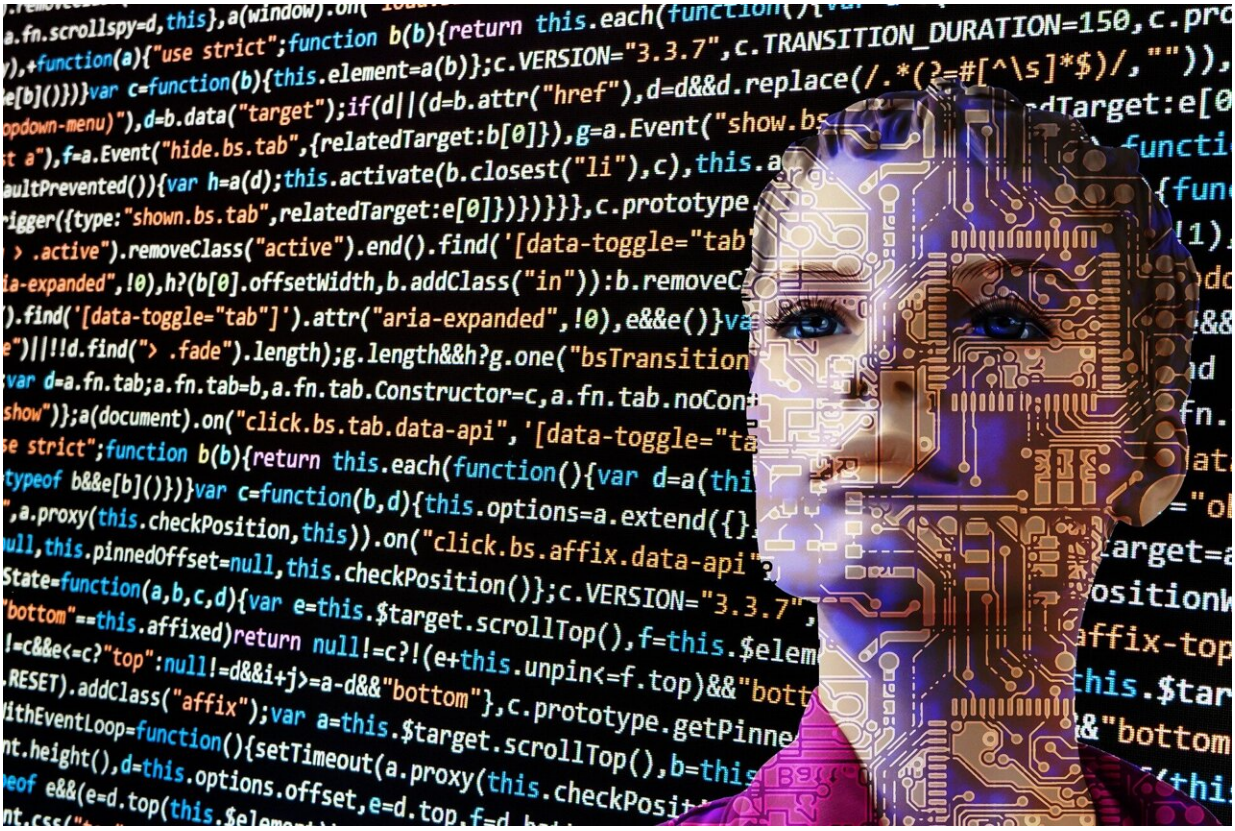
How AI could take over elections—and undermine democracy

June 5 2023, by Archon Fung and Lawrence Lessig



Datasets used to train AI algorithms may underrepresent older people. Credit: Pixabay/CC0 Public Domain

Could organizations use artificial intelligence language models such as ChatGPT to induce voters to behave in specific ways?

Sen. Josh Hawley asked OpenAI CEO Sam Altman this question in a May 16, 2023, U.S. Senate hearing on artificial intelligence. Altman replied that he was indeed concerned that some people might use language models to manipulate, persuade and engage in one-on-one interactions with voters.

Altman did not elaborate, but he might have had something like this scenario in mind. Imagine that soon, political technologists develop a machine called Clogger—a political campaign in a black box. Clogger relentlessly pursues just one objective: to maximize the chances that its candidate—the campaign that buys the services of Clogger Inc.—prevails in an election.

While platforms like Facebook, Twitter and YouTube use forms of AI to get users to spend more time on their sites, Clogger's AI would have a different objective: to change people's voting behavior.

How Clogger would work

As a political scientist and a legal scholar who study the intersection of technology and democracy, we believe that something like Clogger could use automation to dramatically increase the scale and potentially the effectiveness of behavior manipulation and microtargeting techniques that political campaigns have used since the early 2000s. Just as advertisers use your browsing and social media history to individually target commercial and political ads now, Clogger would pay attention to you—and hundreds of millions of other voters—individually.

It would offer three advances over the current state-of-the-art algorithmic behavior manipulation. First, its language model would generate messages—texts, social media and email, perhaps including images and videos—tailored to you personally. Whereas advertisers strategically place a relatively small number of ads, language models

such as ChatGPT can generate countless unique messages for you personally—and millions for others—over the course of a campaign.

Second, Clogger would use a technique called reinforcement learning to generate a succession of messages that become increasingly more likely to change your vote. Reinforcement learning is a machine-learning, trial-and-error approach in which the computer takes actions and gets feedback about which work better in order to learn how to accomplish an objective. Machines that can play Go, Chess and many video games better than any human have used reinforcement learning.

Third, over the course of a campaign, Clogger's messages could evolve in order to take into account your responses to the machine's prior dispatches and what it has learned about changing others' minds. Clogger would be able to carry on dynamic "conversations" with you—and millions of other people—over time. Clogger's messages would be similar to ads that follow you across different websites and social media.

The nature of AI

Three more features—or bugs—are worth noting.

First, the messages that Clogger sends may or may not be political in content. The machine's only goal is to maximize vote share, and it would likely devise strategies for achieving this goal that no human campaigner would have thought of.

One possibility is sending likely opponent voters information about nonpolitical passions that they have in sports or entertainment to bury the political messaging they receive. Another possibility is sending off-putting messages—for example incontinence advertisements—timed to coincide with opponents' messaging. And another is manipulating voters'

social media friend groups to give the sense that their social circles support its candidate.

Second, Clogger has no regard for truth. Indeed, it has no way of knowing what is true or false. Language model "hallucinations" are not a problem for this machine because its objective is to change your vote, not to provide accurate information.

Third, because it is a black box type of artificial intelligence, people would have no way to know what strategies it uses.

Clogocracy

If the Republican presidential campaign were to deploy Clogger in 2024, the Democratic campaign would likely be compelled to respond in kind, perhaps with a similar machine. Call it Dogger. If the campaign managers thought that these machines were effective, the presidential contest might well come down to Clogger vs. Dogger, and the winner would be the client of the more effective machine.

Political scientists and pundits would have much to say about why one or the other AI prevailed, but likely no one would really know. The president will have been elected not because his or her policy proposals or political ideas persuaded more Americans, but because he or she had the more effective AI. The content that won the day would have come from an AI focused solely on victory, with no political ideas of its own, rather than from candidates or parties.

In this very important sense, a machine would have won the election rather than a person. The election would no longer be democratic, even though all of the ordinary activities of democracy—the speeches, the ads, the messages, the voting and the counting of votes—will have occurred.

The AI-elected president could then go one of two ways. He or she could use the mantle of election to pursue Republican or Democratic party policies. But because the party ideas may have had little to do with why people voted the way that they did—Clogger and Dogger don't care about policy views—the president's actions would not necessarily reflect the will of the voters. Voters would have been manipulated by the AI rather than freely choosing their political leaders and policies.

Another path is for the president to pursue the messages, behaviors and policies that the machine predicts will maximize the chances of reelection. On this path, the president would have no particular platform or agenda beyond maintaining power. The president's actions, guided by Clogger, would be those most likely to manipulate voters rather than serve their genuine interests or even the president's own ideology.

Avoiding Clogocracy

It would be possible to avoid AI election manipulation if candidates, campaigns and consultants all forswore the use of such political AI. We believe that is unlikely. If politically effective black boxes were developed, the temptation to use them would be almost irresistible. Indeed, political consultants might well see using these tools as required by their professional responsibility to help their candidates win. And once one candidate uses such an effective tool, the opponents could hardly be expected to resist by disarming unilaterally.

Enhanced privacy protection would help. Clogger would depend on access to vast amounts of personal data in order to target individuals, craft messages tailored to persuade or manipulate them, and track and retarget them over the course of a campaign. Every bit of that information that companies or policymakers deny the machine would make it less effective.

Another solution lies with elections commissions. They could try to ban or severely regulate these machines. There's a fierce debate about whether such "replicant" speech, even if it's political in nature, can be regulated. The U.S.'s extreme free speech tradition leads many leading academics to say it cannot.

But there is no reason to automatically extend the First Amendment's protection to the product of these machines. The nation might well choose to give machines rights, but that should be a decision grounded in the challenges of today, not the misplaced assumption that James Madison's views in 1789 were intended to apply to AI.

European Union regulators are moving in this direction. Policymakers revised the European Parliament's draft of its Artificial Intelligence Act to designate "AI systems to influence voters in campaigns" as "high risk" and subject to regulatory scrutiny.

One constitutionally safer, if smaller, step, already adopted in part by European internet regulators and in California, is to prohibit bots from passing themselves off as people. For example, regulation might require that campaign messages come with disclaimers when the content they contain is generated by machines rather than humans.

This would be like the advertising disclaimer requirements—"Paid for by the Sam Jones for Congress Committee"—but modified to reflect its AI origin: "This AI-generated ad was paid for by the Sam Jones for Congress Committee." A stronger version could require: "This AI-generated message is being sent to you by the Sam Jones for Congress Committee because Clogger has predicted that doing so will increase your chances of voting for Sam Jones by 0.0002%." At the very least, we believe voters deserve to know when it is a bot speaking to them, and they should know why, as well.

The possibility of a system like Clogger shows that the path toward human collective disempowerment may not require some superhuman artificial general intelligence. It might just require overeager campaigners and consultants who have powerful new tools that can effectively push millions of people's many buttons.

Provided by The Conversation