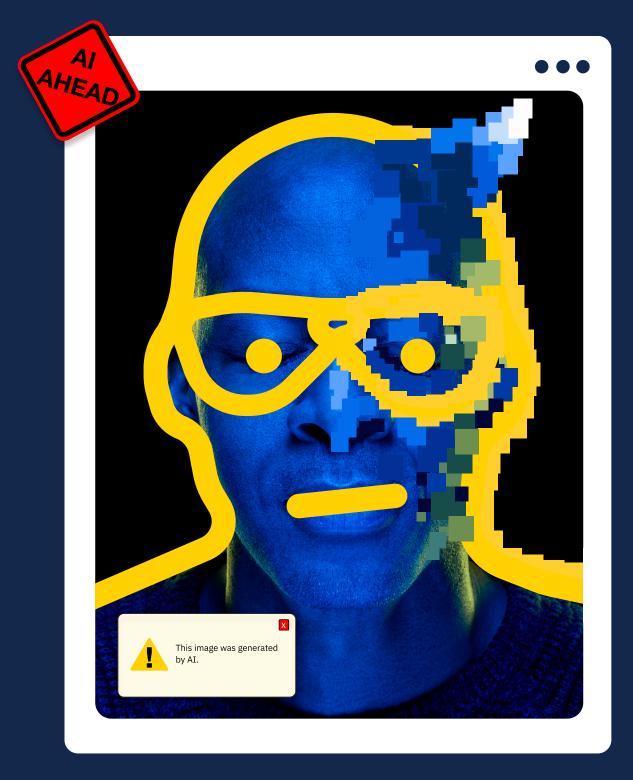
#### **In Disclaimers We Trust:**

# The Effectiveness of State-Required AI Disclaimers on Political Ads

Scott Babwah Brennen, Shelby Lake, Allison Lazard, and Amanda Reid





#### **About the Center on Technology Policy**

NYU's Center on Technology Policy seeks to craft public policy for a better internet. Utilizing an interdisciplinary academic framework, CTP works to identify knowledge gaps and develop actionable policy frameworks that will enable us to realize the potential benefits of technology while minimizing its harms. By working closely with students and expanding the University's offerings in technology policy analysis, we seek to cultivate and train the field's future practitioners.

#### **About the Author**

Scott Babwah Brennen, Ph.D., is the director of the Center on Technology Policy at NYU.

**Shelby Lake** is a PhD candidate in the Hussman School of Journalism and Media at the University of North Carolina at Chapel Hill.

**Allison J. Lazard**, Ph.D., is an associate professor in the Hussman School of Journalism and Media at the University of North Carolina.

Amanda Reid, J.D., Ph.D., is an associate professor at the UNC Hussman School of Journalism and Media.

We would like to thank Tara Queen for her assistance in cleaning and analyzing data and Mirian Avendano-Galdamez for her administrative support. We would also like to thank Atef Mohamed for his work designing this brief, and Le Ha for her work on the cover design.

### **Executive Summary**

The upcoming U.S. general election has been <u>widely described</u> as the first major U.S. "Al election." But it will also be the first major election with mandatory Al labels on political ads. <u>Twenty States</u> have enacted requirements that political ads contain disclaimers if they used generative Al. <u>Policymakers</u> have described these requirements as important tools to protect elections from disinformation while increasing trust in political communication.

But do they work? Do they increase trust or transparency? And what impact might disclaimers have on how viewers assess electoral candidates and ads?

Despite many states enacting new regulations, to our knowledge there have been no empirical studies addressing these questions.

In response, we designed and conducted an online experiment to provide insight into the potential impacts of AI labels on political ads. We tested two potential effects: first, the impact of two different AI labels that are now required by states. Second, we tested whether the impact of labels is influenced by the political alignment between a respondent and a candidate. We also asked respondents to respond to different label policy proposals.

Our experiment revealed six key findings:

- 1. Al labels hurt candidates who used generative Al. When ads contained Al disclaimers, respondents generally rated candidates less trustworthy and less appealing, candidates' ads less accurate, and indicated that they were less likely to like and more likely to flag or report the candidate's ads on social media. This pattern held across both deceptive and more harmless uses of generative Al. These results are broadly in line with the observed effects of funding disclosures on political ads and Al labels on news and social media content.
- 2. Al labels resulted in a "backfire effect" for the candidate who created the attack ad. While Al labels hurt the candidate that created attack ads, labels had no impact on assessments of candidates targeted in attack ads. Respondents ultimately expressed a lower opinion of the creator of the attack ad than of the candidate being attacked.
- 3. Al labels diminished respondents' assessments of candidates that created ads and that either belonged to their own political parties or had no political affiliation. In contrast, in most conditions labels had no impact on viewers' assessments of candidates from opposed political parties, mostly likely because their assessments of those candidates were already low.
- **4.** In most conditions, disclaimer effects were small, and a notable minority of respondents didn't notice disclaimers. As many as 37% of respondents did not remember seeing an Al label.
- 5. Al label wording matters, but it's hard to predict how people will interpret disclaimers. Unless a label specifically says generative Al was employed, respondents may assume other methods (e.g. video editing) were employed.

**6. Respondents were least supportive of the policy approach most commonly adopted by state governments.** When asked about their support for various policy options, respondents expressed the least support for states requiring disclaimers only on political ads that contain deceptive uses of generative AI. This policy option is the one that has been enacted by the greatest number of states.

Our initial results indicate that AI disclaimers on political ads may have some counterproductive effects, and may not reliably increase trust in political communication. For now, the costs of requiring disclaimers may outweigh the benefits. At the same time, our findings support the <a href="mailto:emerging consensus">emerging consensus</a> from studies on AI labels that the wording and design of labels matter. Policymakers may be well served in supporting additional research on the design and impact of AI disclaimers before rushing to require them on political ads.

### **Policy Background**

Over the last year, federal and state policymakers have been considering new laws regarding disclaimers on political ads that contain generative AI. At the state level, policymakers have enacted more than a dozen new laws requiring disclaimers on certain political ads.

#### **State**

As of October 1, 2024, 20 states have passed 29 bills that concern generative AI in election communication. States have passed 20 bills in 2024 alone.

These bills take a few different approaches. First, about half of the laws require disclaimers on deceptive uses of generative AI in ads. There is somewhat wide variation in these laws. Some, such as the one enacted in New Mexico, define deceptive as "depicting an individual" doing something they did not do without consent. Others, such as the bills in New Hampshire or Alabama, do not include a consent provision, but do require that "a reasonable viewer or listener" would believe the depicted person did what they are depicted doing. Finally, some laws, such as the one enacted in Alabama, limit the provision to material "produced by artificial intelligence." Others, like the laws enacted in Hawaii and New Hampshire, apply to any media that has "been created or intentionally manipulated with the use of generative adversarial network techniques or other digital technology."1

Six other laws are more expansive and require disclaimers on ads containing generative AI even if the content is not "deceptive." The law enacted in Florida, for example, requires disclaimers when the ad is created "to injure a candidate or to deceive" the voter. Laws in Indiana and Utah, require a disclaimer on any ad that is "intended to influence" an election, however neither specifies exactly what that entails. The laws enacted in Oregon and Washington require disclaimers on ads that show a "depiction...that didn't actually occur in reality." Finally, the law enacted in Wisconsin requires a disclaimer on any "content that is substantially produced in whole or in part by means of generative artificial intelligence."

Finally, four states, <u>Mississippi</u>, <u>Minnesota</u>, <u>Texas</u>, and <u>New Hampshire</u> enacted laws that ban any use of generative AI in political communication that is intended to deceive voters. The Minnesota legislature passed an <u>amendment</u> this summer to specify that the ban applies only when the perpetrator "acts with reckless disregard about whether the item" is fake.

#### **Federal**

Federal legislators recently debated at least two bills that would have imposed new restrictions or requirements on generative AI in political ads. Most notably, in May 2023 Representative Yvette Clarke (D-NY) introduced The REAL Political Ads Act. The bill would establish a broad labeling requirement for any political ad that "contains an image or video footage which was generated in whole or in part with the use of artificial intelligence (generative AI)." Another bill, introduced in September, 2023 in the Senate, would have banned any political ad that contains "materially deceptive Algenerated audio or visual media" with the intent to "(1) influence an election; or (2) solicit funds."

Two federal agencies have also considered new labeling rules on political ads. Following the <u>robocall</u> that included AI generated audio of President Biden suggesting supporters not vote in the upcoming New Hampshire primary, the Federal Communications Commission (FCC) <u>adopted a Declaratory Ruling</u> that AI generated voices in robocalls is illegal under existing law.

In July 2024, the FCC voted to move forward on a <u>rulemaking</u> that would require political advertisers to disclose if their ads contain generative AI. If enacted, this rule will not go into effect for the 2024 election. FCC rules apply only to TV and radio advertisements, not to digital or print advertisements.

Over the last year, the Federal Election Commission (FEC) has also debated imposing new labeling requirements on some political ads that contain generative Al content. After the advocacy group <a href="Public Citizen submitted a petition">Public Citizen submitted a petition</a> for rulemaking, the FEC opened a public comment on a proposed rule to

<sup>1</sup> Oddly, even though the New Hampshire law would require a label for false content even if it was not created by AI, the label must state: "This [image/video/audio] has been manipulated or generated by artificial intelligence technology and depicts speech or conduct that did not occur."

prohibit "deliberately deceptive AI-produced content in campaign communication." The FEC has not yet ruled on the petition, however, commissioners have <a href="mailto:expressed doubt">expressed doubt</a> that the FEC's existing mandate would permit such an action.

#### **Literature Review**

In the last section, we reviewed some of the new laws passed by U.S. states requiring disclaimers on certain political ads that contain generative AI.

To the best of our knowledge, there have been no studies of the effect or effectiveness of these newly enacted AI disclaimers. We simply do not know what sort of impact the labels in the states may have on how voters assess candidates or ads.

That being said, there are at least two useful literatures that provide insight into the likely effect of these labels: labels: (1) research on AI content labels on non-political content, and (2) research on political ad funding disclaimers. In a policy brief recently released by one of us at the Center on Technology Policy, we provide an extensive review of these literatures. Here, we review the insights most relevant to generative AI labels on political ads.

First, several experiments observed that both AI labels and political funding disclaimers, in one way or another, matter. That is to say, existing work shows that AI content labels "can meaningfully shape viewers' attitudes and behavior." In one notable example, Yin et al. found that, while some respondents felt more "heard" by AI chatbots than a human respondent when discussing "a complex situation that they were dealing with," when AI chatbots included a disclaimer label, that assessment flipped, and respondents reported feeling less heard by the chatbot.

That being said, the potential impact of labels is likely limited by respondents failing to notice disclaimers, or not not fully understanding them. One study that tracked eye motion while showing participants a social media feed containing political ads found that even when respondents spent additional time looking at funding disclaimers, it "did not affect the likelihood of participants remembering they had seen a political ad at all, and only in certain conditions did it help them recall the source of the ad correctly."

Existing research also suggests that how labels are worded and designed can impact not only if they are noticed but also their impact. A series of experiments on political and AI labels have shown that disclaimers with different designs, text, or prominence have different impacts on subjects. Interestingly, Dobber et al. observed that including a funding disclaimer at the beginning of an ad increased trust compared to including it at the end. This may be the result of participants viewing a funding disclaimer as a demonstration of a candidate's honesty. Wittenberg et al. found that AI labels on social media content that combined both information about the "process by which media were made" (i.e. "Al-generated") and the "harm" content poses (i.e. "manipulated") had the largest impact on the believability and credibility of images. Epstein et al. similarly found that there was a wide difference in how respondents understood the terms "AI generated," "AI manipulated," and "deepfake." Notably, respondents did not reliably conclude that content labeled "deepfake" or "manipulated" was Algenerated.

While design and word choice meaningfully moderate label impact, both political funding disclaimers and Al content labels can lower viewers' trust in the labeled content. A series of experiments have observed subjects consistently rating Al labeled content as less credible, trustworthy and accurate than content that is not labeled, regardless of if the news content was true or false. Similarly, several experiments have observed that when subjects view funding disclaimers on political ads, they tend to rate the candidate that sponsored the ad as less credible and less trustworthy. For the most part, the observed effect on trust has been "relatively small" both in absolute terms and in comparison to other interventions.

For AI content labels, some have hypothesized that the impact of labels on trust itself can be lessened by increasing transparency. While <u>Toff and Simon</u> observed a reduction in user trust for content labeled as AI generated, they noted that the "negative effects

associated with perceived trustworthiness [of Allabeled content] are largely counteracted when articles disclose the list of sources used to generate the content." Feng et al. also saw that including incomplete provenance information resulted in a larger reduction in user trust than more comprehensive disclosures.

Beyond the impact on trust, there is evidence that including a funding disclaimer on a political ad mutes or lowers the effect of that ad. For example, including a funding disclaimers on an <u>attack ad lessened</u> the impact that the ad otherwise had on viewers' assessment of candidates. For example, <u>Ridout et al.</u> found that while

attack ads created by unknown groups can be more effective than those created by candidates, including a funding or donor disclaimer "level[ed] the playing field" so that group and candidate-sponsored ads were equally effective. There is also evidence that including a funding disclaimers reduces the favorability boost of positive ads. This effect, however, is likely dependent on political congruence: the impact on favorability was less when members of one party saw ads for candidates from opposed parties that included funding disclaimers.

### **Our Experiment**

Observing that more than a dozen states now require disclaimers on political ads that contain some generative Al content yet there remains no research on the effectiveness of those labels, we designed and executed an experiment to test two different disclaimers now required by states.

Please see the <u>Methodological Appendix</u> for a full description of our methods and links to stimuli.

We completed a 3x3 between subjects online experiment replicated across two ads (N=1051) using a sample from the online service Prolific. We produced two 30-second ads for fictional candidates for real county commission races. In <u>previous work</u>, some of us suggested that generative AI may have greater impact on local and down ballot races, and we were curious how state-required labels would impact ads for candidates that voters did not already know.

The experiment tested three different label conditions: the label now required in Michigan, the label required in Florida, and a control with no label. Michigan requires that political ads that contain deceptive generative Al include the label: "This video has been manipulated by technical means and depicts speech or conduct that did not occur" to appear throughout a video ad. Florida requires that any political ad that contains generative Al display: "This video was created in whole or in part with the use of generative artificial intelligence," throughout the ad.

Because we hypothesized that partisanship would moderate the impact of labels, we tested three separate

conditions related to political affiliation. We created conditions where the subjects saw ads supporting the party to which they belonged (congruent), saw ads supporting the opposite party (incongruent), and where the ads had no clear party indicators (non-partisan). To create these conditions, we made variations of each ad where the sponsoring candidate stated their political affiliation. We combined this with knowledge of each respondent's political affiliation.

The literature on political ads <u>suggests</u> that voters often respond differently to attack ads than to more positive ads. We therefore replicated the experiment across two different ads, one attack ad and one positive ad.

After respondents saw each ad, we asked them to answer a series of questions regarding how they would respond to the ad if they saw it on social media. We also asked respondents to rate how trustworthy and appealing they found each candidate featured in the ad, as well as how likely they would be to vote for the candidate if they lived in that district. We also asked respondents how truthful and accurate they found the ad overall, and if they believed the ad was trying to manipulate them. Finally we asked several questions regarding their policy preferences regarding labeling on political ads that contain generative AI.

### **Findings**

#### 1. Al labels hurt candidates who used generative Al.

In nearly all conditions, when subjects saw a label on an ad, they rated the candidate who made the ad as less appealing and less trustworthy (see Table 1). We observed this both when respondents in the conditions without labels had higher opinions of the candidate--in the positive ad--as well as lower opinions of the candidate--in the attack ad.

In addition to lowering respondents' assessment of the candidate, labels lowered respondents' perception of the accuracy of ads as a whole. Again, this occurred for both the attack ad, which without labels subjects rated as less accurate, as well as for the laudatory ad, which without labels subjects rated as more accurate. This finding concurs with previous experiments that observed AI content labels reduced the perceived accuracy of news headlines and social media content.

Finally, we also asked subjects a series of questions about how they would likely behave in response to each ad if they saw it on social media. Again, broadly speaking, each label increased the chance that

respondents would flag or report the ad, and decreased the chance they would like it. Across these, the Michigan label had a stronger effect than the Florida label, but the trend held for both the attack and positive ads.

We do not know exactly how respondents interpreted each ad when they saw the labels (see below). However, the appearance of a label--especially the MI label--on the attack ad implied that the candidate had used generative AI to create fake audio of their opponent saying something they did not say. In contrast, the appearance of a label on the positive ad implied that the candidate had used generative AI to create or modify two images of the candidates' past. The fact that we observed labels associated with decreases in candidate and ad assessments suggests that the impact of labels may be somewhat independent of ad content.

### 2. Al labels resulted in a "backfire effect" for the candidate that created the attack ad.

While both labels decreased trust in and the appeal of the candidate who produced each ad, neither label had a significant impact on the trust and appeal of the candidate being attacked in the negative ad. Interestingly, without labels, the two candidates in the attack ad--the one doing the attacking, and the one being attacked--were rated as about equally trustworthy and appealing (see Table 2). But while the labels decreased trust and appeal for the candidate doing the attacking, they had no effect on the candidate being attacked.

Political funding disclaimers have been seen to mute the impact of attack ads, suggesting that labels--especially the MI label--might result in a better assessment of the candidate being attacked, as subjects might conclude that the attack was fake. This was not the case. Besides casting doubt on a comparable muting effect here, this finding also implies that labels did not appear to meaningfully correct the claims made in the attack ad.

Ultimately, the fact that labels lowered respondents' opinion of the candidate that sponsored the attack ad, but had no effect on the candidate featured in an attack ad, suggests that labeled attack ads created something of a "backfire" effect. As noted above, in the conditions without labels, respondents rated the candidate sponsoring and the candidate attacked in the ad roughly the same. However, in the labeled conditions, respondents rated the trustworthiness and appeal of the attacked candidate as higher than the producing candidate.

Table 1: Mean<sup>2</sup> evaluations of candidates and candidate ads.  $\star$  indicates result is significantly different from the no label condition at p<0.05. Please see our methodological appendix for additional description and data.

	Attack Ad			Positive (bio) ad		
	No label	Michigan- required label <sup>3</sup>	Florida- required label <sup>4</sup>	No label	Michigan- required label <sup>5</sup>	Florida- required label <sup>6</sup>
Trust in sponsor candidate	2.67	2.44*	2.51*	3.66	3.33*	3.48*
Appeal of sponsor candidate	2.61	2.43*	2.48	3.59	3.27*	3.38*
Trust in attacked candidate	2.58	2.51	2.58			
Appeal of attacked candidate	2.57	2.44	2.48			
Accuracy of Ad	2.89	2.48*	2.77	3.79	3.33*	3.62*
Report ad	1.40	1.93*	1.69*	1.29	1.51	1.35
Like ad	1.64	1.46	1.50	2.42	2.16	2.20

<sup>2</sup> Respondents were asked to evaluate each item using a 5 point likert-scale where, for example 1 = extremely not trustworthy; 2 = somewhat not trustworthy; 3 = neutral; 4 = somewhat trustworthy; 5 = extremely trustworthy.

 $<sup>{\</sup>bf 3}\ \, {\hbox{This video has been manipulated by technical means and depicts speech or conduct that did not occur.}$ 

<sup>4</sup> This video was created in whole or in part with the use of generative artificial intelligence.

 $<sup>5\ \, \</sup>hbox{This video has been manipulated by technical means and depicts speech or conduct that did not occur.}$ 

<sup>6</sup> This video was created in whole or in part with the use of generative artificial intelligence.

# 3. Al labels diminished respondents' assessments of candidates that created ads and that either belonged to their own political parties or had no political affiliation.

The degree of political and affective polarization in the United States and previous research on the impact of political funding disclaimers suggested that congruence between a subject and a candidate (do they belong to the same party, or different parties) would moderate the impact that labels have on how respondents assessed ads. We saw some important differences in how subjects assessed candidates when they shared a party with the candidate (congruent), when they belonged to different parties (incongruent), and when the candidate had no clear party affiliation (nonpartisan). But while political congruence had some impact on the effect of labels, it was limited, and there is indication that other factors may have had a stronger impact on how subjects assessed candidates.

We observed that the strongest and most consistent impact of labels was in nonpartisan conditions. In both ads, when subjects saw labels on ads for candidates with no clear political affiliation, they rated the candidates making the ads as less appealing, less trustworthy, and less likely to win their vote.

We also observed that labels were associated with subjects lowering their assessments of candidates from the same political party (the "congruent" condition). This effect appeared stronger in the positive ad, where there were reductions in trust, appeal, and likelihood of voting. In the attack ad, we observed only a reduction in trust. In the conditions without labels, subjects generally had higher opinions of congruent candidates in the positive ad than in the attack ad. These assessments,

therefore, had the farthest to fall. This is also inline with previous research that found AI labels on politically congruent ads reduced the favorability boost of positive ads, but had no impact on incongruent ads.

That Al disclaimers appear to damage users' assessments of members of their own party in positive ads may mean Al disclaimers have an outsized impact on campaigns. Political ads rarely have strong effects on voters. In particular they have little or even no reliable effect on persuasion. One meta review of 49 field studies concluded "the best estimate of the effects of campaign contact and advertising on Americans' candidate choices in general elections is zero."

However, ads can have more impact on mobilization, such as voter turn-out, fundraising, data donations, and volunteering. This means that ads are likely to be most effective when they are targeted to members of one's own political party.

In contrast, labels appeared to have no impact on how subjects assessed the candidates from opposing parties (the condition we call "incongruent"). Across conditions with and without labels, subjects rated incongruent politicians as less trustworthy, less appealing, and reported that they were less likely to vote for them. Those assessments were not significantly impacted in either direction by disclaimers. In the conditions without disclaimers, respondents already had very low assessments of candidates from opposing parties. It is possible that these assessments of incongruent politicians were so low that there wasn't enough positive sentiment to lower with labels.

Table 2: Moderation of candidate and ad evaluations by political congruence<sup>7</sup>. Numbers represent mean<sup>8</sup> evaluation. \* denotes significance at p < 0.05.<sup>9</sup>

		Attack Ad		Positive (bio) ad	
		No label	Michigan- required label <sup>10</sup>	No label	Michigan- required label <sup>11</sup>
Trust in source candidate	Congruent	3.00	2.72*	3.97	3.54*
	Incongruent	2.13	2.11	3.22	2.91*
	Non-partisan	2.88	2.41*	3.78	3.46*
Appeal of source candidate	Congruent	2.93	2.78	3.90	3.61*
	Incongruent	2.02	1.97	3.02	2.80
	Non-partisan	2.88	2.44*	3.84	3.30*
Trust in attacked candidate	Congruent	2.15	2.11		
	Incongruent	2.97	2.93	Ī	
	Non-partisan	2.68	2.59		
Appeal of attacked candidate	Congruent	2.06	1.96		
	Incongruent	3.02	2.96		
	Non-partisan	2.68	2.55		

<sup>7</sup> Respondents in congruent conditions saw ads that aligned with their political affiliation; those in incongruent conditions saw ads that did not. In the non-partisan conditions, there were no discernable indicators of party affiliations in ads.

<sup>8</sup> Respondents were asked to evaluate each item using a 5 point likert-scale where, for example 1 = extremely not trustworthy; 2 = somewhat not trustworthy; 3 = neutral; 4= somewhat trustworthy; 5 = extremely trustworthy.

<sup>9</sup> We did not analyze the moderation effect of congruence for the Florida-required label. The Michigan label had the strongest impact, and we wanted to understand an upper bound for potential moderation effect.

<sup>10</sup> This video has been manipulated by technical means and depicts speech or conduct that did not occur.

<sup>11</sup> This video has been manipulated by technical means and depicts speech or conduct that did not occur.

## 4. In most conditions, the effect of AI disclaimers was small, and a notable minority of respondents did not notice the disclaimers.

While AI disclaimers had a significant impact on how users assessed candidates and ads, that impact was generally small.

One potential explanation for the small impact was that a notable minority of users did not recall seeing disclaimers. Across conditions, both labels were the same size, font, and color, and were displayed throughout the duration of each ad. However, 26% of respondents shown the Florida disclaimer and 37% of respondents shown the Michigan disclaimer did not remember seeing the disclaimers.

It is notable that a significant percentage of respondents did not see the disclaimer even though the experiment likely overestimates the attention

that viewers will give to labels and to ads in the wild. In the experiment, we asked subjects to watch advertisements. In the wild, most viewers will likely give ads they encounter far less attention. In addition, while our ads displayed disclaimers throughout the entire run of the ad as required by the Michigan and Florida laws, other states require labels to appear only for part of an ad.

# 5. Al disclaimer wording matters, but it's hard to predict how viewers will interpret disclaimers.

We observed important differences in the effect of the two labels we tested: "This video has been manipulated by technical means and depicts speech or conduct that did not occur," and "This video was created in whole or in part with the use of generative artificial intelligence." Most importantly, across many outcomes, the Michigan disclaimer resulted in a larger difference from the unlabeled condition than the Florida label.

But we also observed differences in how users interpreted the labels. We asked respondents to identify the specific technologies that might have been used to produce the ads they saw. Respondents were able to select multiple choices. In the conditions without labels, the most common technology selected was video editing (selected by 43%), while only 6% of users chose generative AI.

In contrast, in the conditions with the Florida label, which specifically said the ad was created by generative AI, only 65% reported that generative AI might have been used.

In the conditions with the Michigan label, which asserts that the ad was created through "technological means," most people selected video or audio editing (59% and 41% respectively) and only 21% selected generative AI. This suggests that unless disclaimers specifically

identify that generative AI was used, respondents may first assume other, more traditional methods were employed.

At the same time, these findings imply that subjects may have limited ability to detect differences in editing or production techniques. They did not seem to identify that ads were produced through generative AI without a disclaimer.

# 6. Respondents were least supportive of the policy solution most commonly adopted by state governments.

In addition to asking subjects to assess candidates and ads, we also asked them to provide their agreement to three policy interventions regarding AI and political ads: requiring labels on all political ads with generative AI, requiring labels only on political ads with deceptive uses of generative AI, and banning all generative AI in political ads. Although most respondents likely had little context for the legal or policy issues around each possibility, the questions provide a look into how subjects are inclined to evaluate these specific policy options.

Subjects indicated that they were most supportive of disclaimers on all political ads that contain generative AI, and least supportive of requiring disclaimers only on ads that contain deceptive uses of the technology. This finding suggests that there was the least support for the solution that has been most commonly adopted

by state governments. The majority of state laws on generative AI and political ads impose disclaimer requirements only on deceptive generative AI, rather than all political ads that contain generative AI.

We also broke policy opinions down by political affiliation. Republicans and Democrats equally supported the two options requiring disclaimers. However, they disagreed on outright bans of generative AI in political ads, with Republicans, on average, being less supportive of bans, and Democrats being more supportive.

#### Conclusion

This report details results of one of the first analyses of the effect and effectiveness of Al disclaimers on political ads. Other work recently published by the Center on Technology Policy provides a reading of the literature on the potential effectiveness of Al content labels in general. That piece concludes that while labels can in some contexts be effective at achieving intended goals, their effect is usually small, heavily dependent on context and design, and may carry significant costs, such as decreasing trust even when content is not less trustworthy. Our findings here largely confirm those conclusions.

Our experimental results underscore that AI disclaimers on political ads not only are unlikely to increase trust in political communication but can also have other unintended consequences. Given this, for now, the costs of requiring AI disclaimers on political ads may outweigh the benefits. Policymakers should ensure we understand better what impact disclaimers have across political communication before enacting further laws requiring them.

At the same time, our findings support the emerging consensus that the design and wording of disclaimers can make a notable difference in their effects and effectiveness. When enacting new disclaimer requirements, policymakers should also provide rigorous, empirically based best practices in label wording and design.

Taken together, our findings also indicate that many respondents may still be crystallizing their views on generative AI in elections and may have unclear or conflicting views. For now, respondents appear somewhat skeptical of the technology, and many may associate the tools most with deceptive uses. While policymakers may intend for labels to be informational, many voters are treating them more as warning labels. Policymakers should ensure that any disclaimer requirement neither restricts campaigns' ability to experiment with generative AI, nor disproportionately harms campaigns exploring new ways to deploy these tools in elections.