THE BROOKINGS INSTITUTION

THE CURRENT: What role is AI playing in election disinformation?

June 25, 2024

Host: Adrianna Pita, Office of Communications, Brookings
Guest: Valerie Wirtschafter, Fellow, Artificial Intelligence and Emerging Technology Initiative, Brookings

---

PITA: You're listening to The Current, part of the Brookings Podcast Network. I'm your host, Adrianna Pita.

The increasing availability of artificial intelligence tools has raised many concerns among voters and policymakers about how AI will affect jobs, copyright issues, privacy, equity, and security. Not least among these are concerns over how generative AI tools could fuel mis- and disinformation in elections, as some 2 billion voters in over 60 countries go to the polls this year.

Leading up to the U.S. elections in November, Brookings aims to bring public attention to consequential policy issues confronting voters and policymakers. You can find explainers, policy briefs, other podcasts and more, plus, sign up for the biweekly email by going to Brookings.edu slash election 24. In today's episode, we're talking to Valerie Wirtschafter about the challenges of AI-fueled mis- and disinformation in elections. Valerie, thanks for talking to us today.

WIRTSCHAFTER: Thanks so much for having me.

PITA: So concerns about misuse of generative AI tools have been building for a while. But mis- and disinformation in elections, of course, has existed and flourished without it. And sometimes when we talk about AI, we talk about how some fears are overblown, or maybe at least premature. So maybe start us off with, with a lay of the land. When we talk about the dangers of AI in elections, what exactly are we talking about and how is it different from what came before?

WIRTSCHAFTER: I think it's a mix of both. We've seen very legitimate concerns about the potential for AI generated content to drown out truthful information, of course. But there's also definitely a recognition that there are risks of overhyping the threat as well. So in the U.S. context, which I've looked a bit at already, but some places as well, we've seen really, I think, limited evidence of widespread adoption of these tools in the information space. They're not exactly necessary, actually, because the information space is already a challenging one. Even absent AI generation, people use decontextualized clips, video, etc. so it's really, you know, it's kind of a, an added bonus, we'll say.

What we're really seeing so far is the potential for AI generated content to actually amplify existing trends. For example, by increasing the persuasiveness, removing some of the grammatical errors, for example, for spear phishing efforts, or allowing for greater personalization. I think it definitely can draw more actors into a space by lowering sort of the technical threshold required for some of these things. You know, language barriers,

things like that, were common challenges of the past, which are now, I think, a little bit more tractable.

We have seen a few high-profile cases. So of course, everybody cites the the Joe Biden robocall in the New Hampshire primaries, trying to get people to not show up for the primaries, to save their votes for the general election. Those were debunked really quickly. That, however, may not always be the case. That's a really high-profile race. That's a race where there's a lot of attention, a lot of eyes watching what's going on. At local level elections, things like that, where media is maybe less prominent --

PITA: Or nonexistent, in some areas.

WIRTSCHAFTER: Yeah, or nonexistent. In countries where the media is not free, right, where there isn't an independent media, I think that is hugely challenging, because they're the ones who ultimately will be able to explore this space and push back on some of these, the generated content that circulates, the real, fully fabricated content using some of these tools. And so, I think it definitely is context-by-context dependent. It's depending on the sort of level of the race, particularly the kind of eyeballs that are watching the space.

And then I think, most crucially, where AI really does matter, and this to me is, I think the, the kind of most important space, is just by being on the radar of the public. And so this is where that balance is really important. Because in some sense, you know, knowing that something could be AI generated gives people permission to deny the truth, potentially, when it is uncomfortable, when it may be harmful to their preferred preferences, when it casts something in doubt that they like. And so you can now have this sort of cover to just be able to brush anything off as AI generated if it's inconvenient. And so I think that by not, you know, maybe not even being there physically, but just existing in public consciousness, it can undermine the credibility of the information space, which I think is a huge challenge.

PITA: So there have already been several big elections this year, including Taiwan, India, Mexico, EU just had their big parliamentary elections. You just cited the example of the New Hampshire primaries. Have we seen any other examples of AI-powered disinformation in any of these elections or others?

WIRTSCHAFTER: Yeah, so I mean, we've definitely seen scattered incidences in Taiwan. We saw some deepfake activity designed to try and scandalize the candidate who ultimately won, drive some wedges in terms of fears about U.S. relationship with Taiwan. In India, we saw deepfake activity around the hypothetical future and the sort of political context, dragging in celebrities for endorsements, things like that, resurrecting dead people, as well, to make endorsements from beloved figures.

One area that I will highlight, because I do think it's an interesting, sort of putting it on the positive side or the flip side of where we did see AI recently in an election was in Belarus, actually, where an AI candidate actually did run for office, despite a ban on opposition candidates. And so I thought that was really interesting because it's a really high-risk environment. The candidate was able to challenge the authoritarian status quo in a way that I think a real opposition candidate wouldn't have been able to necessarily. That candidate’s not real. It can't be identified. It can't be imprisoned, but it can spread a message, right? And so I think that that that's sort of a complicating space in thinking about the way that AI is being deployed in elections, both for the malicious and harmful to

sort of amplify existing disinformation trends, drive those wedges in public opinion, etc., but also maybe in contexts where otherwise there wouldn't be any political debate.

PITA: That's really interesting, yeah. So so let's start talking about what some of the possible solutions are. What are some possible defenses that can be set up for, for disinformation? And maybe, maybe you could distinguish for us some possible defenses specifically for the AI disinformation versus what are some of the broader, disinformation protections that exist out there?

WIRTSCHAFTER: Yeah, so I think it depends on the medium, right? There's generated text, there's generated audio, video, images, and so I think there are challenges to each type of medium, and some of the detection possibilities are different based on the medium. But if we take, like, images, for example, AI generated images, you can probably run them through a tool, like image generator detection tools. They vary in quality tremendously, and some of the better ones are actually not publicly available, which I think is really unfortunate. But they can give you some sort of, looking at the signal in the pixels, some level of predicted probability as to whether an image is generated. There's also things that some of the leading AI labs are doing around watermarking where they're, you know, if we think about from the biggest watermark that sort of slaps a big word or something across an image, these would be maybe like tiny, tiny signals in the pixels, but it allows people to, especially who are, who are more technically sophisticated to dig in and find these signals, to be able to determine if an image has been generated. There's content provenance standards. So, some of the leading tech companies who are producing AI image generators have signed up for these content provenance standards to basically kind of build in what would be the equivalent of a nutrition label, telling you this image was created here, it was edited here, the kind of metadata about the life cycle of the image.

Of course, all these things are flawed in some ways, right? You know, it takes a bit of effort to find that tiny little signal in the watermarking. Maybe it won't be quick enough. I could take a screenshot and remove some of that content provenance information. And so there's all kinds of different approaches, especially on the technical side. They have their drawbacks, they have their benefits. And so I think that all of those have been in play, and we're thinking about similar things for other mediums as well. But images, I think, are among the furthest along.

And then, you know, thinking about real images that are used in broader disinformation efforts, we saw actually, a lot of this, in the early days of the Israel-Gaza conflict. Right, recycled images: those are real images from a time and place, but they were maybe ten years ago, or from a different conflict. And so using things like reverse Google image searches to be able to understand the origins of images when they're taken out of context, I think are other, other kinds of tools that are available too.

PITA: In terms of who's responsible for, for doing this, a lot of this sounds like it's kind of on the end user to go, hmm, let me check this image. Are the responsibilities that lie in the role of government or in the role of the tech companies, you mentioned, like the some of these AI image firms, that it's their responsibility to try and come up with some of these watermarking tools. Are there responsibilities that lie with the media and how they report things and try to fact check? How does that responsibility disperse?

WIRTSCHAFTER: Yeah, I mean, I think that, you know, there's tons of responsibility to go around. Some people, I think, or some groups are more active in this

space than others. Right now, existing regulatory frameworks are fairly poorly equipped to manage this election-specific challenge tied to AI but may be better equipped for other challenges tied to AI. We've seen a little bit of enforcement, so I think really promising actually was, the FCC immediately stepped up in the aftermath of that Joe Biden robocall. They said, no, this is illegal. This is a spoofing violation. We are going to find you and we are going to prosecute, and I think recently they proposed something like a $6 million fine for the person who was responsible for the call. So definitely those kinds of enforcements, I think are critical. But there's been in recent weeks, some reporting around kind of competition between the FCC and the FEC, the Federal Election Commission, around who, who is in charge, like, whose domain is this? Which I think is a little bit unfortunate.

And then on the legislative side, there's been a lot and lot a lot of conversation, a lot of recognition, that this is an issue. We haven't seen a ton by way of legislation. The Rules Committee in the Senate did move a slate of bills forward around AI and election concerns. It's unclear if, when, and how those bills will get pushed forward. I think there's one of them that is really looking at, thinking about how AI is actually being used in elections. That may move forward, and it does have bipartisan support, broader bipartisan support. Others maybe are less likely to get through.

And so I think in the absence of legislation and legislative action, tech companies have been sort of at the forefront of this. It's a challenge, of course, with respect to self-governance. Thinking about competition, profit incentives, etc., who's going to be the first mover? And who's going to have, you know, the latest flashy toy? And so I do think there are a lot of challenges in, in that space. But, you know, we've seen some collaboration, particularly I mentioned around these content provenance standards. And so, you know, I think that's really important. More collaboration where possible.

We've seen some companies putting in guardrails around what they will generate. So, an example of that -- and this is this is real, we tried this -- is that, you could ask an image generator to generate an image of Justin Trudeau, and it won't do that. It'll say “I cannot generate that image of a real person.” Yet if you ask it to generate somebody who looks quite a bit like the Canadian leader, it will generate an image of Justin Trudeau, basically, or somebody that looks a lot like Justin Trudeau. And so I think that those are, you know, there are a lot of challenges and a lot of issues with some of these guardrails that are in place, but I do think they are important. But they're flawed, much like a lot of the technical solutions in the space.

Where I do think there has been some really, really important work from the tech companies’ side is, we've actually seen some holding back of tools. So OpenAI has a really, really impressive voice cloning tool. Basically it needs 15 seconds of audio from this podcast basically, or any other audio, to be able to generate more things in anyone's voice. And they've been holding that back for, for a lot of safety testing. And I think that's really important, you know, they can say they've done this and they put out a teaser, and congratulations, flag planted. But, you know, in terms of the kind of in-the-wild impact, I think it's really important to better understand that. And then we've seen also transparency reports coming out from some of these tech companies. How is AI actually being used as part of foreign influence operations? And so really giving people a better understanding of scope and purpose and impact, I think are really, really valuable because right now, we don't have as much of a corpus of work in that space.

PITA: You participated in an event earlier this year that look specifically at how some other countries have been doing on establishing guardrails in their elections. Like, I

think they looked at Taiwan because their election was right at the beginning of the year. Any good examples from that that maybe the U.S. space could learn from?

WIRTSCHAFTER: Yeah, so I think, you know, the Taiwan example is really interesting. You know, that's a that's a context that should, should be and was flooded. You know, it's a huge interest for China, in the way that that election could have shaken it out. And so it was a real prime focus for, for a lot of disinformation. And, you know, they built out a huge apparatus of fact checkers, built up nationwide awareness campaigns, really, really focused on trust in media. So sort of like a building up of armor for the population in some respects, and, and they were able to generate consensus around that, which I do think is really important.

Another country that is a little more contested because they have, you know, there's been a lot of conversation about some of the ways that the government has looked at political speech, but is that is Brazil. The Brazilian context, the electoral court has basically implemented a ban on the use of deepfakes around elections. And so I think that is, you know, just kind of like drawing that line is really critical as well. They have very different laws with respect to speech than we do. And so I think that there are there are regulatory challenges, given the country context, but I think those are just a few that have done, you know, whether it's clarity of what is permissible or not, or the hardening of society approach, I think are are really useful.

PITA: Lastly, there's only a little over four months before our elections in November. Are we out of time to make any improvements like you mentioned, trying to move some of these legislative bills up? But do you have any recommendations or priorities for things that could be done to make a difference in this short term before this year's elections?

WIRTSCHAFTER: Yeah, I mean, you know, I think it's hugely, hugely important for conversations like these, both to build awareness of these challenges without kind of creating that overarching sense of doom. So really, the kind of scoping of the challenge, I think is really important. And then armoring people are with knowledge about how and where, especially to get information about voting, how to vote, if there's an issue on voting day, where to check that. Because one of the big concerns is that, you know, there will be a targeted robocall on Election Day that, you know, it's too late to correct that, right? But so people know, wait, just because I saw that thing or I heard that thing and it sounded credible, I should check again.

I think all of that is really, really important because one of the things that I think this AI moment and AI future creates is that, you know, we've had this sort of mantra, however trite it is that seeing is believing, right? And we sort of have to rewire our brains a little bit to know that seeing is not necessarily now believing, right? We have to be a little bit more skeptical in the way people approach information and look for alternative sources. And I think that that is, that's an adjustment that's going to definitely take some time. But certainly, should be starting, is starting, it needs to continue through the election.

I think there's also been some really important efforts to actually equip election administrators with tools, thinking about basic cybersecurity hygiene, particularly given the ways in AI can impact elections around, spear phishing, right. All of that, I think is really important just to kind of build that educational muscle. You know, hopefully we can maybe see a bill around elections, sort of very narrow. But I wouldn't count on it at this point. And so I do think the companies are going to have to continue to play this really vital role, whether it's around the transparency side, thinking about further ways to increase those

barriers, right? Like maybe, you know, the, the candidate example I gave, like thinking about all of the other different sort of workarounds, continuing to iterate in that space, I think is really important. And then fostering the kind of deeper collaboration. Content provenance is great, but, you know, maybe there can be a sort of coalition around if an AI generated image is shared in in one, on one platform, it can enter into a repository that then can get picked up by other platforms. And so there's examples a little bit of this happening around terrorist content. But thinking about other ways to expand those types of collaborations. Maybe that's not a short-term solution, that might be more of a medium-term solution. But I do think there are there are more opportunities to be able to, to collaborate in the in the coming months.

PITA: All right. Well, you had an excellent report covering some more of these details that we'll link to in the show notes, as well as to like the event that I mentioned, and some of our other content around these issues. Valerie, thanks so much for talking to us today.

WIRTSCHAFTER: Thank you for having me.