Mark Isakowitz
VP, Government Affairs &
Public Policy, US and Canada
Google LLC
25 Massachusetts Ave., NW -
9th Floor
Washington, DC 20001

June 3, 2024

The Honorable Mark R. Warner
United States Senate
703 Hart Senate Office Building
Washington, D.C. 20510

Dear Senator Warner:

Thank you for your letter dated May 14, 2024 inquiring about the measures Google is taking to implement the Tech Accord to Combat Deceptive Use of AI in 2024 Elections ("Tech Accord"). We share your commitment to democratic processes globally and, earlier this year, were proud to build on our long track record of collaboration in this space by signing onto the Tech Accord.

Every day, we get billions of searches from people who place their trust in Google as they look for information. We take this responsibility seriously and are committed to maintaining this trust. When voters turn to Google for information, they see the result of our long-standing work to support the integrity of democratic elections. These efforts involve safeguarding our platforms from abuse, helping people access authoritative voting information, and equipping campaigns and election entities with best-in-class security.

As artificial intelligence (AI) evolves rapidly, we are continuing this work with an increased focus on both the challenges and opportunities it can create. This means expanding our work with trusted, authoritative partners, leveraging our growing AI capabilities to prevent abuse, and providing context and transparency in support of protecting the integrity of democratic processes around the world.

Please see below for responses to your specific questions and more information about our work to support elections. More details regarding our efforts related to the U.S. election, EU Parliamentary elections, the general election in India, and the UK election are available online.

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?

With more people using AI to create content, we are building on the ways in which we help our users identify AI-generated content through several new tools and policies. Our efforts toward this end include the following:

- **Marking and detecting content provenance:** Synthetic content is an emerging field of study, and there are technical limitations to authenticating, labeling, detecting, testing, and auditing synthetic content. While we recognize the limitations of current approaches, we are committed to developing and implementing state of the art watermarking capabilities with Google DeepMind's SynthID, which currently watermarks images generated by Gemini Chat, ImageFX, and other consumer services. We recently expanded SynthID's capabilities to watermarking AI-generated text in the Gemini app and web experience, as well as to video in Veo, our recently announced and most capable generative video model. SynthID's innovative technology embeds an imperceptible watermark without impacting the quality, accuracy, creativity or speed of the text or video generation process. Additionally, we have long invested in research on AI classifiers capable of detecting the provenance of AI-generated content, such as developing a classifier capable of recognizing the outputs of our AudioLM model or by contributing data to help others train video or audio synthetic content.

- **Ads disclosures:** Last year, we were the first tech company to launch new disclosure requirements for election ads containing synthetic content. Specifically, we require that verified election advertisers prominently disclose when their ads contain synthetic content that inauthentically depicts real or realistic-looking people or events. This disclosure must be clear and conspicuous, and must be placed in a location where it is likely to be noticed by users. This policy applies to image, video, and audio content. Ads that contain synthetic content altered or generated in such a way that is inconsequential to the claims made in the ad are exempt from these disclosure requirements. This includes editing techniques such as image resizing, cropping, color or brightening corrections, defect correction (for example, "red eye" removal), or background edits that do not create realistic depictions of actual events. With this policy, as with all of our ads policies, we use a combination of machine learning classifiers and human reviewers to ensure compliance.

- **Content labels on YouTube:** Similar to the objective of the ads disclosure policy, on YouTube we seek to provide viewers as much context as possible about the content they watch. Earlier this year, we added labels to the description of content created with YouTube generative AI features, like Dream Screen. In mid-March, YouTube also began requiring creators to disclose when the content they're posting to YouTube includes realistic altered or synthetic content. We then apply transparency labels to this content to signal to users that the content is altered or synthetic. For most videos, the label appears in the expanded description, but for videos that touch on more sensitive topics—like health, news, elections, or finance—we also show a more prominent label on the video itself.

- **A responsible approach to Generative AI products**: Out of an abundance of caution on such an important topic, we have begun to roll out restrictions on the types of election-related queries for which Gemini and AI Overviews on Search will return responses. We take our responsibility for providing high-quality information for these types of queries seriously, and are continuously working to improve our protections.

- **Providing users with additional context:** The About this image feature in Search helps people assess the credibility and context of images found online. Our double-check feature in Gemini evaluates whether there is content across the web to substantiate the responses it provides to user queries.

- **AI-assisted red teaming and expert feedback.** To improve our models, we combine cutting-edge research with human expertise. This year, we are working to enhance red teaming — a proven practice where we proactively test our own systems for weakness and try to break them — through a new research technique we are calling "AI-Assisted Red Teaming." This draws on Google DeepMind's gaming breakthroughs like AlphaGo where we train AI agents to compete against each other to expand the scope of their red teaming capabilities. Our efforts will help develop AI models with these capabilities to help address adversarial prompting and limit problematic outputs. We continuously improve our models with feedback from thousands of internal safety specialists and independent experts from sectors ranging from academia to civil society. Combining this human insight with our safety testing methods will help make our models and products more accurate and reliable. We recognize this is a particularly important area of research for us, as new technical advances evolve how we interact with AI.

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?

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?

Because the answers to these questions are related, we have combined our responses to Questions 2 and 3.

Our initiatives and engagements designed to make it easier for people, including media and civil society, to use our tools and obtain accurate information about elections include the following:

- Google works with authoritative partners, such as state election offices in the U.S. and government election commissions in India and the European Parliament, to surface authoritative information at the top of Search results when people seek out topics including how to vote, ID requirements, voting abroad, and more. We are doing this in more countries than ever before this year, providing more users with direct links to official resources for voting information.

- Google's Civic Outreach team regularly holds trainings on our tools - including those related to AI - for political campaigns, election officials, journalists, civil society organizations and others.

- For news and information related to elections, YouTube's recommendation system prominently surfaces content from authoritative sources on the YouTube homepage, in search results, and the "Up Next" panel. Reliable information becomes especially critical as news is breaking. Because it can take time to produce high-quality videos containing verified facts as events are unfolding, we provide short previews of text-based news articles in search results, along with a reminder that breaking and developing news can rapidly change. We also highlight high-quality content from authoritative news sources during key moments, through our Top News and Breaking News shelves, as well as the news watch page. Moreover, we have information panels that indicate funding sources from publishers that receive public or government funding, as well as information panels giving topical context for topics prone to misinformation.

    - In India, for example, under videos related to electronic voting machines, we show an information panel directing viewers to the Election Commission of India's FAQs on the technical and administrative safeguards on electronic voting.

    - An information panel directing viewers in the European Union to Google's "how to vote" and "how to register to vote" appears in response to searches or videos related to voting. On election day, viewers will see an election results information panel above search results or below videos related to the European elections. This information panel will link to Google's election results feature, allowing people to track election results in real time.

Google

    - In the U.S., our 2020 election information panels, with relevant context from voting locations to live election results, were collectively shown more than four billion times, while during the 2022 midterms our election-related information panels and public service announcements were shown more than two billion times. Ahead of 2024, we have triggered our "vote by mail" information panel on videos related to the topic. The panel points users to information about the safety and integrity of vote by mail processes provided by the Bipartisan Policy Center. This year, in addition to providing these panels to users in English and Spanish, we are providing them in Vietnamese and Chinese.

- We have launched a European Union-specific hub with resources and trainings to help campaigns connect with the 400 million voters expected to go to the polls this year and manage their security and digital presence. In advance of the European Parliamentary elections in 2019, we conducted in-person and online security training for more than 2,500 campaign and election officials, and in 2024 we aim to build on these numbers.

- Our Priority Flagger Program provides channels for participating organizations — government agencies and non-governmental organizations — to notify us of potentially harmful content on certain products that may violate specific policies, including misinformation on Search and YouTube. Content reported by Priority Flaggers is not automatically removed or subject to any different policy treatment - the same standards apply for all user flags. However, because of the high degree of trust and expertise of Priority Flaggers, our teams prioritize their flags for review. In the European Union, for instance, we onboarded NASK, a Polish National Research Institute under the supervision of the Chancellery of the Prime Minister of Poland, to our Google Priority Flagger program. And we recently enrolled a government agency in Slovakia and an NGO in Lithuania as Priority Flaggers for YouTube.

- Since our inaugural contribution of €25 million to help launch the European Media & Information Fund, an effort designed to strengthen media literacy and fight misinformation across Europe, 70 projects have been funded across 24 countries so far, covering topics ranging from fact-checking during elections and critical events, to improving media literacy of populations who are typically harder to reach.

- We support the Global Fact Check Fund, as well as numerous civil society, research and media literacy efforts from partners, including Google.org grantee TechSoup Europe, as well as Civic Resilience Initiative, Baltic Centre for Media Excellence, The Central European Digital Media Observatory, and more.

- Together with Moonshot and local partners, Google and Jigsaw are running a prebunking campaign in the European Union to address some of the most prevalent tactics used to manipulate people online. We have run similar campaigns in other countries, such as in advance of the Indonesian elections earlier this year.

- In the EU, YouTube is running a public service announcement on its homepage linking to its existing HitPause media literacy campaign, reminding voters that literacy skills can help users stay informed.

- Ahead of the General Election in India, Google is supporting Shakti, India Election Fact-Checking Collective, a consortium of news publishers and fact checkers in India that are working together to aid the early detection of online misinformation, including deepfakes, and to create a common repository that news publishers can use to tackle the challenges of misinformation at scale. The project will also provide news organizations and fact-checkers essential training in advanced fact-checking methodologies, deepfake detection, and the latest Google tools like the Fact Check Explorer, to streamline verification processes.

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?

We engage extensively with candidates for office and election officials in the context of cybersecurity. For all high-risk individuals, including elected officials, candidates, campaign workers, journalists, and election workers, we have developed our Advanced Protection Program. This program provides our strongest set of cyber protections, including the introduction earlier this year of passkeys. In the U.S., we partner with Defending Digital Campaigns (DDC) to provide U.S. campaigns with these and other security tools they need to stay safe online, including new tools to rapidly configure Google Workspace's security features in a way that is customized to meet the needs of high risk users.

Last year, Google also announced a new Online Safety and Security Partnership with the International Foundation of Electoral Systems (IFES) to provide free security training and tools to high-risk users. IFES is an expert non-partisan non profit with over 35 years of experience helping democracies in 145 countries. With more than 100 engagements to date, IFES has conducted security trainings for high-risk individuals in Bosnia and Herzegovina, Libya, Serbia, Sri Lanka, Nepal, Ukraine, Georgia, Iraq, Lebanon, Ecuador, and Uganda. IFES and Google are partnering to provide valuable online safety resources to high-risk users across the world, including:

- **Security training:** In collaboration with Google's Safety Engineering Center, IFES will expand its curriculum to address the evolving threat environment, incorporate account security best practices, and create content for specialized audiences. Cybersecurity content will be built into a range of trainings for journalists, activists, civil society members, and election officials. IFES will also highlight content for high-risk users that have been underserved by online safety and security education.

- **Raising awareness on security tools across 30 countries:** Google's tools and resources for high-risk individuals will be shared to IFES' international networks, helping more people use tools like the Advanced Protection Program, Google Titan Security Keys, Project Shield, a free tool to help protect sites from Distributed Denial of Service (DDoS) attacks, and more. To date Google has distributed over 150,000 Titan Security Keys at no cost to high-risk individuals around the world, doubling down on this commitment to provide 100,000 of our new Titan Keys through partnerships this year.

- **Industry thought leadership and case studies:** IFES will highlight evolving trends and share insights to help Google contextualize how IFES' partners engage with its products, to keep improving.

This work is designed to help shore up the defenses of democracies that work for all and to combine Google's account security tools and IFES' history in supporting high risk users around the world like journalists, activists, and elections management bodies.

Our teams help identify, monitor and tackle emerging threats, ranging from coordinated influence operations to cyber espionage campaigns against high-risk entities. For example, on any given day, we are tracking more than 270 targeted or government-backed attacker groups from more than 50 countries. We publish our respective findings consistently to keep the public and private sector vigilant and well informed. Our teams also help organizations build holistic election security programs and harden their defenses with comprehensive tools, ranging from proactive compromise assessment services to threat intelligence tracking of information operations. Our intelligence teams' insights are also used to help internal teams understand evolving trends from threat actors trying to undermine the democratic process.

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?

Please see response to question (1), above, for information about our tools to identify AI-generated content.

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 do these measures depend on collaboration or contributions from generative AI vendors?

Please see response to question (1), above, for information about our tools to identify AI-generated content.

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?

Content intended to impersonate a person or channel is prohibited on YouTube. Examples of content that is not permitted under YouTube's impersonation policy include the following:

- Channels with the same identifier (channel name or handle) and image as another channel, with the only difference being a space inserted into the name or a zero replacing the letter "O."
- Using someone else's real name, user name, image, brand, logo, or other personal information to trick people into believing you are that person.
- Setting up a channel using the same identifier (channel name or handle) and image of a person, and then pretending that person is posting content to the channel.
- Setting up a channel using the name and image of a person, and then posting comments on other channels as if they were posted by the person.
- Channels claiming to be a 'fan account' in the channel description, but not stating so clearly in the channel name or handle, or posing as another's channel and reuploading their content.
- Channels impersonating an existing news channel.

YouTube community members may flag content that they believe violates our impersonation policy via our reporting tools.

In addition to prohibiting impersonation under its Community Guidelines, YouTube provides a mechanism for individuals to report AI-generated or other synthetic content that looks or sounds like them. In order to qualify for removal, the content should depict a realistic altered or synthetic version of the individual's likeness. YouTube considers a variety of factors when evaluating the

complaint, including whether the person can be uniquely identified; whether the content is realistic; whether the content contains parody, satire, or other public interest value; and whether the content features a public figure or well-known individual engaging in a sensitive behavior such as criminal activity, violence, or endorsing a product or political candidate.

In the advertising context, Google Ads updated the Unacceptable Business Practices portion of its Misrepresentation policy to include enticing users to part with money or information by impersonating or falsely implying affiliation with or endorsement by a public figure, brand, or organization. We are now able to take account-level (versus ad-level) action for violations involving an advertiser falsely implying a celebrity endorsement or affiliation. Notably, anyone can report these issues through our standard bad ads reporting form.

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?

Our AI Prohibited Use Policy clearly prohibits the generation and distribution of content intended to misinform, misrepresent or mislead, including the "generation of content that impersonates an individual (living or dead) without explicit disclosure, in order to deceive." Users can submit feedback on both Gemini and Google Search Labs using the in-product feedback mechanisms available on both. For legal issues such as potential copyright violations, users can also submit removal requests through our Legal Removals form.

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?

Artificial intelligence innovation raises complex questions that neither Google, nor any other single company, can answer alone. Getting it right will require continued collaboration among companies, academic researchers, civil society, governments, and other stakeholders. To this end and to promote the responsible development of frontier AI models, Google is working with groups like Partnership on AI and ML Commons and, along with other leading AI labs, is a founding member of the Frontier Model Forum.

In addition, as mentioned in response to question 1. above, on February 8th, Google became a steering member of the C2PA coalition and standard, a cross-industry effort to help provide more transparency and context for people when it comes to AI-generated content. C2PA has the benefit of being tamper evident and highly interoperable, making it an excellent vehicle for cross-ecosystem technical collaborations to signal the provenance of content at scale - whether that content is authentic or AI-generated. We are committed to playing our part in furthering the technical development and adoption of the standard, and look forward to continuing to collaborate with its other members towards more transparency.

It is against the backdrop of our long track record of collaboration in this area that we signed onto the Tech Accord at the Munich Security Conference–signaling our commitment to working on technical and educational solutions to combat the deceptive use of AI in elections, and that we continue to engage in industry fora to share and learn from best practices and advances in research across the industry..

***

Thank you again for the opportunity to respond to your inquiries. We look forward to continuing to work with you and your staff on these important issues.

Sincerely,

Mark Isakowitz
Vice President
Government Affairs and Public Policy, US and Canada