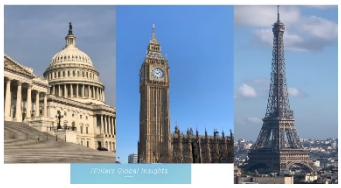


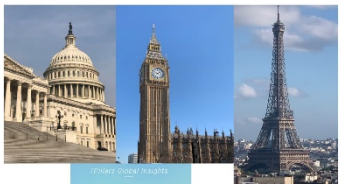
ADDRESSING THE THREAT OF GENERATIVE AI ON UPCOMING ELECTIONS

- The upcoming elections provide a target-rich environment for malign actors to weaponize generative AI tools to disrupt those elections and mislead voters, either to keep them from voting or to sow distrust in the results, or potentially to incite political violence.
- As it is early days for detection tools, it is imperative that election officials and election workers are prepared for the likely deployment at scale of deepfake audio, images, videos and texts and that voters are equipped to navigate a world where what appears to be convincingly real may in fact not be.
- Tabletop exercises and contingency planning will be critical for election administrators. Catchy public service announcements will need to be deployed at scale to educate voters, journalists and law enforcement.

For those tasked with ensuring the integrity of our upcoming elections and reducing the likelihood that they will be disrupted, there is little doubt that generative AI tools pose a significant threat to election processes, election offices and election infrastructure. The challenge is that many election officials still have little awareness of the threats, let alone how best to respond to them and how best to educate voters to not be influenced by them. Civil society groups focused on election security, as well as the Cybersecurity and Infrastructure Security Agency ("CISA"), are prioritizing efforts to ensure that election officials and others involved in the administration of elections are aware of the risks and have a roadmap for responding to the expected threats.

A report recently published by the Brennan Center for Justice, together with the Institute for the Future and The Elections Group, highlights the magnitude of the threats and sets out one such roadmap for responding to the threats. The report describes a crisis planning tabletop exercise on responding to specific scenarios in which generative AI tools are deployed to disrupt the elections. That exercise was conducted by the Office of the Arizona Secretary of State in collaboration with the three groups. While there are an increasing number of articles and reports on the nature of the threats, this report has the added benefit of providing a comprehensive overview of potential scenarios and responses.

Condensing the key messages from the report, as well as two notices recently issued by CISA on the threats posed by generative AI to the upcoming elections: Risks of Generative AI and Foreign Malign Influence Operations, and themes covered in two of my recent briefing notes (Deepfakes Update and Deepfakes and Incitement to Violence), I set out below what voters should understand about election-related synthetic media.



How is generative AI expected to be weaponized for the elections?

Generative AI is software that can reorganize existing data or create new data using statistical models. The software can create new computer code, create new text or develop synthetic media such as fake or manipulated videos, images or audio files. These tools can be deployed to target election processes, election offices, election officials and election vendors. While the tactics are not new, generative AI allows disinformation and other harmful content to be spread at scale, in real time, for nominal cost and in a far more sophisticated format that makes detection far more difficult than the clunky versions deployed in earlier contests. Deepfakes are a form of synthetic media.

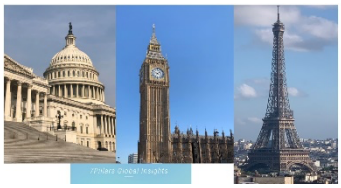
How can deepfake technology be deployed?

- ***Audio*** – malign actors can use audio technology to clone voices and to manipulate those cloned voices via text prompts to say whatever they want to be said. The impersonations can be overt, mimicking, for example, election officials or politicians, or covert, mimicking trusted voices of senior election officials to prompt election workers to act in violation of standard operating procedures.
- ***Images*** – high quality images can be produced from text prompts, allowing malign actors to modify people's faces, modify backgrounds and add/remove people or objects from the images. Deepfake tools can be used to create fake websites or to create deepfake images, for examples of election workers allegedly destroying ballots.
- ***Text*** – malign actors can generate false or misleading content for phishing attacks, to impersonate politicians or to create fake profiles, in each case to create and spread falsehoods about the elections.
- ***Video*** – malign actors can deploy generative AI tools to create fake content by uploading videos or creating fake videos based on text prompts. The technology would likely be used to impersonate trusted voices in the community or politicians.
- ***Malware*** – malign actors can use generative AI tools to create malware that could be deployed to attack election infrastructures or voter databases.

How are election-related deepfakes expected to be deployed? Some scenarios:

Malign actors seek to spread disinformation to misleading voters

- Malign actors impersonate trusted election officials or election workers using AI-generated audio clips or videos to mislead voters, including as to how, when or where to vote.
- Malign actors spread at scale AI-generated, personalized and interactive chatbot-based disinformation about the elections via encrypted messaging platforms, such as WhatsApp, email, social media, text or print, warning for example that ICE agents, police or army personnel will be deployed in polling stations.



- Malign actors spread at scale disinformation via robocalls or robotexts, for example, that voter identification will be cross-checked against outstanding warrants, unpaid tax assessments or child support obligations.
- Malign actors spoof election websites to spread disinformation about the electoral process or post fake election results.
- Malign actors use AI capabilities to increase the scale and persuasiveness of foreign influence operations targeting election processes.

Malign actors seek to mislead election workers to aid cyberattacks or disrupt elections

- Hackers undertake spear-phishing attacks to infect election infrastructure with malware or gain access to sensitive election administration or security information.
- Hackers use voice cloning tools to pretend to be from the office of the Secretary of State or from the Department of Homeland Security to gain access to sensitive election administration or security information.
- Malign actors use AI coding tools to develop malware to evade detection systems.
- Malign actors use voice-cloned audio clips to misinform election workers about polling places or polling hours.

Malign actors harass or seek to overwhelm election workers to disrupt their operations

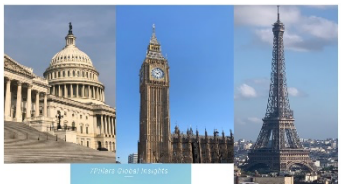
- Malign actors flood election administration offices with robocalls to overwhelm election officials or with AI-generated freedom of information requests that vary slightly in wording to make these mass-produced requests appear to be from multiple sources.
- Malign actors overwhelm election administration websites through denial-of-service attacks.
- Malign actors generate social media posts calling for armed protests at polling places.
- Malign actors harass, impersonate or otherwise delegitimize election officials using AI-generated content such as compromising deepfake videos.

Who is best placed to address deepfakes and what can they be expected to do?

State election officials, typically reporting to the Secretary of State, are the best placed to lead the effort to counter election-related disinformation, regardless of the form in which the disinformation is deployed. Election officials have support from state and federal agencies, including the Federal Bureau of Investigation and CISA. The Office of the Director of National Intelligence, together with the FBI and CISA, provide support on foreign influence operations targeting elections.

The advice provided to election officials as set out in the Brennan Center report includes the following:

- Understanding the breadth and depth of the AI threat.



- Taking control of the online presence.
- Preparing for rapid-response communications.
- Adopting online (cyber) and offline (physical) security protocols.
- Establishing and strengthening relationships with local media and other trusted partners, including faith leaders and other community leaders, sports figures and other celebrities.
- Creating incident response escalation protocols that would include the FBI, the Elections Infrastructure Information Sharing and Analysis Center (EI-ISAC), state election officials and social media platforms.
- Engaging with legal counsel to understand the legal remedies available based on the various potential scenarios.

Are there technological solutions available to combat deepfakes?

Generative AI is evolving at lightning speed, but regrettably it remains early days for deepfake detection technology. According to shirin anlen and Raquel Vázquez Llorente, of the civil society organization WITNESS ("Spotting the deepfakes in this year of elections: how AI detection tools work and where they fail"), disclosure techniques such as visible and invisible watermarking, digital fingerprinting, labelling and embedded metadata need more refinement to be fully effective. They note that today detection tools should not be considered "one-stop solutions" and should be used with caution. In particular, they found that publicly available software they tested has led to confusion, especially if used without the proper expertise to interpret the results. AI detection tools may find no evidence that content was AI-generated even though the content is synthetic.

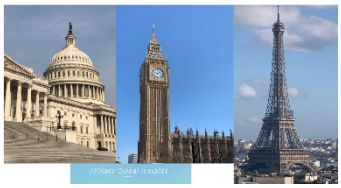
Detection tools are data-driven, trained using specific datasets, including pairs of verified and synthetic content. Accuracy depends on the quantity, quality and type of training data employed, as well as the algorithmic function the tool was designed for. In short, AI detection software can be fooled by the AI techniques they are designed to detect.

How should voters synthesize election-related information and assess sources?

In a separate article, the Brennan Center has provided the following suggestions for voters. These suggestions are premised on the likely absence of effective action by the platforms and effective federal legislation, as well as the rapid evolution of deepfake technology.

- Do not spend too much time trying to detect deepfakes through visual clues or reliance on detection software. Generative AI tools are far more sophisticated, rapidly outgrowing the early visual flaws that made it easier in the pasty to detect video deepfakes. There are gaps in the use of markers to trace provenance, and it is relatively easy to remove the markers. This means there may be fewer clues and the potential for false positives.

- Approach the legitimacy and credibility of election-related information and the sources thereof critically. Employ proven practices for evaluating content, such as



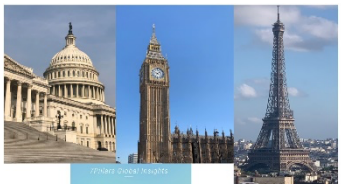
seeking out authoritative context from credible independent fact-checking sites for images, video and audio. Be wary of unfamiliar or unverified websites. Avoid search engines as a first step due to cached history.

- Assess emotionally charged, sensational and surprising content with caution.
- Avoid getting election information from generative AI chatbots and search engines that consistently integrate generative AI, and instead go to authoritative sources such as election office websites.
- Act responsibility when sharing political content that *may* be generated by AI.

Voters should also familiarize themselves ahead of the election with the relevant voting rules. The Brennan Center reported last week that in more than half the states (28 states) voters will face new voting restrictions that were put in place after the 2020 election.

Finally, voters should be mindful of the following:

- The date of the election (November 5) cannot be changed – announcements that the election has been delayed or that ballots can be submitted a day later or by phone, email or text (*e.g.,* due to alleged hacking of voting systems) are false.
- Short of a well-publicized event, a last-minute notice that a polling station has moved will be false.
- Warnings that voter IDs at polling stations and mail-in ballots will be checked against other governmental databases are false. Law enforcement will not be cross-referencing voter IDs or mail-in ballots with outstanding arrest warrants, unpaid parking tickets, unpaid child payment obligations, unpaid tax assessments or other legal obligations. (Two conservative activists who targeted Black voters during the run-up to the 2020 election were found guilty and recently sanctioned in an action brought by the New York Attorney General under New York State civil rights legislation and the federal Ku Klux Klan Act for seeking to suppress Black votes by making false warnings about arrest warrants and debts. They were also sanctioned in Ohio and by the Federal Communications Commission. First Amendment rights are not absolute.)
- Warnings that if you voted in a primary you cannot vote in the general election are false.
- Warnings that if you are registered with one political party you cannot vote in the general election for a candidate of another party are false.
- Claims that early votes, mail-in votes and provisional ballots only count in close elections are false.



Concluding Thoughts

WIRED has reported that more than 50 million robocalls were made in the two months leading up to the start of the election period in India and millions more were made during the election period. Harbingers of things to come in the United States.

WIRED also has called attention to the fact that the only federal agency that appears to be acting to address the use of generative AI tools to disrupt the US elections is the Federal Communications Commission ("FCC"). As I reported previously, the FCC reacted to the January New Hampshire robocall incident by banning the use of generative AI to clone voices for use in robocalls.[1] Yesterday, FCC Chair Jessica Rosenworcel proposed that the FCC adopt rules requiring on-air and written disclosure of AI-generated content in political ads aired on radio and television. If adopted, the rules would cover both candidate and issue ads, and would apply to broadcasters and outlets that engage in origination programming, including cable operators, satellite TV and radio providers. The FCC does not have jurisdiction over social media ads or streaming services.

The proposed FCC rules would not prohibit AI-generated content in political ads; they are disclosure rules only. This is all the more reason to ensure that society is best prepared for the onslaught of synthetic media that is certain to figure prominently in the remaining 166 days before the election.

* * *

Mark S. Bergman
7Pillars Global Insights, LLC
Washington, D.C.
May 23, 2024

[1] Today, the FCC proposed the imposition of a $6 million fine on the consultant behind the New Hampshire robocalls. The basis for the proposed fine is spoofing (inaccurate caller ID) of a prominent New Hampshire Democrat. The consultant reportedly is facing a slew of criminal charges as well.