Deepfakes threaten upcoming elections, but 'responsible AI' could help filter them out before they reach us

June 11 2024, by Shweta Singh

Credit: AI-generated image

Earlier this year, thousands of Democratic voters in New Hampshire received a telephone call ahead of the state primary, urging them to stay home rather than vote.

The call supposedly came from none other than President Joe Biden. But the message was a "deepfake". This term covers videos, pictures, or audio made with artificial intelligence (AI) to appear real, when they are not. The fake Biden call is one of the most high profile examples to date of the critical threat that deepfakes could pose to the democratic process during the current UK election and the upcoming US election.

Deepfake adverts impersonating Prime Minister Rishi Sunak have reportedly reached more than 400,000 people on Facebook, while young voters in key election battlegrounds are being recommended fake videos created by political activists.

But there may be help coming from technology that conforms to a set of principles known as "responsible AI". This tech could detect and filter out fakes in much the same way a spam filter does.

Misinformation has long been an issue during election campaigns, with many media outlets now carrying out "fact checking" exercises on the claims made by rival candidates. But rapid developments of AI—and in particular generative AI—mean the line between true and false, fact and fiction has become increasingly blurred.

This can cause devastating consequences, sowing the seeds of distrust in the political process and swaying election outcomes. If this continues unaddressed, we can forget about a free and fair democratic process. Instead, we will be faced with a new era of AI-influenced elections.

Seeds of distrust

One reason for the rampant spread of these deepfakes is the fact that they are inexpensive and easy to create, requiring literally no prior knowledge of artificial intelligence. All you need is a determination to influence the outcome of an election.

Paid advertising can be used to propagate deepfakes and other sources of misinformation. The Online Safety Act may make it mandatory to remove illegal disinformation once it has been identified (regardless of whether it is AI-generated or not).

But by the time that happens, the seed of distrust has already been sown in the minds of voters, corrupting the information they use to form opinions and make decisions.

Removing deepfakes once they have already been seen by thousands of voters is like applying a sticking plaster to a gaping wound—too little, too late. The purpose of any technology or law aimed at tackling deepfakes should be to prevent the harm altogether.

With this in mind, the US has launched an AI taskforce to delve deeper into ways to regulate AI and deepfakes. Meanwhile, India plans to introduce penalties both for those who create deepfakes and other forms of disinformation, and for platforms that spread it.

Alongside this are regulations imposed by tech firms such as Google and Meta, which require politicians to disclose the use of AI in election adverts. Finally, there are technological solutions to the threat of deepfakes. Seven major tech companies—including OpenAI, Amazon, and Google—will incorporate "watermarks" into their AI content to identify deepfakes.

However, there are several caveats. There is no standard watermark, allowing each company to design their own watermarking technology and making it harder to track deepfakes. The use of watermarks is only a voluntary commitment by tech firms and failure to comply carries no penalty. There are also smart and simple ways to remove the watermark. Take the case of DALL-E, where a quick search reveals the process for removing its watermark.

On top of this, platforms are not the only means of online communication these days. Anyone who is intent on spreading misinformation can easily email deepfakes direct to voters or use less restrictive platforms, such as encrypted messaging apps, as a preferable outlet for dissemination.

Given these limitations, how can we protect our democracies from the threat posed by AI deepfakes? The answer is to use technology to combat a problem that technology has created, by harnessing it to break the transmission cycle of misinformation across the internet, emails, and online chat platforms.

One way to do this is to design and develop a new "responsible AI" mechanism to detect deepfake audio and video at the point of inception. Much like a spam filter, it would remove them from social media feeds and inboxes.

Some 20 leading technology companies, including Adobe, Amazon, Google, IBM, Meta, Microsoft, OpenAI, TikTok, and X have pledged to work together to detect and counter harmful AI content. This combined effort to combat the deceptive use of AI in the 2024 elections is known as the Tech Accord.

But these are first steps. Moving forward, we need responsible AI solutions, which go beyond simply identifying and eliminating deepfakes to finding methods for tracing their origins and ensuring transparency and trust in the news users read.

Developing these solutions is a race against time, with the UK and US already preparing for elections. Every effort should be made to develop and deploy effective counter measures to guard against political deepfakes in time for the US Presidential election later this year.

Given the rate at which AI is progressing, and the tensions that are likely to surround the campaign, it is hard to imagine that we will be able to hold a truly fair and impartial election without them.

Until effective regulations and responsible AI technology are in place to uphold the integrity of information, the old adage that "seeing is believing" no longer holds true. That leaves the current general election in the UK vulnerable to being influenced by AI deepfakes.

Voters must exercise extra caution when viewing any advertisement, text, speech, audio, or video with a political connection to avoid being duped by deepfakes that seek to undermine our democracy.

Provided by The Conversation

Citation: Deepfakes threaten upcoming elections, but 'responsible AI' could help filter them out before they reach us (2024, June 11) retrieved 17 July 2024 from https://techxplore.com/news/2024-06-deepfakes-threaten-upcoming-elections-responsible.html