Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

Senator Mark Warner
703 Hart Senate Office Building
Washington, DC 20510

Dear Senator Warner,

Thank you for your letter and for your leadership in addressing the critical issue of the use of deceptive AI in elections, as well as your ongoing support of the Tech Accord. We share your commitment to ensuring that technology serves to protect elections, empower voters and ensure the security of elections around the world in 2024.

As a signatory of the Tech Accord, we are actively participating in a multi-stakeholder approach to counteract the misuse of AI targeting democracies around the world. We appreciate the opportunity to discuss these matters further and look forward to collaborating with you and other stakeholders to promote safe and secure elections.

Please see below for responses to the questions outlined in your letter:

1. *What steps is your company taking to attach content credentials, and other relevant provenance signals, to any media created using your products? To the extent that your product is incorporated in a downstream product offered by a third-party, do license terms or other terms of use stipulate the adoption of such measures? To the extent you distribute content generated by others, does your company attach labels when you assess – based on either internal classifiers or credible third-party reports – to be machine-generated or machine-manipulated?*

Microsoft is committed to transparency related to AI Generated content that includes credentials. As you know, we were a founding member of the Coalition for Content

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Provenance and Authenticity (C2PA)[1] .To achieve transparency and empower our users, we are leveraging C2PA's "**content credentials**"[2] open standard across several products.

Content credentials are added to all images created with our most popular consumerfacing AI image generation tools, including Bing Image Creator, Microsoft Designer, Copilot, as well as in our enterprise API image generation tools via Azure OpenAI.

Content Credentials are a powerful tool to provide users information about an image's authenticity. To enable that in this critical global election year, we are piloting a tool called "**Content Integrity Certify**[3]" that allows users to add content credentials to their own authentic content in an easy-to-use tool. This pilot is currently available to political campaigns in the US and the EU, as well as elections authorities and select news media organizations around the world.

This tool includes a partnership and collaboration with fellow Tech Accord signatory, TruePic. Announced in April[4], this collaboration leverages TruePic's mobile camera SDK enabling campaign, election, and media participants to capture authentic images, videos and audio directly from a vetted and secure device.

Additionally, we have released a public tool called 'Content Integrity Check'[5] that allows any member of the public or the media to check for the existence of content credentials and see provenance details.

[1] Content Credentials (c2pa.org)
[2] Content Credentials
[3] Expanding our Content Integrity tools to support global elections - Microsoft On the Issues
[4] Truepic's Secure Camera Enhances Microsoft's Content Integrity Tools - Truepic
[5] Microsoft Content Integrity

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

In addition, content is starting to be automatically labeled on LinkedIn[6] on images and videos that contain C2PA metadata. This includes human- or AI-generated content uploaded with C2PA credentials. The first place you'll see the Content Credentials icon is

on the LinkedIn feed, and we'll work to expand our coverage to additional surfaces, including to ads.

2. *What specific public engagement and education initiatives have you initiated in countries holding elections this year? What has the engagement rate been thus far and what proactive steps are you undertaking to raise user awareness on the availability of new tools hosted by your platform?*

**Societal Resilience Grants**

Microsoft, in partnership with OpenAI, recently announced[7] the launch of a series of societal resilience grants aimed at furthering AI education and literacy amongst voters and communities particularly vulnerable during an election year. These grants will impact people globally with specific implementations in the United States, Europe, Latin America, Africa and Southeast Asia.

Grants delivered from the fund will enable several organizations, including Older Adults Technology Services from **AARP** (OATS), **C2PA**, International Institute for Democracy and Electoral Assistance (**International IDEA**) and Partnership on AI (**PAI**) to deliver AI education while supporting their work to create better understanding of AI capabilities.

**Media Literacy**

These grants build on an existing effort[8] by Microsoft to support media, AI and information literacy globally. Last year we announced ongoing partnership with leading news and media literacy non-profits, including the News Literacy Project (NLP), a collaboration led by The Trust Project on the Trust Indicators and Verified to develop

[6] (1) LinkedIn Adopts C2PA Standard | LinkedIn
[7] Microsoft and OpenAI launch Societal Resilience Fund - Microsoft On the Issues
[8] In the digital age, democracy depends on information literacy - Microsoft On the Issues

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

campaigns built on industry research and best practices. Microsoft provided funding for the research and development of the campaigns as well as threat intelligence insights, technical expertise and in-kind ad space on Microsoft platforms to promote the programs. Together with these partners, Microsoft ran an information literacy campaign across several of our platforms.

This work extends to more traditional education spaces as well, particularly with kids. A couple of recent examples of how Microsoft's AI and media literacy has reached children (and their parents!):

- Supported the development of a PBS Kids video series: *Ruff Ruffman: Humble Media Genius*[9], focused on educating children about AI.
- Developed *The Investigators*[10] with Minecraft Education, a game on the Minecraft platform that helps students build information literacy through scenarios that build skills identifying potentially untrustworthy information and what to do.

**Deepfake Awareness Campaign**

Empowering the public by educating them to recognize warning signs and critically evaluate digital content is the first line of defense against deepfakes. Recently, we launched a public awareness campaign in all 27 EU Member States called "Check. Recheck. Vote." It aims at educating EU citizens about deepfakes, the importance of critically scrutinizing voter information, and the role we all have in curbing the spread of disinformation online. The campaign also promotes trusted official EU sources for voting information, as well as other government and NGO-led public education initiatives focused on the potential impact of misinformation and disinformation on elections. The campaign also includes the promotion of an interactive "Real or Not"[11] quiz built to help people understand the sophistication of these tools and the challenge people will have in identifying AI generated content just by looking at an image

In advance of the EU election Bing Search now includes a search experience tailored to the 24 official EU languages with links to official European Parliament sources and banners dedicated to the European elections, making sure that EU citizens have the

[9] Ruff Ruffman, Humble Media Genius | Compilation | PBS KIDS (youtube.com)
[10] InvestiGators | Minecraft Education
[11] Real Or Not (realornotquiz.com)

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

right information on when, where, and how to vote. As we approach the U.S. Presidential election in November we will be rolling out a similar awareness campaign across the U.S. aimed at giving American voters the information they need to be aware and resilient to attempts to use deceptive AI.

3. *What specific resources has your company provided for independent media and civil society organizations to assist in their efforts to verify media, generate authenticated media, and educate the public?*

**True Media Partnership**

In April we announced[12] that we were joining up with AI researcher, Oren Etzioni[13] and his new non-profit, True Media[14]. True Media provides governments, civil society and journalists with access to free tools that enable them to check whether an image or video was AI generated and/or manipulated.

Microsoft's contribution includes providing True Media with access to Microsoft classifiers, tools, personnel, and data. These contributions will enable True Media to train AI detection models, share relevant data, evaluate and refine new detection models as well as provide feedback on quality and classification methodologies.

Additional partnerships within this space are underway, we will share more soon.

**Media Support**

As mentioned in the above section related to content credentials, last month we announced[14] the availability of our content integrity tool[15] to select news and media organizations[16].

[12] TrueMedia.org to Enhance Deepfake Detection Capabilities · TrueMedia
[13] An A.I. Researcher Takes On Election Deepfakes - The New York Times (nytimes.com) [14] TrueMedia.org
[14] Expanding our Content Integrity tools to support global elections - Microsoft On the Issues
[15] Microsoft Content Integrity
[16] Application to Preview Content Integrity Tools for Newsrooms (office.com)

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

4. *What has been your company's engagement with candidates and election officials with respect to anticipating misuse of your products, as well as the effective utilization of content credentialing or other media authentication tools for their public communications?*

**Political Campaign Engagement**

Political campaigns are in many ways the frontlines of likely deceptive uses of AI in elections. That is why Microsoft is engaged in a broad effort to increase campaign staff and candidate awareness of the possible deceptive uses of AI that they may face this cycle. This awareness effort includes sharing best practices and potential interventions that campaigns can put in place to mitigate some of these risks.

Soon after signing the Tech Accord, Microsoft teams began a global campaign to educate political parties, candidates and campaigns in countries around the world to raise awareness about the risks and mitigations to combat deceptive AI in elections. Our teams have delivered these briefings across Europe, India, South Africa, the United States, and Mexico over recent months. Microsoft is scheduled to conduct many more of these sessions in particular across the EU and the UK in the coming weeks ahead of their June and July elections respectively. In total, from the time of the signing of the Tech Accord through the US Presidential Election Microsoft anticipates leading hundreds of sessions with political entities across four continents with nearly one thousand participants.

These sessions are focused on helping political entities understand how deceptive AI may be used to target them, the steps Microsoft is taking to help manage those risks and the steps they can take including use of content integrity tools Microsoft is making available to them, use of our content integrity check tool, and engagement with Microsoft's Campaign Success team. Where legally allowable, these services are being offered free of charge to political campaigns and election officials.

In addition, Microsoft has created a reporting portal for candidates and campaigns to use to report deceptive AI targeting their campaign present on Microsoft platforms. We

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

have launched a website – Microsoft-2024 Elections – where a political candidate can report to us a concern about a deepfake of themselves present on Microsoft platforms. In essence, this empowers political candidates around the world to act if they believe they've been the target of a deepfake.

**Election Official Support**

We have also taken a number of steps to educate and support election officials around the world. Microsoft has created "Election Communications Hubs" to support democratic governments around the world as they build secure and resilient election processes. This hub provides election authorities with access to Microsoft security and support teams in the days and weeks leading up to their election, allowing them to reach out and get swift support if they run into any major security challenges. These hubs build on existing security programs such as the Azure for Elections offering available to state and local election agencies and their partners in the U.S.

Microsoft is also working with election officials in the U.S. and Europe to provide voters authoritative election information on Bing. In partnership with organizations that provide information on authoritative sources, ensuring that queries about election administration will surface reputable sites. For instance, Bing has joined forces with the National Association of State Election Directors[17] (NASED) to receive the websites for authoritative election information for all U.S. states and territories.

Microsoft also continues to offer briefings and transparent information regarding the current threat landscape from our adversaries through our Microsoft Threat Analysis Center (MTAC). For instance, MTAC's recent report on elections[18] outlined activity from China, Russia, Iran and others and their continued exploration of AI in their information operations.

Microsoft is also offering the same access for election authorities to content integrity tools as we are to political parties and candidates. This means that election authorities will be able to sign their content with provenance information in order to provide trusted information to voters in their states and localities.

[17] NASED

[18] Russian US election interference targets support for Ukraine after slow start - Microsoft On the Issues

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

We also continue to engage directly with election officials through briefings and exercises. For instance, we participated in the recent election's tabletop exercise in Arizona focused on the risks of deceptive AI in elections. In addition, Microsoft had the opportunity to talk with members of the National Association of Secretaries of State (NASS) about the risk environment and possible impacts of AI.

In addition to these steps we are continuing our work to safeguard both election authorities and campaigns from cyber threats through M365 for Campaigns and Elections[19] and AccountGuard[20].

5. *Has your company worked to develop widely-available detection tools and methods to identify, catalogue, and/or continuously track the distribution of machine-generated or machine-manipulated content?*

(Also included in answer #3)

**True Media Partnership**

In April we announced[21] that we were joining up with AI researcher, Oren Etzioni[22] and his new non-profit, True Media[24]. True Media provides governments, civil society and journalists with access to free tools that enable them to check whether an image or video was AI generated and/or manipulated.

Microsoft's contribution includes providing True Media with access to Microsoft classifiers, tools, personnel, and data. These contributions will enable True Media to train

[19] Microsoft 365 for Campaigns
[20] Microsoft AccountGuard
[21] TrueMedia.org to Enhance Deepfake Detection Capabilities · TrueMedia
[22] An A.I. Researcher Takes On Election Deepfakes - The New York Times (nytimes.com) [24] TrueMedia.org

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

AI detection models, share relevant data, evaluate and refine new detection models as well as provide feedback on quality and classification methodologies.

Additional partnerships within this space are underway, we will share more soon.

6. *(To the extent your company offers social media or other content distribution platforms) What kinds of internal classifiers and detection measures are you developing to identify machine-generated or machine-manipulated content? To what extent to these measures depend on collaboration or contributions from generative AI vendors?*

---

Please see our response to questions 3 & 5 above related to the creation of detection tools and classifiers. Microsoft will leverage these tools, as needed, when reviewing content to determine whether it violates our policies. In general terms, where content is abusive and violates our policies, we will take steps to moderate that content and enforce our policies, regardless of whether that content is machine-generated or machine-manipulated.

As discussed above, LinkedIn has begun labeling GenAI-created images and videos that contain C2PA metadata. LinkedIn also develops its own AI models to detect machine-generated content, particularly around deepfakes of human faces, as discussed here: Finding AI-generated (deepfake) faces in the wild (linkedin.com)[25]. Our models utilize a new concept that can detect AI-generated images produced by a variety of different generative algorithms. This new concept can tell the difference between real profile photos on LinkedIn and those generated by different types of AI models, such as adversarial-based StyleGANs, generated.photos and EG3d, and Generative AI-based Stable Diffusion, DALL-E 2, and Midjourney.

7. *(To the extent your company offers social media or other content distribution platforms) What mechanisms has your platform implemented to enable victims of impersonation campaigns to report content that may violate your Terms of Service? Do you maintain separate reporting tools for public figures?*

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

The Code of Conduct in the Microsoft Services Agreement prohibits activity that is fraudulent, false, or misleading (e.g., impersonating someone else). We also specifically prohibit deceptive AI election content – our policy specifies that: "Microsoft prohibits the creation or dissemination of deceptive generative AI election content. This includes AI-generated audio, video, and images that deceptively fake or alter the appearance, voice, or actions of political candidates." More information on all our policies can be found here.

Microsoft offers in-product reporting options in many of our services (e.g., in Xbox, Skype, LinkedIn or Bing). If a user cannot find an option to report digital safety abuse in the context in which it was encountered, we also provide a centralized report a concern website, which allows reports to be channeled for review and consideration. This is available to all users.

[25] Finding AI-generated (deepfake) faces in the wild (linkedin.com)

In support of our commitments in the Tech Accord and recognizing the unique risks that generative AI may pose to the integrity of the 2024 elections, Microsoft has also established a dedicated reporting portal for deceptive AI-generated election content. This portal provides a point at which candidates (or their campaigns) can report deceptive AI election content that depicts that candidate. It also provides a place where users can report deceptive AI-generated content about how, when or where to vote where such content has been shared on Microsoft services.

8. *(To the extent your company offers generative AI products) What mechanisms has your platform implemented to enable victims of impersonation campaigns that may have relied on your models to report activity that may violate your Terms of Service?*

Please see the answer to question 7 above. Moreover, additional policies apply to the use of our consumer generative AI services: for example, the Code of Conduct in the Terms of Use for Copilot in Bing prohibits attempts to create or share content that could mislead or deceive others, including for example creation of disinformation, content enabling fraud, or deceptive impersonation. Concerns can be reported to Bing. Copilot also enables users to provide feedback in the product. We have also

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

implemented guardrails in our products that limit to the risk of misuse to generate deceptive AI content about candidates for office.

Additionally, GitHub is a code collaboration platform that is distinct from generalpurpose social media platforms. Content shared on GitHub is primarily software code, so our challenge is addressing tools used to generate synthetic media, rather than synthetic media itself. GitHub recently implemented a policy change process to clarify that we do not allow any projects that are designed for, encourage, promote, support, or suggest in any way the use of synthetic media tools for the creation of nonconsensual intimate imagery and disinformation. This policy change process was developed in order to encourage vital research on synthetic media tools including deepfake detection to happen in the open where others can learn from it, while disallowing tools configured for harmful misuse.

9. *(To the extent your company offers social media or other content distribution platforms) What is the current status of information sharing between platforms on detecting machine-generated or machine-manipulated content that may be used for malicious ends (such as election disinformation, non-consensual intimate imagery, online harassment, etc.)? Will your company commit to participation in a common database of violative content?*

Microsoft has a long-standing commitment to collaboration to address online harms, reflecting the need for whole-of-society approaches to address complex and multifaceted risks. We work closely with a range of stakeholders, including others in industry, to address harm – many of these existing partnerships provide robust frameworks through which we can take action to address known, well-defined harms, regardless of content provenance. Existing hash-sharing initiatives, such as to address child sexual abuse material and through the Global Internet Forum to Counter Terrorism, rely on robust and clear shared taxonomies for harmful content that is likely to violate all member policies (while noting that members take any action on hashes in line with their own policies). Microsoft has also recently donated PhotoDNA hash-matching technology to StopNCII, a UK-based initiative that supports victims of non-consensual intimate imagery to report their imagery in a privacy-preserving way and facilitates cross-industry hash-sharing.

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

We appreciate your support for the work of the Tech Accord and this opportunity to share with you the progress we have made. While we have accomplished a lot towards meeting the commitments within the Accord we know there is much more to do. Should you require any further clarification on any of the answers above, please do not hesitate to reach out to us.

Regards,

Teresa Hutson
Microsoft, Corporate Vice-President, Technology for Fundamental Rights