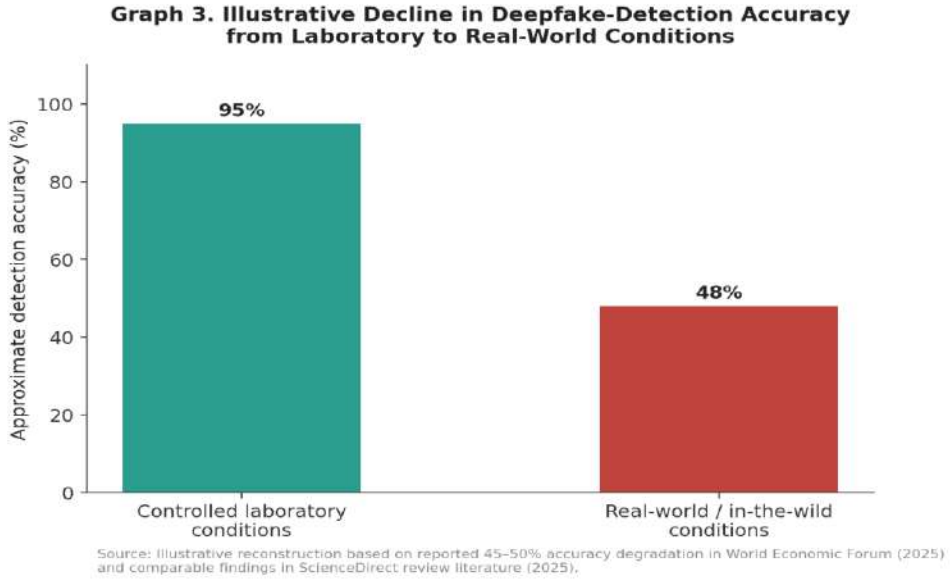
When it comes to actual evidence of the harm that deepfakes can cause in political settings, the most pervasive and significant variety is the non-consensual sexualisation of women, including women in political roles. The American Sunlight Project conducted a first-of-its-kind study that found that over 35,000 instances of non-consensual intimate imagery of 26 members of Congress—25 of them women—circulating on deepfake pornography websites, or approximately one in six women serving in Congress (American Sunlight Project, 2024; The Markup, 2024). This phenomenon is not unique to the U.S. In Pakistan, researchers have noted that information minister of Punjab and other women politicians, including Maryam Nawaz Sharif, the Chief Minister of Punjab, have been subjected to fabricated sexual imagery and doctored videos that researchers claim to be exploiting conservative social norms to severely impact the political viability of women (Binding Hook, 2026). Similar incidents have been reported with women politicians in the United States, Italy and the United Kingdom (International Knowledge Network of Women in Politics, 2025; France24, 2025).

This gendered trend is also a long-standing characteristic of deepfakes: As an example, independent studies have revealed that the overwhelming majority of deepfake content being shared online, not just in the context of politics, is non-consensual sexual imagery and that the content is distributed in unequal fashion with dramatically higher proportions targeting women (France24, 2025). This is referred to as "technology-enabled gender-based violence" and was found to have a substantial impact on women's political participation, with men facing a similar situation to a lesser degree and some analysis suggesting that the volume of deepfakes targeting women and girls is increasing by 245% year-on-year in 2024 (Binding Hook, 2026).

The policy steps taken to address this aspect of the issue have been more limited and delayed than those taken to address campaign-disinformation deepfakes. In the U.S., the bipartisan DEFIANCE Act of 2024 put in place a federal civil right of action for individuals whose personal information is subject to non-consensual intimate digital forgery, and the federal TAKE IT DOWN Act of 2025, which was signed into law in May 2025 with almost unanimous support from Congress, imposes takedown obligations on platforms hosting such content, and will begin to be enforced in May 2026 (TechPolicy.Press, 2024; Stackcyber, 2026). Despite this, legal experts and advocates point to the fact that the majority of countries still have a dearth of comprehensive protection, and that the disproportionate focus on women in public life has repercussions for democratic representation beyond the direct harm to those depicted, such as the potential to deter women from running for office at all in the future (TechPolicy).Press, 2024).

**The Detection Arms Race**

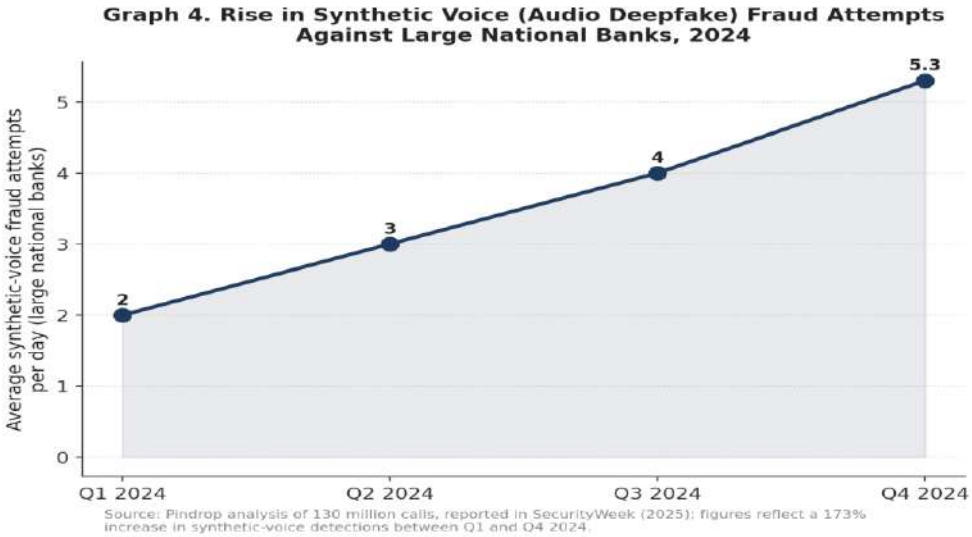
A central technical reality shapes every policy response discussed in this paper: deepfake detection technology has consistently lagged behind deepfake generation technology, and the gap between laboratory performance and real-world performance of detection systems remains substantial. Controlled benchmark evaluations of commercial and open-source detection models report accuracy rates approaching or exceeding 90%, but researchers have found that state-of-the-art automated detection systems experience accuracy drops of 45 to 50 percentage points when applied to deepfakes actually circulating in the wild rather than to curated benchmark datasets (World Economic Forum, 2025). Graph 3 illustrates this pattern using figures drawn from recent detection research.




*Graph 3. Illustrative Decline in Deepfake-Detection Accuracy from Laboratory to Real-World Conditions*

Automated systems perform even poorer than humans in terms of detection. According to a 2,000-consumer study released in 2025, only 0.1% of those who were shown a series of images and videos were able to correctly identify all the deepfakes and authentic stimuli; 999 out of 1,000 participants were misled by at least one fabrication (Adaptive Security, 2026). Likewise, a study published in one of the top engineering and detection science journals revealed that specialized architectural enhancements that increased detection accuracy by 37% under controlled conditions, still performed poorly in a nearly 38-hour real-world instance of celebrity and political speech, indicating that the benefits of the laboratory experiments do not necessarily carry over into the operational environment (Optics & Photonics News, 2025).

The arms race phenomenon applies to voice cloning, too, and it is now a favored weapon for both political and financial scams, such as the New Hampshire robocall. Between the first and fourth quarters of 2024, the volume of synthetic voice calls grew by 173% across all 130 million calls analyzed by fraud detection firm Pindrop, with large national banks seeing a spike from less than two deepfake voice attacks per day in the first quarter to over five per day in the fourth. (SecurityWeek, 2025). This trajectory is as shown in graph 4.



*Graph 4. Rise in Synthetic Voice (Audio Deepfake) Fraud Attempts Against Large National Banks, 2024*

The necessity of retraining models on samples from newly released generative systems can lead to an accuracy of around 99% within a few weeks, but this reactive cycle also results in the fact that a detection system is always tuned to the previous generation of generative systems and never to the current one, which gives an advantage to the attackers over the defenders (SecurityWeek, 2025). Deep fake generation costs have also continued to drop: there are reports of deep fake videos of politicians being created in less than 45 minutes, using publicly available consumer software; and voice cloning, using just 20-30 seconds of audio from any politician, can reportedly take less than 45 minutes (World Economic Forum, 2025). As the trajectory suggests, detection measures alone are unlikely to provide adequate protection, and should be complemented by measures that rely on the principle of provenance, such as cryptographic standards for content authentication, and by regulatory and literacy-based measures to be discussed later.

**Platform Governance and the Retreat of Centralized Fact-Checking**

As the 2024-2026 election cycle ramped up, the institutional apparatus that has traditionally been tasked with detecting and labeling fake information online got significantly smaller. In January 2025, Meta made the decision to end its collaboration with third-party fact-checking groups in the United States, replacing the centralized editorial process with a "Community Notes" system similar in concept to X's (now Meta's) Community Notes system, which uses a crowd-sourced approach where users can note their own corrections and add additional context (X, 2025; Poynter, 2025; NBC News, 2025). Before the switch, fact checkers' partnerships with Meta have represented about 45% of the revenue of the professional fact checking sector worldwide in 2023, making the move a major financial blow to the fact checking ecosystem beyond Meta's platforms (Time, 2026). The company saw the move as a return to "free expression" after a "political and cultural 'tipping point' following the 2024 U.S. presidential election," according to its chief executive, and it also relaxed its content-moderation practices as it relates to political speech more generally (NBC News, 2025).

There has been significant research indicating that crowdsourced moderation may not be a replacement for professional fact checking – especially in the high-stakes, time-sensitive setting of a live election. But systems similar to Community Notes have been shown to make corrections to inaccurate claims too late to help stop the spread of a false post, and they are susceptible to manipulation via coordinated networks or by automated accounts, because the same generative AI system that can be used to create deepfakes can also be used to write many superficially plausible Community Notes (Tufts Now, 2025). In the wake of that change, Meta subsequently reported that it had made about half as many enforcement mistakes, but that metric doesn't necessarily reflect a lower amount of harmful content (Meta, 2026; Berkeley Technology Law Journal, 2025).

The European Union has shifted regulatory direction while platform self-regulation has taken a setback in the United States. Under Article 50 of the EU's Artificial Intelligence Act, which kicks in for transparency obligations on AI-generated content in August 2026, content created or manipulated using AI systems that deceptively mimics a person, a thing or an event would be considered fake if presented as authentic to a viewer, even if the creator did not intend to mislead others (Greenberg Traurig, 2026). It's not only the developers of generative AI systems who are required to apply machine-readable markings to their outputs but also to "deployers" (businesses, campaigns, and individual professionals) publishing content generated by AI, except for content that has been substantively edited by a human (Weventure, 2025). The Code of Practice which is a multilevel approach, consisting of embedded technical watermarking, machine-readable

detection signals, and human-observable labels, has been developed with industry participation by the European Commission.

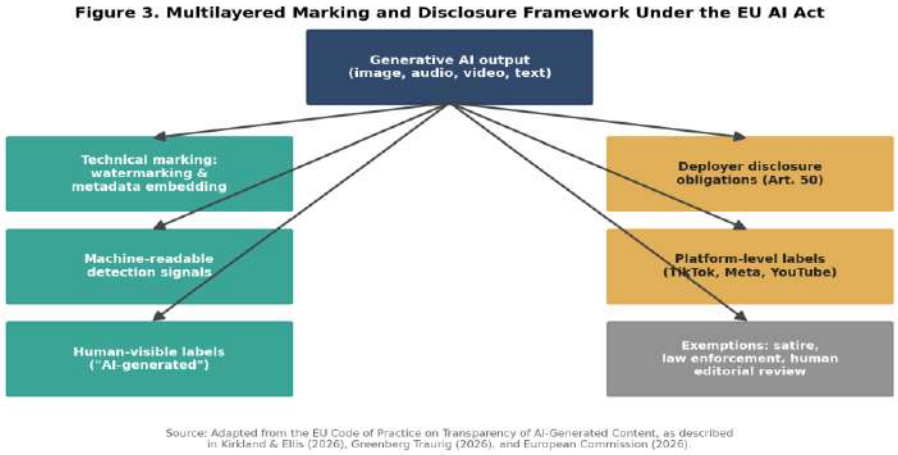


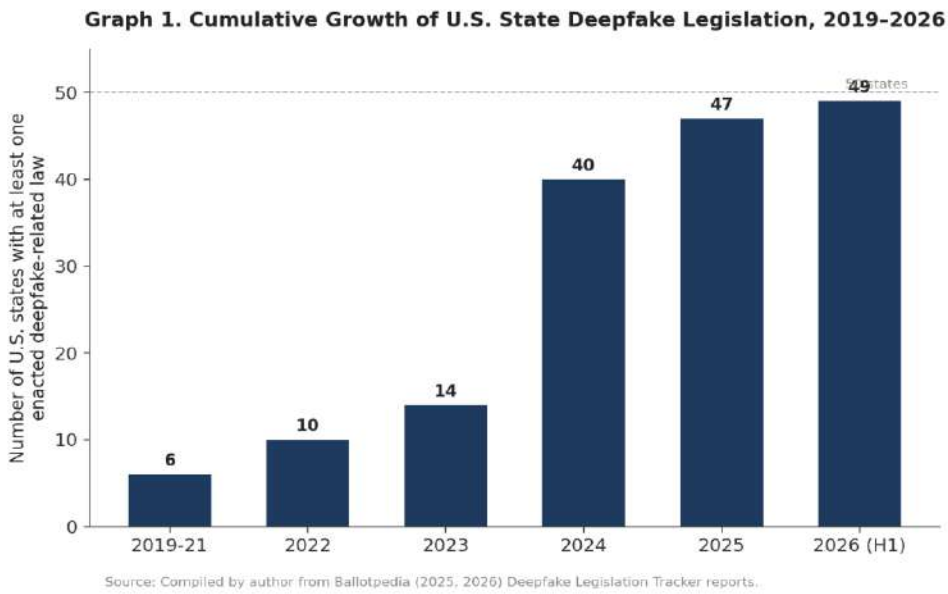

*Figure 3. Multilayered Marking and Disclosure Framework Under the EU AI Act*

Analysis of the Meta transition also point to an issue that is particularly relevant to elections: notes that address highly polarized claims, which are the very type in which most election-related disinformation is found, are structurally less likely to meet a threshold of agreement from ideologically divergent contributors and become publicly visible, than notes that address uncontroversial factual errors are more likely to meet the threshold and become publicly visible ( Tufts Now, 2025). It has also been noted that the model is only effective when it has access to a wide and representative community of contributors for extended periods of time, during which they are expected to provide good faith engagement (Grawitch, 2025).

Under this relatively ambitious set of rules, however, experts warn that technical and coordination shortcomings can threaten enforcement. No single marking method has been found to be effective by themselves: Watermarks can be removed from image and audio files with re-encoding and screenshotting, and there is no common interoperability standard among the dozens of generative AI providers currently on the market (Euronews, 2026; Resemble.ai, 2026). Certain platforms have also added their own requirements on top of the EU guidelines: TikTok has implemented "AI-generated" labels, which are automatically generated with support from the automated detection system, and Meta has automatically labeled AI content on Instagram and Facebook, as well as allowing manual tags (Weventure, 2025). These interwoven but not perfectly congruent systems will undoubtedly have an impact on voter deception, but whether that impact is meaningful, as opposed to merely complying with the disclosure requirements in form without altering how users interact is an open empirical question as the system moves into its first full year of enforcement in the 2026 election cycle.

**Regulatory Responses: State, Federal, and International Law**

As there is no broad federal legislation in the United States, state legislatures are now the main point for legal action on election-related deepfakes, and the number of states passing laws has increased dramatically. By mid-2025, states had passed 64 deepfake-related laws, up 23% from the same period in 2024 but to 49 laws by mid-2026 (Ballotpedia, 2025, 2026a, 2026b). Since tracking began in 2019, 174 state deepfake laws were enacted, with 82% of these laws passed in a single legislative session in 2024–2025, indicating how quickly state legislators have felt the need to enact a deepfake law (Ballotpedia, 2025). This cumulative growth trend is shown in graph 1.




*Graph 1. Cumulative Growth of U.S. State Deepfake Legislation, 2019–2026*

In 2026, 33 states had laws that specifically discussed the creation or use of deepfakes in political communications, the majority of which called for disclosure rather than a ban on deepfakes, and 48 states had laws that specifically addressed the creation or use of a non-consensual sexually explicit deepfake (Stackcyber, 2026; Ballotpedia, 2026b). The federal-level takedown requirements, which are part of the TAKE IT DOWN Act that passed the House of Representatives 409-2 in May 2025 and was unanimously passed by the Senate, are limited to non-consensual intimate imagery and will go into effect in May 2026 (Stackcyber, 2026). In 2025, a federal district judge invalidated California's political deepfake disclosure law on First Amendment grounds, and the state's case is pending before the Ninth Circuit Court of Appeals as of mid-2026 (Ballotpedia, 2026b); and at least one other state election-deepfake law has been invalidated for similar reasons (Stackcyber, 2026). A comparison of the main regulatory strategies in place as of 2026 is provided in Table 2.

***Table 2. Comparative Regulatory Approaches to AI-Generated Election Content, as of 2026***

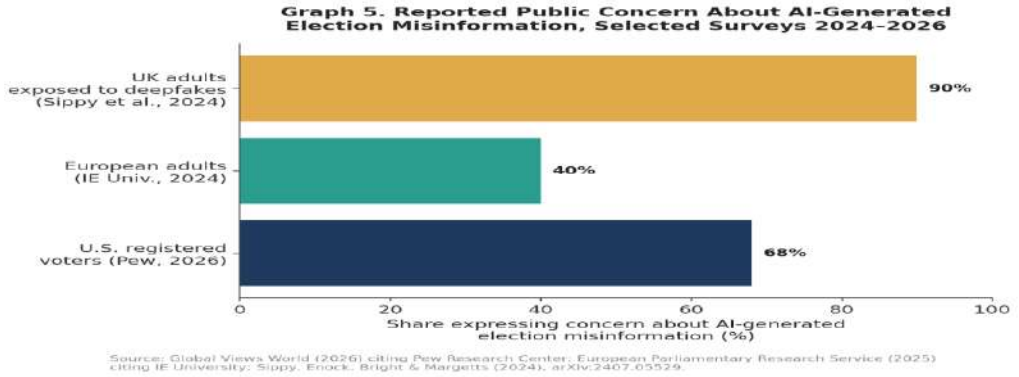
| Jurisdiction / Level | Primary Instrument | Core Approach | Status / Notable Limitation |
|---|---|---|---|
| U.S. states (aggregate) | State deepfake statutes (49 of 50 states) | Disclosure requirements for political ads; criminal/civil penalties for non-consensual explicit content. | Fragmented by state; two laws struck down on First Amendment grounds as of 2026. |
| United States (federal) | TAKE IT DOWN Act (2025) | Takedown obligations for non-consensual intimate imagery, including AI-generated. | Enforcement began May 2026; no comprehensive federal deepfake-in-elections law. |
| European Union | AI Act, Article 50 + Code of Practice | Multilayered technical marking, deployer disclosure, human-visible labeling. | Entered force August 2026; interoperability and enforcement gaps remain. |

| Jurisdiction / Level | Primary Instrument | Core Approach | Status / Notable Limitation |
|---|---|---|---|
| China | National AI content labeling mandate | Mandatory labeling of AI-generated content at platform level. | Enforcement began September 2025; limited independent verification available. |
| Platform self-regulation (global) | TikTok, Meta, YouTube labeling policies | Automated detection plus creator self-disclosure at upload. | Inconsistent enforcement; Meta ended third-party fact-checking in U.S. (2025). |

The legal developments depict a common conflict in deepfake lawmaking: Rapidly drafted legislation in response to salient incidents, like the New Hampshire robocall, is required to meet constitutional safeguards for political speech, while disclosure-based methods that are more likely to pass constitutional muster than bans may be less effective at stopping the underlying misinformation, because a disclosure label added after the deepfake has already gone viral will not reach the individuals who saw it in the first place. The situation is also very diverse around the world. In addition to this comprehensive framework, each nation has implemented a set of measures in a piecemeal fashion, from the labelling obligations that began in China in September 2025, to those that are currently under discussion in the United Kingdom and elsewhere; no formal international coordination mechanism is currently in place, given the cross-border nature of AI-generated content distribution (Adaptive Security, 2026; Profile News, 2026).

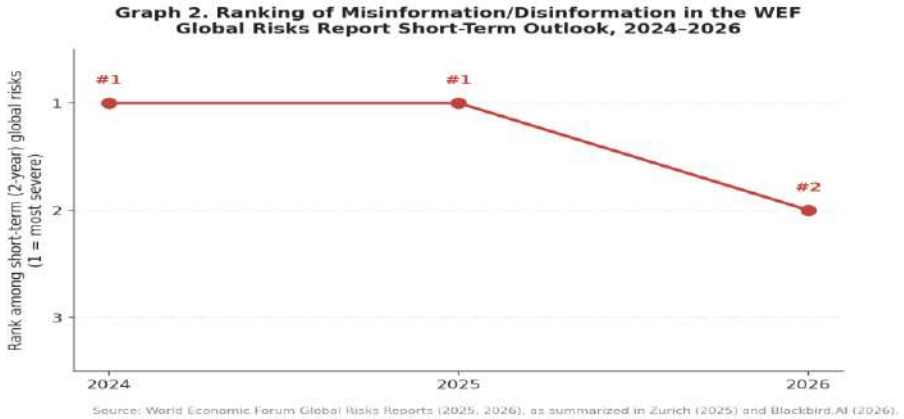
**Public Perception, Trust, and the Liar's Dividend**

The public has consistently expressed high levels of concern about AI-generated election disinformation across the surveys conducted from 2024 to 2026, although the exact percentages differ based on the type of survey, the design of the question(s) asked, and the country. According to a Pew Research Center analysis, 68% of registered voters in the U.S. are worried about the use of AI-generated misleading content having an impact on elections (Global Views World, 2026). Although this was lower than the U.S. estimate (45%), it was still a significant portion of the European population, and the European Parliament's own research service used this as justification to speed up their regulation process, citing the survey results (European Parliamentary Research Service, 2025). In people who report direct exposure, there is a correspondingly high level of concern, and concern is almost universal: Only 8% of a survey-based study in the United Kingdom report having created a deepfake themselves, but 90% of respondents express concern about the technology's implications (cited in arXiv preprint 2509.21135, discussed in relation to UK survey research). The results of these are compared in graph 5.



*Graph 5. Reported Public Concern About AI-Generated Election Misinformation, Selected Surveys 2024–2026*

This high concern aligns with the consistently high level of misinformation and disinformation on the World Economic Forum's annual Global Risks Report, which polls more than a thousand global experts, policymakers and business leaders. Misinformation and disinformation was the most severe of all short-term risks in the world in both the 2024 (Zurich, 2025) and 2025 (World Economic Forum, 2026a, 2026b) editions of the report, and it was the second most severe risk after geoeconomic confrontation in 2026 (Blackbird.).AI, 2026). This three-year trend is illustrated in the graphs.




*Graph 2. Ranking of Misinformation/Disinformation in the WEF Global Risks Report Short-Term Outlook, 2024–2026*

These statistics should be viewed with a grain of salt as evidence is present that public concern is not equal across all age and levels of digital literacy. The findings from this survey research on Americans' general digital literacy reveal that approximately half of U.S. adults are not aware of what a deepfake is, while 60% of Americans under 30 say they know what a deepfake is and what a large language model is, and perhaps this means that digital literacy interventions ought to be designed differently for older and younger generations of voters. Similarly, the same research highlights individual instances that can influence opinion, even among relatively informed audiences: In one instance, a fake video of Clinton endorsing a Republican candidate for president was widely shared on platforms with AI-generated content labelling policies in place – non-existent as it was, but still formally in place – and viewer comments often suggested the video was at least likely to be true.In another, a fabricated video of Clinton endorsing a Republican presidential candidate was widely shared on platforms that have an AI-generated content labelling policy – although it was not, users' comments were frequently suggesting the video was at least plausible, if not actually authentic, pointing to a gap between formal disclosure policy and audience understanding.

A growing awareness of deepfakes among the public does not automatically translate into greater electoral resilience, however, and could even set the stage for the liar's dividend (as outlined in section 2). This is illustrated in Figure 2, where increased public awareness that seemingly credible evidence can be manufactured becomes a weapon that can be wielded by politicians who are under the pressure of genuine, damaging evidence, and who can use it to label that evidence as an "AI-generated fake.

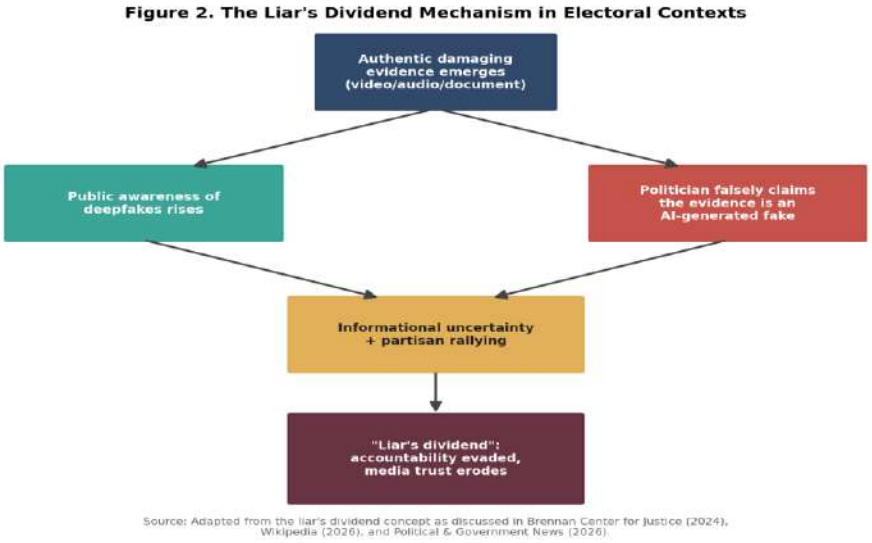



*Figure 2. The Liar's Dividend Mechanism in Electoral Contexts*

The theoretical discussion is complemented with experimental research on this dynamic, providing a more accurate picture than the theory alone. False allegations of "fake news" or AI-generated deepfakes increased the politician's standing over remaining silent or denying the claims—although the information-uncertainty channel was more effective at a general level, with allegations of "fake news" being more effective at raising the politician's standing than an explicit claim of the content being an AI generated deepfake—(Brennan Center for Justice, 2024). Meanwhile, another large-scale study of over 2,000 respondents revealed discernible negative impacts on trust and voting-related assessments after deepfake exposure to political scandals was countered by a journalistic reference check statement, highlighting the ongoing importance of proper verification and the current threats from the explosive contraction of journalistic fact-checks on major platforms (Dan, 2025).

**Generative AI Chatbots as an Emerging Vector**

Another key, and growing, avenue of election-related misinformation is not from intentional deepfakes, but rather from inaccuracies created by AI-based chatbots that are now a frequent source of basic election information – like registration deadlines, polling locations, and candidate positions – for voters. Though these systems were not designed for, or reliably accurate in the electoral information domain, this risk has exacerbated as conversational AI has been adopted as a general information tool. Table 3 presents the results of some independent studies that took place from 2024 to 2026.

***Table 3. Reported Accuracy of AI Chatbots on Election-Related Queries, Selected Studies 2024–2026***

| Study / Year | Models Tested | Reported Accuracy / Error Rate | Key Additional Finding |
|---|---|---|---|
| GroundTruthAI (2024) | Gemini 1.0 Pro, GPT-3.5 Turbo, GPT-4, GPT-4 Turbo, GPT-4o | 73% average correct; range 57%–81% by model. | 2,784 responses generated from 216 unique questions on the 2024 U.S. election. |

| Study / Year | Models Tested | Reported Accuracy / Error Rate | Key Additional Finding |
|---|---|---|---|
| Secretaries of State / X Corp. (2024) | Grok | False information on ballot deadlines in 9 states. | Chatbot amended to redirect queries to Vote.gov after official complaint. |
| Institute for Strategic Dialogue (2026) | Gemini 3.5 Flash, Grok 4.3, Sonnet 4.6, DeepSeek, Muse Spark, others | Accuracy range approx. 61%–84% across six chatbots. | Accuracy fell an additional 16 percentage points when queried in Spanish. |
| NewsBench / Forum AI (2026) | ChatGPT, Gemini, Claude, Grok, others | ~90% of midterm-election answers contained a material flaw. | Over 70% of errors attributed to retrieval failures rather than reasoning failures. |

An initial study during the 2024 general election cycle by analytics company GroundTruthAI, which tested some of the biggest LLMs with 216 questions, found a general average of 27% incorrect responses, ranging from 57% for one iteration of Google's Gemini model to 81% for the latest iteration of the OpenAI model GPT-4o at the time (NBC News, 2024b). In separate news, Grok was discovered to be sharing misleading information about ballot deadlines in nine states shortly after President Biden pulled out of the race, leading to a formal joint letter from secretaries of state from five states asking for a correction, while X itself changed its policies on the chatbot to route queries on ballots to nonpartisan Vote.gov, but not its policy of disseminating other misleading information using real people's names (Axios, 2024; TechRadar, 2024; Fortune, 2024). Later assessments indicate the issue has not been addressed by later model years. In a 2026 study by the Institute for Strategic Dialogue, five of the most prominent chatbots were tested with 15 standardized prompts and came up with an accuracy range of 84 to 60 percent, with all models revealing outdated information like 2024 election dates for a 2026 cycle, and accuracy dropping by another 16 percentage points when the same questions were posed in Spanish, which has direct implications for equitable access to accurate information among Spanish-speaking voters (NewsNation Now, 2026; The Hill, 2026). An independent research effort from a media project has concluded that the AI's response to the midterm election questions was inaccurate in some significant way (whether due to missing information, inaccurate facts, or some other error) in approximately 90% of the instances, with most of those inaccuracies stemming from flaws in how the source information was retrieved and prioritized, not from the models' flaws in reasoning (TechRadar, 2026).

The results of this research as a whole suggest that, while people have been using AI chatbots, as a source of information, they have been an underrecognized but large threat to electoral information integrity separate and apart from the threats of deliberately created “deepfakes.” Errors in election-related chatbots are more likely to be caused by retrieval and training data limitations rather than by malicious intent, which could make it more difficult to address these errors under the disclosure-based regulatory frameworks intended for deepfakes, and may need targeted accuracy benchmarking, authoritative sourcing to election-information portals and independent auditing as the tools become more mainstream in how voters seek information in 2026.

**The 2026 Election Outlook: Risks and Early Warning Signs**

As the 2026 election cycle continues through the U.S. midterm elections in November and many additional subnational and national elections across the globe, there are some common trends emerging that are just as – if not more – likely to increase the risk landscape for disinformation caused by AI than the 2024 cycle. Each of the following developments further strengthens the general causal sequence of accessible generative AI leading to democratic harm summarized in Figure 1.Each of these developments further reinforces the general causal sequence, accessible generative AI leading to democratic harm, as summarized in Figure 1.

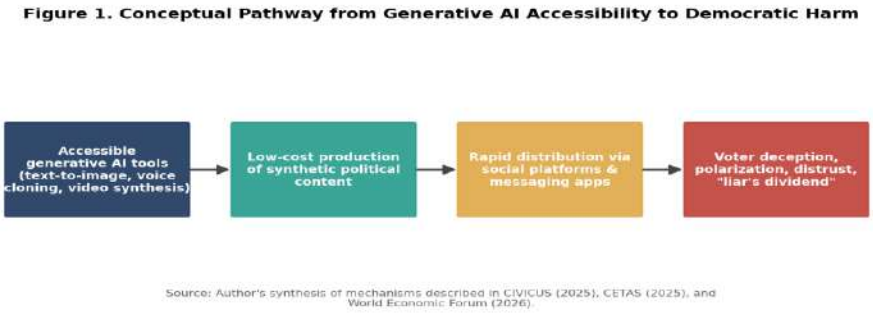



*Figure 1. Conceptual Pathway from Generative AI Accessibility to Democratic Harm*

In some key jurisdictions, the ability to identify and counter disinformation has been diminished rather than enhanced. In addition to Meta's withdrawal from third-party fact-checking, researchers at the Center for Democracy and Technology point out that federal infrastructure in the United States that has been developed to monitor and identify foreign influence operations has been reduced, which will make it harder to detect and respond to coordinated disinformation campaigns during the mid-term elections in 2026 (CDT, 2026).

Second, the way disinformation campaigns operate is shifting: they do not simply target the voters, but instead exploit the retrieval shortcomings documented in Section 10 to get to the voter via a trusted-looking intermediary that obfuscates their true source, enabling what could be called "narrative laundering" (CDT, 2026). Large-scale credential exposure linked to large political fundraising sites, and the cloning of trusted media brands via look-alike domains, are also noted by cybersecurity researchers who are monitoring the 2026 cycle, indicating that AI-powered disinformation risks are merging with more traditional cyber threats (Check Point Blog, 2026).

Third, importantly, the social and professional norms that have stigmatised the use of AI-generated content in political campaigns during the 2024 cycle are increasingly disappearing by 2026, as candidates and political consultants increasingly use the content as a routine campaign instrument without the same level of reputational risks (CDT, 2026). The relatively low documented prevalence of harmful deepfakes in the 2024 and 2025 cycles may not serve as a guide to what might be expected in the 2026 cycle, as much of that moderation seems to be on the decline as well.

Fourth, the use of generative AI is evolving in international disinformation operations, and European intelligence services report that state-backed influence networks are exploiting deep fake technology particularly to attack European politicians, their aides, and political and military backing for Ukraine, including the work of fake “whistleblowers” and bogus local news outlets (European Parliamentary Research Service, 2025). The threat picture for 2026 elections is not just domestic political, but also geopolitically motivated disinformation campaigns targeting the same techno-institutional weaknesses.

Putting these developments together, it does not appear that the size and impact of election disinformation during the 2024 election cycle were as large as earlier predicted by the alarmists, but rather that the path of institutional, technological, and normative evolution in the interval 2024-2026 has been more one-way than two-way, and could reasonably lead to an increase in the size and effectiveness of AI-generated election disinformation in the 2026 election cycle, absent any single game-changing breakthrough.

The Brennan Center for Justice (2026), which has been closely tracking U.S. election administration ahead of the midterm elections in November, places the threat of AI-enhanced misinformation in the context of other concurrent pressures on electoral institutions, such as increased federal-state conflict over election administration, worker shortages among local election officials, and increased election security concerns for poll workers, noting that the threat of AI-generated misinformation does not stand alone. This escalation effect is important because it indicates that the marginal effect of any specific deepfake or AI-chatbot error could be magnified by a sense of distrust of institutions and administrative pressure that is not limited to the scope or impact of a single technology. Election officials who have been surveyed for related research have repeatedly expressed worry about both the immediate risks posed by AI-mediates disinformation as well as the compounding threat of diminished capacity among institutions to counter disinformation, further emphasizing the need for a multi-layered response to AI instead of any one agency or platform taking responsibility for the risk (Brennan Center for Justice, 2026).

**Toward Resilience: A Multi-Tier Policy Framework**

No single intervention will likely be game-changers in the fight against the threat posed by AI-generated disinformation to the 2026 election cycle and beyond, given the arms-race dynamics outlined in Section 6, the institutional retrenchment outlined in Section 7, and the fragmented regulatory landscape outlined in Section 8. The evidence presented in this paper rather suggests a framework that integrates technical, regulatory and societal interventions, all working together in parallel, as shown in Figure 4.

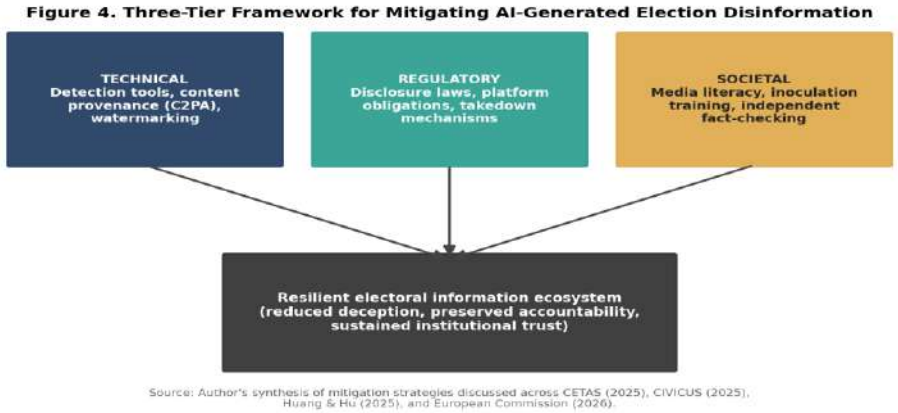



*Figure 4. Three-Tier Framework for Mitigating AI-Generated Election Disinformation*

At the technical level, the continued disparity between lab and field detection precision described in Section 6 implies that content-detection cannot be a front-line defense; it is important to invest more in the development of content-provenance approaches like cryptographically verifiable metadata standards that establish a chain of custody for authentic media at the point of capture than in simply detecting false content after it is created. The EU's multi-layered marking system (discussed in Section 7 and illustrated in Figure 3) takes steps in this direction, but will rely heavily on cross-platform interoperability, which is not yet fully complete (Resemble.ai, 2026).

At the regulatory level, the growing speed of state-level action detailed in Section 8 is a sign of the real urgency of the situation, but it has created an uncoordinated and constitutionally challenged mix that makes it difficult for national campaigns and platforms to comply with state-level regulations in multiple jurisdictions. This fragmentation can be mitigated by a more coordinated approach to regulation, possibly a hybrid of federal disclosure rules for political AI-generated content, and the methods that already exist for non-consensual intimate imagery under the TAKE IT DOWN Act, but any such scheme will need to be drafted carefully to avoid the constitutional challenges that have already been brought against at least two state-level deepfake laws (Ballotpedia, 2026b; Stackcyber, 2026).

The media literacy research highlighted in Section 2 provides one of the most easily implementable and evidence-based interventions that can take place before the 2026 cycle is over at the societal level. Media literacy interventions should focus on specific technique-based training that explicitly includes the development of diagnostic skills to enhance the accuracy of AI media detection, the perceptions of AI media literacy within the information environment, and self-efficacy in detecting AI in media, as these methods have been demonstrated to be more effective than warnings at boosting detection accuracy, perceived AI media literacy, and self-efficacy in detecting AI in media, with the latter being especially effective in the video format through which AI media literacy is taught and acquired (Huang & Hu, 2025). The specific harms to women and girls reported in Section 5 should also be a focus of complementary societal investments, as the scope of the legal and platform response to campaign-disinformation deepfakes is comparatively limited, even though there is significantly more reported non-consensual sexualised content.

Lastly, election officials and civil society groups should ensure that authoritative sources of election information, which are easily found and accessible, such as official government election information portals, are included as preferred sources, available in the information retrieval of AI systems, particularly as most inaccuracies from AI chatbots are retrieval failures, not reasoning failures, as noted in Section 10 (TechRadar, 2026).

These four levels are not likely to be effective on their own and the findings from the case studies examined in this paper indicate that they are best used as complementary rather than alternative. It was not one, but a series of enforcement actions by the Federal Communications Commission, criminal prosecutions at the state level, and a good deal of investigative reporting that ultimately addressed the New Hampshire robocall, rather than any single approach, and at the same time, the outcome of the case (an acquittal on the primary state charges) proved the limits of applying existing law to a truly new form of technological harm (The Register, 2024; CBS News, 2025). Likewise, researchers have credited the relatively low rate of the use of harmful deepfakes in the 2025 Canadian election to a mix of proactive platform detection, growing public and journalistic awareness after the many high-profile incidents of harmful deepfakes in other elections last year, and the ongoing, though reduced, availability of professional fact-checking capacity (CETAS, 2025). This seems to indicate that the marginal benefit of any single component of the intervention—whether a new detection algorithm, a new disclosure law, or a new literacy curriculum—will probably be much lower than the benefit of coordinated investment in all three components at the same time, and this has significant implications for the allocation of time in the relatively short window that lies ahead before the peak of the 2026 election season.

**Limitations**

There are a number of restrictions in this paper. It is a desk-based synthesis of secondary sources, meaning that it does not provide original empirical data; it depends on the accuracy and methodological rigor of the sources and reports it includes, some of which are preliminary or yet to be fully peer-reviewed, such as some cybersecurity industry reports and news-based summaries

of the results of various academic studies. Second, because of the changing nature of both generative AI technology and regulations, some of the statistics presented here are subject to change from the time of the original publication of the sources to the time of publication or reading. For example, the number of bills introduced and the number of bills that passed are subject to change. Also, policy details on the platforms are subject to change. Third, the geographic distribution of available empirical research is uneven as well, with much more written about elections in the United States, the European Union, and a handful of prominent cases in Asia and Latin America, and less on many other parts of the world, which could underestimate the extent or nature of the issues in less-studied settings. Lastly, as many of the figures and data on survey and detection-accuracy comes from industry sources with possible commercial interests in either alarming or promoting reassuring statistics, for many of the specific figures, the reader should assume some ambiguity in the accuracy of the data, where possible, this paper has tried to triangulate such figures against other independent academic and journalistic sources.

**Conclusion**

The evidence considered in this paper presents a picture of the threat of AI-generated disinformation to democratic elections as real, measurable, and evolving, but not one that could be neatly characterized as technological apocalypse vs. hyped-up panic since 2023 as so much public debate on the subject has done. Not only was the number of recorded deepfakes in the 2024 global election cycle smaller than originally anticipated, empirical evidence shows that the limiting factors for disinformation's impact remain its distribution and audience attention, rather than content production. But the fact that the volume of AI-generated content can underestimate the extent of harm to certain groups (such as the high number of non-consensual sexualised images of women in public life) and to certain moments (such as the high number of AI-generated audio fabrications designed to discourage voters from participating in the election) is clear.

More importantly, however, the last three years (2024–2026) have shown that the indirect impact of AI disinformation – such as by triggering general public uncertainty, the liar's dividend, or the withdrawal of institutional fact-checking infrastructure – may have a more lasting effect on democratic accountability than any single viral piece of disinformation. The professional fact-checking capability, which can allow someone to rule on such conflicts, has been significantly diminished on the main platforms in a political and media environment where it's plausible to dismiss as fake any evidence that might be real.

As AI becomes more integrated into election campaigns and its usage of AI-generated content grows, as centralized content moderation becomes less predominant, and as AI chatbots are now a less reliable source of information but are increasingly consulted, the risk environment has changed in a way that demands ongoing monitoring even if the volume of deepfakes observed in 2024 and 2025 doesn't significantly rise. The multi-level approach suggested in this paper, as well as a coordinated, constitutionally durable regulatory framework and evidence-based media literacy programs, is not a guarantee of resilience, but it is the best synthesis of what the extant research suggests can usefully combat the negative effects of AI-driven disinformation on democracy. In the end, ensuring the integrity of the 2026 elections and subsequent ones will probably rely less on any one technology or legal solution and more on an ongoing, coordinated effort by platforms, regulators, election officials and citizens to maintain a verifiable, shared factual basis for democratic decision-making.

**References**

Adaptive Security. (2026, July 10). AI deepfake trends 2025–2026: Threats, detection & defense. https://www.adaptivesecurity.com/blog/ai-deepfake-trends-the-complete-2025-2026-guide-to-statistics-threats-detection-and-defense-stra

Al Jazeera. (2024, February 20). Deepfake democracy: Behind the AI trickery shaping India's 2024 election. https://www.aljazeera.com/news/2024/2/20/deepfake-democracy-behind-the-ai-trickery-shaping-indias-2024-elections

Al Jazeera. (2024, August 27). Elon Musk's X tweaks chatbot after warning over US election misinformation. https://www.aljazeera.com/economy/2024/8/27/elon-musks-x-tweaks-chatbot-after-warning-over-us-election-misinformation

American Sunlight Project. (2024, December 11). Deepfake pornography targeting members of Congress. https://www.americansunlight.org/updates/deepfake-pornography-targeting-members-of-congress

Axios. (2024, August 5). Musk's AI chatbot spread election misinformation, secretaries of state say. https://www.axios.com/2024/08/05/elon-musk-grok-2024-election-ballot-misinformation

Ballotpedia. (2025, July 30). Press release: State deepfake laws hit record pace. https://ballotpedia.org/Press_release_State_Deepfake_Laws_Hit_Record_Pace

Ballotpedia. (2026a, April 3). 15 deepfake bills enacted so far this year, number of states with deepfake laws remains the same. https://news.ballotpedia.org/2026/04/03/15-deepfake-bills-enacted-so-far-this-year-number-of-states-with-deepfake-laws-remains-the-same/

Ballotpedia. (2026b, July 30). 49 states have passed at least one deepfake law since 2019. https://news.ballotpedia.org/2026/07/30/49-states-have-passed-at-least-one-deepfake-law-since-2019/

Berkeley Technology Law Journal. (2025, May 24). Meta's fact-checking rollback: Governance, free speech, and user safety. https://btlj.org/2025/05/metas-fact-checking-rollback-governance-free-speech-and-user-safety/

Binding Hook. (2026, February 3). How deepfakes and gendered disinformation exclude women from public and political life. https://bindinghook.com/how-deepfakes-and-gendered-disinformation-exclude-women-from-public-and-political-life/

Blackbird.AI. (2026, June 10). WEF Global Risks Report: Narrative attacks are an increasingly critical global threat in 2026. https://blackbird.ai/blog/wef-global-risks-report-narrative-attacks-critical-global-threat-2026/

Brennan Center for Justice. (2024, January 23). Deepfakes, elections, and shrinking the liar's dividend. https://www.brennancenter.org/our-work/research-reports/deepfakes-elections-and-shrinking-liars-dividend

Brennan Center for Justice. (2026). Midterms 2026. https://www.brennancenter.org/midterms-2026

CBS News. (2025). New Hampshire jury acquits consultant behind AI robocalls mimicking Biden on all charges. https://www.cbsnews.com/amp/boston/news/new-hampshire-jury-acquits-consultant-ai-robocalls-biden

Center for Democracy and Technology. (2026, June 15). When the algorithmic odds are not in your favor: Algorithmic poisoning and the 2026 U.S. midterms. https://cdt.org/insights/when-the-algorithmic-odds-are-not-in-your-favor-algorithmic-poisoning-and-the-2026-u-s-midterms/

Centre for Emerging Technology and Security. (2025, November 17). From deepfake scams to poisoned chatbots: AI and election security in 2025. https://cetas.turing.ac.uk/publications/deepfake-scams-poisoned-chatbots

Check Point Blog. (2026, June 2). The 2026 U.S. midterms have a cyber problem, but it's not at the ballot box. https://blog.checkpoint.com/exposure-management/the-2026-u-s-midterms-have-a-cyber-problem-but-its-not-at-the-ballot-box/

CIVICUS & Digital Democracy Initiative. (2025). Future-proofing elections against deepfake disinformation. https://www.civicus.org/documents/ddi/final-deepfakes-elections-report-civicus-ddi-research-2025.pdf

Context by TRF. (2024, January 3). Deepfakes deceive voters from India to Indonesia before elections. https://www.context.news/ai/deepfakes-deceive-voters-from-india-to-indonesia-before-elections

Dan, V. (2025). Deepfakes as a democratic threat: Experimental evidence shows noxious effects that are reducible through journalistic fact checks. International Journal of Press/Politics. https://doi.org/10.1177/19401612251317766

Dawn.com. (2024, January 4). Deepfakes deceive voters in India, Pakistan before elections. https://www.dawn.com/news/1803005

DiResta, R., & Goldstein, J. A. (2024). How spammers and scammers leverage AI-generated images on Facebook for audience growth. arXiv. https://arxiv.org/abs/2403.12838

Drolsbach, C., & Pröllochs, N. (2025). Characterizing AI-generated misinformation on social media. arXiv. https://arxiv.org/abs/2505.10266

Euronews. (2026, July 28). The EU is forcing tech companies to label deepfakes. Will it work? https://www.euronews.com/my-europe/2026/07/28/the-eu-is-forcing-tech-companies-to-label-deepfakes-will-it-work

European Commission. (2026). Code of practice on transparency of AI-generated content. Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/policies/code-practice-ai-generated-content

European Parliamentary Research Service. (2025). Briefing: AI, deepfakes and the integrity of elections. European Parliament. https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/779259/EPRS_BRI(2025)779259_EN.pdf

Federal Communications Commission. (2024, May 23). FCC proposes $6 million fine for illegal robocalls that used deepfake, generative AI voice message [Press release]. https://docs.fcc.gov/public/attachments/DOC-402762A1.pdf

Fortune. (2024, August 26). Elon Musk's X changes Grok chatbot after states say it spread election misinformation. https://fortune.com/2024/08/26/elon-musk-x-grok-ai-chatbot-presidential-election-misinformation-joe-biden

France24. (2025, January 6). 'Form of violence': Across globe, deepfake porn targets women politicians. https://www.france24.com/en/live-news/20250106-form-of-violence-across-globe-deepfake-porn-targets-women-politicians

Frontiers in Political Science. (2024). AI-generated misinformation in the election year 2024: Measures of the European Union. Frontiers in Political Science. https://doi.org/10.3389/fpos.2024.1451601

Frontiers in Political Science. (2025). Disinformation, AI and regulation in Ecuador's 2025 presidential election. Frontiers in Political Science. https://doi.org/10.3389/fpos.2025.1624206

Global Views World. (2026). AI's impact on 2026 elections: What campaigns must know. https://globalviewsworld.com/2026-elections-ai-policy-s-impact-on-campaigns/

Grawitch, M. (2025). Fact-checking in the age of Community Notes: What Meta's move to crowdsourcing reveals about misinformation and truth. https://mattgrawitch.substack.com/p/fact-checking-in-the-age-of-community

Greenberg Traurig. (2026, June 8). Deepfakes, chatbots, AI-generated text: European Commission details transparency obligations under the AI Act. https://www.gtlaw.com/en/insights/2026/6/deepfakes-chatbots-ai-generated-text-european-commission-details-transparency-obligations-under-the-ai-act

Huang, G., & Hu, B. (2025). "A warning is not enough. Teach me how to spot deepfakes.": Testing media literacy interventions for combating deepfakes. Journalism & Mass Communication Quarterly. https://doi.org/10.1177/10755470251382889

International Knowledge Network of Women in Politics. (2025, January 7). AI-generated deepfakes targeting women politicians around the world. https://iknowpolitics.org/en/news/ai-generated-deepfakes-targeting-women-politicians-around-world

Kirkland & Ellis. (2026, February 17). Illuminating AI: The EU's first draft code of practice on transparency for AI-generated content. https://www.kirkland.com/publications/kirkland-alert/2026/02/illuminating-ai-the-eus-first-draft-code-of-practice-on-transparency-for-ai

Knight First Amendment Institute. (2024, December 13). We looked at 78 election deepfakes. Political misinformation is not an AI problem. https://knightcolumbia.org/blog/we-looked-at-78-election-deepfakes-political-misinformation-is-not-an-ai-problem

LiveNOW from FOX. (2025, April 4). Meta to stop fact-checking in the US on April 7 — Community Notes to take over. https://www.livenowfox.com/news/meta-ends-us-fact-checking-2025

Malay Mail. (2024, February 10). AI-generated party video has Khan claiming victory in Pakistan election. https://www.malaymail.com/amp/news/world/2024/02/10/ai-generated-party-video-has-khan-claiming-victory-in-pakistan-election/117337

Momeni, M. (2025). Artificial intelligence and political deepfakes: Shaping citizen perceptions through misinformation. India Quarterly. https://doi.org/10.1177/09732586241277335

NBC News. (2024a, March 18). New Hampshire voters sue Biden deepfake robocall creators. https://www.nbcnews.com/politics/2024-election/new-hampshire-voters-sue-biden-deepfake-robocall-creators-rcna143662

NBC News. (2024b, June 3). AI chatbots got questions about the 2024 election wrong 27% of the time, study finds. https://www.nbcnews.com/tech/tech-news/ai-chatbots-got-questions-2024-election-wrong-27-time-study-finds-rcna155640

NBC News. (2025, January 7). Meta is ending its fact-checking program in favor of a 'community notes' system similar to X's. https://www.nbcnews.com/tech/social-media/meta-ends-fact-checking-program-community-notes-x-rcna186468

Meta. (2026, February 6). More speech and fewer mistakes. About Meta. https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/

NewsNation Now. (2026). Almost 30% of AI chatbots show mixed accuracy on election information: Study. https://www.newsnationnow.com/business/tech/ai/ai-chatbot-replies-election-information-inaccurate/

Optics & Photonics News. (2025, February 1). Generating and detecting deepfakes: A 21st-century arms race. https://www.optica-opn.org/home/articles/volume_36/february_2025/features/generating_and_detecting_deepfakes_a_21st-century_arms_race/

Political & Government News. (2026, July 16). Deepfakes and the paradox of the liar's dividend. https://politics-government.news-articles.net/content/2026/07/16/deepfakes-and-the-paradox-of-the-liar-s-dividend.html

Poynter. (2025, January 8). Meta is ending its third-party fact-checking partnership with US partners. Here's how that program works. https://www.poynter.org/fact-checking/2025/meta-ends-fact-checking-community-notes-facebook/

Profile News. (2026, February 16). AI in global elections 2026: Disinformation risks and new regulations. https://www.profilenews.com/en/ai-in-global-elections-2026-explainer/

Public Citizen. (2026, May 13). 30 states now have laws to regulate election deepfakes. https://www.citizen.org/news/30-states-now-have-laws-to-regulate-election-deepfakes/

Resemble.ai. (2026, July 6). The EU AI Act: What generative AI companies need to know in 2026. https://www.resemble.ai/resources/the-eu-ai-act-what-generative-ai-companies-need-to-know-in-2026

SecurityWeek. (2025, June 13). The AI arms race: Deepfake generation vs. detection. https://www.securityweek.com/deepfakes-and-the-ai-battle-between-generation-and-detection/

Stackcyber. (2026, May 14). Deepfake legislation tracker: Federal, state laws. https://stackcyber.com/posts/ai-deepfake-laws

TechPolicy.Press. (2024, February 29). Deepfakes and elections: The risk to women's political participation. https://www.techpolicy.press/deepfakes-and-elections-the-risk-to-womens-political-participation/

TechPolicy.Press. (2026, January 7). What the EU's new AI code of practice means for labeling deepfakes. https://www.techpolicy.press/what-the-eus-new-ai-code-of-practice-means-for-labeling-deepfakes/

TechRadar. (2024, August 26). X pledges to fix AI chatbot after it spread misinformation about US election. https://www.techradar.com/pro/x-fixes-ai-chatbot-after-it-spread-misinformation-about-us-election

TechRadar. (2026, May 22). AI chatbots got election information wrong 90% of the time in a new study — including ChatGPT rivals. https://www.techradar.com/ai-platforms-assistants/ai-chatbots-got-election-information-wrong-90-percent-of-the-time-in-a-new-study-including-chatgpt-rivals

The Hill. (2026). Study: Nearly one-third of AI chatbots show mixed accuracy on election information. https://thehill.com/policy/technology/6071757-ai-chatbots-election-info-inaccuracy/

The Markup. (2024, December 11). 1 in 6 Congresswomen targeted by AI-generated sexually explicit deepfakes. https://themarkup.org/artificial-intelligence/2024/12/11/1-in-6-congresswomen-targeted-by-ai-generated-sexually-explicit-deepfakes

The Register. (2024, May 25). Man behind deepfake Biden robocall indicted, faces $6M fine. https://www.theregister.com/2024/05/24/biden_robocall_charges/

Time. (2026, April 13). Meta ends fact-checking, prompting fears of misinformation. https://time.com/7205332/meta-fact-checking-community-notes/

Tufts Now. (2025, January 23). The end of fact-checking increases the dangers of social media. https://now.tufts.edu/2025/01/23/end-fact-checking-increases-dangers-social-media

Weventure. (2025, July 14). AI labeling requirement starting now: What you need to know. https://weventure.de/en/blog/ai-labeling

Wikipedia. (2026). Liar's dividend. https://en.wikipedia.org/wiki/Liar%27s_dividend

World Economic Forum. (2025, July). Detecting dangerous AI is essential in the deepfake era. https://www.weforum.org/stories/2025/07/why-detecting-dangerous-ai-is-key-to-keeping-trust-alive/

World Economic Forum. (2026a, January 14). Top 10 risks in 2026: Geoeconomic confrontation tops the list. https://www.weforum.org/stories/global-risks/global-risks-2026-top-10-two-and-ten-year-horizon/

World Economic Forum. (2026b, March 12). How cognitive manipulation and AI will shape disinformation in 2026. https://www.weforum.org/stories/2026/03/how-cognitive-manipulation-and-ai-will-shape-disinformation-in-2026/

Zurich. (2025, February 24). Global Risks Report spells out top risks for 2025 and beyond. https://www.zurichna.com/knowledge/articles/2025/02/global-risks-report-spells-out-top-risks-for-2025-and-beyond