CSMaP | CENTER FOR SOCIAL MEDIA AND POLITICS

July 7, 2023

Director Arati Prabhakar
Office of Science and Technology Policy
Executive Office of the President
Eisenhower Executive Office Building
1650 Pennsylvania Avenue
Washington, D.C. 20504

Dear Director Prabhakar:

Thank you for the opportunity to comment on the Administration's National Artificial Intelligence (AI) Strategy.

We lead NYU's Center for Social Media and Politics, an academic institute that works to strengthen democracy by conducting rigorous research, advancing evidence-based public policy, and training the next generation of scholars. Part of our research concerns studying how information and misinformation spread online, and the impact that spread has on politics, policy, and democracy.

AI has the potential to revolutionize society in so many ways, impacting everything from the economy to national security to civil rights. As researchers who study online information flows, we are submitting comments focusing on topic area 15, specifically discussing how the United States should "address the challenges that AI-generated content poses to the information ecosystem."

Our comments will focus on three areas: (1) the evolution of the modern information ecosystem; (2) how AI-generated content could be used and abused; and (3) what can be done.

1.) The Evolution of the Modern Information Ecosystem

The information environment has transformed radically over the past half century. For much of the 20th century, the American media landscape was fairly limited. There were three major broadcast networks, a few national newspapers, and hundreds of local papers. Cable news fractured this information environment in the 1980s, and social media deepened the cracks in the 2000s.

When social media first burst onto the political scene, it was hailed as a "Liberation Technology" that would accelerate the spread of democracy around the world. Gone were the days of Walter Cronkite as a singular news source for the American public. Gone, too, were the gatekeepers: the television bookers, newspaper editors, and others who had kept out so many voices from the conversation.

Yet, in the aftermath of the 2016 U.S. elections, the mood shifted. Experts instead asked, "Can Democracy Survive the Internet?" The speed at which social media turned from savior to foil of democracy in under five years was head-spinning. The rise of hate speech, echo chambers, and filter bubbles, and, perhaps most of all, the spread of false information online, led to serious re-evaluations of the technology's relationship to politics.

Enter generative AI, which took less than six months to go from a marvel of technological sophistication to quite possibly the next great threat to democracy. While many would-be threats have been ascribed to the rise of Large Language Model (LLM) chatbots, in terms of the information environment, the primary concern is that generative AI will turbo-charge the spread of political misinformation ahead of the 2024 U.S. elections and beyond.

There is good reason to be concerned. In the aftermath of the 2016 elections, misinformation became both an object of concern and inextricably linked to social media because social media drove down the cost of *spreading* misinformation. Misinformation has long been part and parcel of the political world in both democratic and non-democratic political systems. However, spreading misinformation has traditionally required real resources, such as access to a printing press centuries ago or, in the modern era, access to print media, radio, or television.

Social media changed this calculation. Now, anyone could share information that — at least in theory — had the potential to go viral and reach millions of people with little more than the cost of setting up an account on one of many social media platforms that were free for all to use. Furthermore, there were economic benefits to be had from producing such viral content. And even when content did not go viral, it could still be seen by many in one's online networks.

And yet, the content that would hypothetically go viral still needed to be produced by someone. Even if it was easier than ever to spread fake news stories, someone still had to write those stories or Photoshop the pictures in them.

In the past six months, though, we have learned that AI can now write high-quality text and produce (largely) high-quality images; video is likely not too far behind. In other words, just as social media reduced barriers to the *spread* of misinformation, AI has now reduced barriers to the *production* of misinformation.

Throughout these decades of changes — from three TV networks to cable news to social media — we have also seen troubling evidence of a democracy in decline. Poll after poll finds declining trust in institutions (government, media, science), in each other, and in democracy itself. Research also finds increasing levels of political polarization and sectarianism.

Inserting AI into an already fractured democracy risks further eroding trust in our society and degrading the information environment. At times it already seems as if Republicans and Democrats live in completely different worlds. What happens if we truly can't tell what's real and what's not?

2.) How AI-Generated Content Could be Used and Abused

When considering the challenges AI poses to the information ecosystem, it is useful to distinguish between the various content types and consider how each may be used or abused by various actors.

a. Text

Text is likely to be the most ubiquitous form of AI-generated content online. Because it can be copy and pasted and edited, it will be very difficult to tell what content has been produced by humans and what content has been produced by AI.

Determining whether this is a problem varies greatly by how it is used, however. For example, you can imagine a scenario where a good actor harnesses generative AI to advance democracy. A political campaign may use it to create more and better targeted fundraising materials. Policymakers might use it to quickly synthesize research in a given field.

This is very different from bad actors using it to sow chaos. In 2016, for example, a group of Macedonian teenagers created a network of fake news sites and social media accounts to help spread disinformation. This took human labor, but generative AI dramatically lowers the costs for this type of influence operation.

Reports already indicate there are hundreds of fake news sites created using LLMs. It's easy to imagine a world where false AI-generated content proliferates across the web, and is promoted relentlessly by social media accounts created with the help of AI tools.

b. Images

AI-generated images are another area of concern, perhaps more so than text. We've already seen these tools create convincing fake images that circulate rapidly online, ranging from silly and light-hearted (the Pope's puffy coat) to potentially problematic (fake Trump arrest photos, fake Pentagon explosion).

Again, determining whether this is a problem varies greatly by how it is used. Using campaigns as an example once again, you could imagine local political candidates, with very few resources, using AI to generate high-quality (or, as it turned out for one Toronto mayoral candidate who released an ad featuring an attentive constituent with three arms, low-quality) images for campaign materials and social media. Businesses or nonprofits might use AI image generators or editors for promotional materials.

You could also imagine how bad actors could use it to sow chaos. The Pentagon explosion image, though easy to spot, is a case in point. Imagine images circulating on Election Day showing ballots being shredded or discarded. Then imagine prominent politicians sharing those images online and calling for their supporters to rush to polling places. One fake image could very quickly lead to real-world violence.

c. Video

Quality AI-generated video is challenging to produce in its entirety. We've seen a few examples, such as a video-based pro-China disinformation campaign, discovered by Graphika. But we most often see deepfakes, when people use AI and other tools to doctor the content of real videos. A few prominent examples include a video of Nancy Pelosi, which was edited to make it seem like she was slurring her words, and one of Ukrainian President Volodymyr Zelenskyy manipulated to say he was surrendering to Russia.

Again, it depends on how those are used. Already, the Republican Party released a 2024 campaign ad featuring AI generated imagery, but the ad included a disclaimer indicating it featured AI. Florida Governor Ron DeSantis also recently released an ad using AI, which featured a fake image of Donald Trump hugging and kissing Anthony Fauci. While these were not entirely harmless, one could imagine dozens of ways nefarious actors could use deepfakes and other AI generated videos to undermine democracy, i.e. editing a clip of one of the major candidates to say something offensive, etc.

d. Real Content

Finally, one of the greatest risks of AI-generated content flooding the information environment is that it lowers our trust in *real* content. It also makes it easier to dismiss pretty much anything as fake. Already, during a lawsuit about Tesla's self-driving technology, Elon Musk has claimed some of his past statements could be deep fakes. This could very well happen in the political realm, making it increasingly challenging for citizens to hold politicians accountable for their actions.

3.) What Can Be Done?

What can be done to mitigate the challenges to our information environment posed by AI-generated content? There are two main questions to consider:

First, can we identify AI-generated content?

Several companies now offer tools to detect AI-generated text, photos, and videos. As AI advances, these tools should also get better. But if possible, we should try to label AI-generated content during the production process. Several major companies are already working on this through groups such as the Coalition for Content Provenance and Authenticity. The goal is to create a technical standard "for certifying the source and history (or provenance) of media content," using metadata, unalterable watermarks, or other technical systems to label content as real or AI-generated. Another idea, currently under consideration by the Federal Election Commission, is to require disclosures when AI is used in political ads.

There are also some roadblocks for reliably identifying AI-generated content. Right now, most of this content comes from a few large companies, such as OpenAI and Google. Those companies could self-impose certain guidelines or precautions related to content provenance. But other companies, notably Meta, want to make it easier for creators to use AI. Earlier this year Meta open-sourced its LLM, meaning it can run locally on a user's computer, without any restrictions or guardrails in place. Given this dynamic, stricter regulation around who can build and operate these systems, disclosure requirements regarding their use, or standards for outputs may be necessary.

Second, how can we integrate these tools with the rest of the web?

Ideally, AI-generated images or videos should be marked automatically by search engines or on social media platforms. Indeed, Google is already working on such a tool. Likewise, if text detection tools advance, platforms should consider labeling posts that have a high likelihood of being AI-generated.

This level of transparency will be critical to creating a more secure and trustworthy online information environment. In fact, without an approach that requires we *know* what is generated by AI and what is not, the debate around AI could become politicized, just as it has with misinformation.

Today, although the goal of reducing misinformation should be nonpartisan, often the left and right do not agree on what information is false and what is true. While the left calls for labeling or taking down harmful content, the right frames the debate as a witch hunt designed to suppress or censor conservative voices.

Since addressing AI-generated content in any form — removing it, down weighting it in feeds, or attaching labels — requires someone (or an algorithm) to decide what is true, the same opportunities for claims of bias could happen with AI. However, focusing on transparency — simply labeling content as AI-generated or not — could, hopefully, avoid this fight. Otherwise the decisions around how to deploy these tools could be just as politically charged as the content they are designed to identify.

Sincerely,

Zeve Sanderson
Executive Director
NYU's Center for Social Media and Politics

Joshua A. Tucker
Co-Director
NYU's Center for Social Media and Politics

Solomon Messing
Research Associate Professor
NYU's Center for Social Media and Politics

Jonathan Nagler
Co-Director
NYU's Center for Social Media and Politics