



U.S. DEPARTMENT *of* STATE

Democratic Roadmap:

# Building Civic Resilience to the Global Digital Information Manipulation Challenge



March 2024

# **Democratic Roadmap: Building Civic Resilience to the Global Digital Information Manipulation Challenge**

March 2024

*"For all of the challenges that these disruptive technologies present, no system of government is better equipped to drive the forces that they represent in improving our people's lives than democracy. We excel at innovation. We're nimble. We encourage a multiplicity of voices and perspectives to find solutions. We let the best ideas rise to the top, rather than assuming that the best ideas come from the top. We believe our people have a vital role to play in the ongoing process of making our system better, of fixing its flaws. We embrace vigorous and open debate within and across our democracies."*

**—Secretary Antony Blinken, the Second Summit for Democracy**

## Purpose

The United States is advancing a global response to tackle the information integrity challenge in ways that are consistent with democratic values, freedom of expression, and international human rights law. As part of that effort, the Bureau of Cyberspace and Digital Policy (CDP) at the U.S. Department of State has developed a democratic roadmap for global policymakers, civil society, and the private sector to:

- **Step 1:** Highlight the importance of the digital information manipulation challenge as a threat to the functionality and vitality of society
- **Step 2:** Recognize that building information integrity can be consistent with freedom of opinion and expression
- **Step 3:** Reinforce private sector digital platforms' ability to strengthen civic resilience to promote information integrity
- **Step 4:** Prioritize efforts to address generative AI (GAI), particularly in the context of global 2024 elections

## Preamble

Democracy depends on open, free, public debate and consistent access to diverse sources of fact-based information. These features are essential for citizens to inform their opinions and exercise their human rights, including freedom of expression and freedom of peaceful assembly and association, and the right to vote. Citizens' access to and trust in accurate information is necessary to effectively participate in open, democratic societies. The functionality and vitality of democratic societies, as well as the ability to address transnational challenges such as climate change and pandemics, depend on the integrity of the information realm.

An evolving lexicon has been developed to describe different dimensions of the information manipulation challenge – foreign information manipulation, including disinformation and propaganda, misinformation, mal-information, and “fake news.” The essential problem all of these terms point to, but don’t fully capture, is that the erosion of trust and integrity in the digital information realm has become a threat to the future of democracy. While the term “disinformation,” often defined as false and intentionally misleading content, is often used as shorthand for the entire spectrum of information challenges, we are using the term “digital information manipulation” to capture the range of challenges related to the erosion of integrity in the digital information realm. In a rapidly evolving information ecosystem, the global response to digital information manipulation has not matched the gravity of the threat, and the consequences are playing out, very visibly, as a dimension of the democratic recession around the world.

Challenges to information integrity are not new, but digitization of the information realm has exponentially increased the speed, scale, and reach of all forms of misleading and manipulated content. Digital platforms have had a dramatic globalizing effect on the information realm and provided new modes of amplification. State and non-state actors alike now have the means and mechanisms to instantaneously manipulate civic discourse within communities and across borders, eroding the quality of the information citizens depend on to form opinions and voters use to make decisions in democratic societies around the world.

Many democracies are struggling to address the digital information manipulation challenge without falling into the trap of unintentionally undermining freedom of expression. Some nations have opted for content-based regulations that criminalize disinformation or manipulated content in ways that are not consistent with their international human rights law obligations regarding restrictions on freedom of expression.

Meanwhile authoritarian regimes leverage concerns about “disinformation” and “fake news” online as a guise when passing regulations intended to stifle political dissent and censor individuals, including human rights defenders. Some authoritarian states, most notably the People’s Republic of China (PRC), have embarked on a strategy to advance their global “discourse power,” including by exploiting global digital platforms.

New capacities to manipulate civic discourse across borders via global information platforms have exposed asymmetries between open democratic societies and closed authoritarian ones, where the digital information realm is closely controlled by the state. Democracies will need a more sophisticated understanding of how to protect individuals’ freedom of expression online to showcase democratic resilience.

Simultaneously, GAI has burst onto the scene and into the global digital information realm. While AI and algorithmic filtering and promotion of content have long been features of global digital platforms, new GAI-enabled tools are distinctive in that they provide the means to *generate* seemingly credible content for both constructive and malign purposes. GAI-created content risks further impacting the information realm, making it harder for citizens to assess the validity of sources they rely on to form opinions. Importantly, GAI also has the potential to enhance the ability to detect digitally manipulated content. That said, in the short term, it is not hard to imagine that GAI-enabled tools may favor malign actors who seek to erode the quality of civic discourse within democracies. Simply put, GAI could turbocharge the digital information manipulation challenge for democracies.

We are at an inflection point: globally, 2024 could be one of the most consequential election years in history; national elections are anticipated in 40 countries representing about 40 percent of the world’s population. Citizens need accurate sources of information to form opinions and participate in free and fair elections. Without reliable information, citizens cannot participate in government in an informed manner. Citizens

also need skills to critically assess the digital information that will influence the exercise of their fundamental freedoms. Democratic stakeholders must continue to actively engage to mitigate potentially harmful effects of GAI on the digital information realm in the context of our elections.

Over the past decade, a great deal of time and energy has been invested in monitoring how social media, encrypted chat apps, and other digital information platforms have been misused and exploited to undermine the quality of civic discourse within democracies. We must now focus on developing an affirmative, action-oriented plan to build civic resilience to digital information manipulation and to strengthen the integrity of the digital information realm in support of democracy. This must start with greater collective understanding of how digital information platforms work, how algorithms shape users' information context, and how to prepare for the potential onslaught of GAI-generated content.

This roadmap offers affirmative steps to strengthen civic resilience to digital information manipulation.

## **Step 1: Highlight the Importance of the Digital Information Manipulation Challenge as a Threat to the Functionality and Vitality of Society**

The first step in addressing the digital information manipulation challenge is to emphasize that it poses a threat to the future of global democracy. Degradation in the integrity of the global digital information realm has already eroded citizen trust in democratic values and institutions around the world. Digital information manipulation has also impaired a shared sense of reality, making responding to global threats such as climate change or pandemics more complex and difficult to address.

Authoritarians and other malign actors have capitalized on digital platforms to shape discourse internally and across borders, and to exacerbate social divisions and increase polarization within and across democratic societies.

If these trends continue, the future of democracy will be further undermined, which itself would constitute a grave national security threat. Building civic resilience to digital information manipulation is a national security and foreign policy imperative.

## Step 2: Recognize that Building Information Integrity Can Be Consistent with Freedom of Opinion and Expression

Digital information manipulation should be recognized as a threat to the exercise of freedom of opinion and expression, rather than as an unfortunate byproduct of protecting free expression. The right to freedom of opinion and expression is set forth in both Article 19 of the Universal Declaration of Human Rights (UDHR) and Article 19 of the International Covenant on Civil and Political Rights (ICCPR).

### Article 19, UDHR:

*“Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.”*

### Article 19, ICCPR:

*“1) Everyone shall have the right to hold opinion without interference. 2) Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.”*



A key step in addressing the digital information manipulation challenge is to recognize that the tension between a commitment to protect freedom of expression and the goal of preventing the spread of “harmful” content is a false dichotomy. A binary framing misses the real issue: digital information manipulation undermines one’s ability to exercise the freedom to seek and receive information necessary to form opinions.

Access to fact-based information, especially in the context of elections, is essential for citizens to form opinions and exercise their rights, including their right to vote. We must understand all aspects of freedom of opinion and expression – not only the freedom to impart information, but also the freedom to seek and receive information, as well as the freedom to hold opinions – especially in the digital context. All of these elements are fundamental to citizens in democratic society.

Freedom of expression is further impacted when technology is misused to facilitate gender-based violence to undermine the full and active participation of women, girls, and LGBTQI+ persons in political and other civic spaces. When more than half of a national community is forced to retreat from participation in civic engagement online, it undermines full participation in democratic processes and debate.

### **Step 3: Reinforce Private Sector Digital Platforms’ Ability to Strengthen Civic Resilience to Promote Information Integrity**

Private sector platform efforts to respect freedom of expression can entail more than a single-minded focus on platform “users’” ability to express themselves or impart information. It is important to acknowledge that platform community guidelines and other policies play a role in shaping civic discourse. Digital platforms can play a role in building healthy information ecosystems in the societies in which they operate.

Private sector platforms also can contribute to building civic resilience to digital information manipulation. An essential first step is enhanced platform transparency.

Lack of understanding of how their digital information feeds are shaped has made citizens less resilient to digital information manipulation. The addition of GAI content, including deep/cheap fakes, audio fakes, and more, has magnified this challenge.

Supported by policies and coordination with governments and civil society, where appropriate, private sector digital platforms can enhance transparency and communication about the following:

1. Algorithmic promotion and demotion of content
2. Privacy policies
3. Use and sharing of user data
4. Political advertising
5. Labeling of GAI content

## **Step 4: Prioritize Efforts to Address GAI, Particularly in the Context of Global 2024 Elections**

Generative AI (GAI) has captured the world's attention and democratic stakeholders must capitalize on that attention to quicken the pace of progress in building civic resilience to digital information manipulation. GAI-enabled tools have been released to the public and "into the wild" without a full understanding of how these new tools might be applied, which could have significant implications for the quality and integrity of the information realm. New capacities to generate synthetic content are now widely available to everyone, for both beneficial and malign purposes. GAI will allow bad actors to more effectively micro-target and automate the creation of misleading information at scale, in ways that are cheaper and faster than pre-existing means. Citizens will need to quickly develop new skills to adapt to the potential effects of GAI.

The blockbuster 2024 election year context has inspired new urgency around the responsibility to mitigate risks posed by GAI to freedom of expression and election integrity. GAI-manipulated content already has been deployed in Slovakia, Pakistan, Indonesia, and Bangladesh, demonstrating the risk of AI-generated disinformation to dramatically alter civic discourse, public opinion, and narratives around the integrity of democratic elections. It is not hard to imagine the global digital information realm flooded with deepfakes and synthetic media, designed to deceive citizens about candidates or impact voter turnout with misleading information about election processes. GAI could fuel public concern about election fraud.

A big part of the solution to mitigating GAI-related risk is incentivizing private sector investment in beneficial applications of GAI that build civic resilience to digital information manipulation and enhance the quality of the digital information realm. For example, much greater investment must be made in the development and use of GAI-enabled methods to detect AI-generated content. GAI also could support higher quality civic debate online, by alerting users when they are using vitriolic and incendiary language.

## **Best Practices for Mitigating Risks of GAI in the Context of 2024 Global Elections**

### **What Can Governments Do?**

- Prioritize investment in initiatives that strengthen civic resilience to digital information manipulation and improve digital, media and information literacy
- Support multistakeholder, whole-of-society approaches to build information integrity

- Develop the capacity of local election officials to become trusted sources for citizens about election administration information and communication
- Ensure that accurate information about elections is easily accessible to all citizens
- Promote fact-based public education about how GAI works and how it affects information integrity to increase citizen understanding
- Support research to monitor and track the impacts of information manipulation and GAI on global elections
- Encourage digital platform governance that is transparent and consistent with protection of freedom of opinion and expression and the right to vote

## What Can Private Sector Companies Do?

- Develop and make available tools for identifying/labeling AI-generated content and detecting manipulated content.
- Engage in multistakeholder table-top and red teaming exercises to anticipate potential information manipulation relevant to elections.
- Evaluate models for risks they may present regarding AI content to better understand and mitigate potential for abuse
- Involve users and external stakeholders to develop safeguards for election-related information manipulation
- Monitor and publicly report on the coordinated misuse of chatbots and other GAI tools to increase polarization, encourage violence, or to discourage participation in elections

- Monitor and report on the use and impact of GAI created deepfakes which aim to undermine democratic processes or human rights and make that information public
- Invest in systematic public messaging about how GAI-enabled tools are being used around elections
- Exchange best practices and explore sharing information across the industry, including in the context of mitigating risks posed by deceptive AI-generated election content

## What Can Journalists Do?

- Engage in table-top and red-teaming exercises to anticipate and prepare for significant election-related information manipulation including deepfakes of candidates
- Support AI literacy with coverage of AI's impacts on elections, and encourage skepticism about election manipulation claims that are not supported by evidence
- Partner with civil society to evaluate the provenance of widely distributed election-related content

## What Can Civil Society, Researchers, and Academics Do?

- Study the effects of labeling AI-generated content on users'/citizens' trust in the information realm and democratic processes
- Work with industry to develop AI products that can support fact checkers with research and verification tasks, lowering costs and increasing speed

- Participate in private sector development of safeguards for election-related information manipulation
- Educate citizens on civic processes and engagement and to beware of potential use of microtargeting and deepfakes to manipulate voter opinions and trust

## Additional Resources

- [White House Voluntary Commitments Ensuring Safe, Secure, and Trustworthy AI \(July 2023\)](#)
- [Hiroshima Process International Guiding Principles and Code of Conduct for Organizations Developing Advanced AI Systems \(October 2023\)](#)
- [Tech Accord to Combat Deceptive Use of AI in 2024 Elections \(February 2024\)](#)

## Conclusion

As articulated in the White House Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, the United States must lead key global conversations and collaborations to ensure that AI benefits the whole world, rather than exacerbating inequities, threatening human rights, and causing other harms.

As part of the same broad effort, the United States is committed to leading efforts to build civic resilience to digital information manipulation and strengthening information integrity as intrinsic to our core national interests, recognizing that tech innovation driven within the United States has been enabled by our commitment to human rights, including freedom of expression.

Degradation in integrity of the digital information realm has already significantly eroded citizen trust in democratic institutions and elections around the world. If this ominous

trend continues, the future of democracy is at risk. This in turn will constitute a grave national security threat. It's time for democracies to come together more effectively to take on the information integrity challenge together, as called for in the Global Declaration on Information Integrity Online. In the context of the blockbuster year of elections 2024, we must find effective, practical ways to address the information integrity challenge.



U.S. DEPARTMENT *of* STATE

**Democratic Roadmap:  
Building Civic Resilience to the Global Digital Information  
Manipulation Challenge**



@StateCDP



@StateCDP