

## COMMENTARY

# Watch out for false claims of deepfakes, and actual deepfakes, this election year

Daniel S. Schiff, Kaylyn Jackson Schiff, and Natália Bueno

May 30, 2024

- 
- We find evidence across several studies that politicians and other public figures may receive a “liar’s dividend” in the form of a public support boost for falsely claiming that *real* scandals about them are fake.
- 
- These effects occur across the political spectrum, not just for supporters of certain politicians or political parties. However, claims of “deepfakes” in response to video evidence are currently less believable than attempts to discredit *text-based* reporting.
- 
- Such impacts warrant increased attention to fact-checking; watermarking or labeling synthetic content; initiatives aimed at increasing media literacy, digital literacy, and AI literacy; and efforts to further track the political impacts of deepfakes.
- 

Concerns are growing about the social and political impacts of deepfakes and generative AI, leading many to consider the [risks during a global election year](#)

[\(https://www.brookings.edu/articles/the-impact-of-generative-ai-in-a-global-election-year/\)](https://www.brookings.edu/articles/the-impact-of-generative-ai-in-a-global-election-year/) . Deepfakes, generative AI, and synthetic content are also receiving policy attention in the form of Biden’s [Executive Order on AI](#) ↗ and the National Institute of Standards and Technology’s [AI Safety Institute Consortium](#) ↗. This growing attention has resulted in decisions such as the Federal Communication Commission’s [ban on AI-generated voices in robocalls](#) ↗ and an [agreement](#) ↗ amongst Big Tech companies to label and limit, although not ban, AI-generated content on their platforms.

As researchers and as directors of the [Governance and Responsible AI Lab \(GRAIL\)](#) ↗ at Purdue University, we have followed these developments closely but are also concerned about more extensive *indirect* impacts of AI-generated content on individuals’ trust and belief in the broader informational and political environment. We have studied whether an environment of increasing AI-generated media content and misinformation might enable politicians and other public figures to successfully spread *misinformation about misinformation*, or to credibly “cry wolf” over fake news. Such claims could provide a benefit, or a [“liar’s dividend](#) ↗,” at the expense of accountability and public trust in institutions such as the government and media.

## The liar’s dividend

The liar’s dividend is a concept [coined by legal scholars](#) ↗ Bobby Chesney and Danielle Citron, who suggest that the existence of actual, and increasingly

realistic, deepfakes can make false claims of misinformation more credible. We extend this idea to thinking about political accountability: Can strategic and false claims that news stories are fake news or deepfakes benefit politicians by helping them maintain support after a scandal?

We started to think about the political impacts of generative AI back in 2018 when Jordan Peele released a [deepfake of President Obama ↗](#) for the purpose of entertainment and raising awareness about fake news. At the time, some doubted whether deepfakes would be harmful or have much of an impact at all. However, since then, we have seen deepfakes used to [interfere in elections ↗](#), [commit fraud ↗](#), accuse politicians of [sex-related scandals ↗](#), and to even to launch a [military coup ↗](#). Importantly, in many instances, the authenticity of images and video has been disputed, experts have been unable to come to a definitive conclusion (e.g., whether audio was faked or not), and deepfake detectors have been [found to be unreliable ↗](#).

Given this confusion over falsified content and deepfakes, we set out to experimentally study the liar's dividend to better understand the broader political impacts of generative AI.

## Evidence from our research

In [recently-published research](#) <sup>7</sup>, we surveyed over 15,000 American adults across five studies over several years (2020 to 2022). We presented participants with news stories about real scandals involving politicians from both major political parties, shown via a video or a text transcript. These stories included offensive comments about issues like race and gender (such as [infamous comments](#) <sup>7</sup> made by Todd Akin that were condemned by both parties).

After the scandals, participants were shown a response in which the politician claimed that the story is false and misleading or a deepfake, asserting that “you can’t know what’s true these days with so much misinformation out there.” For the purposes of comparison, some participants instead received an apology, a claim the story was false *without* invoking misinformation, or no response from the politician. We examined whether these false claims of misinformation raised support for the politician (e.g., willingness to vote for, support, or defend the politician) and impacted belief in the story and trust in the media.

We find consistent evidence across years and studies for the existence of a liar’s dividend. That is, people are indeed more willing to support politicians who cry wolf over fake news, and these false claims produce greater dividends for politicians than remaining silent or apologizing after a scandal. Surprisingly, we also find that false claims of misinformation impact people across the political spectrum, not just supporters of certain politicians or political parties. That is, even individuals of the opposite political party to the politician are susceptible to such false claims.

Importantly, false claims of misinformation are more effective when scandals are reported via text, like in a news article, whereas we find (in most cases) no evidence of a liar's dividend for scandals shown via video. For example, in one of our studies, 44% of respondents who learned of the scandal via text disagreed or strongly disagreed that they would "support the politician." But respondents who were subsequently exposed to the politician's claim of misinformation substantially decreased their opposition to the politician to around 32% to 34%, a 10 to 12 percentage point reduction.

In contrast, for respondents who learned about the scandal via video, we fail to see significant differences between those who were exposed to politician claims of misinformation and those who were not. This is interesting because the idea of the liar's dividend originated out of concerns about deepfakes. Still, our most recent study does suggest that as awareness of deepfake images and videos increases, false claims of deepfakes might also become more effective. One optimistic note is that we do not find that crying wolf about misinformation decreases trust in the media.

What do the results mean for citizens in a year with [major elections](#) around the world? First, it's important to be aware of the liar's dividend and that politicians may seek to profit from spreading misinformation *about* misinformation. Watch out for false claims of deepfakes, in addition to actual deepfakes, this election year. Claims of misinformation should warrant additional attention, including searching for confirmation, different forms of evidence, and multiple competing sources.

However, another takeaway is that if audio-visual evidence is indeed so convincing (for example, see [OpenAI's new video model ↗](#)), then politicians may also find it hard to rebut actual deepfakes when the time comes.

## Potential solutions

Given these challenges, what are possible technical and policy solutions?

First, **fact-checking** should be used not only in regards to news stories about public figures, but also in regards to their subsequent claims regarding the story. [Consistent with prior studies ↗](#), we find some evidence in our research that [fact-checking ↗](#) can eliminate the liar's dividend. However, other research points to [challenges with fact-checking ↗](#), including some individuals' limited willingness to seek out fact-checking information or trust certain sources, even though fact-checking organizations tend to offer [consistent assessments of claims ↗](#).

Second, [watermarking ↗](#) and **labeling** synthetic content are possible solutions. This could mean either a visual indicator or logo identifying content produced by generative AI, or information contained in the metadata of an image or video. [OpenAI ↗](#) has committed to using the standards from the Coalition for Content Provenance and Authenticity ([C2PA ↗](#)) for images produced using its DALL-E 3; [Meta ↗](#) has announced that it will not only continue to label images built with its AI generator as "Imagined with AI," but it will also label other AI-generated images

uploaded to its platforms; and [Google](#) will use SynthID to watermark audio generated using its Lyria model. Yet, for these solutions to be effective, people must recognize and understand visual labels, logos, or symbols; they must trust the providers; and the watermarks must be [tamper-proof](#), a technical goal that remains elusive.

Third, we should engage in initiatives to promote [media](#), [digital](#), and [AI](#) **literacy**. For example, individuals need to know how to evaluate the credibility of news sources, how social media platforms and generative AI generally work, and how they might encounter AI in their everyday lives. This should involve K-12 and post-secondary education, and even broader public information campaigns.

Finally, increased attention should be devoted to the **political impacts of generative AI**. Our lab, the Governance and Responsible AI Lab (GRAIL) at Purdue, has created the [Political Deepfakes Incidents Database](#) to track the targets, spread, and strategies employed by deepfakes related to political actors, events, or institutions. The database can be used by members of the public or journalists who are interested in a particular incident, image, or video, and it can also support researchers tracking broader trends in political deepfakes. Moreover, the database can support evidence-based policymaking around generative AI and can inform [legislation on the topic of deepfakes and democracy](#), such as the federal [DEEPFAKES Accountability Act](#) or pending legislation in [states like Indiana](#) that would prohibit the dissemination of “fabricated media” close to an election.

Unfortunately, and ironically, raising awareness about the proliferation and problematic incidents related to deepfakes and generative AI can also exacerbate feelings of uncertainty and distrust. However, we are hopeful that a combination of the solutions above, reflecting joint efforts across researchers, policymakers, journalists, and educators, as well as AI designers and developers, can mitigate the liar's dividend and build our collective resilience.

## AUTHORS



**Daniel S. Schiff** Assistant Professor, Department of Political Science

- Purdue University, **Co-Director** - Governance and Responsible AI Lab

X @Dan\_Schiff



**Kaylyn Jackson Schiff** Assistant Professor, Department of

**Political Science** - Purdue University, **Co-Director** - Governance and Responsible

AI Lab X @kaylynjackson



**Natália Bueno** Assistant Professor, Department of Political Science -

Emory University

---

## Acknowledgements and disclosures

Google and Meta are general, unrestricted donors to the Brookings Institution. The findings, interpretations, and conclusions posted in this piece are solely those of the authors and are not influenced by any donation.

---

**The Brookings Institution is committed to quality, independence, and impact.**

We are supported by a [diverse array of funders \(/about-us/annual-report/\)](/about-us/annual-report/). In line with our [values and policies \(/about-us/research-independence-and-integrity-policies/\)](/about-us/research-independence-and-integrity-policies/), each Brookings publication represents the sole views of its author(s).

Copyright 2025 The Brookings Institution