Written Statement

U.S. Senate AI Insight Forum: Elections & Democracy

Yael Eisenstat

Vice President, Center for Technology & Society

ADL

AI Insight Forum
U.S. Senate

November 3, 2023
Washington, DC

Working to stop the defamation of the Jewish people and to secure Justice and fair treatment to all since 1913.

Leader Schumer, Senators Rounds, Heinrich, and Young:

Since 1913, the mission of ADL (the Anti-Defamation League) has been to "stop the defamation of the Jewish people and to secure justice and fair treatment to all." For over a century, ADL has been a leader in the fight against hate, bigotry, and antisemitism wherever it exists. ADL has unique expertise in fighting hate online because of the organization's work at the intersection of civil rights, extremism, and technology, and because we are rooted in a community that has been relentlessly targeted online by extremists and bigots.

The rise of artificial intelligence (AI) technologies has undoubtedly revolutionized the way we access and consume information. Today, as AI systems become more sophisticated and as generative AI (GAI) becomes popularized, there exists a legitimate concern and lack of trust in the information we consume. For years, we have seen the ways social media's AI-powered systems amplify and recommend engaging content—distorting our idea of public opinion and also shaping it. Compounding this, since last year's generative AI boom, readily available tools make it easier than ever before for bad actors to create and disseminate harmful disinformation. The combination of social media algorithms that amplify incendiary content, new and accessible GAI tools, and gutted trust and safety teams at tech companies creates a perfect storm.

It is crucial to explore how AI can contribute to undermining the very essence of trust upon which our democratic institutions rely. In the hands of bad actors, AI and GAI tools pose significant threats to the integrity of our electoral processes and strength of our democracy. In an era where information has become both ubiquitous and weaponized, we must consider the role AI plays in eroding trust in the information we consume. This submission will discuss: ADL's work fighting extremism and hate to secure democracy; antisemitism and election-related misinformation's inextricable link; considerations for curbing AI-fueled misinformation and hate related to elections; and recommendations for government and industry alike.

ADL's work fighting extremism and hate to secure democracy

ADL is at the forefront of the battle against hate and extremism in the digital age, bringing over a century of expertise in combating these threats. Our Center on Extremism (COE) examines how extremists, spanning the ideological spectrum, manipulate the online landscape to disseminate their messages, recruit followers, finance hatred, and even incite acts of terror. Our Center for Tech & Society (CTS) is a research-driven advocacy center committed to putting an end to the proliferation of online hate, harassment, and extremism. CTS collaborates with industry, civil society, government entities, and targeted communities to achieve this shared objective. CTS works to hold tech companies accountable for their dynamic roles in normalizing and perpetuating hate and harassment online, which is especially critical during election seasons. ADL's PROTECT, COMBAT, and REPAIR plans are policy initiatives dedicated to counteracting violent domestic extremism, antisemitism, and online hate—all of which pose threats to elections integrity and healthy democracy.

I joined ADL to head CTS just over a year ago, after two decades fighting to protect our democracy, beginning with 13 years as a public servant. I was hired in 2018 by Facebook to head its new elections integrity efforts for political advertising, specifically to address some of the very issues we are discussing today. After leaving the company, I wrote and spoke extensively about how social media companies in particular are affecting democracy and have spent the past five years working on bringing accountability to the industry.

Antisemitism and election-related mis- and disinformation are inextricably linked

Antisemitic and election-related disinformation share a troubling connection: both thrive on exploiting existing divisions and spreading discord by sowing extremism and hate. During election periods, there is often an uptick in disinformation campaigns, which frequently incorporate elements of hate and antisemitism. Social media platforms provide fertile ground for these campaigns. Hate and conspiracy theories are amplified across platforms via AI-powered recommendation engines, as bad actors exploit the charged political atmosphere to manipulate public opinion and undermine democratic processes.

Antisemitic disinformation is inextricably linked to conspiracism, anti-democratic attitudes, and escalation into violence. Nearly one in five Americans believes QAnon conspiracy theories, a set of ideas intimately connected to both antisemitic extremism and election denial. Notably, the same percentage of Americans subscribes to six or more false antisemitic tropes. While these narratives start online, they do not stay there. ADL's most recent antisemitism audit revealed that antisemitic incidents surged by 36 percent in 2022, marking the highest number of incidents since ADL started tracking this data in 1979. Time and time again we have seen the ways online antisemitism intersects with anti-democratic misinformation and is linked to deadly harm. There have been too many illustrations of the fatal consequences of antisemitic and antidemocratic extremism, which have a clear nexus to mis- and disinformation spread on social media. From Pittsburgh to Poway, Buffalo to Club Q, mass shooters who have been fed a steady stream of conspiracy theories and hate online engaged in extreme violence offline.

Considerations and insights to curb AI-fueled misinformation and hate in the wake of elections

We know social media is an integral part of American elections, as more than half of all Americans turn to social media for at least some of their news. ADL has seen the ways false or misleading election content—including *misinformation*, spread without malice or coordination, and *disinformation*, purposely created to manipulate or cause harm—runs rampant online. This proliferation of misleading election content, exacerbated by AI systems and further supported by GAI tools, has the capacity to subvert democracy. We must have a shared understanding of AI's influence on the information ecosystem and consider what both government actors and tech companies must understand and prioritize ahead of the 2024 election. Considerations:

1. ***AI can help improve content moderation, but not if scalability is the only thing companies prioritize***

When I was at Facebook, my team put together a plan to ensure that political advertising ahead of the 2018 U.S. midterm election would not include false information about voting—the most fundamental, indisputable election-related information. It was an absolutely achievable solution, involving AI systems to scan political ads. One of the reasons senior leadership said they would not approve it was because it did not "scale globally." Every election has its own unique issues, every country has its own political realities. There is no singular AI solution that addresses all elections, in all languages, across all locations.

2. ***AI-fueled online discourse impacted offline violence long before GAI's surge in popularity***

It is widely recognized that the AI-powered reward systems of several major social media platforms incentivize users to spread misinformation by amplifying incendiary, high-engagement content more than its truthful counterparts. According to a recent study from ADL and TTP (Tech Transparency Project), some of the biggest social media platforms and search engines at times directly contribute to the proliferation of online hate, antisemitism, and extremism through their own AI-powered tools. Social media's amplification of extremism, disinformation, and conspiracy theories—and the complete lack of transparency and accountability about how that amplification takes place—pose a serious threat to democracy in this country, and to the safety of vulnerable individuals and communities worldwide.

Election-related misinformation has already led to offline unrest. The deadly insurrection at our Capitol, which ADL has repeatedly called the most predictable act of political violence in American history, illustrates this connection. In fact, I very publicly warned of exactly the way Facebook's own tools would help lead to post-election violence months before January 6. The insurrection made the harms of online disinformation—and its offline impact—abundantly clear. As verified in leaked internal Facebook documents, insurrectionists' actions were the product of weeks, months, and years of incitement, spread across the social media ecosystem. Importantly, the insurrection pre-dates GAI's availability to the mass market. We are clearly in the midst of discovering the ways in which social media's flawed engagement models can combine with GAI to further incite lawlessness and violence. The ease with which malicious actors can create compelling false narratives about vulnerable communities and the validity of democratic processes is another tool they can employ to find and radicalize susceptible targets.

Interrupting AI-fueled mis- and disinformation and finding effective mitigation strategies to counter election-related disinformation and antisemitism is no longer a marginal issue. It now requires a whole-of-government and society approach. As noted above, there is a clear connection between election-related disinformation and online extremist, antisemitic, misogynist, racist, and hateful images, and tropes. The amplification and normalization of these messages—facilitated by AI systems—has and can continue to lead to offline violence.

3. *Generative AI is now easy-to-use and readily accessible*

In an environment already rife with misinformation and inflammatory rhetoric, the introduction of GAI tools—such as deepfakes and synthetic audio—present a significant threat to public trust. Manipulated videos and images have the capacity to distort reality, spread confusion, and incite violence at unprecedented scale and speed. In a world fraught with tensions, synthetic media can be used to falsely implicate nations or groups in acts of violence, undermine diplomatic initiatives, distort electoral procedures, and escalate conflicts.

The use of fake news and manipulated content during elections—and other global conflicts—is not a new phenomenon. What's profoundly changed is the accessibility of information-generating tools and the volume of information available to consumers. Today, anyone with basic tech skills and internet access can create highly convincing fake images, videos (commonly known as "deepfakes"), and audio. This content includes synthetic speech or video, which uses AI to mimic real voices and images that have been "cloned" from samples of authentic content from prominent figures. These fabricated materials can be disseminated to global online audiences at minimal or no cost. They then flow through algorithmic models tuned to amplify engaging content, independent of accuracy. For example, in October, an AI-generated video circulated purportedly showing First Lady Biden denouncing her husband's support for Israel and calling for a ceasefire. This video, posted on X (formerly Twitter) features the First Lady speaking directly to the camera. It also has voice-over while displaying images and video clips of Gaza war zones and news takes. The content featuring Dr. Biden is fabricated but looks and sounds incredibly real. Even though the video was quickly classified as a "deep fake," discussions about it took place across social media. GAI-generated content fascinates audiences and will undoubtedly be prevalent throughout the 2024 election season.

4. *Extremists are already taking advantage of GAI tools—we should be cautious about how they can do so as it relates to elections*

Extremists and conspiracy theorists routinely use GAI tools to create misleading content on social media. On fringe sites like 4chan, users have shared audio files of celebrities and politicians being made to say hateful or violent rhetoric. Sometimes, these audio files are mapped onto videos of the speaker whose voice is being cloned in order to create a deepfake video. In February 2023, a viral video appeared to show President Biden publicly invoking the Selective Service Act, announcing a draft of U.S. citizens for the war in Ukraine. About 45 seconds in, far-right activist and conspiracy theorist Jack Posobiec tells viewers that the video was a deepfake created by his producers to show his predicted future for America and "nuclear war." While the video was ostensibly created to prove a point, it was misleading enough to warrant a debunk from Snopes. GAI has also been used by affiliates of the white supremacist and antisemitic hate network, Goyim Defense League (GDL). Several deepfake videos shared across GDL Telegram chats seem to have been created by the same user. One of these videos is a deepfake of Nina Jankowicz, researcher and former director of the now-defunct Disinformation Governance Board. Using synthetic speech mapped onto authentic footage of Jankowicz, the video falsely depicts her making a series of disturbing and antisemitic claims.

During election cycles, these strategies can be used to trick the public into believing that a candidate has endorsed an issue when they haven't, alter audio from campaign speeches, or even create fake phone call exchanges. While the results produced by such technology aren't always convincing, they are becoming more realistic as the tools evolve. In an information landscape where even the validity of authentic video has come into question, we expect that GAI content will continue to cause confusion as we approach the 2024 election.

5. The mere awareness of GAI tools can lead audiences to question the authenticity of legitimate content

Beyond the concern that synthetic media will be used nefariously for disinformation campaigns, deepfakes have also made it easier for extremists and conspiracy theorists to publicly dismiss legitimate media content. By employing GAI to spread disinformation, malicious actors aim to not only achieve propaganda victories but also to contaminate the information landscape. The objective is to foster a climate of widespread distrust in any and all online content. Bad actors do not necessarily need to actively use GAI tools. Ultimately, the proliferation of deepfakes and the manipulation of online content at the hands of GAI tools has us question the validity of all the information we consume. The mere awareness of synthetic media can precondition certain audiences to question the authenticity of legitimate content. This unsettling trend is often referred to as "the liar's dividend."

Most recently, after the brutal Hamas attack on October 7, there were unverified reports of decapitated babies and toddlers in the Kfar Aza kibbutz. President Biden referenced these reports in a national address, but later, both the Israeli government and the US State Department could not immediately confirm whether the pictures were authentic. This led to social media condemnation and accusations of propaganda. While the Israeli Prime Minister's office posted graphic photos that were later verified by multiple sources as authentic, the mistrust had already been sowed. This erosion of confidence in the information we see can have profound implications for our elections and trust in democracy. Citizens will continue to find it increasingly challenging to distinguish fact from fiction online and engage in informed, constructive discourse.

Recommendations

Considering the lessons learned from the past decade, one thing is clear: we cannot afford to wait until further harms occur to rein in big tech. While companies that develop AI and machine learning tools are best positioned to create and voluntarily implement safeguards that prevent online harms from occurring in the first place, such safeguards are not sufficient in and of themselves. ADL urges government to regulate AI with a combination of *proactive measures* to support a transparent industry that incentivizes pro-social behaviors, and *responsive measures*, to ensure accountability when AI tools exacerbate the spread of misinformation and cause hate-based or anti-democratic harms.

1. ***Promote Access to Authentic and Verifiable Information:*** ADL urges tech companies to provide clear disclosure mechanisms that help users differentiate between authentic content and artificially generated content. Tech companies should keep records of harmful GAI-created media they detect and the steps they take to mitigate its harms. Additionally, as the Department of Commerce develops standards for content

authentication per the President's Executive Order on AI, tech companies should consider implementing established best practices for AI-generated content authentication. Notably, while these solutions are crucial, they do not solve for the necessary management of mis- and disinformation—which is also exacerbated through AI-powered recommendation engines.

2. ***Mitigate Risks with Proactive Measures:*** As AI evolves, tech companies must ensure they are resourcing and supporting robust trust and safety practices. To establish a responsible ecosystem of AI accountability, legislators must play an active role in ensuring tech companies mitigate risks that threaten civil rights or consumer safety. Government should explore requiring companies to adopt consensus industry standards for proactive measures before the deployment of AI technologies to consumers. This could include red teaming, requiring risk assessments, and implementing appropriate regulatory requirements.

3. ***Require Transparency:*** Despite the significant impact AI can have, both positive and negative, the public often lacks insight into AI systems. Therefore, legislators should mandate AI developers issue regular, public-facing transparency reports. While critics may raise concerns about privacy or trade secrecy, transparency reporting is a flexible process, not an all-or-nothing proposition. Consumers have the right to make informed decisions about AI products they use. In fact, ADL conducted a national survey which found that 84 percent of Americans are worried GAI tools will increase the spread of false or misleading information. Eighty-seven percent want to see action from Congress mandating transparency and data privacy for GAI tools.

4. ***Increase accountability:*** Lawmakers must assess the impact of AI business models to prevent AI companies from enabling election interference, hate crimes, civil rights violations, or acts of terror. Tech companies currently lack sufficient incentive to prioritize public trust and user safety due to inadequate oversight and accountability measures. Without changes to these incentive structures, they will not prioritize protecting democracy.

5. ***Hold bad actors accountable for the malicious use of AI/ML tools:*** Both industry and government must take proactive measures to prevent GAI-generated harm. The ability to generate large volumes of synthetic content with speed and precision creates an opportunity for bad actors to potentially engage in unlawful cyberstalking, doxing, and harassment at an unprecedented scale, resulting in severe consequences for targets. Legislators should update laws concerning election interference and online harassment of election officials to disincentivize bad actors from wielding GAI tools for political gain.

6. ***Develop Public Competence in Identifying AI-Generated Misinformation:*** All stakeholders, including industry, civil society, and government, should enhance public resilience against AI-generated misinformation. For example, industry can promote user vigilance by creating educational resources and incentives, encouraging users to check content sources, conduct reverse image searches, and verify information from multiple sources. Government should assess the integration of media literacy and disinformation resilience into education curricula. Implementing media literacy programs early on can protect communities from harmful effects of misinformation, whether related to elections or online hate.