



January 15, 2024

How OpenAI is approaching 2024 worldwide elections

We're working to prevent abuse, provide transparency on AI-generated content, and improve access to accurate voting information.

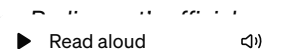


Illustration: Justin Jay Wang × DALL·E

Updated on May 14, 2024

As part of our ongoing work to promote transparency around AI content during this important election year, we recently began providing researchers with early access to a new tool that can help identify images created by OpenAI's DALL·E 3. We also joined the Steering Committee of C2PA – the Coalition for Content Provenance and Authenticity. C2PA is a widely used standard for digital content certification, developed and adopted by a wide range of actors including software companies, camera manufacturers, and online platforms.

Building on our efforts to direct people to authoritative sources of information about voting in the U.S., we've introduced a new experience ahead of the 2024 election for the European Parliament. ChatGPT now directs users to the Europe





In addition to the steps we're taking at OpenAI, we believe there is an important role for governments. Today we are endorsing the "Protect Elections from Deceptive AI Act," a bi-partisan bill proposed by Senators Klobuchar, Hawley, Coons, Collins, Ricketts, and Bennet in the United States Senate. The bill would ban the distribution of deceptive AI-generated audio, images, or video relating to federal candidates in political advertising, while including important exceptions to protect First Amendment rights. We do not want our technology – or any AI technology – to be used to deceive voters and we believe this legislation represents an important step to addressing this challenge in the context of political advertising.

Protecting the integrity of elections requires collaboration from every corner of the democratic process, and we want to make sure our technology is not used in a way that could undermine this process.

Our tools empower people to improve their daily lives and solve complex problems—from using AI to enhance state services to simplifying medical forms for patients.

We want to make sure that our AI systems are built, deployed, and used safely. Like any new technology, these tools come with benefits and challenges. They are also unprecedented, and we will keep evolving our approach as we learn more about how our tools are used.

As we prepare for elections in 2024 across the world's largest democracies, our approach is to continue our platform safety work by elevating accurate voting information, enforcing measured policies, and improving transparency. We have a cross-functional effort dedicated to election work, bringing together expertise from our safety systems, threat intelligence, legal, engineering, and policy teams to quickly investigate and address potential abuse.

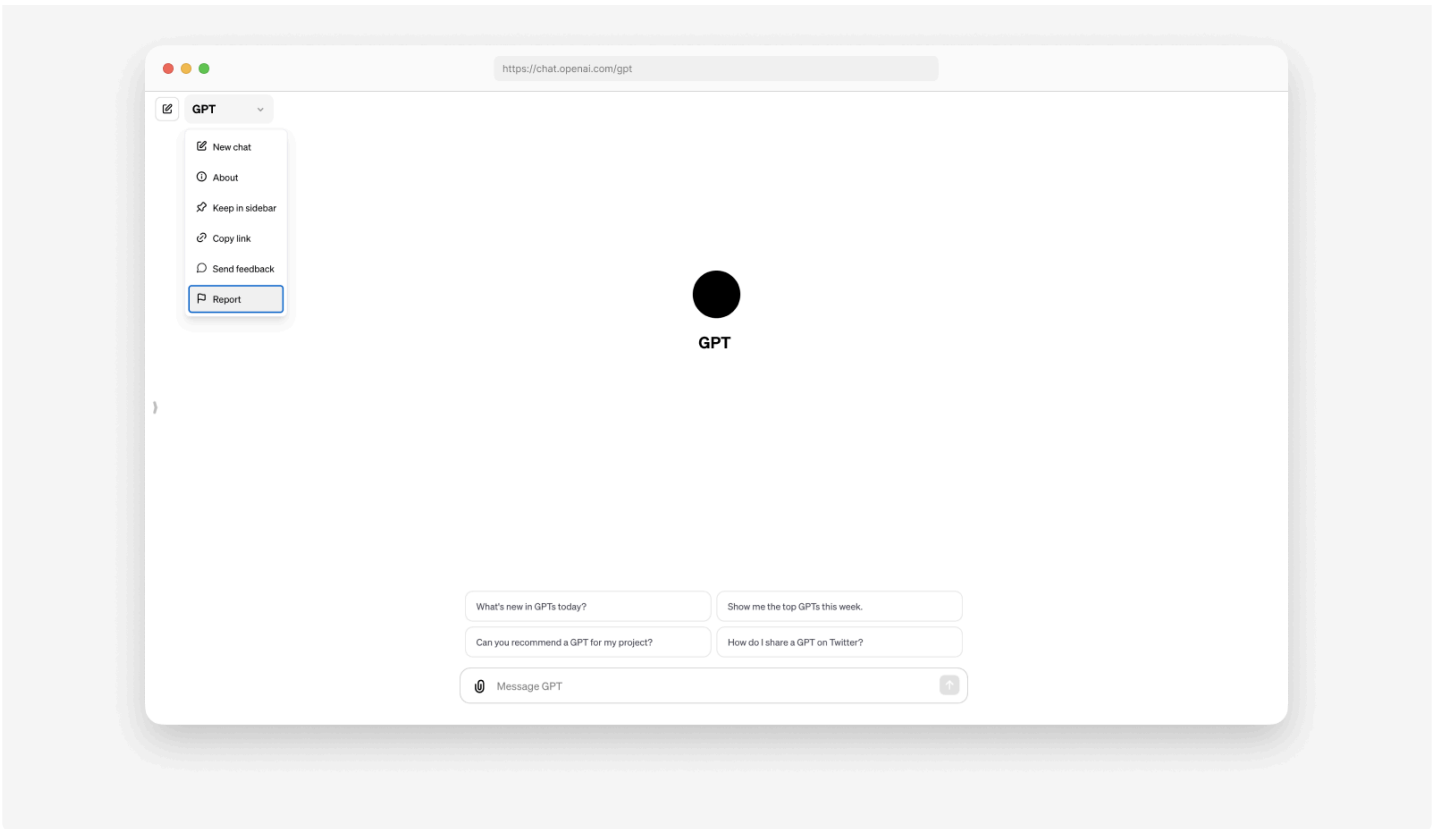
The following are key initiatives our teams are investing in to prepare for elections this year:

Preventing abuse

We expect and aim for people to use our tools safely and responsibly, and elections are no different. We work to anticipate and prevent relevant abuse—such as misleading “deepfakes”, scaled influence operations, or chatbots impersonating candidates. Prior to releasing new systems, we red team them, engage users and external partners for feedback, and build safety mitigations to reduce the potential for harm. For years, we've been iterating on tools to improve factual accuracy, reduce bias, and decline certain requests. These tools provide a strong foundation for our work around election integrity. For instance, DALL-E has guardrails to decline requests that ask for image generation of real people, including candidates.

We regularly refine our Usage Policies for ChatGPT and the API as we learn more about how people use or attempt to abuse our technology. A few to highlight for elections:

- We're still working to understand how effective our tools might be for personalized persuasion. Until we know more, we don't allow people to build applications for political campaigning and lobbying.
- People want to know and trust that they are interacting with a real person, business, or government. For that reason, we don't allow builders to create chatbots that pretend to be real people (e.g., candidates) or institutions (e.g., local government).
- We don't allow applications that deter people from participation in democratic processes—for example, misrepresenting voting processes and qualifications (e.g., when, where, or who is eligible to vote) or that discourage voting (e.g., claiming a vote is meaningless).
- With our new GPTs, users can report potential violations to us.



With our new GPTs, users can report potential violations to us.

Transparency around AI-generated content

Better transparency around image provenance—including the ability to detect which tools were used to produce an image—can empower voters to assess an image with trust and confidence in how it was made. We're working on several provenance efforts. We implemented the [Coalition for Content Provenance and Authenticity's digital credentials](#)—an approach that encodes details about the content's provenance using cryptography—for [images](#) generated by DALL-E 3.

We are also experimenting with a provenance classifier, a new tool for detecting images generated by DALL-E. Our internal testing has shown promising early results, even where images have been subject to common types of modifications. We plan to soon make it available to our first group of testers—including journalists, platforms, and researchers—for feedback.

Finally, ChatGPT is increasingly integrating with existing sources of information—for example, users will start to get access to real-time news reporting globally, including attribution and links. Transparency around the origin of information and balance in news sources can help voters better assess information and decide for themselves what they can trust.

Improving access to authoritative voting information

In the United States, we are working with the [National Association of Secretaries of State \(NASS\)](#), the nation's oldest nonpartisan professional organization for public officials. ChatGPT will direct users to [CanIVote.org](#), the authoritative website on US voting information, when asked certain procedural election related questions—for example, where to vote. Lessons from this work will inform our approach in other countries and regions.



Safety & Alignment

Author

[OpenAI](#)

Related articles

[View all Safety & Alignment >](#)

