Strengthening Democracy in the Digital Age

Democratic Defense and Resilience in the Face of AI-Enabled Threats

Amy Larsen
Director of Global Field Engagement & Strategic Projects, Democracy Forward, Microsoft

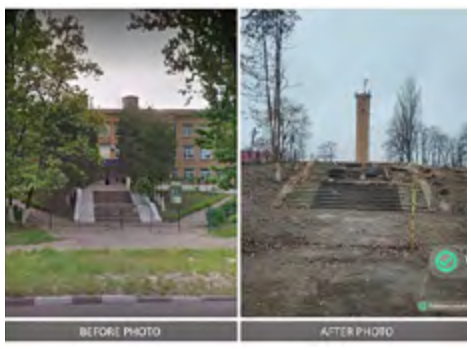
An impromptu scroll of your news feed can quickly remind you that democracy is under threat in the digital age. New technologies pose unprecedented challenges to the integrity of information and elections, which are cornerstones of democratic societies. From deepfakes and disinformation to cyberattacks and hacking, the digital landscape is rife with risks that can undermine public trust, distort political discourse, and manipulate the information environment leading up to and during an election. How can democracies protect themselves from such threats, and what is the proper role for technology companies to play in strengthening democratic resilience in the age of AI?

Nation state threat actors are also upskilling, leading to the potential for increasingly sophisticated and scalable cyberattacks and disinformation campaigns.

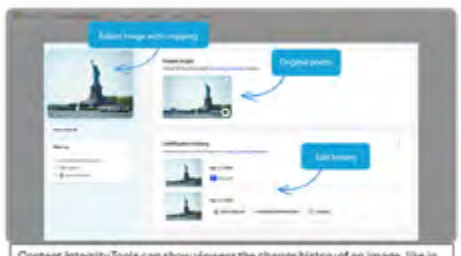
During this Vote-a-Rama year of elections in which over two billion people have the opportunity to vote in nationwide elections, Microsoft is committed to enhancing the security and resilience of elections by developing innovative solutions and tools for voters, candidates, political campaigns, and election authorities. In February 2024 at the Munich Security Conference, Microsoft joined over two dozen technology companies in signing on to the Tech Accord to Combat Deceptive Use of AI in 2024 Elections. This voluntary pledge by technology companies focuses on combatting deepfakes that take the form of video, audio, and images attempting to fake or alter the appearance, voice, or actions of political candidates, campaigns, or election officials. The Tech Accord commitments are intended to make it harder for nefarious actors to use legitimate tools to create deepfakes, to bring the tech sector together to detect and respond to deepfakes in elections, in part by enabling political candidates to immediately report concerns of deepfakes of themselves, and by advancing transparency and societal resilience.

Over the past several months, Microsoft has worked to rise to the occasion and implement our Tech Accord commitments to combat AI-enabled deepfakes and disinformation across various regions in order to safeguard elections taking place around the world. In the United Kingdom, European Union, India, South Africa, France, Belgium, and the United States, we launched initiatives like the "Check. Recheck. Vote." deepfakes public awareness campaign to empower the public to critically evaluate digital election content and sources, and to build their AI literacy skills through tools like the Real or Not deepfake quiz. We also launched Election Communications Hubs to support election authorities with direct access to our security and support teams leading up to an election. These build on existing security programs like Azure for Elections offerings available to state and local election agencies and their partners in the United States. In the first half of this Vote-a-Rama year, Microsoft has also undertaken over 110 deepfakes training sessions for political stakeholders in 20 countries reaching almost 3,300 participants, while our deepfakes public awareness campaigns across EU, UK, and U.S. have helped educate hundreds of millions of people.

We have also expanded our Content Integrity tools available to political candidates in the U.S. to those in the EU and UK, as well as to newsrooms globally, which allow them to add secure "content credentials" to digital content showing who created the content, where and when it was created, whether it was created by AI, and if it has been edited or tampered with since its creation. This tool leverages the open-source industry standard published by the Coalition for Content Provenance and Authenticity (C2PA), of which Microsoft is a founding member. This is the same open standard that underpinned Project Providence, a partnership Microsoft and TruePic developed in Ukraine to document the destruction of cultural heritage sites by allowing for end-to-end operability between the capture, storage, and display of images, ensuring the provenance of such images.



These Content Integrity tools exist alongside our longstanding cybersecurity protections for high-risk, highly targeted stakeholders of democracy. For example, AccountGuard, a threat detection and notification service that protects over 5.4 million inboxes in 35 countries, helps keep election officials, political campaigns, journalists, think-tanks, nonprofits, and human rights organizations safe. These tools also exist alongside new efforts to enhance societal resilience inspired by the Tech Accord, such the Societal Resilience Fund recently launched by Microsoft and OpenAI to advance AI literacy among voters, election authorities, and vulnerable communities in the U.S. and abroad.


Content Integrity Tools can show viewers the change history of an image, like in this example of the Statue of Liberty.

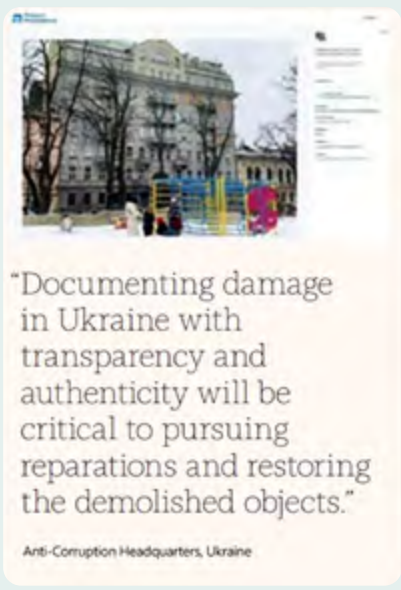
Despite these efforts, democracies still face new and evolving challenges in the digital age. Nation state threat actors are also upskilling, leading to the potential for increasingly sophisticated and scalable cyberattacks and disinformation campaigns. As our Microsoft Threat Analysis Center (MTAC) noted in a recent reports, nation-state threat actors like China and Russia are improving at using generative AI tools as they continue to exacerbate political and social divisions in the United States and other democracies. Even as we work to drive and democratize the benefits of AI and other emerging technologies, which can enhance voter participation and engagement, lower the barriers and costs of voting and campaigning, and provide access to authoritative information to voters, we must be mindful of how antidemocratic actors may deploy these same technologies against free and open societies.

Democracy in the digital age, as in every age, requires constant vigilance, defense, and advocacy from all stakeholders, including governments, technology companies, civil society, media, and citizens.

Beyond this election year, we must continue to look around corners and protect the public broadly from abusive AI-generated content that can undermine trust, spread misinformation, and harm individuals and communities. In a recently published white paper, Microsoft outlined our approach and policy recommendations to address this issue. We recommended requiring providers of AI systems to label synthetic content and use content provenance tooling to track sources and history. We also recommended that governments enact laws and regulations that combat deceptive AI content in elections, fraudulent AI scams, synthetic child sexual abuse material, and non-consensual intimate imagery. Finally, we encouraged governments to publish and update best practices, fund national research programs, and support education campaigns for various audiences and demographics.

Microsoft is committed to playing its part in strengthening democracy, by providing solutions and tools that help enhance the security and resilience of elections; protect vulnerable and targeted members of society such as children, women, the elderly, and individuals running for office; and bolster the integrity of the information ecosystem. Our vision is for technology to support and defend democracy, and to help people reach their potential and achieve their goals.

Democracy in the digital age, as in every age, requires constant vigilance, defense, and advocacy from all stakeholders, including governments, technology companies, civil society, media, and citizens. While nation state threat actors continue to wield cutting-edge technologies against democratic societies through AI, cyber, and other digitally enabled threats, we must each contribute to protecting the democratic systems that enables our individual and collective flourishing. ■



Microsoft and Truepic have partnered to develop and pilot Project Providence, the world's first interoperable system using Truepic's authenticating camera SDK and the Microsoft Azure cloud platform to maintain the provenance of images captured, from storage to display. The Project Providence platform leverages the Coalition for Content Provenance and Authenticity (C2PA) open standard to allow end-to-end interoperability between the capture of documentation, storage, and display. In the first several months following this project's launch, our partner on the ground, the Anti-Corruption Headquarters (ACHQ) has captured over 1,200 images documenting damage to more than 600 unique cultural heritage and civilian infrastructure sites in eight major cities across Ukraine. A subset of these images has been accepted as evidence in ongoing criminal investigations led by District Prosecutor offices in Ukraine, addressing violations of Customary Law and the Law of War.