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

Written Statement of

**Alex Stamos**
Director, Stanford Internet Observatory
Stanford University

November 8, 2023

—

Leader Schumer, Senator Rounds, Senator Heinrich, Senator Young, and distinguished members of the U.S. Senate, thank you for the opportunity to address the use of generative artificial intelligence (AI) technology for elections and our democracy.

My name is Alex Stamos and I am the director of the Stanford Internet Observatory (SIO) at Stanford University. While my statement is provided in an individual capacity, I offer guidance from the research we conduct at the Internet Observatory in addition to my professional experiences, including serving as the Chief Security Officer at Facebook during the 2016 presidential election.

I founded the Internet Observatory as a cross-disciplinary program for research, teaching, and policy engagement to better understand and address abuse in current information technologies. We have a focus on social media and emerging technology, particularly generative AI. My intention in founding this organization was to combine rigorous academic research with practical, impactful engagement with difficult policy and technical challenges.

Generative AI technologies are already here. They are increasingly accessible and powerful, offering vast potential with continued U.S. innovative leadership *and* U.S. policy leadership. It takes both. We need to be prepared for the use of generative AI to create novel and widespread forms of deception in the democratic process. These risks should not be overblown and should not be used to restrict fundamental AI research, but this is also not a hypothetical threat. Deceptive uses of AI-generated content are already occurring in the U.S. presidential election[1] and democratic elections around the world.[2, 3]

[1] Nehamas, N. (2023, June 8). DeSantis Campaign Uses 'Deepfake' Images to Attack Trump, Experts Suggest. *The New York Times*. https://www.nytimes.com/2023/06/08/us/politics/desantis-deepfakes-trump-fauci.html
[2] Zuidijk, D. (2023, October 4). Deepfakes in Slovakia Preview How AI Will Change the Face of Elections. *Bloomberg*. https://www.bloomberg.com/news/newsletters/2023-10-04/deepfakes-in-slovakia-preview-how-ai-will-change-the-face-of-elections
[3] Bristow, T. (2023, October 9). Keir Starmer suffers UK politics' first deepfake moment. It won't be the last. *Politico*. https://www.politico.eu/article/uk-keir-starmer-labour-party-deepfake-ai-politics-elections/

There is no silver bullet. A collaborative approach is needed to safely harness the potential of generative AI — from codes of practice and technical standard setting, to research, auditing, regulation, and education. Stanford Internet Observatory researchers and our colleagues across campus are working on research and engagement across each of these key issue areas.

Ultimately, a government and industry response must protect creative and free expression rights while limiting risks and deception from the use of AI-generated text and media. None of these steps offer a panacea, but there is bipartisan public agreement that generative AI will make elections worse and Congress has the power to act now.[4]

**Background on Generative AI and Online Discourse**

Just as social media democratized content distribution, making it possible for just about anyone to spread anything, generative AI makes it possible for anyone to create human-like writing and realistic media at scale. The widespread ability to create convincing images, audio, or video also creates a dilemma known as the "liar's dividend," in which politicians and government officials abuse a lack of trust in media technology and content to claim that something real is actually digitally manipulated fake content, with the goal of having it removed or discredited.[5]

Generative AI content is increasingly realistic, persuasive and difficult to detect. SIO research finds that existing AI tools and technology can generate photorealistic images,[6] or text that is persuasive,[7, 8] difficult to distinguish from content written by a human,[9] and that can be widely published with limited resources.[10] While technical tools and measures are being developed by industry to track or disclose the origin and manipulation of digital media, these standards are limited by voluntary adoption.[11, 12, 13]

---

[4] Swenson, A., & O'Brien, M. (2023, November 3). Poll shows most US adults think AI will add to election misinformation in 2024. *Associated Press*. https://apnews.com/article/artificial-intelligence-2024-election-misinformation-poll-8a4c6c07f06914a262ad05b42402ea0e

[5] Chesney, R., & Citron, D. K. (2019, December). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, 107(1753). https://dx.doi.org/10.2139/ssrn.3213954

[6] Goldstein, J. A., & DiResta, R. (2022, September 15). Research note: This salesperson does not exist: How tactics from political influence operations on social media are deployed for commercial lead generation. *Harvard Kennedy School (HKS) Misinformation Review*. https://doi.org/10.37016/mr-2020-104

[7] Karinshak, E., Liu, S. X., Park, J. S., & Hancock, J. T. (2023, April 16). Working With AI to Persuade: Examining a Large Language Model's Ability to Generate Pro-Vaccination Messages. *Proceedings of the ACM on Human-Computer Interaction*, *7*(CSCW1), 1-29. https://dl.acm.org/doi/abs/10.1145/3579592

[8] Goldstein, J. A., Chao, J., Grossman, S., Stamos, A., & Tomz, M. (2023, April 8). Can AI Write Persuasive Propaganda?. https://doi.org/10.31235/osf.io/fp87b

[9] Jakesch, M., Hancock, J. T., & Naaman, M. (2023, March 7). Human heuristics for AI-generated language are flawed. *PNAS*, 120(11). https://doi.org/10.1073/pnas.2208839120

[10] Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M., & Sedova, K. (2023, January 10). *Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations*. https://doi.org/10.48550/arXiv.2301.04246

[11] Content Authenticity Initiative. https://contentauthenticity.org

[12] Coalition for Content Provenance and Authenticity (C2PA). https://c2pa.org

[13] *AI & Media Integrity*. Partnership on AI. https://partnershiponai.org/program/ai-media-integrity

Manipulative content and political propaganda have been part of the public discourse for centuries. However, generative AI makes it easier to create propaganda at a lower cost and higher volume, and with more refined targeting. Resource constraints for writing compelling content or generating realistic media are lowered, especially in non-native languages of disinformation actors. The technology also has the potential to transform targeted distribution through the creation of automated, interactive fake personas — a next generation of bots.

Even as it becomes easier to create realistic content, research suggests that people have difficulty identifying whether such text or media was generated by AI. For example, research shows that people struggle to identify computer-generated headshots which are then used to create fake online personas that manipulate public discourse.[14, 15] Generative language models have also improved to the point where text outputs are difficult to distinguish from human-written content.[16]

In a study where participants were asked to distinguish real profiles on dating and similar online biographies from those generated with AI, researchers found that, regardless of background or training, the ability of people to identify AI-generated profiles was close to chance.[17] Even tools created by generative AI companies themselves are inconsistent in identifying AI-generated text.[18]

As generative AI tools and technology become more powerful, they can create more persuasive content across a wider range of topics and formats. My research with colleagues found that an existing large language model can write "highly persuasive" propaganda articles that can be made "even more persuasive" with light editing.[19] This means propagandists — whether Russia, China, or other U.S. adversaries, those seeking a profit, or domestic chaos creators — can create convincing content with limited effort and free tools, scaling their capacity.

Our paper concludes: "Language models offer propagandists a way to create text that is as persuasive as content from existing covert propaganda campaigns, while requiring less human involvement and lower cost than human-written text." In other words, what once took a team of 20 to 40 people working out of Russia or China to create 100,000 pieces of English-language propaganda is now possible with a single person using freely accessible generative AI tools.

Recent methods for detecting influence operations have included examining posts for repetition (or "copypasta") or other evidence of linguistic inauthenticity, as well as examining profiles in large, seemingly-connected networks for the presence of AI-generated profile photos.[20] Improvements in generative

[14] Nightingale, S. J., & Farid, H. (2022, February 14). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *PNAS*, 119(8). https://doi.org/10.1073/pnas.2120481119
[15] Goldstein, J., & DiResta, R. (2022, September 15). Research note: This salesperson does not exist: How tactics from political influence operations on social media are deployed for commercial lead generation. *Harvard Kennedy School (HKS) Misinformation Review*. https://doi.org/10.37016/mr-2020-104
[16] Jakesch, M., Hancock, J. T., & Naaman, M. (2023, March 7). Human heuristics for AI-generated language are flawed. *PNAS*, 120(11). https://doi.org/10.1073/pnas.2208839120
[17] Ibid.
[18] Kirchner, J. H., Ahmad, L., Aaronson, S., & Leike, J. (2023, January 31). *New AI classifier for indicating AI-written text*. OpenAI. https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text
[19] Goldstein, J. A., Chao, J., Grossman, S., Stamos, A., & Tomz, M. (2023, April 8). Can AI Write Persuasive Propaganda?. https://doi.org/10.31235/osf.io/fp87b
[20] Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021, March 17). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592, 590–595. https://doi.org/10.1038/s41586-021-03344-2

AI will enable manipulators to avoid this repetitiveness, and produce content that uses language that resonates with target audiences. Existing mechanisms for identifying coordinated accounts and current media literacy programs will need to be updated.

Much of the work to stop generative AI from misleading the public is falling to technology companies. Most of these online services — such as chatbots, social media platforms, or search engines — have rules for synthetic media, but it is difficult for any online service to detect AI-generated content. Content will spread through many online and traditional media platforms, so provenance and detection efforts will require cross-industry collaboration. Still, an over-reliance on technological solutions for watermarking generative content risks creating a false perception of legitimacy for content created with open source tools that skirt disclosure mechanisms.[21]

There are also efforts to create technical standards marking "real" content, such as photographs. Here too, there will be limitations with industry participation and legitimacy for existing media and devices. Despite important standards development, it will remain challenging to navigate this new reality. Public education about this technology and inauthentic online content and behavior is the best remedy.

While much attention is given to addressing political disinformation, we should acknowledge that this is a thorny policy challenge due to politicization and limitations on government action from conflicts with free speech law and values. Congress must also prioritize legislation that addresses online child sexual abuse material, a growing problem with openly available generative AI tools.[22]

**Recommendations**

Many of the actions that will be necessary to protect the 2024 election and all those afterward from AI-enabled manipulation must be taken by private actors, such as the social media platforms and model developers. However, there are still concrete steps that Congress can take in the coming months.

- **Require Platform Transparency** - Congress should immediately take up the bipartisan Platform Accountability and Transparency Act[23] to require social media platforms of a certain size to provide access to publicly-available content and basic metadata (such as interactions and views). This would help reestablish[24] the capability of external researchers to understand what is happening in the social media space and to find foreign influence campaigns and AI-generated manipulation.

[21] DiResta, R., & Willner, D. (2023, November 1). White House AI Executive Order Takes On Complexity of Content Integrity Issues. *Tech Policy Press*. https://techpolicy.press/white-house-ai-executive-order-takes-on-complexity-of-content-integrity-issues/

[22] Thiel, D., Stroebel, M., & Portnoff, R. (2023, June 24). *Generative ML and CSAM: Implications and Mitigations*. Stanford Digital Repository. https://purl.stanford.edu/jv206yg3793

[23] https://www.coons.senate.gov/imo/media/doc/pata_bill_text_118th_congress1.pdf

[24] The cancellation of API access agreements between Twitter and academic institutions and legal threats against researchers have greatly reduced the ability of outside groups to study foreign influence using what is otherwise public and freely available content.

- **Legal Recourse for Defamatory Fakes -** AI-generated imagery has already been used in the Republican primary to generate fake images of President Trump[25] by one of his rivals. The legal options available to individuals who have been defamed utilizing AI technologies are currently murky. The FEC has taken this matter up in the context of political advertising, but Congress could act now to explicitly grant recourse under existing libel laws against those who create photorealistic imagery using AI to harm the reputation of normal citizens, and to explicitly ban the use of generative AI from regulated political ads.

- **Clarify the Appropriate Role of the Executive Branch** - There is an important role for the federal government to play in preventing foreign influence in U.S. elections. Government coordination with industry is essential for identifying and addressing foreign influence operations, while protections are needed to prevent political coercion in these efforts. Congress should pass an explicit law against jawboning, or government pressure on social media platforms to remove or limit the spread of legal content posted by U.S. citizens. Such a law should also specifically authorize law enforcement to share evidence with platforms of foreign influence campaigns, especially those targeted against elections, and recognize the First Amendment rights of non-government researchers.

- **Support Private Efforts Towards Digital Provenance & Transparency** - Digital provenance standards for identifying the origin and changes to content should be the initial focus for government and industry as an essential first step for developing watermarking, disclosure, and similar transparency measures. Congress can financially support the development of open-source and free tools to detect potential fakes, and create liability shields for companies taking good-faith steps to label or remove content they detected as fake or manipulated.

- **Address Non-Political Abuses of Generative AI** - As I pointed out above, the most impactful abuses of generative AI technologies today are those targeted at harming individuals, such as the creation of fake, non-consensual nudes. Congress must not forget these victims of targeted abuse as it addresses the important issues involving elections and democracy.

**Conclusion**

Current discourse often frames foreign interference in elections as a partisan issue. This framing is stuck in the aftermath of the 2016 election. There is an assumption that the most significant foreign actor targeting U.S. elections is Russia and that foreign propaganda only benefits Republicans and hurts Democrats. This is not only incorrect, but counterproductive to democratic values.

Heading into the U.S. presidential election next year, China and Iran have built up their online propaganda programs and will surely utilize generative AI. These foreign efforts, in addition to Russia's, are complex. They target both political parties and seek to sow chaos. All Americans who value free and fair elections are best served by bipartisan cooperation on this issue.

[25] Nehamas, N. (2023, June 8). *DeSantis Campaign Uses Apparently Fake Images to Attack Trump on Twitter*. The New York Times. https://www.nytimes.com/2023/06/08/us/politics/desantis-deepfakes-trump-fauci.html