



ACTIONABLE DEEPPAKE CONTENT VERSUS PARODY – HAS GENERATIVE AI BLURRED THE DISTINCTION BETWEEN HUMOR AND IMPROPER ATTEMPTS TO INFLUENCE ELECTIONS?

- Elon Musk reposts a purported Harris campaign ad. It is digitally manipulated. But should it be treated as a “deepfake” or a “parody”?
- In the age of generative AI, where is the boundary between humor and manipulation?
- The Federal Communication Commission has published a rule for public comment that mandates labelling of AI-generated content used in political ads.
- The Federal Election Commission seems paralyzed.
- Most States have enacted or are considering labelling requirements for deepfakes. Many have exceptions for satire and parody.
- Just over three months to go before Election Day, and more questions than answers.

A week ago Friday, Elon Musk reposted a purported Harris campaign ad that within 24 hours had been viewed 134 million times. But it was not real.

The original “We Choose Freedom” [campaign video](#) had been digital manipulated to change the voice-over and accompanying text. A voice mimicking the Vice President’s, among other things, claimed that President Biden is senile, that the Vice President “does not know the first thing about running the country” and is a “deep state puppet,” and that she was selected because she is the “ultimate diversity hire.” The manipulated ad also inserted images of President Biden in lieu of Donald Trump and JD Vance. The digital manipulation, when first uploaded, had carried a disclaimer on X and YouTube that it was a “parody,” but when Musk reposted the video on his own X account a few hours later, there was no such disclaimer. Instead, “This is amazing,” and a laughing emoji. Musk eventually clarified on his profile that the ad was a joke.

The Harris campaign responded in a media statement, “We believe the American people want the real freedom, opportunity, and security Vice President Harris is offering; not the fake, manipulated lies of Elon Musk and Donald Trump.”

From the moment the manipulated video was reposted, a number of questions were being considered by the campaign, by regulators, by members of Congress, by election officials, and by the media.

- Was the ad technically a deepfake that should be covered by regulation and now is/should be covered by state law? Or was it a subset of manipulated content that should enjoy different treatment by reason of being a parody?
- Did the ad run afoul of Federal Communications Commission (“FCC”) or Federal Election Commission (“FEC”) rules?
- Did the reposting of the ad by Musk on X run afoul of X’s policies?



- What can/should be done about ads like this in the future? And, if the ad is a parody, what should be done about manipulated content in any event?

The experts I turned to after receiving an inquiry on behalf of the campaign quickly pointed out that this was a “parody,” rather than a “deepfake.” I was reminded that, when first researching the potential threat that deepfakes posed for elections and the views of experts (*see* my briefing note of [January 2024](#)), I learned that the more concerning deepfake ads/texts/robocalls would be designed to keep voters at home, to sow distrust in the electoral process and incite violence, rather than to change minds about a candidate. This was because these latter efforts would be so obvious. And, yes, the ad obviously was manipulated, but at what point would a “parody” be so real as to be treated differently.

And finally let us pause (putting aside the legal, regulatory and other technical questions) and consider the context here: the owner of one of the major social media platforms endorses a candidate for president and then promotes with near unimaginable reach content that has been digitally manipulated and is intended to harm that candidate’s opponent. As Scott Rosenberg writing in Axios (“[Deepfakes' parody loophole](#)”) asked, can those relying on the parody excuse continue to argue that it is “obvious” the manipulation was intended to amuse rather than deceive. Was the ad akin to Saturday Night Live impersonations or Sarah Cooper lip-synching to Trump recordings? And as [Rosenberg concluded](#), while there will always be space for public mockery, AI “is making it increasingly tough to establish the boundary between election-year humor and malicious fraud.”

FCC

After the January robocall incident in New Hampshire in which President Biden’s voice was used in an attempt to depress turnout in the Democratic primary (*see* my prior briefing note of [May 2023](#), and updates of [March 2024](#) and [May 2024](#)), the FCC banned the use of AI-generated voices in robocalls (so-called voice cloning). In May, FCC Chairwoman Rosenworcel [proposed](#) a Notice of Rulemaking, the first step in a process that would culminate in on-air and written disclosure when there is AI-generated content in political ads. The announcement was careful to reiterate that the proposal was intended only to require disclosure and would not prohibit or restrict the use of any AI-generated content. Existing laws ban outright deception in ads generally, regardless of the medium; these truth-in-advertising laws are enforced by the Federal Trade Commission. Also, the FCC has no jurisdiction over streaming services and social media.

In June, the Republican chairman of the FEC, Sean Cooksey, [wrote](#) to Chairwoman Rosenworcel asking that the FCC defer to the FEC on the ground that the FCC proposal described in the May announcement would “invade the FEC’s jurisdiction.”

Chairwoman Rosenworcel ignored the entreaty and, in late July, the FCC [issued](#) its Notice of Proposed Rulemaking. The Notice takes note of the fact that the FEC is also considering rulemaking and that a number of states have enacted, or are considering, legislation regulating AI-generated deepfakes. The proposed FCC action is intended to complement, not replace, these other efforts.



The proposed rule would:

- define “AI-generated content” as “an image, audio, or video that has been generated using computational technology or other machine-based system that depicts an individual’s appearance, speech, or conduct, or an event, circumstance, or situation, including, in particular, AI-generated voices that sound like human voices, and AI-generated actors that appear to be human actors”;
- require radio and television broadcasters that air political ads to inquire whether political ads scheduled to be aired on their stations contain any AI-generated content;
- require the broadcasters, in cases where political ads scheduled to be aired contain AI-generated content, to make an on-air announcement that the ads contain AI-generated content. The station would be required to make the on-air announcement immediately preceding or during the broadcast of any ad by or on behalf of a legally qualified candidate for public office and any issue ad that contains AI-generated content;
- require the broadcasters to use standardized language for the on-air disclosure; and
- require broadcasters to include in their online political files a notice disclosing the use of AI-generated content for each political ad that contains such content.

These requirements would also apply to cable broadcasters, direct broadcast satellite (DBS) providers, satellite digital audio radio service (SDARS) licensees engaged in origination programming and Section 325(c) permitholders (permitting broadcasts to foreign countries for rebroadcast to the United States).

The proposal is subject to a 30-day comment period and a 15-day reply period. It remains unclear if there is sufficient time for the requirements to be in place by Election Day.

FEC

Last August, the FEC [announced](#) it was seeking public comment on a proposal to regulate AI-generated deepfakes. The announcement was a [response](#) to a petition for rulemaking submitted by [Public Citizens](#) in July 2023 asking the FEC to amend its regulation on “fraudulent misrepresentation” (52 U.S.C. §30124) so that it applies to deceptive AI campaign ads. That provision prohibits candidates from fraudulently misrepresenting themselves as speaking or acting for or on behalf of another candidate or political party on a matter damaging to the other candidate or political party.

Also in July 2023, a group of House and Senate Democrats [urged](#) the FEC to move forward with rulemaking based on the Public Citizen proposal. The Democratic National Committee [supported](#) the rulemaking, while predictably the counsel on behalf of the Republican National Committee [opposed](#) the rulemaking, arguing that the FEC lacks statutory authority and that the proposed rulemaking would “function as content-based speech regulation” raising First Amendment issues.

In June 2023, the FEC deadlocked (3-3) on [a previous petition](#) from Public Citizens, with the rejection predicated on concerns over statutory authority and a failure to specify which



regulation was to be amended. In response, the July petition posited that the FEC did have the statutory authority and identified the regulation (52 U.S.C. §30124). (See [AP summary](#).)

Public Citizen has continued to highlight the failure of the FEC to proceed with the proposed rulemaking (see press releases of [February 16](#) and [June 27](#)). The FEC is evenly split between Republican and Democratic members. (See [AP Summary](#) of the partisan divide underpinning the FEC and FCC positions.)

State Legislation

According to the Public Citizen [tracker](#), as of July 29, 20 states had passed bipartisan legislation to address the use of deepfakes (generally referred to as “synthetic media”) intended to influence elections. These statutes vary considerably as to coverage, consequences and remedies. Generally, these prohibit manipulated media content as defined (which vary), with most, but not all, exempting content that bears a disclaimer that the content was manipulated (and in some cases that the content depicts speech or conduct that did not occur). Some apply the restrictions only within 60 or 90 days before an election.

Many exempt satire and parodies, though again the standards vary. The New Hampshire [legislation](#) is an interesting example of a ban on deepfakes for purposes of harassment, which has an exemption for paid ads provided the distributor does not remove or modify a disclaimer provided by the creator or sponsor, as well as for media “that constitutes satire or parody or the production of which is substantially dependent on the ability of one or more individuals to physically or verbally impersonate another person without reliance on artificial intelligence.” Florida’s new [statute](#) does not have a parody exception, but conditions the labelling requirement on creation of the generated content with intent to injure a candidate or deceive regarding a ballot initiative.

California’s currently effective [statute](#) (AB 730 enacted in 2019, and AB 972 amended in 2022) conditions the labelling requirement for audio and visual content on distribution with actual malice and intent to injure reputation or deceive a voter, and also has an exception for satire or parody. Effective January 1, 2027, new restrictions become effective in California, which among other things define “actual malice” (knowledge that the image of a person has been superimposed on a picture or photograph to create a false representation, or a reckless disregard of whether or not the image of a person has been superimposed on a picture or photograph to create a false representation) and has no parody exception, but seems to focus only on visual representations. There are also two proposed new statutory provisions ([AB 2839](#) and [AB 2655](#)), both of which exempt satire or parody.

California Governor Newsom clashed with Musk over the Harris ad, threatening to sign new legislation to address these types of ads. It is not clear which he was referring to if the existing proposals exempt satire or parody.

Federal Legislation

As the Brennan Center noted in December (“[Regulating AI Deepfakes and Synthetic Media in the Political Arena](#)”), a number of bills have been introduced:



- the [AI Disclosure Act of 2023](#) (H.R. 3831) would require any AI-created content to include the disclosure “Disclaimer: this output has been generated by artificial intelligence.”
- the [AI Labeling Act of 2023](#) (S. 2691) would require a similar disclosure and specifies that the disclaimer must be legible and difficult to remove.
- the [REAL Political Advertisements Act](#) (H.R. 3044/S. 1596) would require disclaimers on any political ad using content generated “in whole or in part” by AI.
- the [DEEPFAKES Accountability Act](#) (H.R. 5586) would require a disclaimer on a deepfake of any person, political figure or not.

Since December, the [Protect Elections from Deceptive AI Act](#) (S. 2770) was introduced in May. It would ban materially deceptive audio or visual media created by AI relating to candidates running for federal office. Companion [legislation](#) was introduced in the House shortly thereafter. Both exempt media that constitutes satire or parody.

It is unclear what legislation if any could pass the Senate and the House before Election Day.

Platform Policies

Meta. Over the past six months, Meta has made various announcements (*see* [February 6 statement](#) and [April 5 statement](#)) regarding changes to policies relating to manipulated media on Facebook, Instagram and Threads. These policies were most recently updated effective July 1 (see April 5 statement, as updated). Based on [feedback](#) from its Oversight Board, following Meta’s decision to leave up a video of President Biden that had been edited to make it appear as if he had inappropriately touched his adult granddaughter, and criticism that the existing policy was insufficiently narrow, Meta began adding the label “AI Info” (a change from “Made with AI”) to a wider range of video, audio and image content when it detects industry standard AI image indicators or when people disclose that they are uploading AI-generated content. An earlier Meta policy (dating back to 2020) only covered videos that were created or altered by AI to make a person appear to say something they did not say.

The upshot of the internal policy debates to date is that Meta would keep AI-generated content on its platforms (reversing an earlier removal policy) so it “can add informational labels and contexts” (including the “AI Info” label) unless the content otherwise violates its policies, for example, “policies against voter interference, bullying and harassment, violence and incitement, or any other policy in [its] Community Standards.”

Google. YouTube, owned by Google, adopted a [policy](#) in March 2024 requiring creators “to disclose to viewers when realistic content – content a viewer could easily mistake for a real person, place, scene, or event – is made with altered or synthetic media, including generative AI.”

Microsoft. Microsoft launched a tool that would allow campaigns to insert a digital watermark in their ads, which would allow viewers to know who created the ads and would ensure that that the ads could not be altered digitally without leaving a trail.

X. The day after the manipulated ad was reposted by Musk, Alex Howard, director of the Digital Democracy Project at the Demand Progress Education Fund, noted, in his [post](#) on X,



that Musk had seemingly violated X’s [policy](#) on synthetic and manipulated media (adopted in April 2023), which bans sharing “synthetic, manipulated, or out-of-context media that may deceive or confuse people and lead to harm (‘misleading media’). In addition, [X] may label posts containing misleading media to help people understand their authenticity and to provide additional context.”

Interestingly, the X policy states that memes or satire would not violate the policy, “provided these do not cause significant confusion about the authenticity of the media.” It is also interesting that the X policy bans a practice and then suggests it would resort to labeling that which has been banned. As the New York Times reported ([“Elon Musk Wants People on X to Police Election Posts. It’s Not Working Well.”](#)), the Community Notes [feature](#), which had been intended to allow a group of approved platform users to insert fact-checking labels and vote on whether the labels are helpful (and is the intended replacement for content moderation by platform staff), “has been far from consistent” in suppressing the spread of misinformation on X. Last week, X made it possible for any user to fact-check in Community Notes, merely by signing up (see [FAQ](#)).

Concluding Thoughts

So, what do we do about this?

First, are labels the answer? While the original manipulated ad carried a disclaimer that it was a “parody,” Musk’s reposting carried no such label and it was the Musk reposting that garnered the 134 million views. This is reminiscent of the issue presented when the Republican National Committee (“RNC”) posted an AI-generated dystopian video about President Biden following his April 2023 announcement he was running; the RNC included a disclaimer, but the disclaimer did not follow the YouTube version of the video, accessible incidentally not only directly on YouTube but also via the RNC’s own website.

Second, is it appropriate to draw a line between impermissible deepfake content and deepfake content intended as satire or parody? Kat Tenbarge of NBC News ([“Elon Musk made a Kamala Harris deepfake ad go viral, sparking a debate about parody and free speech”](#)) confirms that the manipulated campaign ad has generated significant debate over where that line should be drawn – should ads of this type be subject to regulation or are they satire/parody, subject to different treatment? And does it matter that the manipulated ad was reposted by the owner of X, garnering now close to 200 million views? Clearly, the robocall urging voters to remain at home during the New Hampshire primary was not a parody.

While the status of parody (and satire) as protected speech is well recognized (*see* First Amendment discussion below), in the context of manipulated content, a key element (is the parody obvious, is the parody not reasonably believable) may no longer be so straightforward. Returning to [Scott Rosenberg’s](#) question, has AI so blurred the boundary between humor and malicious attempts to influence voters?

There are a few other angles to this issue. First, experts have for some time been tracking the increasing use by malign actors of humor, parody and satire to normalize what once would have been considered fringe ideas, including hate speech. Second, partisan divides have the



potential to imperil meaningfully addressing the potential impact of manipulated media on election outcomes. Third, we must remain mindful of First Amendment concerns.

As far back as 2017, Julia Ebner and Jacob Davey, then research fellows at the Institute for Strategic Dialogue (“ISD”), were tracking the use of memes and parody accounts to spread disinformation and normalize far-right tropes and violence. Disinformation experts speak of the use by extremists of memes to disguise hateful or extreme messaging with humor and irony to shift the “[Overton Window](#),” which is a model for analysing how societal ideas shift over time and influence politics. (See “[The Fringe Insurgency](#)”; see also report prepared for the European Commission, [It’s not funny anymore. Far-right extremists’ use of humor](#)” and ISD’s Explainer “[Memes & the Extreme Right-Wing](#).”)

As for the partisan divides, interestingly the splits have a greater impact at the federal level than the state level. As noted above, a number of states have enacted legislation to address manipulated media in the context of elections, and these efforts, as well as proposals pending in most other states, largely are bipartisan.

Finally, there are First Amendment considerations, which may explain why most state statutes call for labelling rather than outright bans on manipulated content. First Amendment considerations also could explain why there are carve out for satire and parody, which are protected speech (see [Hustler v. Falwell](#) – First Amendment protects parodies of public figures, even if they are intended to cause distress, unless actual malice is established – here the parody was labelled as such and no reasonable person could have expected the parody to be true).

That said, the First Amendment landscape is quite complex. For example, while the First Amendment generally protects hate, inflammatory or derogatory speech, making false statements of fact that harm a person’s reputation (defamation) typically is not protected by the First Amendment, though there are higher standards if the false statement of fact concerns a public figure or public official (plaintiff must show negligence versus actual malice, see [New York Times v. Sullivan](#)). And note that, as private business (meaning not the government), social media platforms have their own First Amendment protections, and are free to remove content and users based on content provided they follow their terms of use.

I note as well that the FCC [Notice of Proposed Rulemaking](#) addresses First Amendment issues by addressing the various levels of scrutiny that courts would apply to any challenges to their proposed on-air disclosure and political file requirement and concludes that they would be deemed to comport with First Amendment principles, regardless of the level of scrutiny.

While we have just over 90 days to go before Election Day, we also have far more questions than answers when it comes to regulating deepfake political ads and content.

* * *

Mark S. Bergman
[7Pillars Global Insights, LLC](#)
Washington, D.C.
August 4, 2024