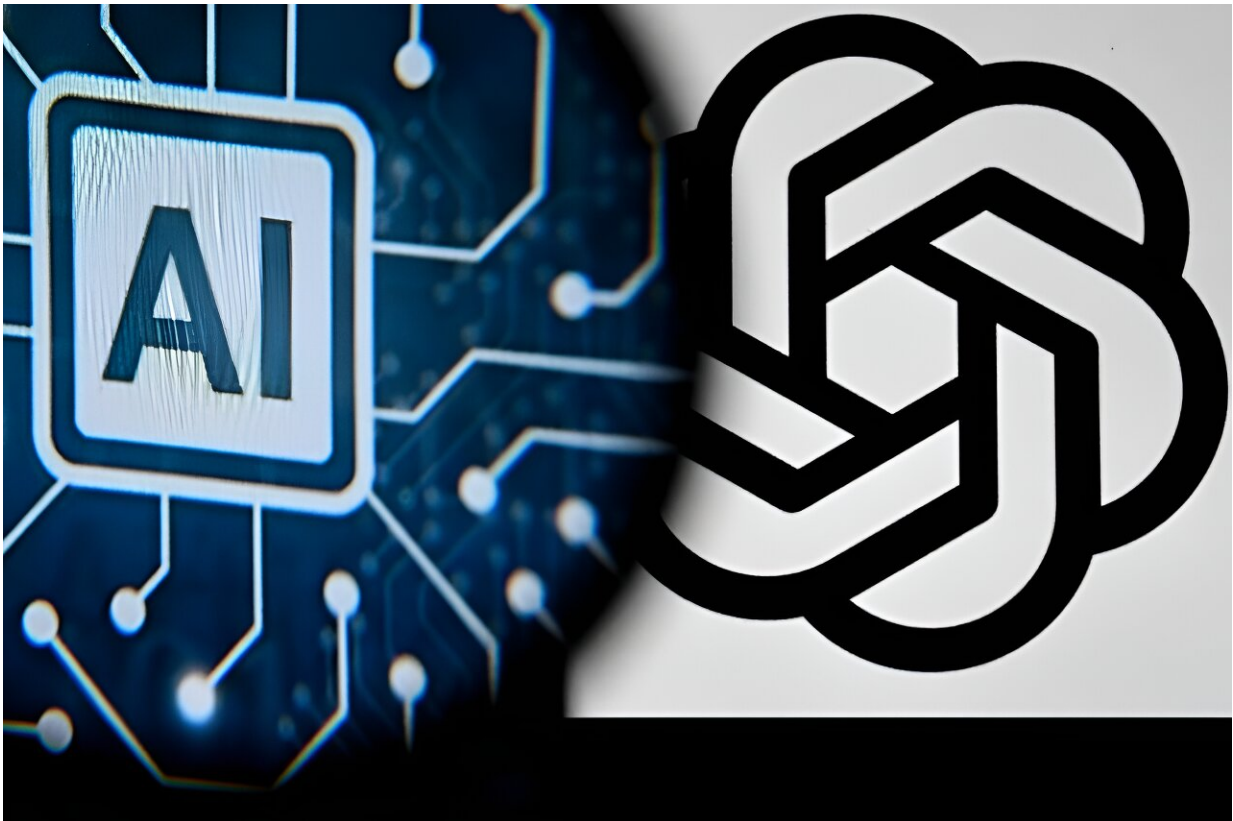# OpenAI to launch anti-disinformation tools for 2024 elections

January 16 2024

OpenAI's ChatGPT is one of the most powerful generative AI tools available to the public.

ChatGPT maker OpenAI has said it will introduce tools to combat disinformation ahead of the dozens of elections this year in countries

that are home to half the world's population.

The explosive success of text generator ChatGPT spurred a global artificial intelligence revolution but also triggered warnings that such tools could flood the internet with disinformation and sway voters.

With elections due this year in countries including the United States, India and Britain, OpenAI said Monday it will not allow its tech—including ChatGPT and the image generator DALL-E 3—to be used for political campaigns.

"We want to make sure our technology is not used in a way that could undermine" the democratic process, OpenAI said in a blog post.

"We're still working to understand how effective our tools might be for personalized persuasion," it added.

"Until we know more, we don't allow people to build applications for political campaigning and lobbying."

AI-driven disinformation and misinformation are the biggest short-term global risks and could undermine newly elected governments in major economies, the World Economic Forum warned in a report released last week.

Fears over election disinformation began years ago, but the public availability of potent AI text and image generators has boosted the threat, experts say, especially if users cannot easily tell if the content they see is fake or manipulated.

OpenAI said Monday it was working on tools that would attach reliable attribution to text generated by ChatGPT, and also give users the ability to detect if an image was created using DALL-E 3.

"Early this year, we will implement the Coalition for Content Provenance and Authenticity's digital credentials—an approach that encodes details about the content's provenance using cryptography," the company said.

The coalition, also known as C2PA, aims to improve methods for identifying and tracing digital content. Its members include Microsoft, Sony, Adobe and Japanese imaging firms Nikon and Canon.

'Guardrails'

OpenAI said ChatGPT, when asked procedural questions about US elections such as where to vote, will direct users to authoritative websites.

"Lessons from this work will inform our approach in other countries and regions," the company said.

It added that DALL-E 3 has "guardrails" that prevent users from generating images of real people, including candidates.

OpenAI's announcement follows steps revealed last year by US tech giants Google and Facebook parent Meta to limit election interference, especially through the use of AI.

AFP has previously debunked deepfakes—doctored videos—of US President Joe Biden announcing a military draft and former secretary of state Hillary Clinton endorsing Florida Governor Ron DeSantis for president.

Doctored footage and audio of politicians were circulated on social media ahead of the presidential election this month in Taiwan, AFP Fact Check found.

While much of this content is low-quality and it is not immediately clear if it is created with AI apps, experts say disinformation is fueling a crisis of trust in political institutions.