GENERATIVE AI AND ELECTORAL COMMUNICATIONS

Evan Chiacchiaro[*]

TABLE OF CONTENTS

[*] Georgetown Law J.D. 2025; Georgetown University M.A. 2017; Tufts University B.A. 2011.

I. INTRODUCTION

There is growing consensus that generative AI (GAI) will affect upcoming elections worldwide.[1] There is less consensus, however, as to what the actual effect of GAI on elections will be. Predictions range from catastrophic "October surprise" deepfakes to less tangible, but equally concerning, warnings about a degraded information ecosystem, where discerning the truth is extremely difficult. Without a common perspective on these risks, there is no agreement on how to regulate the use of GAI by candidates and campaigns or to communicate important election-related information.

Although much has been written about deepfakes, disinformation, and the law,[2] most GAI use by campaigns will likely involve leveraging its capabilities to produce otherwise legitimate electoral communications[3] at scale. These scaled communications can take the form of a chatbot designed to answer questions about a candidate, mass microtargeting of voters with ads tailored to their personalities and able to respond to their feedback, or AI phone bank "volunteers." There are clear risks to the use of GAI for these communications, including the type of lower-probability but high-impact effects—which this paper will refer to as "tail risks"—some warn can threaten democracy itself, especially if malicious actors use GAI tools to mimic legitimate communications. But there are also benefits, such as helping under-resourced candidates level the playing field. As lawmakers and regulators contemplate rules to shape how GAI tools are built and deployed, they can attempt to craft measures that address the harms of GAI for electoral communications without eliminating the positives.

[1] *See, e.g.*, Galen Druke, *2024 is the 1st 'AI Election.' What does that mean?*, ABC NEWS (Dec. 1, 2023, 1:25 PM), https://abcnews.go.com/538/2024-1st-ai-election/story?id=105312571 [https://perma.cc/2MLJ-EYVY]; Jack Nicas & Lucía Cholakian Herrera, *Is Argentina the First A.I. Election?*, N.Y. TIMES (Nov. 15, 2023), https://www.nytimes.com/2023/11/15/world/americas/argentina-election-ai-milei-massa.html [https://perma.cc/R6RH-7MYU].

[2] *See e.g.*, Richard W. Painter, *Deepfake 2024: Will* Citizens United *and Artificial Intelligence Together Destroy Representative Democracy?*, 14 J. NAT'L SEC. L. & POL'Y 121 (2023).

[3] This paper uses the term "electoral communications" to denote communication by candidates or campaigns to voters with the intent of informing them about a candidate for election, persuading them to vote for or against a candidate, and/or informing them about the time, place, and manner of voting. It is intended to be broader than the FEC definition of "electioneering" but not so broad that it encompasses almost any speech that one could classify as "political."

This paper will survey and analyze the benefits and harms of GAI for electoral communications, along with current regulatory and legislative efforts to address its risks. The paper will proceed in three parts. Part II will describe the current state of GAI use for electoral communications and examine potential positive and negative effects of GAI for electoral communication in the near-term, medium-term, and long-term. Part III will look at current legislative and regulatory approaches in the United States and Europe—including self-regulation by AI companies—that either address GAI for electoral communications or can be used as a basis for crafting similar laws and regulations. Separating these approaches into those that regulate the user and those that regulate the system, it will evaluate how well they address the negative effects of GAI while preserving its benefits. Finally, Part IV will provide some recommendations for lawmakers and regulators who wish to enact measures that prevent or mitigate harm from GAI-enabled electoral communications without eliminating all benefits.[4]

II. GAI AND ELECTORAL COMMUNICATIONS: CURRENT STATE, POSITIVE EFFECTS, AND NEGATIVE EFFECTS

GAI and Large Language Models (LLMs) are reshaping nearly every industry, with the launch of ChatGPT by OpenAI in December 2022 cited as a "tipping point" for the nascent technology.[5] GAI refers broadly to a subset of AI capabilities that are able to generate content—images, text, audio, or video—in response to inputs from a user.[6] GAI tools are built on top of LLMs (also sometimes referred to as foundation models), which are "very large deep learning models that are pre-trained on vast amounts of data;"[7] these models can be deployed on their own with general-purpose chatbots

[4] Some lawmakers and regulators may decide that the risks of GAI-enabled electoral communications so far outweigh the benefits that sweeping bans on their use are appropriate. Conversely, some lawmakers and regulators may see the harms as speculative and assess that regulations are not necessary. Both of these approaches are outside the scope of this paper's recommendations.

[5] *See e.g.*, Ethan Mollick, *ChatGPT Is a Tipping Point for AI*, HARV. BUS. REV. (Dec. 14, 2022), https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai [https://perma.cc/ZY87-UCLB].

[6] *See Science and Tech Spotlight: Generative AI*, GOV'T ACCOUNTABILITY OFF., (June 2023), https://www.gao.gov/assets/gao-23-106782.pdf [https://perma.cc/49QR-P55L].

[7] *What is LLM (Large Language Models)?,* AMAZON WEB SERVICES, https://aws.amazon.com/what-is/large-language-model/ [https://perma.cc/JA8W-MKM4] (last visited on March 25, 2024).

such as ChatGPT or similar tools, or they can be further refined and trained—a process known as fine-tuning[8]—to handle specific tasks and use cases. Understanding the different ways GAI tools can be built is critical to understanding options for regulating them.

This Part will proceed in two sections. First, it will look at current GAI tools and use cases of GAI being deployed for electoral communications. Next, it will examine the potential positive and negative effects of GAI for electoral communications, identifying benefits and harms that can occur in the near-term, medium-term, and long-term.

A. CURRENT USES: PUBLICLY AVAILABLE TOOLS, PURPOSE-BUILT TOOLS, AND THIRD-PARTY TOOLS

Political operatives, practitioners, and candidates have three primary choices when deciding to use GAI to communicate with prospective voters. First, they can use existing open-source GAI tools such as OpenAI's ChatGPT or Dall-E to create content such as stump speeches or advertising images. For example, a California congressional candidate used AI to "depict his life story in a campaign launch video," and Sergio Massa's presidential campaign in Argentina used an LLM from the company Stability AI to produce campaign posters.[9] Using an existing open-source GAI tool does not require any software development or "fine-tuning" of any LLMs.

Political actors can also build fit-for-purpose tools on top of open-source LLMs. For example, a Super Political Action Committee (PAC) supporting 2024 Democratic presidential primary candidate Dean Phillips built a tool on top of ChatGPT that enabled individuals to have a conversation with an audio chatbot mimicking Phillips's voice.[10] This violated OpenAI's ban on political campaigning tools[11] and was quickly taken down.[12] These tools leverage existing LLMs and do not require any fine-tuning.

[8] *See* Paul Ohm, *Focusing on Fine-Tuning: Understanding the Four Pathways for Shaping Generative AI*, 25 COLUM. SCI. & TECH. L. REV. 214, 223 (2024).

[9] Mohar Chatterjee & Madison Fernandez, *'An Arms Race Forever' as AI Outpaces Election Law*, POLITICO (Feb. 07, 2024, 10:00 AM), https://www.politico.com/news/2024/02/07/ai-2024-elections-00140005 [https://perma.cc/S2S6-ZFB2].

[10] Meryl Kornfield & Elizabeth Dwoskin, *Silicon Valley Insiders Are Trying to Unseat Biden with Help from AI*, WASH. POST (Jan. 18, 2024, 05:46 PM), https://www.washingtonpost.com/elections/2024/01/18/ai-tech-biden/ [https://perma.cc/9BXB-VCL3].

[11] *See infra* Part III.A and note 57 .

[12] *See* Chatterjee & Fernandez, *supra* note 9.

Finally, rather than building their own tools, political actors can use an AI tool built by a third-party on top of an LLM and designed specifically for campaigns. Higher Ground Labs is a "startup accelerator and venture fund" for technology companies aiming to assist progressive political and advocacy campaigns.[13] Their funding supports AI-enabled political tools in five categories: content creation, data and analytics, media and messaging, research and polling, and voter contact and organizing.[14] Of those five categories, content creation and voter contact and organizing are powered by GAI capabilities rather than "traditional AI," expanding their uses beyond analytics and automation to save campaigns time and effort on traditionally creative or time-intensive tasks.[15] The content creation tools can "draft campaign messages and create engaging social media posts, freeing up precious time to focus efforts in other areas."[16] The Daisychain voter contact and organizing platform, meanwhile, is a tool that "parses incoming text messages from voters, and suggests the correct response—saving staff and volunteers countless hours that can be used on other campaign tasks."[17] These tools may use open-source LLMs or leverage fine-tuned LLMs specifically designed for electoral communications.

B. POTENTIAL POSITIVE AND NEGATIVE EFFECTS

Most uses of GAI for electoral communications will likely be relatively innocuous and value-neutral; a speechwriter with writer's block might use ChatGPT for a first draft or set of potential themes, or a digital communications specialist might turn to Gemini for a list of synonyms for "Donate now!" Candidates, campaigns, and civil society organizations that thoughtfully leverage GAI-enabled tools, however, can create pro-democratic effects, including increased voter outreach capabilities and voter knowledge in the near-term and improved political discourse in the long-term. Conversely, reckless or nefarious actors can misuse GAI tools, hindering voter access to accurate information in the near-term, fueling political polarization

[13] *Mission: We Bridge Tech and Politics*, HIGHER GROUND LABS, https://highergroundlabs.com/mission/ [https://perma.cc/HV2P-7Y4D] (last visited March 23, 2024).

[14] Ali Talib & Leah Bae, *Spotlighting Challenges: The Role of Generative AI in the 24 Election & Beyond*, HIGHER GROUND LABS (Jan. 31, 2024), https://highergroundlabs.com/spotlighting-challenges-the-role-of-generative-ai-in-the-24-election-beyond/ [https://perma.cc/2ALC-WYAC].

[15] *Id.*

[16] *Id.*

[17] *Id.*

through microtargeting in the medium-term, and threatening the information ecosystem necessary for democratic governance in the long-term. Understanding both the benefits and the risks—including the types of future tail risks some warn could have catastrophic effects on democracy—is necessary to attempt to craft legislation or regulation that properly balances both interests.

1. *Positive Use Cases for GAI*

There are several potential near-term and long-term benefits of leveraging GAI for electoral communications, both for campaigns and for democracy itself. Candidates and campaigns can use GAI tools to aid in voter outreach, voters can use these tools to increase knowledge and awareness, and some research suggests GAI tools can actually improve democratic discourse.

a) Aiding Voter Outreach: Near-term

GAI-enabled tools that assist with voter outreach can help campaigns reach more voters with less manual work. This is likely a welcome development for campaign staffers dealing with "the immense workloads, fatigue, and tight deadlines" traditionally found in field offices and campaign headquarters.[18] But beyond improving the working experience for organizers and operatives, these tools could make a difference for underfunded and understaffed candidates. GAI's ability to create different content for different audiences could allow a small campaign to act like a more well-funded incumbent or establishment candidate, potentially centering a race around ideas and policy rather than sheer volume of communications.[19]

[18] *Id.*

[19] *See* ETHAN BUENO DE MESQUITA, BRANDICE CANES-WRONE, ANDREW B. HALL, KRISTIAN LUM, GREGORY J. MARTIN, & YAMIL RICARDO VELEZ, PREPARING FOR GENERATIVE AI IN THE 2024 ELECTION: RECOMMENDATIONS AND BEST PRACTICES BASED ON ACADEMIC RESEARCH (November 2023) ("Generative AI could also help level the playing field with respect to political campaigns. Under-resourced campaigns may face challenges in creating content that appeals to voters. Generative image and text tools can enable these campaigns to draft more compelling speeches, press releases, social media posts, and other materials. These methods can also be used to create tailored materials for different audiences. To the extent that generative AI allows financially constrained campaigns to maintain a veneer of professionalism, it could reduce imbalances between lesser-known candidates and more established politicians.").

Chatbots or other tools enabled by LLMs could especially help candidates trying to connect with non-native English speakers, particularly in diverse districts with several spoken languages. For example, the company Civox built a GAI-powered robocalling tool for Shamaine Daniels, a candidate in the 2024 Democratic primary for Congress in Pennsylvania's 10th Congressional District.[20] "Ashley" could converse with voters in over twenty languages, boosting Daniels' ability to connect with voters in her district.[21] Local races—which often struggle for resources—can also particularly benefit from these tools, which can be used to draw voter attention to elections many voters are not aware of.

Leveraging these capabilities is not without risks. At a micro-level, campaigns using GAI-enabled tools will need to monitor their outputs to ensure the tools are not hallucinating and providing voters with incorrect information.[22] At a macro-level, the increase in communications enabled by GAI-enabled tools could overwhelm voters and degrade the overall information environment.[23] Campaigns and politicians planning to use these tools, as well as policymakers trying to regulate their use, will need to balance the potential benefits of GAI-enabled voter outreach against the potential harmful effects.

b) Increasing Voter Knowledge and Awareness: Near-term

Democracy takes work. Voters must sort through huge amounts of information about just the presidential election every four years, not to mention down-ballot races and off-cycle elections. Sorting

[20] *See* Anna Tong & Helen Coster, *Meet Ashley, the World's First AI-powered Political Campaign Caller*, REUTERS (December 15, 2023, 10:00 PM), https://www.reuters.com/technology/meet-ashley-worlds-first-ai-powered-political-campaign-caller-2023-12-12/. The Federal Communications Commission in February 2024 issued a declaratory ruling barring the use of AI-enabled robocalls that employ an artificial voice unless the individual explicitly consents to such robocalls. In the Matter of Implications of Artificial Intelligence Technologies on Protecting Consumers from Unwanted Robocalls and Robotexts. Fed. Commc'ns Comm'n, Declaratory Ruling in the Matter of Implications of Artificial Intelligence Technologies on Protecting Consumers from Unwanted Robocalls and Robotexts (Feb. 8, 2024), https://docs.fcc.gov/public/attachments/FCC-24-17A1.pdf [https://perma.cc/B2PF-8KG6].

[21] Tong & Coster, *supra* note 20 . Although the FCC decision limiting the use of AI robocalls limits the effectiveness of this approach, these language capabilities could still be leveraged through a text-based chatbot.

[22] *See supra* Section II.B.2.a.

[23] *See supra* Section II.B.2.c.

through, analyzing, and presenting information culled from large amounts of data is exactly what GAI tools are best equipped to do. As Ethan Bueno de Mesquita argues, "training chatbots to understand political issues and the positions of parties and candidates may be an effective way to help Americans become better informed about politics," especially if "conversing through chatbots proves to be a more natural and engaging way for people to consume political information."[24] Chatbots—text-based or audio-based—could be deployed by campaigns, advocacy organizations, or neutral voter awareness organizations with the express goal of conversing with citizens at scale and informing them of upcoming elections, major issues at stake, and the relevant positions of the different candidates. Again, organizations leveraging GAI-enabled tools for this purpose will need to monitor their use to ensure they do not unintentionally spread incorrect information.

c) Improving Democratic Discourse: Longer-term

Rather than degrading the political information environment (*see infra* Section II.B.2.b), some research suggests political conversations enhanced by GAI can actually *improve* the tenor of discourse and "enhance commitment to democratic reciprocity."[25] A 2023 study indicated that using AI-enabled chatbots to suggest potential responses during a conversation on gun policy in the United States led to "higher levels of a willingness to understand and allow the expression of opposing viewpoints in the political system in general, and not just in the context of their single conversation partner."[26] A recent working paper also indicates that GAI tools can be used to meaningfully reduce belief in conspiracy theories.[27] Expanding the use of AI-enabled tools to facilitate better conversations between candidates, campaigns, civil society, and voters with opposing views could therefore foster a healthier political environment. Similarly, Danielle Allen and E. Glen Weyl argue that the best response to a highly atomized information environment

[24] BUENO DE MESQUITA, ET AL., *supra* note 19.

[25] Lisa P. Argyle, Christopher A. Bail, Ethan C. Busby, Joshua R. Gubler, Thomas Howe, Christopher Rytting, Taylor Sorensen, & David Wingate, *Leveraging AI for Democratic Discourse: Chat Interventions Can Improve Online Political Conversations at Scale*, 141 PROC. OF THE NAT'L ACAD. OF SCIS., Oct. 2023, at 1, 6, https://doi.org/10.1073/pnas.2311627120.

[26] *Id.* at 3, 6. Crucially, participants maintained the ability to choose whether they used the AI's suggestions.

[27] Thomas H. Costello, Gordon Pennycook, & David G. Rand, *Durably Reducing Conspiracy Beliefs Through Dialogues with AI*, 385 SCIENCE, Sept. 2024, at 6, https://doi.org/10.1126/science.adq1814.

fueled by GAI and microtargeting is not to attempt—probably unsuccessfully—to ban the technology, but instead to "harness" GAI and turn it toward uses to "enrich the foundations of democratic institutions."[28] If Allen, Weyl, and other proponents of leveraging GAI to improve democracy are correct, efforts to outright ban it are counterproductive.

While these longer-term benefits are speculative, the near-term benefits of GAI represent a clear improvement over current electoral communication capabilities. Legislative and regulatory efforts can acknowledge these benefits and try to create policies that do not foreclose or disincentivize companies from providing their tools for these positive use cases, while still mitigating the significant and very real risks of unregulated GAI for electoral communications.

2. *Negative Effects Of GAI On Electoral Communications*

GAI's ability to generate content at scale threatens to disrupt the information ecosystem necessary for an effective democracy. Incorrect information—whether intentional or unintentional—can change election results by preventing voters from ever reaching the polls. Additionally, microtargeting voters with content designed just for them and only viewed by them can swing close local races in the near-term, increase political polarization in the medium-term, and turn elections into a battle of AI rather than a battle of ideas in the long-term. Finally, voter awareness of the volume and nature of targeted communications risks sowing distrust and creating a 'liar's dividend.'[29] Each of these risks threatens the legitimacy of democratic elections.

a) Incorrect Information and AI Hallucinations: Near-term

A major risk of using GAI for electoral communications is how frequently it "hallucinates" incorrect information.[30] Incorrect

[28] Danielle Allen & E. Glen Weyl, *The Real Dangers of Generative AI*, 35 J. OF DEMOCRACY, Jan. 2024, at 147, 154, https://doi.org/10.1353/jod.2024.a915355.

[29] *See* discussion *infra* Section II.B.2.c.

[30] *See* Cade Metz, *Chatbots May 'Hallucinate' More Often Than Many Realize*, N.Y. TIMES (Nov. 6, 2023), https://www.nytimes.com/2023/11/06/technology/chatbots-hallucination-rates.html [https://perma.cc/WBQ3-AGDP] ("Now a new start-up called Vectara, founded by former Google employees, is trying to figure out how often chatbots veer from the truth. The company's research estimates that even in situations designed to prevent it from happening, chatbots invent information at least 3 percent of the time — and as high as 27 percent."). These hallucinations occur when the underlying model learns incorrect

information is ubiquitous, pervasive, and unavoidable during an election campaign, but it can be particularly dangerous if it comes from a source voters view as authoritative and if it affects a citizen's ability to vote: for example, by providing incorrect information on the time, location, and manner of voting.

Unfortunately, publicly available GAI-enabled chatbots are not proving to be up to the task. In a test ahead of "Super Tuesday" in February 2024, five leading LLMs provided inaccurate responses to queries over half the time in aggregate, including responses that provided incorrect information about same-day voter registration and the location of voting precincts.[31] Voters turning to these chatbots for critical election information could end up being disenfranchised, and the results of smaller races could even hinge on the change in turnout.

b) Microtargeting: Near-term, Medium-term, and Longer-term

The risks to democracy of GAI-enabled microtargeting are relatively small in the near-term, but they increase in the medium-term and could lead to catastrophic tail risks in the long-term. In the near-term, microtargeting voters with individually created ads at scale presents only a small risk of shifting an election's results in the shadows. Research suggests that microtargeted, personalized campaign ads are unlikely to persuade many voters.[32] Campaign

patterns from its source data because the data itself is either incorrect or in some way skewed. *See also What Are AI Hallucinations?*, GOOGLE, https://cloud.google.com/discover/what-are-ai-hallucinations [https://perma.cc/ZWY2-8LVC] (last visited on May 6, 2024).

[31] *See* Garance Burke, *Chatbots' Inaccurate, Misleading Responses about US Elections Threaten to Keep Voters from Polls*, ASSOCIATED PRESS (February 27, 2024, 05:06 PM), https://apnews.com/article/ai-chatbots-elections-artificial-intelligence-chatgpt-falsehoods-cc50dd0f3f4e7cc322c7235220fc4c69 [https://perma.cc/ZVF8-VWRH].

[32] *See* ANTHONY NADLER, MATTHEW CRAIN, & JOAN DONOVAN, WEAPONIZING THE DIGITAL INFLUENCE MACHINE: THE POLITICAL PERILS OF ONLINE AD TECH, DATA & SOC'Y 16 (2018), https://datasociety.net/library/weaponizing-the-digital-influence-machine/ [https://perma.cc/R9GX-UB7J]. *But see* Almog Simchon, Matthew Edwards, & Stephan Lewandowsky, *The Persuasive Effects of Political Microtargeting in the Age of Generative Artificial Intelligence*, 3 PROC. OF THE NAT'L ACAD. OF SCIS. NEXUS, Jan. 2024, at 3, https://doi.org/10.1093/pnasnexus/pgae035 ("[p]olitical microtargeting is an effective technique and can be automated using off-the-shelf generative AI."). Simchon *et al.* examine four studies and estimate that between 2.49-11.405% of individuals "exposed to political messages tailored to their personality" would be persuaded by the messages: a relatively small number, but one that could be decisive in many close elections. They argue

advertising in general has a very low persuasion rate; therefore, GAI-enabled campaign advertising would need to be "substantially more effective than other existing campaign advertising technology" for it to possibly have a major persuasive effect,[33] meaning that "weaponized political ad targeting will rarely, if ever, be effective in changing individuals' deeply-held beliefs."[34]

More finely grained segmentation that targets smaller groups of voters with specific messages is also unlikely to help. While targeting voters based on a single attribute, such as political affiliation, is up to 70% more effective than showing the same ad to all audiences, attempts to further segment audiences through combinations of attributes did not increase their persuasive effect in a 2023 study.[35] Additionally, while GAI can solve the problem of creating misinformation at scale, it does not solve the problem of dissemination.[36] Altogether, the barriers to a successful GAI-driven microtargeting campaign remain high.

One exception may be local elections with smaller electorates and less media attention and campaign spending.[37] A local campaign

that therefore a candidate who could "deliver politically targeted content at scale to very large populations," including "untruthful or manipulative content" that could "exploit[ing] individual vulnerabilities," would be well-positioned to tip the scales in their favor in any close election. *Id.*

[33] BUENO DE MESQUITA ET AL., *supra* note 19; *see also* Felix M. Simon, Sacha Altay, & Hugo Mercier, *Misinformation Reloaded? Fears About the Impact of Generative AI on Misinformation Are Overblown*, HARV. KENNEDY SCH. MISINFORMATION REV., Oct. 2023, at 1, 5, https://doi.org/10.37016/mr-2020-127 (the evidence suggests that micro-targeting by, for example, political actors has mostly limited persuasive effects on the majority of recipients . . . In general, the effects of political advertising are small and will likely remain so, regardless of whether they are (micro)targeted or not, because persuasion is difficult.").

[34] NADLER, CRAIN, & DONOVAN, *supra* note 32, at 2.

[35] *See* Peter Dizikes, *Study: Microtargeting Works, Just Not the Way People Think*, MIT NEWS (June 21, 2023), https://news.mit.edu/2023/study-microtargeting-politics-tailored-ads-0621 [https://perma.cc/56C4-AYWN]; *see also* Simon, Altay, & Mercier, *supra* note 33, at 5 ("the evidence on the effectiveness of political advertising personalized to target, for instance, people with different personalities is mixed, with at best small and context-dependent effects").

[36] Simon, Altay, & Mercier, *supra* note 33, at 5 ("[T]he cost of reaching people with misinformation, rather than the cost of creating it, remains a bottleneck").

[37] Matt Perault & J. Scott Babwah Brennen, *A Policy Framework to Govern the Use of Generative AI in Political Ads*, BROOKINGS (December 11, 2023), https://www.brookings.edu/articles/a-policy-framework-to-govern-the-use-of-generative-ai-in-political-ads/ [https://perma.cc/94LC-RWKE] ("smaller, down-ballot races may be more susceptible to the impact

that successfully targets voters with individually-tailored messages could sway enough voters with misleading or dubious advertisements divorced from actual facts without the ability for meaningful counterspeech from their opponent. These are the same races, however, that may benefit most from GAI tools to mitigate the disparities caused by a typical lack of funding, resources, and volunteers, creating a tension between these benefits and the opportunity for abuse.[38]

GAI-fueled microtargeting may not guarantee electoral victory, but it can contribute to one proven risk to democratic governance in the medium-term: political polarization. Research suggests that microtargeted content delivered via the "Digital Influence Machine" contributes to political polarization for multiple reasons.[39] First, the targeted content most effective for influencing voter behavior also increases polarization through appeals to voter identity—particularly when focusing on threats to that identity.[40] Second, politicians no longer need to champion policies that appeal to a wide swath of voters, instead targeting "policies narrowly to a subset of voters," which "induces an increase in polarization if both moderate and extreme voters can be targeted more precisely."[41] As moderate voters disappear, the room for democratic negotiation and compromise shrinks. This can paralyze institutions and turn every policy debate into a zero-sum issue with no zone of possible agreement. GAI-enabled microtargeting can rapidly increase this process by generating content at scale that targets more microsegments and is responsive to current events and user feedback.

Looking ahead to future tail risks, microtargeting poses the risk of undermining democratic processes by changing elections from a

of political ads, since there's often much less advertising in these races, voters are less familiar with the candidates, and there's less oversight and media attention").

[38] *See supra* Section II.B.1 for further discussion of positive uses in state and local elections.

[39] Nadler, Crain, and Donavan define the term "Digital Influence Machine" as "an infrastructure of data collection and targeting capacities" that "incorporates a set of overlapping technologies for surveillance, targeting, testing, and automated decision-making designed to make advertising – from the commercial to the political – more powerful and efficient." NADLER, CRAIN, & DONOVAN, *supra* note 32, at 4–5. Campaigns can leverage increased data collection via the Internet and social media along with advanced data analytic capabilities to create "voter profiles with commercial data" and then target those voters with specific content based on their known or predicted attributes. *Id.* at 18.

[40] *See id.* at 29–31.

[41] Anja Prummer, 188 *Micro-targeting and Polarization*, J. PUB. ECON., Aug. 2020, at 12–13.

contest between candidates and parties to a contest between AI tools. Harvard professors Archon Fung and Lawrence Lessig describe what they claim is a believable scenario centered around a GAI tool they call 'Clogger.'[42] In their hypothetical, 'Clogger' leverages GAI, automation, and advances in microtargeting to create messages that are tailored to specific individuals, which it can do at scale to "generate countless unique messages for you personally – and millions for others – over the course of a campaign."[43] 'Clogger' would continuously adapt these messages based on what it learns from an individual's responses as well as voter persuasion tactics learned from millions of conversations, allowing it to become highly effective at changing minds or behavior.[44] Once 'Clogger' is created, elections would be decided not by who was the most popular with voters—the essence of democracy—but by whichever campaign had the most effective 'Clogger' machine in its toolkit.

c) A Degraded Information Environment: Longer-term

GAI's ability to deliver electoral communications at scale challenges the quality of information voters receive and voters' trust that what they are seeing is real. One concerning negative effect of microtargeting is that by creating specifically-targeted ads that are only seen by a narrow audience, [45] microtargeting creates what can be called "information microuniverses." Neighbors with different "information microuniverses" could have entirely different perceptions of basic facts about the government and candidates for

[42] Archon Fung & Lawrence Lessig, *How AI Could Take Over Elections – and Undermine Democracy*, CONVERSATION (June 2, 2023, 08:42 AM), https://theconversation.com/how-ai-could-take-over-elections-and-undermine-democracy-206051/ [https://perma.cc/UJ3S-MAYL].

[43] *Id.*

[44] *Id.* Fung and Lessig also raise alarms that if 'Clogger' is trained only to change minds or affect voter turnout, it might not even use political content to do so, having identified "strategies for achieving this goal that no human campaigner would have thought of." *Id.*

[45] *See* Dawn Carla Nunziato, *Misinformation Mayhem: Social Media Platforms' Efforts to Combat Medical and Political Misinformation*, 19 FIRST AMEND. L. REV. 32, 59–60 (2020), https://scholarship.law.unc.edu/cgi/viewcontent.cgi?article=1288&context=falr [https://perma.cc/Y94C-CUFM] ("Unlike political advertising on mass media like broadcast television or radio—in which large national or regional audiences are exposed to the same political advertisement—by employing narrowly cast microtargeted ads on social media, a political advertiser can craft a specific ad to a much narrower intended audience, and to *only* that specific audience, thereby preventing others from accessing and scrutinizing the content of the ad.").

election because of microtargeted electoral communications designed to appeal to their specific needs or concerns. And because microtargeting takes place on personalized platforms such as Facebook or X, individuals often have no ability to see what their neighbor is seeing or to be confronted with facts or alternative opinions.[46] This lack of shared common truth undermines democracy, making it impossible for real debate to occur given the necessity of shared facts to a true debate over policy.[47]

There can also be negative effects when one discovers that their neighbor's "information microuniverse" is different from their own. The term "flooding" is typically associated with deliberate efforts, usually by foreign adversaries, to overwhelm a populace with so much information that it is exceedingly difficult to deduce basic facts about political groups, such as "the goals of those groups, and the level and kind of support that they enjoy."[48] Disinformation actors already deliberately deploy social media bots for "false amplification"—using fake accounts to exploit platform algorithms and make a topic trend[49]—as well as "manufacturing consensus"—amplifying an obscure or otherwise unpopular figure or viewpoint.[50] But in a world where GAI is producing millions of conflicting tailored messages, this flooding can happen by accident, as individuals who learn of the sheer number of conflicting communications are unable to determine which sources they can trust.

Finally, bias present in the LLMs themselves poses the risk of degrading the information environment. LLMs often reflect the biases in the underlying training data.[51] Attempts by developers to "fix" these biases are not necessarily successful and instead may cause LLMs to shift from obvious "overt" biases to more subtle "covert" ones.[52] As the public grows more aware that LLMs and their

[46] *See id.*

[47] *See* HENRY FARRELL & BRUCE SCHNEIER, COMMON-KNOWLEDGE ATTACKS ON DEMOCRACY 3 (2018) ("if the decentralized system of democracy is not to break down into chaos, then citizens and their representatives have to roughly agree about what they disagree about."), https://doi.org/10.2139/ssrn.3273111.

[48] *Id.* at 15.

[49] Madeline Lamo & Ryan Calo, *Regulating Bot Speech*, 66 UCLA L. Rev. 988, 1000 (2019).

[50] *Id.* at 1000.

[51] *See e.g.*, Hadas Kotek, Rikker Dockum & David Q. Sun, *Gender Bias and Stereotypes in Large Language Models*, *in* PROC. OF THE ACM COLLECTIVE INTEL. CONF. 12 (Michael Berstein, Saiph Savage & Alessandro Bozzon eds., 2023), https://doi.org/10.1145/3582269.3615599.

[52] *See* Elizabeth Gibney, *Chatbot AI Makes Racist Judgements on the Basis of Dialect*, 627 NATURE 476 (2024), https://doi.org/10.1038/d41586-

related chatbots are not neutral sources of information, bad faith actors could exploit this issue to cast doubt on GAI tools designed to provide authenticated election information—even when those tools have been carefully vetted by trusted agents.

This degraded information environment has three effects. First, citizens can become disillusioned with the entire democratic process and choose to disengage. Second, domestic or international actors who *are* deliberately trying to sow chaos through misinformation or disinformation can operate in an environment with fewer trusted sources, making it more likely that their efforts are successful or at minimum cause doubt and confusion. Finally, politicians caught doing something unethical or illegal can reap what is known as the "liar's dividend": deny it happened and blame AI.[53] For example, in 2023 a politician in India claimed that politically-damaging audio recordings of him were computer-generated, but deepfake researchers verified at least one of the clips as genuine.[54] Each of these effects can harm democracy by reducing participation and creating distrust.

024-00779-1. Google's Gemini AI has been ridiculed by American conservatives for a perceived liberal or left-wing bias. *See e.g.*, Megan McArdle, *Female Popes? Google's Amusing AI Bias Underscores a Serious Problem*, WASH. POST (Feb. 27, 2024, 06:30 AM), https://www.washingtonpost.com/opinions/2024/02/27/google-gemini-bias-race-politics/ [https://perma.cc/6XKL-SXFC]. A 2023 study—amusingly titled "How would ChatGPT vote in a federal election?"—cautiously concluded that ChatGPT shows "an overall political preference for left-leaning ideology" but that it is potentially misleading to declare the tool as having a left-wing bias, as its responses differed by prompt and policy area. Michaela Sullivan-Paul, *How Would ChatGPT Vote in a Federal Election? A Study Exploring Algorithmic Political Bias in Artificial Intelligence,* at 26–27 (Jun. 15, 2023) (M.P.P. thesis, University of Tokyo) , https://www.pp.u-tokyo.ac.jp/wp-content/uploads/2016/02/10_51218255_SULLIVANPAUL_Michaela.pdf [https://perma.cc/3XCJ-DNJ7].

[53] *See* Danielle K. Citron & Robert Chesney, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CAL. L. REV. 1753, 1785–86 (2019). Citron and Chesney highlight that this risk *increases* as the public becomes aware of the risks of AI: "[t]he liar's dividend . . . flows, perversely, in proportion to success in educating the public about the dangers of deepfakes." *Id.* at 1785.

[54] Nilesh Christopher, *An Indian Politician Says Scandalous Audio Clips are AI Deepfakes. We Had Them Tested*, REST OF WORLD (July 5, 2023), https://restofworld.org/2023/indian-politician-leaked-audio-ai-deepfake/ [https://perma.cc/599H-APGG].

III. REGULATING GAI FOR ELECTORAL COMMUNICATIONS

Balancing the near-term, medium-term, and long-term benefits and risks of GAI for electoral communications will be a difficult task for legislators and regulators that is further complicated by the current GAI ecosystem. The opportunities presented by LLMs and the open-source nature of some leading LLMs have led to an explosion of available models.[55] This open-source nature limits options for regulation, as nefarious actors can retrain these models for their purposes.[56] Lawmaking bodies tackling GAI for electoral communications must take into account the diffuse nature of the market and either craft laws that account for that lack of consolidation or target entities at different levels of the GAI supply chain.

Legislative and regulatory bodies in the United States and Europe have started to try to govern the use of these tools. These efforts broadly fall into two categories: efforts to regulate users of AI and penalize misuse and efforts to regulate the AI systems themselves to mitigate the risk of misuse. Each approach has different implications for preserving the benefits of GAI for electoral communications while addressing the harms.

This Part will examine efforts to regulate use of AI, identify key features of the different approaches, and assess how well these efforts address the near-term, medium-term, and long-term benefits and risks of GAI for electoral communications. First, it will look at the GAI industry's current self-regulation of LLMs for electoral communications and show how companies either favor broad bans on use for this purpose or have a nearly single-minded focus on audio, video, and image deepfakes. Next, it will examine efforts to regulate misuse, analyzing both anti-deepfake state bills in the United States and FCC and FEC actions to clarify that existing laws apply to the use of AI-generated communications. It will then look

[55] The website HuggingFace.co, an AI platform that, among other things, connects AI tool developers with LLMs, as of May 14, 2024 listed 641 LLMs that are available to perform natural language processing for question answering in English. HUGGINGFACE, https://huggingface.co/models?pipeline_tag=question-answering&language=en&sort=trending [https://perma.cc/TW95-7Z2C] (last visited May 14, 2024).

[56] R. MICHAEL ALVAREZ, FREDERICK EBERHARDT & MITCHELL LINEGAR, GENERATIVE AI AND THE FUTURE OF ELECTIONS (2023), https://lindeinstitute.caltech.edu/documents/25475/CSSPP_white_paper.pdf [https://perma.cc/UM8M-78W5] ("[R]estrictions placed on foundational models would be short-lived at best. Given the public accessibility of code and data, it is inevitable that, given time or resources, entities would train uncensored versions of the models.").

at two system-centric approaches to regulating GAI and LLMs: the European Union (EU) AI Act and the Biden administration's Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (AI EO). Finally, it will briefly address how the First Amendment could impact US efforts to regulate GAI for electoral communications, arguing that laws and regulations that limit some GAI-created text-based electoral communications can survive judicial scrutiny.

A. INDUSTRY SELF-REGULATION OF POLITICAL USE OF LLMS

Calls for increased regulation of AI—including from AI researchers themselves[57]—continue to grow, and the technology industry, including companies building LLMs, has taken proactive steps to address the perceived threats to democracy arising from its products. The AI Elections Accord, signed by 27 leading technology companies[58] in February 2024, is designed to "set expectations for how signatories will manage the risks arising from Deceptive AI Election Content."[59] Among other principles, the Accord commits its signatories to a set of principles that include "deploying reasonable precautions" to prevent Deceptive AI Election Content and watermarking content when the company is able to clearly show its provenance.[60] The Accord, however, limits the definition of Deceptive AI Election Content to "convincing AI-generated audio,

[57] *See* Cade Metz & Tiffany Hsu, *An A.I. Researcher Takes On Election Deepfakes*, N.Y. TIMES (Apr. 2, 2024), https://www.nytimes.com/2024/04/02/technology/an-ai-researcher-takes-on-election-deepfakes.html [https://perma.cc/7L7N-UNBE] ("Many artificial intelligence researchers warn that the threat is gathering steam. Last month, more than a thousand people—including Dr. Etzioni and several other prominent A.I. researchers—signed an open letter calling for laws that would make the developers and distributors of A.I. audio and visual services liable if their technology was easily used to create harmful deepfakes.").

[58] The 23 signatories are Adobe, Amazon, Anthropic, Arm, ElevenLabs, Gen, Google, IBM, Inflection, LG, LinkedIn, McAfee, Microsoft, Meta, NetApp, Nota AI, OpenAI, Snapchat, Stability.AI, TikTok, Trend Micro, Truepic, and X. AI ELECTIONS ACCORD, https://www.aielectionsaccord.com/ [https://perma.cc/CH5Y-92WX] (last visited Mar. 23, 2024).

[59] AI ELECTIONS ACCORD, A TECH ACCORD TO COMBAT DECEPTIVE USE OF AI IN 2024 ELECTIONS (2024), https://www.aielectionsaccord.com/uploads/2024/02/A-Tech-Accord-to-Combat-Deceptive-Use-of-AI-in-2024-Elections.FINAL_.pdf [https://perma.cc/UJ4V-QRYY].

[60] *Id.*

video, and images."[61] This definition effectively means the Accord only addresses deepfakes, leaving other uses of GAI unregulated. As Part III.B will show, state laws looking to regulate GAI and electoral communications have taken the same narrow approach.

Several of the AI Elections Accord signatories also have policies governing the use of their LLMs for political purposes. OpenAI announced in January 2024 that it would not allow political campaigns to use its ChatGPT chatbot and other GAI products to create tools for campaigning.[62] Anthropic does not allow its Claude AI LLM to be used for "political campaigning or lobbying;"[63] the company explicitly states this includes a ban on candidates using Claude to create "chatbots that can pretend to be the [candidate]."[64] Claude also directed voters to a tool run by the nonprofit Democracy Works if it received queries about voting information for elections in 2024.[65] Meta bars using the GAI advertising products it released in 2023 for ad campaigns that "qualify as ads for . . . Social Issues, Elections, or Politics"[66]—part of a broader shift away from political

[61] *Id.*

[62] *See* Gerrit de Vynck, *OpenAI Won't Let Politicians Use Its Tech for Campaigning, for Now*, WASH. POST (Jan. 15, 2024, 3:00 PM), https://www.washingtonpost.com/technology/2024/01/15/openai-election-misinformation-disinformation/ [https://perma.cc/NM49-B5LU]. OpenAI forbids using either their API platform or ChatGPT to "perform or facilitate the following activities that may significantly impair the safety, wellbeing, or rights of others, including . . . [E]ngaging in political campaigning or lobbying, including generating campaign materials personalized to or targeted at specific demographics." *Usage Policies*, OPENAI, https://openai.com/policies/usage-policies [https://perma.cc/JZ67-7DNQ] (last visited Nov. 1, 2024). Previous efforts by OpenAI to limit use of ChatGPT for political purposes have had limited success, with users able to use prompt engineering to circumvent bans on creating "materials targeting specific voting demographics." Cat Zakrzewski, *ChatGPT Breaks Its Own Rules on Political Messages*, WASH. POST (Aug. 28, 2023, 12:19 PM), https://www.washingtonpost.com/technology/2023/08/28/ai-2024-election-campaigns-disinformation-ads/ [https://perma.cc/8BD5-V6NT].

[63] *Acceptable Use Policy*, ANTHROPIC, https://www.anthropic.com/legal/aup [https://perma.cc/JNF7-N5V8] (last visited May 9, 2024) (defining political campaigning or lobbying as "creating targeted campaigns to influence the outcome of elections or referendums" or "political advocacy or lobbying").

[64] *Preparing for Global Elections in 2024*, ANTHROPIC (Feb. 16, 2024), https://www.anthropic.com/news/preparing-for-global-elections-in-2024 [https://perma.cc/W64B-QFTV].

[65] *Id.*

[66] Katie Paul, *Meta Bars Political Advertisers from Using Generative AI Ads Tools*, REUTERS (Nov. 7, 2023, 4:21 PM), https://www.reuters.com/technology/meta-bar-political-advertisers-using-

content on its platforms[67]—but has no official policies or stated positions about the use of its Llama LLM for electoral communications that are not paid ads.[68] Google's Gemini AI LLM reportedly did not allow users in a country with a 2024 election to ask any questions about that election,[69] but the terms of service do not prohibit the LLM from being used by candidates or campaigns for electoral communications.[70]

Companies like OpenAI and Anthropic are understandably concerned about their LLMs providing incorrect information to voters in the near-term, as well as the potential for their GAI tools to degrade the information environment.[71] Their self-regulation, however, may be counterproductive from the broader perspective of protecting democratic governance. Given the proliferation of open

generative-ai-ads-tools-2023-11-06/ [https://perma.cc/RW4P-FEXM] (quoting Meta webpages providing resources for advertisers).

[67] *See* Taylor Lorenz & Naomi Nix, *Meta Turns Its Back on Politics Again, Angering Some News Creators*, WASH. POST (Feb. 10, 2024, 8:00 AM), https://www.washingtonpost.com/technology/2024/02/10/politics-meta-threads-instagram/ [https://perma.cc/5ZRG-AX8K] ("Meta announced . . . it would stop proactively recommending political content on Instagram or its upstart text-based app Threads").

[68] *See Use Policy*, META, https://ai.meta.com/llama/use-policy/ (last visited Mar. 23, 2024) [https://perma.cc/9WD4-JJCT] (no policy about the use of Llama for political purposes); Nick Clegg, *How Meta Is Planning for Elections in 2024*, META (Nov. 28, 2023), https://about.fb.com/news/2023/11/how-meta-is-planning-for-elections-in-2024/ [https://perma.cc/ERP3-BMU8] (no statement about use of Llama for political purposes). Meta does require deepfake disclosures for posts on Meta platforms that are images, videos, or "realistic sounding audio" that is digitally created or altered and either depicts a person doing or saying something they did not do, invents a realistic-looking person or event, or is a digitally-created depiction of a real event. *Ads about Social Issues, Elections or Politics*, META, https://transparency.meta.com/policies/ad-standards/siep-advertising/siep [https://perma.cc/53U6-4QJS] (last visited May 9, 2024).

[69] Jagmeet Singh, *Google Won't Let You Use Its Gemini AI to Answer Questions About an Upcoming Election in Your Country*, TECHCRUNCH (Mar. 12, 2024, 8:48 AM), https://techcrunch.com/2024/03/12/google-gemini-election-related-queries/ [https://perma.cc/QN9H-AL67] ("When a query about a particular political party or candidate is asked, Gemini shows a message: 'I'm still learning how to answer this question. In the meantime, try Google Search.").

[70] *Generative AI Prohibited Use Policy*, GOOGLE, https://policies.google.com/terms/generative-ai/use-policy [https://perma.cc/3ZWE-FU5H] (last visited March 23, 2024) (no policy about the use of Gemini for political purposes).

[71] *See supra* Section II.B.2.c.

model LLMs,[72] nefarious actors can likely manipulate existing models or leverage other models provided by companies that are less concerned about trust and safety. Legitimate campaigns, meanwhile, will be unable to reap the democracy-promoting benefits that AI offers during this campaign cycle. Overly strict self-regulation may therefore do little to stop near-, medium-, or long-term harms while halting all benefits.

B. REGULATING THE USERS: PENALIZING MISUSE OF GAI

1. *State Laws Addressing Deepfakes*

While Congress has not passed any federal legislation addressing GAI and electoral communications, as of May 9, 2024,[73] fourteen states have enacted laws designed to address the risk of deepfakes being used to sway an election.[74] Twenty-two other states as of that date have bills before their legislatures or awaiting a governor's signature.[75] These fourteen states have largely coalesced around an approach with four key elements: defining what is prohibited; defining who can be prosecuted or held liable; an intent requirement; and exemptions for properly marked materials. Each element has pros and cons if incorporated into legislation designed to substantively address GAI for electoral communications.

a) What is Prohibited?

Each of the fourteen state laws addressing deepfakes in elections limit the law's application to video, audio, and/or image outputs from a GAI system or LLM. The Texas law applies exclusively to videos, specifically those "created with the intent to deceive, that appear[] to depict a real person performing an action that did not occur in

[72] *See supra* Section II at note 49.

[73] Five additional states have enacted deepfake laws between May 9 and the publication date that are not included in this analysis: New Hampshire (August 2, 2024), Hawaii (July 5, 2024), Alabama (May 16, 2024), Arizona (May 21, 2024), and Colorado (May 24, 2024). *See Tracker: State Legislation on Deepfakes in Elections*, PUB. CITIZEN (Oct. 15, 2024), https://www.citizen.org/article/tracker-legislation-on-deepfakes-in-elections/ [https://perma.cc/PY7U-E3GY].

[74] California, Florida, Idaho, Indiana, Michigan, Minnesota, Mississippi, New Mexico, New York, Oregon, Texas, Utah, Washington, and Wisconsin. *See Tracker: State Legislation on Deepfakes in Elections*, *supra* note 73.

[75] *See Tracker: State Legislation on Deepfakes in Elections*, *supra* note 73.

reality."[76] The other laws limit their application to "deepfakes,"[77] "synthetic media,"[78] "fabricated media,"[79] or "materially deceptive" media[80] that share two common features: they are created

[76] Tex. Elec. Code Ann. § 255.004.

[77] H.F. 1370, 93rd Sess., (Minn. 2023) ("[A]ny video recording, motion-picture film, sound recording, electronic image, or photograph, or any technological representation of speech or conduct substantially derivative thereof . . . ."); S.B. 2577, 2024 Reg. Sess. (Miss. 2024) ("'Digitization' means to alter an image or audio in a realistic manner utilizing an image or audio of a person, other than the person depicted, computer-generated images or audio, commonly called deepfakes. 'Digitization' also includes the creation of an image or audio through the use of software, machine learning artificial intelligence or any other computer-generated or technological means."); Tex. Elec. Code Ann. § 255.004 ("[A] video, created with the intent to deceive, that appears to depict a real person performing an action that did not occur in reality.").

[78] S.B. 5152, 68th Leg., Reg. Sess. (Wash. 2023) ("'[S]ynthetic media' means an image, an audio recording, or a video recording of an individual's appearance, speech, or conduct that has been intentionally manipulated with the use of generative adversarial network techniques or other digital technology in a manner to create a realistic but false image, audio, or video . . . ."); S.B. 131, 2024 Gen. Sess. (Utah 2024) ("[A]udio content . . . image or video that was substantially produced by generative artificial intelligence."); H.B. 664, 67th Leg., 2d. Reg. Sess. (Idaho 2024) ("[A]udio recording or a video recording of an individual's speech or conduct that has been created through the use of generative adversarial network techniques or other digital technology in a manner to create a realistic but false audio or video . . . ."); S.B. 1571, 82d. Leg. Assemb., Reg. Sess., (Or. 2024) ("[A]n image, audio recording or video recording of an individual's appearance, speech or conduct that has been intentionally manipulated with the use of artificial intelligence techniques or similar digital technology in a manner to create a realistic but false image, audio recording or video recording . . . ."); A.B. 664, 2023-2024 Leg., Reg. Sess., (Wis. 2024) ("'[S]ynthetic media' means audio or video content that is substantially produced in whole or in part by means of generative artificial intelligence.").

[79] H.B. 1133, 123rd Gen. Assemb., 2d Reg. Sess., (Ind. 2024) ("Media that includes audio or visual recording of an individual's speech, appearance, or conduct that has been altered without the individual's consent . . ." "in which an artificially generated audio or visual imitation of an individual that . . . has been created without the individual's consent . . ." or "depicting the speech, appearance, or conduct of an artificially generated person, the appearance or speech of which is not a recognizable imitation of an identifiable individual." This last clause could potentially be used to stop robocall chatbots like 'Ashley,' even if the campaign had obtained consent as required by the FCC declaratory ruling. *See supra* note 91.).

[80] A.B. 730, 2019-2020 Leg., Reg. Sess. (Cal. 2019) ("[A]n image or an audio or video recording of a candidate's appearance, speech, or conduct that has been intentionally manipulated in a manner such that both of the

and manipulated with technology, and they are limited to audio, video, or image depictions of candidates or other political individuals. None capture purely text outputs; even the laws that include "text" in their definitions of GAI limit the application of the law to content that is visually or audibly deceptive.[81] None of them would have covered, therefore, a text-based chatbot called HarrisBot supporting abolishing the Supreme Court or one called TrumpBot supporting a repeal of the Second Amendment.[82]

This shared element of the state laws is thus a poor fit to address many of the near-term or longer-term harms of text-based GAI-created content being used for electoral communications.

b) Who is Responsible?

None of the enacted laws place responsibility on the maker of the GAI tool or the underlying LLM, instead implicating individuals or campaigns that create and/or disseminate the content. Some also include something akin to Section 230 protection,[83] preventing tech

following conditions are met: (1) The image or audio or video recording would falsely appear to a reasonable person to be authentic. (2) The image or audio or video recording would cause a reasonable person to have a fundamentally different understanding or impression of the expressive content of the image or audio or video recording than that person would have if the person were hearing or seeing the unaltered, original version of the image or audio or video recording."); H.B. 182, 2024 Reg. Sess. (N.M. 2024) ("[A]n image, video or audio that: (1) depicts an individual engaged in conduct or speech in which the depicted individual did not engage; (2) was published, disseminated, distributed or displayed to the public without the consent of the depicted individual; and (3) was produced in whole or in part by using artificial intelligence . . . ."); A.B. 8808, 2023-2024 Gen. Assemb. (N.Y. 2024) ("'Materially deceptive media' means any image, video, audio, text, or any technological representation of speech or conduct fully or partially created or modified that: (1) exhibits a high level of authenticity or convincing appearance that is visually or audibly indistinguishable from reality to a reasonable person; (2) depicts a scenario that did not actually occur or that has been altered in a significant way from how they actually occurred; and (3) is created by or with software, machine learning, artificial intelligence, or any other computer-generated or technological means, including adapting, modifying, manipulating, or altering a realistic depiction.").

[81] *See* A.B. 8808, 2023-2024 Gen. Assemb. (N.Y. 2024).

[82] This may be due to First Amendment concerns. For a discussion of why some government regulation of false, text-based electoral communications should survive First Amendment scrutiny, *see infra* Section III.D.

[83] Section 230 "generally precludes providers and users from being held liable . . . for information provided by another person . . . Courts have

companies from being held responsible when deceptive content is shared on their platforms.[84]

Targeting the users rather than the underlying systems helps prevent campaigns' misuse of GAI. Further, it preserves the benefits of GAI for electoral communications, as it removes incentives for companies to keep their products out of the electoral space. However, targeting users requires state governments to deter, detect, and prosecute misuse, which may be less likely when dealing with skilled disinformation actors who are unlikely to be deterred by state laws and may have the resources and capabilities to avoid being detected.

c) Intent is Required

Almost every state law has some requirement of intent to disseminate the deepfake for the purposes of affecting an election, with the most common statutory language being "intent to injure a candidate or influence the result [or outcome] of an election."[85]

---

interpreted Section 230 to foreclose a wide variety of lawsuits and to preempt laws that would make providers and users liable for third-party content. For example, the law has been applied to protect online service providers like social media companies from lawsuits based on their decisions to transmit or take down user-generated content." Valerie C. Brannon & Eric N. Holmes, CONG. RSCH. SERV., R46751, *Section 230: An Overview* (Jan. 4, 2024).

[84] The Washington state law, for example, makes liable the "sponsor of the electioneering communication" and not the "medium disseminating the electioneering communication," unless the medium either removes a disclosure or alters the communication in such a way that it falls under the definition of synthetic media. S.B. 5152, 68th Leg., Reg. Sess. (Wash. 2023). It goes on to state that "[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." *Id. See also* A.B. 664, 2023-2024 Leg., Reg. Sess., (Wis. 2024) ("No liability for a violation of this subsection shall attach to any person who is a broadcaster or other host or carrier of a video or audio communication . . . that contains synthetic media, unless the person is a committee responsible for the communication.").

[85] Tex. Elec. Code Ann. § 255.004.; H.F. 1370, 93rd Sess., (Minn. 2023). *See also* H.B. 1133, 123rd Gen. Assemb. 2d Reg. Sess., (Ind. 2024). Wisconsin's statute includes language defining issue advocacy, which the statute covers, as "a communication that provides information about political or social issues and is made to influence the outcome of an election." A.B. 664, 2023-2024 Leg., Reg. Sess., (Wis. 2024). Idaho and New York are the exceptions, with no intent requirements. Idaho's law applies to all "electioneering communications, H.B. 664, 67th Leg., 2d. Reg. Sess. (Idaho 2024), and New York's applies to any "person, firm,

Requiring intent addresses some concerns about First Amendment violations.[86] Most of the laws also provide First Amendment protection for parody and for news organizations showing content while clearly stating its provenance. However, adopting an intent requirement for regulation of all electoral communication-related harms would not mitigate against the near-term harm of AI hallucinations regarding the time, place, and manner of voting because that is the result of unintentional model outputs. Nor would it mitigate against unintentional flooding and the resulting degraded information environment, which is an effect of the overall scale of GAI-generated messages rather than an individual's intent to cause harm.

d) Exemptions for Appropriate Disclosures

Finally, most of the laws are either explicitly aimed at disclosure or contain exemptions stating the law does not apply if the content contains appropriate disclosures.[87] Researchers have not made definitive findings on the effectiveness of disclosures and watermarks. One concern is that even if watermarks and disclosure statements are initially placed into electoral communications, they can be easily removed or evaded by third-parties.[88] Additionally, a 2023 study examining perceptions of labeled AI-generated content demonstrated that the specific label used matters; for example, labeling a video as a "deepfake" helped individuals effectively

association, corporation, campaign, committee, or organization that distributes or publishes any political communication that was produced by or includes materially deceptive media . . ." A.B. 8808, 2023-2024 Gen. Assemb. (N.Y. 2024).

[86] *See infra* Section III.B for discussion on intent and the First Amendment.

[87] For example, the California version of the exemption requires a disclosure stating: "This _____ has been manipulated" with either "image," "video," or "audio" filling in the blank, and it provides specific standards for how the disclaimer should be presented to ensure it is readable or hearable by the consumer. A.B. 730, 2019-2020 Leg., Reg. Sess. (Cal. 2019).

[88] *See* Metz & Hsu, *supra* note 57 (image watermarks are easily removed; "detection tools will struggle to surpass new generative A.I. technologies."); ALICE DAWSON & JAMES BALL, GENERATING DEMOCRACY: AI AND THE COMING REVOLUTION IN POLITICAL COMMUNICATIONS, DEMOS 23-24 (Jan. 2024), https://demos.co.uk/wp-content/uploads/2024/01/Generating-Democracy-Report-1.pdf [https://perma.cc/BS5P-WSUT] ("tests have suggested that [AI provenance tools] are currently imperfect systems that can be broken, evaded or even corrupted.").

identify misleading content but not effectively identify that the content was AI-generated.[89]

Watermarking and disclosure statements are also not a substitute for critical thinking about election-related advertising; as political scientist Bruce Bueno de Mesquita writes, "nothing about watermarking will tell you whether or not you should believe the claims and information in a piece of content. There is ample misleading content that is not AI-generated, and there will be plenty of perfectly accurate AI-generated content."[90] Disclosure and watermarking requirements are therefore not a panacea against the short-, medium-, and long-term harms posed by GAI for electoral communications.

2. *United States Regulatory Agencies*

The FCC and FEC have both taken steps to extend existing rules and regulations to cover GAI-based risks. Like the state laws in Section III.B.1, they attempt to prevent some of the most serious misuses of GAI for electoral communications, but they each only address some parts of the problem; the FCC regulation is limited to deceptive audio and /or video content, while the FEC interpretive rule appears to leave huge loopholes that can be exploited by nefarious actors. The FCC's sweeping approach may also prevent candidates from deploying these tools for good uses. While this may be justified given the potential negative effects of GAI-enabled electoral communications, the combined effect of the FCC and FEC rules may be to foreclose some potential benefits of GAI without sufficiently addressing the potential harms.

a) FCC Ban on AI-Generated Voices in Robocalls

The FCC in February 2024 adopted a new Declaratory Ruling clarifying that the Telephone Consumer Protection Act's restrictions on artificial voice covers AI-generated voices, immediately banning the use of AI-generated voices in robocalls and robotexts without obtaining prior express consent from the called party.[91] The

[89] *See* Ziv Epstein, Cathy Mengying Fang, Antonio A. Arechar, & David G. Rand, *What Label Should Be Applied to Content Produced by Generative AI?*, PSYARXIV, July 2023, at 15-16, https://doi.org/10.31234/osf.io/v4mfz.

[90] BUENO DE MESQUITA ET AL., *supra* note 19.

[91] Implications of Artificial Intelligence Technologies on Protecting Consumers from Unwanted Robocalls and Robotexts, FCC 24-17, CG Docket No. 23-362 (Feb. 8, 2024) (Declaratory Ruling),

accompanying statements from Chairwoman Jessica Rosenworcel and Commissioner Geoffrey Starks made it clear that the harms they envisioned included the use of AI-generated voices to deceive voters.[92]

Preventing reckless and nefarious actors from deploying AI-generated voices that can trick voters and contribute to a degraded information environment is an effective use of government regulatory power to address potential harms. However, this broad ban is both overinclusive and underinclusive, foreclosing all effective use of GAI-created audio by campaigns while continuing to leave text-based communications unregulated. Effective tools like Shamaine Daniels' campaign's "Ashley" chatbot can no longer be used to conduct outreach to voters, including voters who do not speak English, even if the audio-based chatbot immediately and clearly disclaims that the called party is speaking to an AI-generated voice. At the same time, the FCC's ruling only applies to phone calls or to AI-generated voice messages sent by text, therefore missing a critical way that campaigns now communicate with voters: text-based messages. Political campaigns in 2022 increased text-based communications by 158% while decreasing phone calls by 57%.[93] The Declaratory Ruling thus protects voters from being deceived by GAI-created voices but does not sufficiently address either the near-term harms of text-based chatbot-disseminated false information, the medium-term harms of microtargeting, or the longer-term harms of a degraded information environment.

b) FEC Rule on Fraudulent Misrepresentation and AI

The FEC in September 2024 released a new Interpretive Rule in response to a petition by the civil society organization Public Citizen asking the FEC to clarify that its regulation prohibiting fraudulent misrepresentation of candidates or political parties extends to "deliberately deceptive" campaign advertisements generated using AI.[94] The FEC chose to clarify that the Federal Election Campaign Act's prohibition against fraudulent misrepresentation is technology

https://docs.fcc.gov/public/attachments/FCC-24-17A1.pdf [https://perma.cc/6VBG-HUZU].

[92] *Id.*

[93] Sareen Habeshian, *Why Political Campaigns Won't Stop Texting You*, AXIOS (Feb. 22, 2024), https://www.axios.com/2024/02/22/political-campaign-texts-election-2024 [https://perma.cc/5QTP-XE2J].

[94] Statement by Vice Chair Ellen L. Weintraub on the Disposition of the Rulemaking Petition Regarding Fraudulent Misrepresentation and Artificial Intelligence, REG 2024-02 (Artificial Intelligence in Campaign Ads) (Sept. 19, 2024).

agnostic and therefore covers the use of AI (and GAI) tools for the purpose of "speaking or writing or otherwise acting for or on behalf of any other candidate or political party or employee or agent thereof on a matter which is damaging to such other candidate or political party or employee or agent thereof."[95]

The Interpretive Rule provides one significant benefit that the state laws and FCC regulation do not; it clearly extends beyond audio and video misrepresentation to cover text-based communications, as the statute explicitly covers "writing." However, because the FEC chose to issue the interpretive rule rather than conduct notice-and-comment rulemaking, there remain significant grey areas about what constitutes "speaking or writing or otherwise acting for or on behalf of" a political candidate. For example, it is unclear whether the Interpretive Rule covers the type of deepfakes that led Public Citizen to submit its petition in the first place, as an ad could use AI to deceptively portray a candidate without pretending to speak on behalf of that candidate.[96] These gaps invite the type of misuse by political candidates and campaigns that can cause near-term harms through GAI-enabled false information and longer-term harms of a degraded information environment.

There is also a major gap in *whom* the interpretive rule covers that the FEC could not address because it lacked authority. As FEC Vice Chair Ellen L. Weintraub explained in her statement on the disposition of Public Citizen's petition, "the [Federal Election Campaign] Act's fraudulent misrepresentation prohibition covers only misrepresentations made by a federal candidate or the candidate's employee or agent or misrepresentations in solicitations. In other words, the prohibition does not generally prohibit non-candidate committees, such as Super PACs, or other individuals from using AI to misrepresent a candidate's image or statement."[97] In the 2020 election cycle, PACs spent roughly the same amount as presidential candidates, congressional candidates, and party

[95] FEC Fraudulent Misrepresentation of Campaign Authority, 11 C.F.R. § 110 (2024) (Interpretive Rule).

[96] *See* Weintraub, *supra* note 94 ("The petition posited that '[a] deepfake audio clip or video by a candidate or their agent that purports to show an opponent saying or doing something they did not do would violate . . . the law. . . Specifically, by falsely putting words into another candidate's mouth, or showing the candidate taking action they did not, the deepfake would fraudulently speak or act "for" that candidate in a way deliberately intended to damage him or her.' The Commission has not addressed this interpretation, which is at the crux of the petition, and it is not at all clear that a majority of commissioners agrees.").

[97] Weintraub, *supra* note 94.

committees combined,[98] meaning that the FEC's interpretive rule likely only covers around half of all electoral communications. Super PACs are free to continue to misuse GAI-enabled tools for electoral communications, thus contributing to the longer-term harm of a degraded information environment.

C. REGULATING THE SYSTEMS

1. *The EU AI Act*

The EU AI Act takes a very different regulatory route, placing the onus on AI systems' creators to sufficiently safeguard their technology against abuse. The AI Act designates AI systems with certain functions as high-risk, requiring its creators to follow a stricter set of protocols to have their product on the market. High-risk systems include those involved in the "administration of justice and democratic processes," including "AI systems intended to be used for influencing the outcome of an election or referendum or the voting behavior of natural persons in the exercise of their vote in elections or referenda."[99] This clearly designates GAI systems used for electoral communications as high-risk AI systems.

The act also classifies any general-purpose LLM with a sufficient "cumulative amount of computation used for its training" or otherwise determined to have "high impact capabilities evaluated on the basis of appropriate technical tools and methodologies, including indicators and benchmarks"[100] as high risk. These broad criteria will almost certainly classify any LLM that would be used for electoral communications as high-risk. These high-risk LLMs are required to submit to strict transparency requirements and cooperate with their downstream providers to enact strict, preventative safety measures against unexpected harm (although it is unclear how enforcement will take place). Thus, the risk is passed on to downstream AI product providers leveraging the LLM.[101] Whichever

[98] Press Release, Fed. Election Comm'n, Statistical Summary of 24-Month Campaign Activity of the 2019-2020 Election Cycle, (Apr. 2, 2021), https://www.fec.gov/updates/statistical-summary-24-month-campaign-activity-2019-2020-election-cycle/ [perma.cc/HGE4-YDKT].

[99] *Annex III*, E.U. AI ACT, 2024 O.J. (L 2024/1689), https://artificialintelligenceact.eu/annex/3/ [perma.cc/V4BE-L6GC] (last visited on Mar. 25, 2024).

[100] *Article 51*, E.U. AI ACT, 2024 O.J. (L 2024/1689), https://artificialintelligenceact.eu/article/51/ [perma.cc/V4BE-L6GC] (last visited on Mar. 25, 2024).

[101] *See* Josephine Wolff, William Lehr, & Christopher S. Yoo, *Lessons from GPDR for AI Policymaking*, 27 VA. J.L. & TECH. 1, 4-5 (2024).

system is high-risk and bears the responsibility then needs to comply with a very strict set of documentation, risk management, data governance, and oversight procedures.[102]

Regulating systems, rather than users, allows the AI Act to tackle a broader set of harms over time, making it a better fit to mitigate the longer-term harms posed by GAI for electoral communications. However, although this approach does more to fight harms at the source than the US state laws or federal regulations, it risks foreclosing many GAI-enabled electoral communication tools. Penalties are strict: up to 15,000,000 EUR or "3% of its total worldwide annual turnover for the preceding financial year."[103] Like OpenAI and Anthropic, many companies may choose to instead ban political use cases, thus cutting off the benefits of GAI for electoral communications.

2. *The Biden Administration's AI Executive Order*

Like the EU AI Act, the Biden administration's 2023 AI Executive Order (AI EO) also seeks to regulate the underlying AI systems and LLMs to mitigate risk. The EO is a sweeping attempt to use the whole of government to both capitalize on AI's benefits and address AI's risks by directing administrative agencies to take a variety of actions within their delegated powers.[104]

Although the AI EO does not specifically address the use of GAI for electoral communications, one provision can be used by relevant administrative agencies to bring this use case under the EO's purview. The EO orders the Secretary of Commerce to use Defense Production Act authorities to require "[C]ompanies developing or demonstrating an intent to develop potential dual-use foundation models" to provide regular reports to the government, including information on the LLM training process and red-team testing results.[105] The EO definition of a dual-use foundational model

[102] *See Title III,* E.U. AI Act, (L 2024/1689), https://artificialintelligenceact.eu/chapter/3/ [perma.cc/7HNR-FCZ8] (last visited Mar. 25, 2024).

[103] *Article 99*, E.U. AI Act, (L 2024/1689), https://artificialintelligenceact.eu/article/99/ [perma.cc/V4DP-WG2R] (last visited Mar. 25, 2024).

[104] *See Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence,* BRIEFING ROOM (Oct. 30, 2023), https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/ [perma.cc/9G4A-SMZ6].

[105] Exec. Order No. 14,110, 88 Fed. Reg. 75191 (Oct. 30, 2023). The U.S. AI Safety Institute at the National Institute of Standards and

includes "an AI model that is trained on broad data . . . and that exhibits, or could be easily modified to exhibit, high levels of performance at tasks that pose a serious risk to security, national economic security, national public health or safety."[106] The Secretary of Commerce could clarify that an AI model that can be used, or modified for use, to negatively affect electoral processes is considered a "dual-use foundational model" due to the serious risks to national and economic security posed by threatening the legitimacy of elections.

On its own, this would not mitigate the near-term nor long-term harms presented by GAI for electoral communications. However, requiring LLMs to meet a minimum set of security standards is the first step towards creating a set of vetted GAI models that citizens can trust are not being manipulated by malicious actors. Combined with provenance requirements that let people know which LLM created the material,[107] this could mitigate intentional efforts to exploit a degraded information environment by helping voters identify information coming from trusted, verified sources.[108]

3. *Surviving First Amendment Scrutiny*

As this Part has demonstrated, state and federal legislative approaches in the United States have avoided tackling the issue of GAI-created text that impersonates a candidate or campaign.[109] While this may be partially because the impersonation risks are greater with video, audio, or image, it also probably reflects concerns among the legislatures that purely text-based electoral communications are afforded greater First Amendment protections.

Academics are divided as to whether GAI-created content is speech under the First Amendment.[110] Even if it is protected speech,

Technology defines AI red-team testing as a "structured testing effort to find flaws and vulnerabilities in an AI system, often in a controlled environment and in collaboration with developers of AI." NAT'L INST. OF STANDARDS AND TECH., MANAGING MISUSE RISK FOR DUAL-USE FOUNDATION MODELS, NIST AI 800-1 IPD (2024).

[106] *Id.*

[107] *See infra* Section IV.B.2

[108] *See supra* Section II.B.2.c.

[109] *See supra* Section III.B.

[110] *See* Lamo & Calo, *supra* note 49, at 1005 ("There is no requirement that speech be original, creative, or well-reasoned in order to qualify for First Amendment protection. Therefore, even bots that do not generate any kind of original content might receive protection under the First Amendment. . . . the fact that a Twitter bot creator may not know what her creation will tweet next should not place the bot outside the protection of the First Amendment. . . Finally, the First Amendment protects not only the

carefully written laws and regulations addressing some harms should be able to survive judicial scrutiny as long as they are narrowly tailored to "preserve the institutional function of election administration."[111] Intentional misinformation about the time, place, and manner of voting probably meets that description; in *Minnesota Voters Alliance et.al. v. Mansky* the Supreme Court stated that "we do not doubt that the State may prohibit messages" in a polling place that are "intended to mislead voters about voting requirements and procedures,"[112] suggesting that the Court is comfortable limiting some false speech that would affect someone's ability to vote.[113] Limiting restrictions to a certain period of time before an election—as some of the state deepfake laws do—can help demonstrate that the laws are narrowly tailored and that counterspeech would not be an effective alternative to the restrictions.[114] Requiring intent—again like the state deepfake laws—will also bring any new statute in line with most interpretations of the *mens rea* needed for false speech to be prosecuted.[115] Additionally, dicta in the Eastern District of New York's decision in *United States v. Mackey* suggests that deliberate misinformation about the time, place, and manner of voting should not be afforded First Amendment protections because it constitutes fraud; [116] impersonating a candidate can also be considered fraud.

speaker's right to speak but the right of those who wish to read or listen to bot speech."). *But see* DAN L. BURK, ASEMIC DEFAMATION, OR, THE DEATH OF THE AI SPEAKER 20 (2023) ("[W]here there is no speaker, there is unlikely to be speech . . . The First Amendment safeguards communication from undue governmental intrusion, but no communication occurs in the case of LLM text generation, in the sense that no speaker is attempting to convey any message to the recipient.").

[111] Deborah Pearlstein, *Democracy Harms and the First Amendment*, KNIGHT FIRST AMEND. INST., Oct. 2022, at 628.

[112] Minnesota Voters Alliance v. Mansky, 585 U.S. 1,18 n.4 (2018).

[113] *But see* Eugene Volokh, *When Are Lies Constitutionally Protected?*, KNIGHT FIRST AMEND. INST., Oct. 2022, at 686– 87 ("[T]hat case [*Mansky*] focused on speech in a nonpublic forum (polling places), and it's not clear that the Court meant to authorize such prohibitions in public speech more generally").

[114] *See* United States v. Mackey, 652 F. Supp. 3d 309, 347 (E.D.N.Y. 2023) (counterspeech "is unlikely to be of much use in the context of tweets spread across the far reaches of the internet in the days and hours immediately preceding an election").

[115] *See, e.g.*, Richard L. Hasen, *A Constitutional Right to Lie in Campaigns and Elections?,* 74 MONT. L. REV. 53, 71 (2013) (interpreting Court decisions as requiring an actual malice standard to prosecute false speech).

[116] *Mackey,* 652 F. Supp. 3d at 347. The court also argues that the evidence in *Mackey* supports the creation of a new category of unprotected speech: "false speech injuring the 'integrity of Government processes.'" *Id.*

Finally, although the First Amendment protects against compelled speech in many circumstances, compelling campaigns to provide watermarks or provenance markings on GAI-created content is consistent with other electoral communications laws, where "the courts have upheld reasonable accountability measures to preserve the sanctity of elections that require attribution for advertisements."[117] Such a requirement also does not restrict content, viewpoints, or capacity for anonymity, making it likely that such a regulation would survive a First Amendment challenge even if a court decided that GAI-created content is speech under the First Amendment.[118]

IV. RECOMMENDATIONS FOR REGULATING GAI FOR ELECTORAL COMMUNICATIONS

GAI presents tangible benefits for political actors, voters, and potentially democracy itself, but it poses major risks to democratic governance. It is also not going away; as LLMs become more accurate and technologists discover more use cases, GAI's inclusion in the electoral communications toolkit is almost inevitable. Lawmakers, regulators, and advocates looking to effectively regulate the use of GAI can acknowledge these realities and attempt to craft solutions that mitigate risks, allow benefits, and evolve as the technology and its use cases change.

This section will provide some suggestions on how policymakers can try to achieve these goals, drawing on both the analysis in the previous sections as well as recommendations from academics and advocates in the space. It proposes two sets of recommendations, one for the near-term and one for the long-term, that would be significant steps toward mitigating the harms of GAI and electoral communications while maintaining its benefits.

A. NEAR-TERM RECOMMENDATIONS

1. *Target Uses and Users*

The EU Act and the Biden AI EO both show that system-centric efforts to regulate LLMs and GAI tools require time, both for

at 347 (citing *U.S. v. Alvarez*, 567 U.S. 709, 720–21 (2012)). Mackey's appeal is currently before the Second Circuit; however, these issues may not be addressed on appeal, as the court first held in *Mackey* that intermediate scrutiny was the proper standard and that the government's prosecution met this standard. *Id.* at 347.

[117] Lamo & Calo, *supra* note 49, at 1010.

[118] *Id.* at 1014, 1018–19.

regulatory agencies to develop policies and standards and for companies to demonstrate that they are complying with those policies. Hasty efforts to target underlying systems risk being so sweeping that they either eliminate all benefits of GAI for electoral communication through broad bans or induce excessive self-regulation (like Anthropic or OpenAI).[119] Potential bans would do little to stop nefarious actors from leveraging open-source LLMs. And although regulating users will not deter nefarious actors who are determined to misuse GAI despite the law, it may deter political campaigns and candidates from testing the limits of using GAI for misinformation and strike a better balance by allowing for positive uses by campaigns and candidates.

2. *Enact an Amended Version of the Protect Elections from Deceptive AI Act*

Congress can pass a version of the bipartisan *Protect Elections from Deceptive AI Act* introduced in the Senate on September 12, 2023,[120] but with amendments to cover text-based electoral communications created by GAI systems. The current language mirrors much of the state legislation, applying the terms of the law only to "image, audio, or video that appears authentic."[121] Adding "and text-based communications that purport to be written by a candidate, candidate's staff, or an organization working on behalf of a candidate" would directly address the near-term risk of an AI chatbot representing itself as a campaign and spreading misinformation at scale about the time, place, and manner of voting. It could also address medium and longer-term concerns about a degraded information environment, where any chatbot tool could pretend to speak on behalf of any candidate.

To address First Amendment concerns and allow campaigns to leverage some of the benefits of GAI, the federal law should adopt the timing restrictions featured in some state laws. Five states with laws enacted at the time of this writing have timing provisions

[119] *See supra* Section III.A.

[120] *See Klobuchar, Hawley, Coons, Collins Introduce Bipartisan Legislation to Ban the Use of Materially Deceptive AI-Generated Content in Elections*, U.S. SEN. AMY KLOBUCHAR WORKING FOR THE PEOPLE OF MINN. (Sept. 12, 2023) https://www.klobuchar.senate.gov/public/index.cfm/2023/9/klobuchar-hawley-coons-collins-introduce-bipartisan-legislation-to-ban-the-use-of-materially-deceptive-ai-generated-content-in-elections. [https://perma.cc/C39U-68WD]

[121] Protect Elections from Deceptive AI Act, S. 2770, 118th Cong. (2023).

limiting when the laws kick in based on proximity to an election; Texas (30 days),[122] California (60 days),[123] and Michigan, Minnesota, and Mississippi (90 days).[124] This will have negative effects as well; banning GAI-enabled communications during the critical get-out-the-vote phase of a campaign will hinder smaller, under-resourced campaigns using these tools. Lawmakers and AI platforms will also need to communicate effectively about the timing of these bans to avoid disinformation about the reasons for the restrictions.[125] But given the potential effects of an AI-enabled chatbot giving out incorrect election information when early voting in many states has already begun, the risk calculus favors caution as the calendar moves toward Election Day.

3. *Require GAI Disclosures by Campaigns*

Candidates and campaigns could be required to disclose to voters when they are communicating with GAI, including via a text-based tool. This requirement would be similar to the disclosure exemptions in state deepfake laws and the FCC requirement that voters consent to receiving calls from an AI-generated voice, aligning with some calls for "professional users" to disclose when content is AI-generated.[126] Although watermarks are not always effective,[127] people are responsive to some information disclosures.[128] As more citizens become aware of the risks of AI hallucinations, these disclosures can prompt some to check authoritative sources to ensure the information they are receiving is accurate, mitigating some

[122] Tex. Elec. Code Ann. § 255.004. (West 2024).

[123] Cal. Elec. Code § 20010 (West 2019).

[124] H.B. 5144, 102d Leg., Re. Sess. (Mich. 2023); H.F. 1370, 93rd Sess., (Minn. 2023); S.B. 2577 2024 Reg. Sess. (Miss. 2024).

[125] *See* Megan Shahi, *Protecting Democracy Online in 2024 and Beyond*, CTR. FOR AM. PROGRESS (Sept. 14, 2023) https://www.americanprogress.org/article/protecting-democracy-online-in-2024-and-beyond/ [https://perma.cc/4ZFA-U4AW] ("Ahead of plans to enact a restriction period for political ads and/or disable the creation of new political ads in an election period, platforms should provide advertisers and users sufficient notice of any restrictions.").

[126] Philipp Hacker, Andreas Engel, & Marco Mauer, *Regulating ChatGPT and other Large Generative AI Models*, *in* 2023 ACM CONF. ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 1112, 1119 (2023).

[127] *See supra* Part III.B.1.d.

[128] *See* Simchon, Edwards, & Lewandowsky, *supra* note 32 , at 3 ("The effectiveness of advertisements can significantly diminish when individuals become aware of unacceptable targeting practices, such as using external data or inferred information without their consent.") (internal citations omitted).

concerns about misinformation regarding the time, place, and manner of voting. This requirement will also not negatively affect the positive benefits of GAI for electoral communications, as all of those use cases are compatible with clear messaging that a voter is engaging with AI.

4. *Encourage Partnerships to Provide Voters with Authenticated Election Information*

Anthropic's partnership with Democracy Works to redirect voting-related queries to the Democracy Works tool TurboVote is an excellent example of a GAI company allowing its tools to be used as a resource for voters while providing them with the option to receive information from a trusted, authenticated source.[129] It mitigates the risk of AI hallucinations spreading incorrect information about the time, place, and manner of voting while still allowing voters to query the Claude chatbot and become more informed citizens. These partnerships thus mitigate against the most plausible near-term harms without infringing on any benefits.

B. MEDIUM AND LONGER-TERM RECOMMENDATIONS

1. *Categorize LLMs that Can Be Used for Electoral Communication Tools as Relevant to National Security and Critical Infrastructure in Accordance with the Biden administration's AI EO*

The Secretary of Commerce could clarify the definition of "dual-use foundational model" in the AI EO to include LLMs and other AI models that can be used for or modified for use for electoral communications, due to the security and economic security risks of those tools being misused.[130] This will require the companies who own those models to submit testing and compliance documents under the Secretary's Defense Production Act authority. Although this could lead some companies to prevent their products from being used for electoral communications, the broad definition of "dual-use foundational model" should lead enough LLMs to submit to this process to provide candidates and campaigns with suitable tools. Combined with watermarking and provenance requirements, the EO can create an ecosystem where citizens can check to see if the AI-

[129] *Preparing for Global Elections in 2024*, ANTHROPIC (Feb. 16, 2024), https://www.anthropic.com/news/preparing-for-global-elections-in-2024 [https://perma.cc/W64B-QFTV].

[130] *See supra* Section III.C.2.

generated electoral communication they received was generated by a red-teamed LLM.

The Department of Homeland Security can also release guidance explicitly stating that the designation of election security as a critical infrastructure subsector[131] means those same models are relevant to critical infrastructure as defined in the AI EO, allowing appropriate federal government agencies to develop guidelines for their use in electoral communications.[132] Agencies can use this authority to issue regulations such as watermarking and provenance requirements.

2. *Require Provenance Markings on AI-Generated Electoral Communications*

As leading industry companies continue to develop digital watermarking and provenance tools,[133] future legislation or regulation can include a requirement that GAI tools used for electoral communications show both users and regulators what underlying LLM was leveraged. When enacted in concert with the safety and compliance testing requirements in the EU AI Act and Biden AI EO,[134] these measures will enable civil society and voters to authenticate that the content was provided by systems that have met standards promulgated by the government, providing one layer of validation.

It is critical, however, that digital watermarks and provenance notes are only viewed as part of the solution. They have clear benefits when it comes to knowing the origins of GAI-created media campaigns, enforcing violations of statutes against the correct parties, and providing voters and civil society with information to understand the source of electoral communications and counter misuses. But given the questionable effectiveness of watermarks,[135] it will take civic education on AI, voter literacy efforts, and

[131] *See* Election Security, *Department of Homeland Security Cybersecurity and Infrastructure Security Agency*, CISA, https://www.cisa.gov/topics/election-security [https://perma.cc/9S58-XEAA] (last visited May 13, 2024).

[132] Exec. Order No. 14,110, 88 Fed. Reg. 75,191 (Oct. 30, 2023).

[133] *See, e.g.,* Susan Jasper, *How We're Approaching the 2024 U.S. Elections*, GOOGLE: THE KEYWORD (Dec. 19, 2023), https://blog.google/outreach-initiatives/civics/how-were-approaching-the-2024-us-elections/ [https://perma.cc/D5BV-VFPA] ("SynthID, a tool in beta from Google DeepMind, directly embeds a digital watermark into AI-generated images and audio.").

[134] *See supra* Section III.C.

[135] *See supra* Section III.B.1.d.

additional work by civil society to ensure that GAI is a benefit, not a detriment, to our democracy.

3. *Hold LLM and AI System Providers Responsible for Meeting Standards, Not for User Misuse*

Holding LLM and AI system providers responsible for ensuring their models meet foundational safety standards that can help prevent some of the worst-case scenarios, such as a widely-used LLM being manipulated by a nefarious actor to spread false information, can mitigate tail risks of GAI-enabled electoral communications. However, if companies are meeting those standards, they should not be held liable for harms caused by users deploying their models or tools for electoral communications. Otherwise, regulations run the risk of pushing reputable LLM providers to make the same decision as OpenAI and Anthropic in the 2024 election cycle: not allowing any political campaign use at all.[136] This is likely to foreclose many beneficial uses while not preventing harms from malicious actors who are able to leverage and manipulate open-source LLMs for their activities.

Whenever possible, GAI systems designed for electoral communications should be regulated at the fine-tuning phase.[137] This approach—which is broadly the approach taken by the EU AI Act—means that rather than requiring GAI companies to go through costly LLM retraining that could lead companies to pull their products from the electoral communications space, companies like Civox and Daisychain would be instead required to go through a much less expensive process in order to meet statutory or regulatory requirements. It would also allow lawmakers and regulators to craft rules targeting electoral communications "more narrowly, precisely, and surgically," rather than needing to think about the collateral effects of a rule on an entire model.[138]

4. *Increase Civic Education on AI, Voter Literacy Programs, and Civil Society Partnerships to Address the Harms of Microtargeting and a Degraded Information Environment*

Ultimately, the risks of microtargeting-fueled polarization and a degraded information environment will not be mitigated through GAI regulation alone. Nefarious actors will continue to 'flood' the

[136] *See supra* Section III.A.

[137] *See* Ohm, *supra* note 8, at 236–237 (arguing for focusing policy interventions on the fine-tuning stage).

[138] *Id.* at 25.

US information ecosystem with false and conflicting information, and some unscrupulous candidates and campaigns will find loopholes to leverage GAI to create dubious content that skirts the line of legality. Civic education on how to communicate with AI tools, "digital literacy programs that are focused on detecting and contextualizing false online content,"[139] and partnerships between industry, government, and civil society to counter misinformation and promote better democratic discourse will be necessary to combat the damage enabled by technology platforms and AI tools. Efforts such as the MediaWise Voter Project, a Poynter Institute initiative in 2020 aimed at providing digital literacy education to first-time voters, could serve as models for such efforts.[140]

The good news is that many leading AI companies recognize the role they should play in this work. The 23 signatories to the AI Election accord have committed to "engage with a diverse set of global civil society organizations, academics, and other relevant subject matter experts" and to "foster public awareness and all-of-society resilience regarding Deceptive AI Election Content," including through education campaigns and partnerships with "organizations and communities engaging in responding to these risks."[141] Policymakers should provide grant money to civil society organizations interested in creating these partnerships and leading these education and awareness campaigns.[142]

These efforts can include education on how to use GAI to become a more informed voter. Civil society and campaigns can leverage GAI tools to reap the potential longer-term benefits discussed in Section II.B.1.c, including debunking conspiracy theories spread through microtargeting and reaching more voters with relevant, authenticated information on candidates and on the time, place, and manner of voting. The best way to counter the harms of GAI may in some instances be to harness its benefits.

[139] Perault & Scott Babwah Brennen, *supra* note 37.

[140] *See* Tina Dyakon, The Poynter Institute Announces Investment from Facebook to expand MediaWise Digital Information Literacy Program to First-Time Voters, POYNTER INST. (Jan. 22, 2020), https://www.poynter.org/from-the-institute/2020/the-poynter-institute-announces-investment-from-facebook-to-expand-mediawise-digital-information-literacy-program-to-first-time-voters/ [https://perma.cc/ZM3E-TZCJ].

[141] *A Tech Accord to Combat Deceptive Use of AI in 2024 Elections*, AI ELECTIONS ACCORD (Feb. 16, 2024), https://www.aielectionsaccord.com/uploads/2024/02/A-Tech-Accord-to-Combat-Deceptive-Use-of-AI-in-2024-Elections.FINAL_.pdf [https://perma.cc/FCD7-9TL2].

[142] *See* Perault & Scott Babwah Brennen, *supra* note 37.

V. CONCLUSION

GAI-enabled political tools will almost certainly continue to proliferate, aided by the open-source nature of many LLMs. These tools can provide benefits for political candidates and campaigns, and research suggests they can even be leveraged to strengthen democracy. Using GAI-enabled tools for electoral communications, however, also poses significant risks, including tail risks that can threaten democracy itself.

Some policymakers may conclude that these risks—especially the tail risks—so far outweigh the benefits that sweeping regulations, or even outright bans, on GAI-enabled electoral communications are required to protect democratic institutions. Others may dismiss these very real risks as speculative and conclude that strict regulation is not needed. Policymakers who want to mitigate these risks but also preserve the benefits, however, have options. Some of them build on work done by state legislatures, federal regulatory bodies, the Biden administration, and the EU; others will require entirely new approaches.

In the near-term, enacting laws and/or regulations that target users and uses, defend against deepfakes and text-based fraudulent representation, require disclosures when campaigns use GAI, and encourage partnerships with civil society to disseminate authenticated information can mitigate against the near-term harms of GAI-enabled electoral communications. In the medium and longer-term, categorizing LLMs that could be used for electoral communications as relevant to U.S. national security and critical infrastructure, requiring provenance markings on GAI-enabled electoral communications, holding LLM and AI system providers accountable to a set of standards, and increasing civic education and voter literacy can mitigate the longer-term harms of microtargeting and a degraded information environment. These solutions would still allow political candidates, campaigns, and civil society to leverage GAI tools for positive uses, including to counter some of the harms caused or accelerated by GAI.