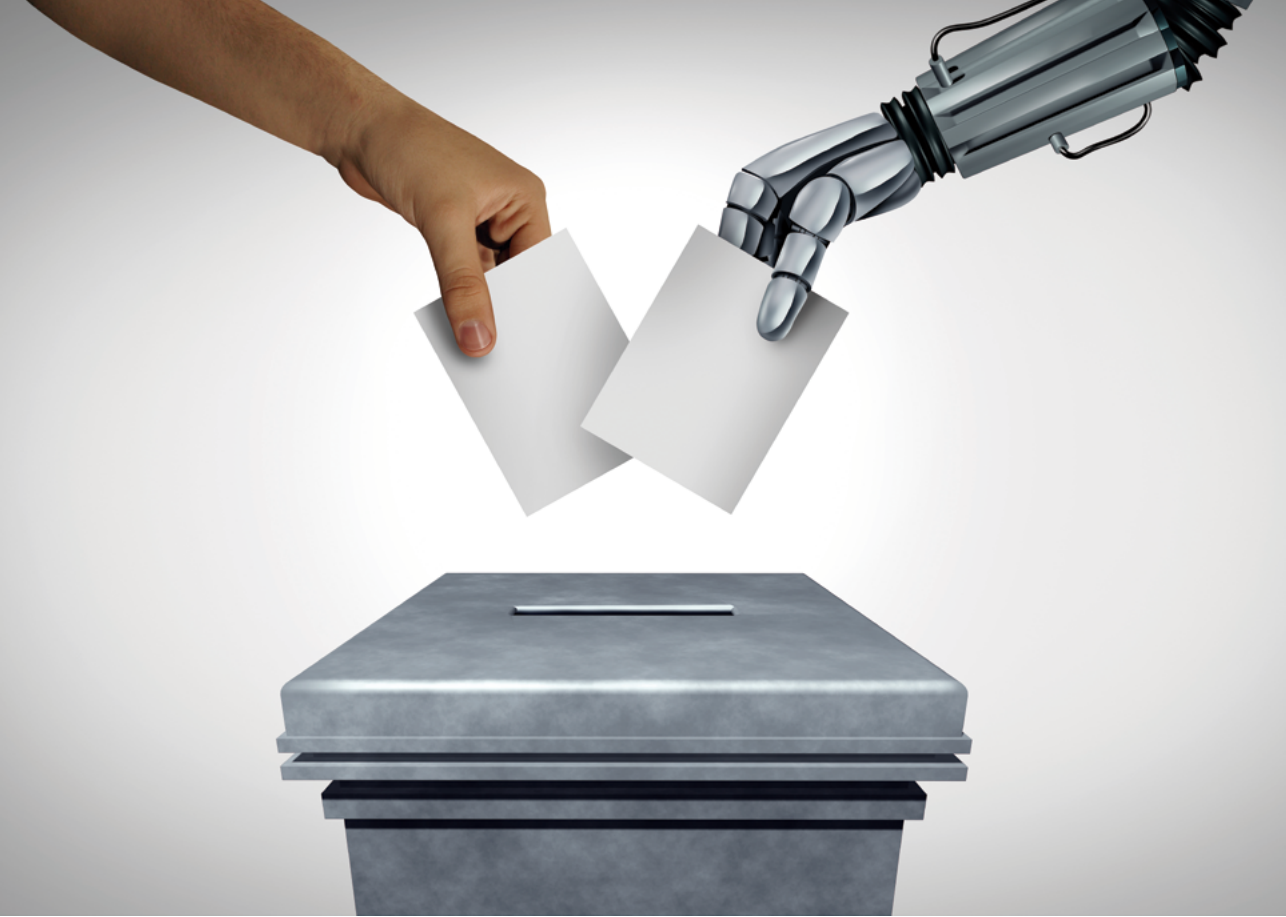




Artificial Intelligence and elections: premature threats?

SARAH E. KREPS

The new technological paradigm and deepfakes influence democratic processes. Manipulated products can mislead people into believing that certain contents are real. Democracies need to foster media literacy among their citizens to ensure that the founding values of democracy among their citizens endure in the face of new challenges.

In 2020, one of the major parties in India, the Bharatiya Janata Party (BJP), used deepfake technology to create videos of one of its [politicians](#), Manoj Tiwari. The videos portrayed Tiwari speaking languages that he does not speak, such as Haryanvi and English, with a goal of targeting different linguistic demographics. The video was relatively benign, an attempt to portray the individual in a positive light, but made manifest the potentially less benign consequences were the technology to be misused. The prospect was not just hypothetical. Just a year before in Gabon, the Gabon presidential office [released](#) a video of its infirm leader, Ali Bongo, who had suffered a stroke, aiming to dispel rumors about his health and political stability. Skepticism about its authenticity fueled political unrest and led to an attempted coup among military officers seeking to restore democracy and stability in the country.

With the rise of artificial intelligence and the growing examples of deepfakes insinuating themselves into the democratic process, researchers have warned about more extreme misuse cases. In 2023, the warnings about deepfakes mounted as 2024, a historic year for elections—with “more voters than ever in history” voting, 64 countries and a combined population

of 49% of the world—[approached](#). *Newsweek* [warned](#) that “deepfakes could destroy the 2024 election.” More generally, the proliferation of generative Artificial Intelligence such as ChatGPT means that it’s not just images that can be inauthentic, but also text such as the news stories individuals read about politics ([Kreps et al 2020](#)). This has prompted scholars to warn about the democracy-eroding effects of AI-generated text ([Kreps and Kriner 2023](#)).

And yet, election after election in 2024 showed that these warnings were overblown or at least premature. In most of the elections, deepfakes or AI-generated content was largely absent, seen by relatively few individuals, and certainly not consequential enough to sway elections.

The question then is why? With such potential to destroy elections in ever-different ways—given that the number and use of deepfakes or other AI-generated content is almost infinite—why has AI either not been created or not had consequential impacts on the elections consistent with the pessimistic predictions?

This essay first defines deepfakes and AI-generated content and the reason the technology has been predicted to undermine democracy, particularly elections. It

then takes stock of how AI has been used in different 2024 elections, pointing to the dearth of significant impacts relative to the theoretical prospects. The essay then offers ideas for why AI has not had the consequential impact in line with its potential and suggests why these past experiences may not be prologue. It closes with reflections about potential future misuses and how democratic polities must remain vigilant and digitally literate.

What is Generative AI and What is the Potential Threat to Democracy?

Generative Artificial Intelligence [refers](#) to a subset of artificial intelligence that is capable of creating new content, such as text, images, audio, and video. The technology relies on *machine learning* models, particularly those involving deep learning, to generate outputs that mimic real-world data. Although the rise of consumer-facing ChatGPT has led to the proliferation of text-based generative AI and concerns about democratic disruptions, deepfakes have already created realistic but fake videos of political candidates or public figures and been used for misinformation.

Deepfakes [are](#) synthetic media in which a person in an existing image or video is replaced with someone else's likeness using artificial intelligence (AI) and machine learning techniques. The term "deepfake" comes from the combination of "deep learning," which is a subset of machine learning, and "fake," which implies something that is inauthentic. Deepfakes are created using a type of artificial intelligence called Generative Adversarial Networks (GANs). Think of GANs as a pair of digital artists. One artist, the "generator," tries to create fake images or videos that look real, while the other artist, the "discriminator," tries to spot the fakes. Through this back-and-forth

process, the generator gets better at making realistic-looking fake content. Another type of AI, called autoencoders, helps by learning the patterns in real images or videos and then using that knowledge to recreate similar, but fake, content.

The fundamental concern with deepfakes is that their verisimilitude between reality and the manipulated audio, video, or images can mislead individuals into thinking that the content is real. Deepfakes stand to pose a theoretical threat to democratic elections due to their potential to manipulate public perception, spread misinformation, and undermine trust in the electoral process.

For example, they could be used to create false narratives by making it appear that political figures have said or done things they never actually did. This can mislead the public and sway public opinion based on fabricated information. The [Gabon deepfake](#) is a good example, having both eroded trust in public figures and the media and undermining the integrity of the democracy.

Indeed, scholars have warned and shown that people cannot discern AI-generated news content from actual news content ([Kreps et al 2020](#)), that AI-generated propaganda is persuasive to individuals in ways that could accomplish the goal of psychological manipulation (Goldstein et

» The fundamental concern with deepfakes is that their verisimilitude between reality and the manipulated audio, video, or images can mislead individuals into thinking that the content is real. «

» [Deepfakes] could be used to create false narratives by making it appear that political figures have said or done things they never actually did. «

al [2024](#)), and that members of Congress cannot distinguish between AI and human-generated constituent messages, which can potentially affect the legislative agenda given the potential to write advocacy messages at scale ([Kreps and Kriner 2023](#)). Although these studies are about AI-generated content in general rather than specifically deepfakes, the points are the same. Individuals cannot tell the difference between AI and human-generated content, the content can manipulate individuals, and these technologies have the capacity to proliferate and scale.

Although propaganda and misinformation have long been available, generative AI is different in several ways. For example, deepfakes can create highly realistic audio, images, and videos that are much more convincing than traditional methods that might seek to misrepresent reality, which makes it more difficult for individuals to flag the inauthenticity. Relatedly, the auditory or visual nature of some types of generated content can have a stronger emotional impact on individuals compared to text-based misinformation, which makes the message more memorable and persuasive. The proliferation of generative AI technology also makes it faster, cheaper, and easier—even in real-time—for anyone to produce sophisticated content that was previously only possible for well-resourced organizations or state actors. These differences make deepfakes a particularly

potent tool for [disinformation and manipulation](#), posing new challenges for maintaining the integrity of public discourse and democratic processes.

Beyond the direct manipulative potential of AI-generated content, another concern about deepfakes is not even deepfakes themselves but the way in which the proliferation of deepfakes creates dubiousness about information truth more generally. If people come to think that anything could be untrue, in other words, they might not trust that anything is true, leading to an erosion of information sources. The so-called “[liar’s dividend](#),” in which inauthentic information is so prevalent that deniability of reality becomes more plausible, becomes more of a problem as the existence of deepfakes undermines trust in genuine media and information.

Government officials across the world have taken the risks of generative AI seriously. The United States’ Cybersecurity and Infrastructure, for example, has cautioned that generative AI “may amplify existing risks to election infrastructure.” Wired magazine, which covers tech, called 2024 “[the Year of the Generative AI Election](#).” While the theoretical risk for harm exists, whether through text or, more likely, through images, have the experiences actually validated the concerns? The next section offers cautious optimism, suggesting that the use or rather misuse of generative AI has been limited despite the widespread availability of the technology.

Taking Stock of 2024 Elections

Warnings about the potential influence of generative AI and in particular deepfakes in the 2024 elections come from both the potential use case but also the rapid advancement of AI technology in the

last couple of years that has made the technology more accessible in previous major election cycles. Those warnings have not been entirely inflated.

Deepfakes have been used to target specific political figures in the 2024 elections in several different ways. AI-generated audio and video have been used to create fake recordings of politicians. In the Democratic primary in [New Hampshire](#), AI-generated audio was used in robocalls, with Biden's voice urging voters not to vote. Biden still won the New Hampshire primary handily despite the deepfake attempt to discourage voting.

In [Slovakia](#), a deepfake audio falsely attributed plans to rig an election to a political leader in advance of the 2023 parliamentary election. AI-generated images and videos have also been used to misrepresent politicians. AI-generated pictures [showed](#) Donald Trump with Black voters in ways intended to endear the constituency to Donald Trump. During the primary season in 2023, a political action committee associated with Florida Governor Ron DeSantis [used](#) AI-generated audio of Donald Trump to portray him attacking the Iowa Governor, intended to paint him as disrespectful of the caucus state. Other deepfakes have been used to create embarrassing or compromising content. For example, in the UK, an investigation [uncovered](#) 400 instances of digitally altered deep fakes showing 30 high-profile UK politicians in compromising situations.

In [Poland](#), the opposition party Civil Platform (Platforma Obywatelska) created a deepfake video imitating the voice of the Prime Minister. The content was based on leaked emails from the prime minister's chief of staff and alternated between showing real video clips of the Prime Minister speaking and AI-generated audio clips that read sections of the leaked emails.



» The auditory or visual nature of some types of generated content can have a stronger emotional impact on individuals compared to text-based misinformation. «



The apparent motivation was to contrast the Prime Minister's public statements about unity within the ruling coalition with private messages that acknowledged tensions in the government. Only after skepticism and criticism did the platform acknowledge that the content was AI-generated.

In the UK, the first day of the Labour Party conference in Liverpool saw the release of a deepfake audio clip showing

Keir Starmer verbally abusing and using profanity towards his staff members and another showing him criticizing the city of Liverpool. The clips [posted](#) on X by an account with less than 3,000 followers, receiving 1.4 million views. One deepfake detection company, Reality Defender, [indicated](#) that the audio was 75% likely manipulated and the British government's analysis also [confirmed](#) the inauthenticity of the content.

While voices across the political spectrum criticized the audio, it raised concerns about the threat of deepfakes to democracy and highlighted the challenges with debunking content.

The creation of deepfakes is certainly not limited to domestic politics. Russia has repeatedly been accused of creating deepfakes of the pro-Western Moldovan President Maia Sandu to ridicule the leader and undermine her credibility. CopyCop, a suspected Russian-aligned influence network, has used AI and inauthentic sites to create and distribute disinformation. Reports [suggest](#) little engagement or amplification on social media, however.

Although these examples point to ways that groups or individuals have used AI-generated deepfakes, the 2024 elections have also been notable for the absence of deepfakes. In Mexico's 2024 election, deepfakes did not feature at all and indeed, the election appears to have been conducted [without](#) major disruptive incidents, although outside actors had little probable cause.

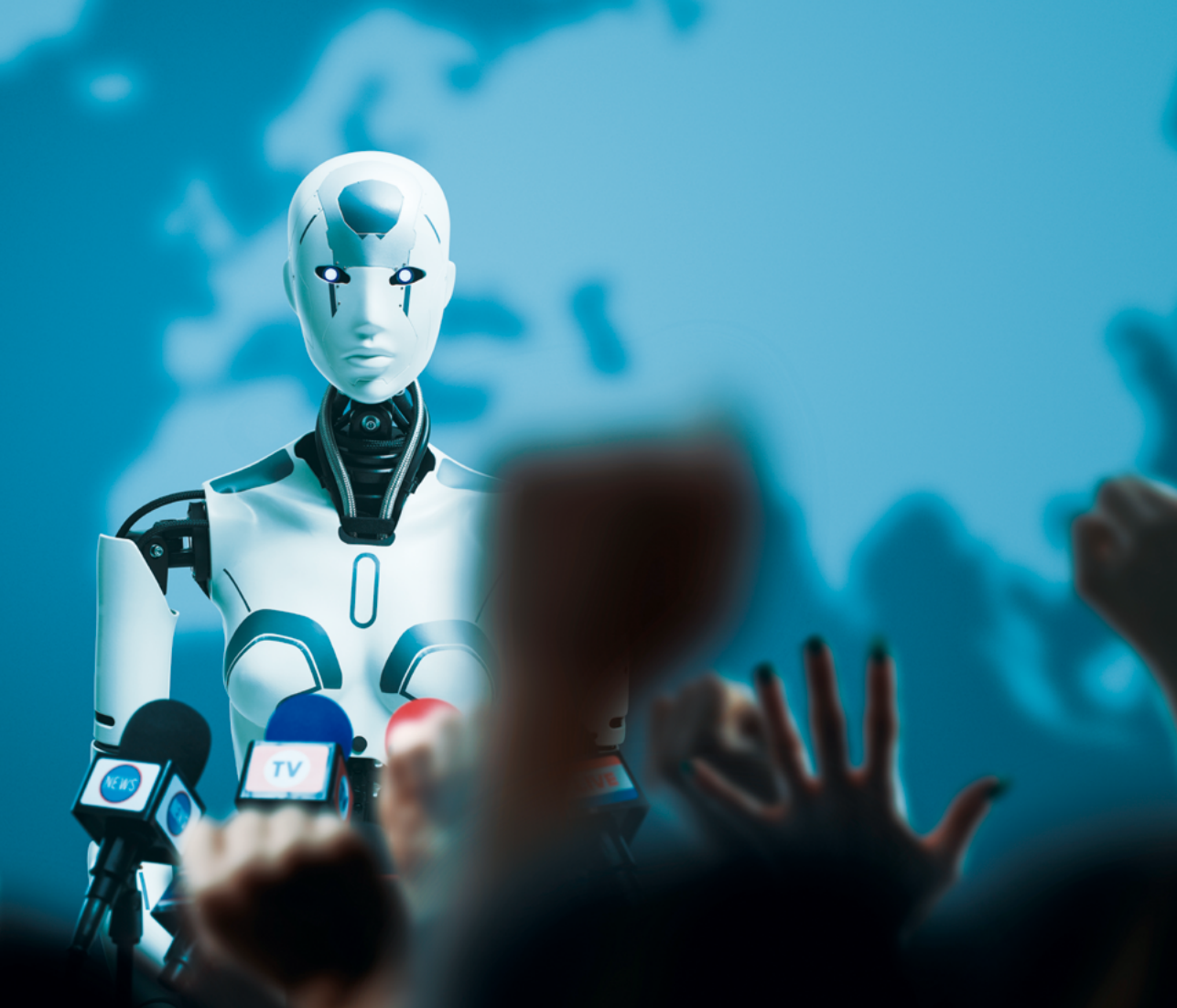
Despite the availability of deepfakes and tensions between Taiwan and China, its use was limited. China appeared to use AI-generated audio clips to [target](#) the Democratic Progressive Party presidential candidate Lai Ching-te. The AI-generated content included manipulated video content of the candidate, with audio appearing to be Lai talking about scandals that had not occurred and supporting a coalition with Kuomintang (KMT), the Chinese nationalist party that ruled until it was defeated by the Communist party of China on the mainland.

But more than deepfakes, China relied on other misinformation techniques. China pushed false or misleading stories on social media such as portraying the



» If people come to think that anything could be untrue, in other words, they might not trust that anything is true. «

United States as an unreliable ally that would abandon Taiwan, framing the election as a choice between “peace” (unifying with China) and “war” (continued independence), spreading false claims about U.S. biological labs in Taiwan,



promoting conspiracy theories about CIA interference in the elections, and promoting racist narratives against migrant workers.

China has become associated with techniques such as “[spamouflage](#)”, in which Chinese government-affiliated groups use accounts to actively promote particular narratives on social media. For example, these accounts aim to portray the United States in a negative light by highlighting urban decay, police brutality, and deteriorating infrastructure. These accounts are particularly active during events such as a natural disaster or election. In

April 2023, the US Department of Justice [charged](#) 40 employees of the Chinese Ministry of Public Security’s 912 Special Projects Working Group for their involvement in an influence campaign that appeared to be Spamouflage.

In addition to the continued spamouflage efforts, China not only distributed these messages but then relied on [Taiwanese proxies](#) to spread disinformation to make it more difficult to trace back to China. Beyond the immediate election outcomes, China appeared focused on eroding trust in Taiwan’s democracy and sovereignty over time.

But in no way were the deepfakes prevalent or did they appear to be effective. The same is true in the 2024 Indian election. Despite the concerns about widespread deepfake use, the actual number of verified AI-generated misinformation cases was relatively low. Out of 258 election-related fact-checks conducted by [Boom Live](#), only 12 involved AI-generated misinformation. Isolated cases included using deepfake technology to “resurrect” dead politicians for campaign purposes. Both major parties, BJP and Congress, both created and shared AI-generated content such as memes and satirical videos or AI-translated speeches or personalized AI robocalls. One study showed that of about 2000 viral WhatsApp messages, only 1% were generated by AI, a small footprint according to [Nature](#). The prevalence and impact of misused AI-generated content was limited, certainly less than initially feared.

Why Has AI-Generated Content Been Low Impact?

Scholars have shown proofs of concept for how malicious actors could use AI-generated misinformation at scale to disrupt democratic elections and yet there is a dearth of evidence to suggest that actors are either producing such content or having an impact on elections, which raises the question of why the predictions have been at odds with reality (Pawelec 2022).

One possibility for the limited effectiveness is that the technology is still nascent, particularly with respect to deepfake videos or audio. Users quickly identify and debunk the images or video in part because deepfakes are still discernible. The incorporation of video and audio sometimes creates mismatched synchronization in how words are projected,

which means attention to the lip movements can highlight inconsistencies. Relatedly, there are often subtle inconsistencies between the AI and the human likeness, much as were obvious with the Tom Cruise [deepfakes](#). Individuals viewing these images might experience a skeptical emotional response referred to as “[the uncanny valley](#)” when viewing a robotic image that is subtly not lifelike.

Another set of factors is that while the technology is nascent, individuals, political leaders, states, and social media platforms are actually [prepared](#) for these deepfakes, helping to neutralize the effect. Government officials, in some cases, have conducted simulations and tabletop exercises to respond to deepfakes. Some states have [passed](#) laws regulating the use of deepfakes in political campaigns, which may deter some potential bad actors. Algorithmic detection is [improving](#) in ways that allow social media platforms to flag and remove deepfakes, many of which have been banned by these platforms.

Another consideration, as the example of China in the 2024 Taiwanese election suggests, actors may simply see other forms of disinformation more practical or effective, such as appearing in online forums with particular perspectives or political valences to shape what people read.

Further, political persuasion is difficult. Research suggests that the effect of misinformation or disinformation (with the intention to mislead) tends to have limited impacts because people’s views are fairly entrenched. Indeed, studies of whether or how misinformation has affected political behavior often [produce](#) null findings because people tend not to change their minds, even when faced with the more viscerally powerful images (versus text).

» Democracies must foster a culture of critical thinking and media literacy among their citizens. «

Conclusion

Despite their potential for disruption, deepfakes and AI-generated content more generally have not yet been as prevalent or consequential during the 2024 election waves. Efforts to both identify and remove deepfakes have grown on the part of individuals, governments, and platforms. Impact on public opinion has been limited. The measures that society has adopted to guard against deepfakes have mitigated their impact but past success in mitigating the influence of deepfakes does not guarantee future immunity.

Nonetheless, as AI evolves, so will the efforts to manipulate democratic politics for electoral advantage. Public awareness and media literacy will continue to play an important role in reducing the impact. As the technology changes, so too should the campaigns by governments, non-profit organizations, and media outlets to highlight the type of critical thinking and skepticism needed to consume digital content to protect against manipulation and misinformation. News outlets will need to continue rigorous fact-checking to verify the authenticity of visual or audio content before publication to prevent the spread of deepfakes and prop up trust in the media. Regulatory and legal frameworks will need to stay current and should continue to evaluate the new technologies and ask what types of deepfakes are free speech and which ones should be prohibited.

This analysis produces an additional

cautionary note. As the Chinese online influence in the Taiwanese election suggests, actors seeking to influence public opinion will continue to use other means such as armies of online trolls or groups that engage in spamouflage, creating messages aiming to manipulate political opinion. Preoccupation with deepfakes may obscure those approaches and worse, divert resources and attention not just from those more established mechanisms that include phishing, ransomware, and other cyber threats. Security measures, funding, and research efforts might become disproportionately skewed towards combating deepfakes at the expense of broader cybersecurity initiatives, an opportunity cost in terms of time and effort invested in deepfake detection and prevention. Overemphasis on deepfakes may contribute to a self-fulfilling process, however, in the form of eroding trust in media and information by inculcating skepticism in the public and casting doubt on legitimate and authentic content.

Ongoing advancements in detection technologies, increased public awareness, and robust legal frameworks have proven effective in mitigating many of the threats posed by deepfakes. However, the dynamic nature of technology and the ever-evolving tactics of malicious actors require continuous vigilance and adaptation. Democracies must foster a culture of critical thinking and media literacy among their citizens while maintaining transparency and accountability in their institutions. By doing so, they can safeguard their integrity and continue to thrive in the digital age, turning potential vulnerabilities into strengths and ensuring that their foundational values endure against the challenges of modern technology. ♦

References

- 40 [Officers of China's National Police Charged in Transnational Repression Schemes Targeting U.S. Residents.](#) (2023, April 17). Office of Public Affairs.
- ADAM, D. (2024, June, 18). [Misinformation might sway elections — but not in the way that you think.](#) *Nature*.
- BICKERTON, J. (2023, March 24). [Deepfakes Could Destroy the 2024 Election.](#) *Newsweek*.
- CAHLAN, S. (2020, February 13). [How misinformation helped spark an attempted coup in Gabon.](#) *The Washington Post*.
- DE VYNCK, G. (2024, April 5). [The AI deepfake apocalypse is here. These are the ideas for fighting it.](#) *The Washington Post*.
- [Deepfake audio of Sir Keir Starmer released on first day of Labour conference. \(2023, October 9\).](#) *Sky News*.
- DEVINE, C., O'SULLIVAN, D., & LYNGAAS, S. (2024, February 1). [A fake recording of a candidate saying he'd rigged the election went viral. Experts say it's only the beginning.](#) *CNN*.
- ELLIOT, V. (2024, May 30). [2024 Is the Year of the Generative AI Election.](#) *Wired*.
- ELLISON, S., & WINGETT SANCHEZ, Y. (2024, May 8). [In Arizona, election workers trained with deepfakes to prepare for 2024.](#) *The Washington Post*.
- EWE, K. (2023, December 28). [The Ultimate Election Year: All the Elections Around the World in 2024.](#) *Time*.
- FISHER, M. (2022, July 21). [How I Became the Fake Tom Cruise.](#) *Hollywood Reporter*.
- GARIMELLA, K. & CHAUCHARD, S. (2024, June 5). [How prevalent is AI misinformation? What our studies in India show so far.](#) *Nature*.
- GILLIS, A. (2024, February). [uncanny valley.](#) *Techtarget*.
- GOLDSTEIN, J., CHAO, J., GROSSMAN, S., STAMOS, A., & TOMZ, M. (2024, February 20). [How persuasive is AI-generated propaganda?](#) *PNAS Nexus*, 3(2).
- [Here's How Deepfakes, Like the One BJP Used, Twist the Truth.](#) (2020, February 20). *VICE*.
- HUNG, CH-L., FU, W.-CH., LIU, CH-C., & TSA, H-J. (2024). [AI Disinformation Attacks and Taiwan's Responses during the 2024 Presidential Election.](#) Thomson Foundation.
- INSIKT GROUP. (2024, June 24). [Russia-Linked CopyCop Expands to Cover US Elections, Target Political Leaders.](#) *Recorded Future*.
- ISENSTADT, A. (2023, July 17). [DeSantis PAC uses AI-generated Trump voice in ad attacking ex-president.](#) *Politico*.
- IYENGAR, R. (2024, January 23). [How China Exploited Taiwan's Election—and What It Could Do Next.](#) *FP*.
- JACKSON, K., SCHIFF, D., & BUENO, N. (2024, February 20). [The Liar's Dividend: Can Politicians Claim Misinformation to Evade Accountability?](#) *American Political Science Review, First View*, pp. 1-20.
- JACOB, N. (2024, June 3). [2024 Elections Report: Fake Polls, Cheap Voice Clones, Communal Claims Go Viral.](#) *Boom*.
- KREPS, S., et al. (2020, November). [All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation.](#) *Journal of Experimental Political Science*, 9(1), 1-14.
- KREPS, S., & KRINER, D. (2023, October). [How AI Threatens Democracy.](#) *Journal of Democracy*.
- KREPS, S., & KRINER, D. L. (2023). [The Potential Impact of Emerging Technologies on Democratic Representation: Evidence from a Field Experiment.](#) *New Media & Society*.

MARTIN, A. (2023, October 9). [UK opposition leader targeted by AI-generated fake audio smear.](#) *The Record*.

MARTINAU, K. (2023, April 20). [What is generative AI?](#) *IBM*.

MORGAN, L. (2024, July 2). [Deepfake pornography is being used to humiliate and silence powerful female politicians like Angela Rayner and Penny Mordaunt. Why doesn't the law protect them?](#) *Glamour*.

[Opposition criticised for using AI-generated deepfake voice of PM in Polish election ad](#) (2023, August 25). *Notes from Poland*.

PAWELEC, M. (2022, September). Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions. *Digital Society*, 1(2).

POLGÁR, J., & WEN, A. (2024, October 10). [Deceptive Audio or Visual Media \('Deepfakes'\) 2024 Legislation.](#) NCSL.

PRUNEDA, P., & SALAZAR UGARTE, P. (2024, May 29). [Elections in Mexico: Beyond "Deepfakes".](#) *Wilson Center*.

SAINATO, M. (2024, March 4). [AI-generated images of Trump with Black voters being spread by supporters.](#) *The Guardian*.

SWENSON, A., & WEISSERT, W. (2024, January 23). [New Hampshire investigating fake Biden robocall meant to discourage voters ahead of primary.](#) *AP*.

Tsu, T. (2024, February 15). [Chinese Influence Campaign Pushes Disunity Before U.S. Election, Study Says.](#) *The New York Times*.

Verma, P. (2023, December 17). [The rise of AI fake news is creating a "misinformation superspreader".](#) *The Washington Post*.

YASAR, K. (2024). [What is deepfake technology?](#) *Techtarget*.



Sarah E. Kreps

US political scientist. Air Force veteran. Policy analyst on US foreign and defence policy. Professor of government at Cornell University and adjunct professor at West Point's Modern Warfare Institute. Author of books on drones, US military interventions and how the US funds its wars. Life member of the Council on Foreign Relations.

X: @sekreps

LI: sarah-kreps-51a3b7257