Testimony of Alexandra Reeve Givens
President & CEO, Center for Democracy & Technology

For the U.S. Senate Committee on the Judiciary Subcommittee on Human Rights and the Law
Hearing Entitled "Artificial Intelligence and Human Rights"

June 13, 2023

Chair Ossoff, Ranking Member Blackburn and other members of the Subcommittee, thank you for inviting me to testify today on the important issue of AI and human rights. The world's attention is rightly focused on the possibilities and the risks of AI systems. As policymakers look to address potential harms and promote responsible innovation, it is essential that they do so with a focus on human rights – and in particular, with the conviction that fundamental rights and freedoms belong inalienably to all people, including the rights to liberty, privacy, freedom of expression and opinion, peaceful assembly, and equal treatment before the law.[1]

AI systems are already being used in ways that threaten these rights, and rapid advancements in generative AI and text and image analysis will exacerbate the risks. Today I will focus on two distinct areas where AI harms are already being felt: the use of face recognition and biometric surveillance capabilities by law enforcement, and the impact of generative AI on elections and democratic discourse. For reasons I will explain in my testimony, these applications of AI are vastly different from one another, with different considerations at stake as Congress considers appropriate policy interventions.

Of course, these areas are not the only ways in which AI is impacting human rights. In previous testimony before the U.S. Senate Committee on Homeland Security and Government Affairs, I described risks posed by AI systems to people's civil rights and access to economic opportunities – for example when people are applying for jobs, housing, or credit – and potential policy responses.[2] I also described how AI is being used in ways that jeopardize the fair administration of public benefits programs, and steps the government should take to protect people's access to basic services and due process rights. Those issues are ripe and important priorities for government intervention.

At a time when many are discussing the long term existential risks of AI systems, there are concrete issues on which Congress and the U.S. government can act *today* — and, in doing so, demonstrate what it means to ensure AI is developed in a manner that centers democratic values and human rights.

[1] United Nations General Assembly. The Universal Declaration of Human Rights (UDHR). New York: United Nations General Assembly, 1948.
[2] Alexandra Reeve Givens, "Press Release: In Senate Testimony, CDT CEO Alexandra Givens Calls For Cross-Society Effort in Addressing Risks of AI", Center for Democracy & Technology, March 8, 2023, https://cdt.org/insights/press-release-in-senate-testimony-cdt-ceo-alexandra-givens-calls-for-cross-society-effort-in-addressing-risks-of-ai/.

AI & Government Surveillance

Last fall, many of us were inspired by the images of brave Iranian women protesting the death of 22-year-old Mahsa Amini after she was arrested for allegedly improperly wearing the hijab. But we were not the only ones watching those protests. In Iran today, face recognition technology allows the government to identify protestors and take action against them. Demonstrators have received text messages from local police stating that they were observed at a protest and should not join further demonstrations.[3] Iranian officials also announced that they would use face recognition in public spaces to detect and identify women who were not "correctly" wearing a hijab.[4] A member of parliament explained that women who dress improperly would receive text message warnings, followed by penalties such as their bank accounts being blocked. In Iran, citizens must use biometric national identity cards to receive pensions and food rations, open bank accounts and access the domestic internet – making these threats of automated punishments all too real. In this context, AI systems are enabling a repressive regime to identify dissenters, subject them to pervasive surveillance, and then automate their punishment.

Face recognition technology has been used in similar ways by the Chinese government, to promote social control through mass enforcement and public shaming of minor offenses such as jaywalking,[5] as well as for its notorious treatment of China's Uyghur minority.[6] Face recognition has also been used to identify protestors in Russia, Hong Kong and Uganda, among other countries.[7]

Such examples may feel far from the United States, but the technical capabilities exist here, and we do not have adequate legal frameworks to address them. In the U.S. there have already been abuses: In 2020, police in multiple Florida cities used facial recognition to identify and catalog activists engaging in peaceful civil rights protests supporting the Black Lives Matter movement.[8] In Baltimore, face recognition technology was used in real time to target people who were protesting after the death of Freddie Gray, with law enforcement scanning the crowd to identify individuals with outstanding warrants for unrelated offenses, and arresting them on site.[9] When

---

[3] Sam Biddle and Murtaza Hussain, "Hacked Documents: How Iran Can Track And Control Protesters' Phones", *The Intercept*, Oct. 28, 2022, https://theintercept.com/2022/10/28/iran-protests-phone-surveillance/.

[4] Khari Johnson, "Iran to use facial recognition to identify women without hijabs", *Ars Technica*, Jan. 11, 2023, https://arstechnica.com/tech-policy/2023/01/iran-to-use-facial-recognition-to-identify-women-without-hijabs/.

[5] Alfred Ng, "How China uses facial recognition to control human behavior", *CNET*, Aug. 11, 2020, https://www.cnet.com/news/politics/in-china-facial-recognition-public-shaming-and-control-go-hand-in-hand/ ("The punishing of these minor offenses is by design, surveillance experts said. The threat of public humiliation through facial recognition helps Chinese officials direct over a billion people toward what it considers acceptable behavior, from what you wear to how you cross the street").

[6] Paul Mozur, "One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority", *The New York Times*, Apr. 14, 2019, https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html.

[7] Paul Mozur, "In Hong Kong Protests, Faces Become Weapons", *The New York Times*, July 26, 2019, https://www.nytimes.com/2019/07/26/technology/hong-kong-protests-facial-recognition-surveillance.html; Lena Masri, "Facial recognition is helping Putin curb dissent with the aid of U.S. tech", *Reuters*, Mar. 28, 2023, https://www.reuters.com/investigates/special-report/ukraine-crisis-russia-detentions/; Stephen Kafeero, "Uganda is using Huawei's facial recognition tech to crack down on dissent after anti-government protests", *Quartz*, Nov. 27, 2020, https://qz.com/africa/1938976/uganda-uses-chinas-huawei-facial-recognition-to-snare-protesters.

[8] Joanne Cavanaugh Simpson and Marc Freeman, "South Florida police quietly ran facial recognition scans to identify peaceful protestors. Is that legal?", *Sun Sentinel*, June 26, 2021, https://www.sun-sentinel.com/2021/06/26/south-florida-police-quietly-ran-facial-recognition-scans-to-identify-peaceful-protestors-is-that-legal/.

[9] Kevin Rector and Alison Knezevich, "Social media companies rescind access to Geofeedia, which fed information to police during 2015 unrest", *The Baltimore Sun*, Oct. 11, 2016, https://www.baltimoresun.com/news/crime/bs-md-geofeedia-update-20161011-story.html.

face recognition is used in this way, it violates people's rights to freedom of expression and peaceful assembly. Congress must act to rein it in.

Facial recognition technology is becoming more widely available and cheaper to use. A study by Georgetown's Center on Privacy and Technology published in 2016 showed that at least one in four state and local law enforcement agencies had access to facial recognition – and that was seven years ago.[10] Research suggests that the FBI conducts thousands of scans per month, matched against reference databases of hundreds of millions of photos.[11] Several years ago, Americans were shocked to learn about the practices of the private company Clearview AI, which claims to have scraped over 20 *billion* photographs from the internet to power its face recognition systems.[12] Clearview has now been used by over 3000 federal, state and local law enforcement agencies in the United States to provide facial recognition services.[13]

Policymakers should treat facial recognition as a priority because it is a double-edged sword: Facial recognition is dangerous when it works poorly, and dangerous in an entirely different way when it works well. States have begun to respond to this threat, with over a dozen enacting meaningful limits and some jurisdictions banning the technology.[14] It is critical that Congress act as well. As our nation considers its approach to governing AI, this is an area where Congress could draw a clear contrast to autocratic regimes, demonstrating America's commitment to human rights.

The urgent need for regulation of facial recognition technology is clear. Facial recognition misidentifications have already caused numerous innocent people to be wrongfully arrested and jailed. Most recently, Randel Reid was held for six days in a Georgia jail because a facial recognition system misidentified him,[15] the latest in a series of known cases.[16] Because of police overreliance on AI, these individuals faced indignity, deprivation of liberty, and lasting harms such as loss of employment, steep legal fees, and mental trauma.[17] And since police use of facial recognition is often hidden,[18] these incidents likely represent just the tip of the iceberg.[19]

---

[10] The Perpetual Line-Up: Unregulated Police Face Recognition in America, Georgetown Law Center on Privacy and Technology, Oct. 18, 2016, https://www.perpetuallineup.org/.
[11] *Id.*; *see also* Charlie Osborne, "FBI, ICE plunder DMV driver database 'gold mine' for facial recognition scans", *ZDNET*, July 8, 2019, https://www.zdnet.com/article/fbi-and-ice-are-using-dmv-gold-mine-for-facial-recognition-scans/.
[12] Kashmir Hill, "Your Face is Not Your Own", *The New York Times Magazine*, Mar. 18, 2021, https://www.nytimes.com/interactive/2021/03/18/magazine/facial-recognition-clearview-ai.html.
[13] *Id.*
[14] Jake Laperruque, "Limiting Face Recognition Surveillance: Progress and Paths Forward", Center for Democracy & Technology, Aug. 23, 2022, https://cdt.org/insights/limiting-face-recognition-surveillance-progress-and-paths-forward/.
[15] Kashmir Hill and Ryan Mac, "'Thousands of Dollars for Something I Didn't Do'", *The New York Times*, Mar. 31, 2023, https://www.nytimes.com/2023/03/31/technology/facial-recognition-false-arrests.html.
[16] Khari Johnson, "How Wrongful Arrests Based on AI Derailed 3 Men's Lives", *WIRED*, Mar. 7, 2022, https://www.wired.com/story/wrongful-arrests-ai-derailed-3-mens-lives/.
[17] *Id.*; *see also* Elaisha Stokes, "Wrongful arrest exposes racial bias in facial recognition technology", *CBS News*, Nov. 19, 2020, https://www.cbsnews.com/news/detroit-facial-recognition-surveillance-camera-racial-bias-crime/; Kashmir Hill, "Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match", *The New York Times*, Dec. 29, 2020, https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html.
[18] Khari Johnson, "The Hidden Role of Facial Recognition Tech in Many Arrests", *WIRED*, Mar. 7, 2022, https://www.wired.com/story/hidden-role-facial-recognition-tech-arrests/; Jennifer Valentino-DeVries, "How the Police Use Facial Recognition, and Where It Falls Short", *The New York Times*, Jan. 12, 2020, https://www.nytimes.com/2020/01/12/technology/facial-recognition-police.html.
[19] Disturbingly, some facial recognition misidentifications likely have resulted in prison time for innocent persons, either wrongfully convicted or pressured to accept a plea bargain out of fear of long sentences or extended time in pretrial detention.

Misidentification stems from a range of causes. Most facial recognition systems display algorithmic bias; studies have repeatedly shown propensity to misidentify people of color and women at higher rates than white people and men.[20] Software settings and nature of use impact accuracy as well. Many law enforcement agencies, including the FBI, set their systems to return several potential matches for *every* facial recognition scan even if the "confidence threshold"—meaning the required level of certainty to list an individual as a possible match—is unreliably low.[21] Law enforcement also regularly uses dubious methods to alter or replace images before scanning, from using CGI to artificially fill in uncaptured portions of a face, to replacing photos entirely with a composite sketch or celebrity look alike.[22] Finally, accuracy can vary significantly based on image quality: Lighting, photo resolution, distance, camera angle, and facial obstructions can all have a major impact on whether facial recognition returns accurate matches.[23] This is critical because even if algorithmic bias were solved, and responsible settings and use parameters were employed, varying image quality will always cause misidentification risk.

Just as serious as misidentifications are the dangers of accurate facial recognition being used for surveillance. The examples I shared previously from Iran, China, Russia, Uganda – and at least three U.S. cities – shows how easily face recognition technology can impinge on people's rights to express themselves through protest and to peacefully assemble. Facial recognition could be employed to monitor, catalog, and engage in disparate targeting of individuals participating in a variety of sensitive or constitutionally protected activities, such as attending a political rally, going to a house of worship, purchasing a firearm from a licensed shop, or visiting a medical clinic. Absent strong limits, law enforcement authorities could misuse AI technology to track and catalog individuals' most sensitive activities with little effort, and on an unprecedented scale. The U.S. must show leadership by curtailing such a direct assault on civil liberties.

Given the range of risks facial recognition poses to civil rights and civil liberties, there is not a silver bullet policy solution: lawmakers need to enact a broad set of safeguards to prevent harm,

---

[20] Joy Buolamwini and Timnit Gebru (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, *Fairness, Accountability and Transparency, Proceedings of Machine Learning Research* 81:77-91. http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf; Patrick Grother, Mei Ngan, and Kayee Hanaoka (Dec. 2019). *Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects*, National Institute Of Science and Technology. https://doi.org/10.6028/NIST.IR.8280.

[21] Kimberly J. Del Greco, "Facial Recognition Technology: Ensuring Transparency in Government Use", Federal Bureau of Investigation, June 4, 2019, https://www.fbi.gov/news/testimony/facial-recognition-technology-ensuring-transparency-in-government-use; Drew Harwell, "Oregon became a testing ground for Amazon's facial-recognition policing. But what if Rekognition gets it wrong?", *The Washington Post*, April 30, 2019, https://www.washingtonpost.com/technology/2019/04/30/amazons-facial-recognition-technology-is-supercharging-local-police/ ("But deputies here are not shown that search-confidence measurement when they use the tool. Instead, they are given five possible matches for every search, even if the system's certainty in a match is far lower").

[22] James O'Neill, "How Facial Recognition Makes You Safer", *The New York Times*, June 9, 2019, https://www.nytimes.com/2019/06/09/opinion/facial-recognition-police-new-york-city.html; Clare Garvie, "Garbage In, Garbage Out | Face Recognition on Flawed Data", Georgetown Law Center on Privacy & Technology, May 16, 2019, https://www.flawedfacedata.com/ ("One detective from the Facial Identification Section (FIS), responsible for conducting face recognition searches for the NYPD, noted that the suspect looked like the actor Woody Harrelson …. A Google image search for the actor predictably returned high-quality images, which detectives then submitted to the face recognition algorithm in place of the suspect's photo.")

[23] The Constitution Project's Task Force on Facial Recognition Surveillance and Jake Laperruque, "Facing the Future of Surveillance", Project on Government Oversight, Mar. 4, 2019, https://www.pogo.org/report/2019/03/facing-the-future-of-surveillance.

both from misidentifications and misuse.[24] The Center for Democracy & Technology views the following measures as key to effectively regulating law enforcement use of facial recognition:

1) **A warrant rule:** Law enforcement use of facial recognition should require obtaining a warrant from a judge, based on probable cause that the individual to be scanned committed a crime.[25] Warrants are a fundamental privacy safeguard and key to preventing abuse, notably using facial recognition to identify, catalog, and target individuals engaged in lawful and sensitive activities, such as protests.
2) **A serious crime limit:** Face recognition technology should be restricted for use only in investigating serious offenses.[26] Such limitations would prevent selective targeting and prosecution, as well as prevent misidentifications in scenarios least likely to receive due scrutiny: the investigation and prosecution of low-level crimes.
3) **Notification for arrested individuals:** Law enforcement should not be allowed to routinely hide their use of facial recognition from defendants and the broader public.[27] This common practice undermines defendants' due process rights, and prevents examination of errors and other meaningful oversight.
4) **Prohibiting overreliance on matches:** Police should not be permitted to use facial recognition as the sole basis for arrests or other police actions. Given that the technology's accuracy varies significantly based on a range of factors, independent investigative work is essential.
5) **Prohibiting untargeted scans:** Facial recognition technology may soon focus on untargeted scans—whereby every individual passing through a video feed is identified with facial recognition—but this method is far too unreliable for law enforcement use. Pilot programs have produced false positives of 81 to 96 percent.[28] Even if these extreme error rates were to improve, such a use of face recognition technology would constitute unacceptable dragnet surveillance that should not be deployed.
6) **Testing and accuracy standards:** Any law enforcement use of facial recognition should require that software be subject to independent testing and meet accuracy standards. Testing should focus on live field conditions that replicate investigative use, and accuracy standards should limit use to algorithms with highest overall accuracy and that display no variance based on demographic traits.

---

[24] While our recommendations focus on safeguards and limits for law enforcement use of facial recognition, it is important to acknowledge that many privacy, civil rights, and civil liberties groups—including CDT—have called for a moratorium on facial recognition, or for its use by law enforcement to be banned entirely. Some local face recognition laws have taken this approach. CDT supports enacting a moratorium while evaluating proper restrictions and safeguards as providing the strongest protections for civil rights and civil liberties. *See, e.g.*, LDF Letter re: July 13, 2021 Subcommittee on Crime, Terrorism, and Homeland Security Hearing on Law Enforcement Use of Facial Recognition Technology, https://www.naacpldf.org/wp-content/uploads/2021.07.20-LDF-Statement-on-Law-Enforcement-U_Emily-Fisher-1.pdf.

[25] This should include sensible limited exceptions, such as identifying victims and incapacitated persons.

[26] A serious crime limit has been used for over 50 years to prevent wiretap surveillance from becoming pervasive. *See* 18 U.S.C. § 2516.

[27] Khari Johnson, "The Hidden Role of Facial Recognition Tech in Many Arrests", *WIRED*, Mar. 7, 2022, https://www.wired.com/story/hidden-role-facial-recognition-tech-arrests/; Jennifer Valentino-DeVries, "How the Police Use Facial Recognition, and Where It Falls Short", *The New York Times*, Jan. 12, 2020, https://www.nytimes.com/2020/01/12/technology/facial-recognition-police.html.

[28] Lizzie Dearden, "Facial recognition wrongly identifies public as potential criminals 96% of time, figures reveal", *The Independent*, May 7, 2019, https://www.independent.co.uk/news/uk/home-news/facial-recognition-london-inaccurate-met-police-trials-a8898946.html; Rachel England, "UK police's facial recognition system has an 81 percent error rate", *Engadget*, July 4, 2019, https://www.engadget.com/2019-07-04-uk-met-facial-recognition-failure-rate.html.

The adoption of face recognition laws by over a dozen states[29] demonstrates an emerging consensus for regulating this surveillance. Unfortunately, thus far Congress has placed no limits on facial recognition, leaving this powerful technology unrestricted. Last year a bill was introduced in the House, H.R. 9061, The Facial Recognition Act, that included many of the recommendations listed above, and that the Center for Democracy & Technology endorsed.[30] We encourage Congress to act with urgency to place safeguards on this form of AI surveillance, and focus on the policies described above.

Generative AI, Elections & Democratic Discourse

Turning to my second area of focus, rapid advances in generative AI are spurring creativity and innovation, but also raise significant threats for human rights. Already there have been instances showing the professional, reputational and potential physical harms that may arise when people rely on generated results as accurate, not accounting for the likelihood of "hallucinations", or mistaken results.[31] Generative AI tools are likely to exacerbate fraud, as tools make it easier to quickly generate massive amounts of convincing text, as well as personalized scams, or to trick people by impersonating a familiar voice.[32] Deepfakes – videos or images that have been digitally manipulated to misrepresent the voice and likeness of another person – can misrepresent public figures or events in a way that threatens elections, national security, and general public order.[33] Deepfakes can also be used to defraud, harass, and extort people.[34] None of these harms is new, but they are made cheaper, faster, and more effective by the ease, speed and widespread accessibility of generative AI tools.

The threats to elections and democratic discourse are particularly worth highlighting. In previous elections, operatives used robocalls to spread incorrect information about mail-in voting in an effort to suppress Black voter turnout,[35] and deceptive text messages to spread intentionally misleading voting instructions for a Kansas ballot initiative in 2022.[36] It is easy to imagine bad actors using AI to exponentially grow and personalize voter suppression or other targeting efforts, increasing their harmful impact. Today, consumers can often spot a scam email, text or robocall because it uses non-personalized language and there may be grammatical

---

[29] Jake Laperruque, "Limiting Face Recognition Surveillance: Progress and Paths Forward", Center for Democracy & Technology, Aug. 23, 2022, https://cdt.org/insights/limiting-face-recognition-surveillance-progress-and-paths-forward/.
[30] Jake Laperruque, "The Facial Recognition Act: A Promising Path to Put Guardrails on a Dangerously Unregulated Surveillance Technology", *Lawfare*, Nov. 1, 2022, https://www.lawfareblog.com/facial-recognition-act-promising-path-put-guardrails-dangerously-unregulated-surveillance-technology.
[31] Karen Weise and Cade Metz, "When A.I. Chatbots Hallucinate", *The New York Times*, May 1, 2023, https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucination.html.
[32] Steve Mollman, "Scammers are using voice-cloning A.I. tools to sound like victims' relatives in desperate need of financial help. It's working", *Fortune*, Mar. 5, 2023, https://fortune.com/2023/03/05/scammers-ai-voice-cloning-tricking-victims-sound-like-relatives-needing-money/.
[33] Shannon Bond, "Fake viral images of an explosion at the Pentagon were probably created by AI", *NPR*, May 22, 2023, https://www.npr.org/2023/05/22/1177590231/fake-viral-images-of-an-explosion-at-the-pentagon-were-probably-created-by-ai; David Klepper and Ali Swenson, "AI presents political peril for 2024 with threat to mislead voters", *AP News*, May 14, 2023, https://apnews.com/article/artificial-intelligence-misinformation-deepfakes-2024-election-trump-59fb51002661ac5290089060b3ae39a0.
[34] *See e.g.,* Henry Ajder, Giorgio Patrini and Francesco Cavalli, "Automating Image Abuse: Deepfake bots on Telegram", *Sensity*, Oct. 2020 (deepfake bots on Telegram digitally "undress" more than 100,000 women on the platform); Thomas Brewster, "Fraudsters Cloned Company Director's Voice In $35 Million Heist, Police Find", *Forbes*, Oct. 14, 2021, https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/?sh=7d29a3f87559 (audio deepfake of executives' voices used to steal millions of dollars from companies).
[35] Christine Chung, "They Used Robocalls to Suppress Black Votes. Now They Have to Register Voters.", *The New York Times*, Dec. 1, 2022, https://www.nytimes.com/2022/12/01/us/politics/wohl-burkman-voter-suppression-ohio.html.
[36] Isaac Stanley-Becker, "Misleading Kansas abortion texts linked to Republican-aligned firm", *The Washington Post*, Aug. 2, 2022, https://www.washingtonpost.com/politics/2022/08/02/kansas-abortion-texts/.

or language errors (or, in the case of robocalls, a notably automated voice). Generative AI tools will make it easier to create tailored, accurate, realistic messages that draw victims in.

Generated images can also twist public understanding of political figures and events. Recordings of public figures' voices have been manipulated to trick senior government officials into thinking they are speaking with government leaders.[37] Videos and images have been digitally altered to make public officials appear incompetent, compromised, or to misrepresent their policy positions.[38] Experts have warned how deepfakes, which are difficult to authenticate or rebut, could impact an election in the closing days of voting, when there is little time to set the record straight, or before a debate.[39] More generally, the growth of inauthentic content makes it harder for people to know what news and content they can trust, such that even authentic content is undermined. Journalists, whistleblowers, and human rights defenders are experiencing these effects already, facing higher hurdles than ever before to establish and defend their credibility.[40]

While the rise of affordable generated content poses new threats to public discourse, policy interventions must be approached with care. This is because there are many legitimate reasons why people use software to generate and alter content: from laypeople and artists using AI to make creative works; to people engaging in parody; actors being de-aged in a movie; voices being sampled for a music track; or researchers altering images of North American and European cities to show what they would look like if they faced the same bombardment as the cities attacked in the Syrian war.[41] Barring or heavily restricting such activities would harm free expression, creativity and innovation, and quickly run afoul of the First Amendment.

Efforts to restrict or condition the distribution of generative images may also suppress protected expressive activities. To give one example, in recent years a number of companies and stakeholders have come together in the Content Authenticity Initiative, an impressive undertaking that allows photographers and other content creators to attach immutable provenance signals showing the authenticity of their work (such as details of the image's creator, date/time/location, tracked edits and more).[42] This is a creative solution to help newspapers, human rights watchdogs and others reassure the public about the authenticity and provenance of images they create and display. But *mandating* the use of such an authenticity standard (or

---

[37] *See e.g.,* Bobby Allyn, "Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn", *NPR*, Mar. 16, 2022, https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia (the minute long deepfake video "shows a rendering of the Ukrainian president appearing to tell his soldiers to lay down their arms and surrender the fight against Russia"); Philip Oltermann, "European politicians duped into deepfake video calls with mayor of Kyiv", *The Guardian*, Jun. 25, 2022, https://www.theguardian.com/world/2022/jun/25/european-leaders-deepfake-video-calls-mayor-of-kyiv-vitali-klitschko.

[38] *See e.g.,* Hannah Denham, "Another fake video of Pelosi goes viral on Facebook", *The Washington Post,* Aug. 3, 2020, https://www.washingtonpost.com/technology/2020/08/03/nancy-pelosi-fake-video-facebook/ (video depicts Pelosi slurring her speech and appearing intoxicated"); Alexandra Ulmer and Anna Tong, "Deepfaking it: America's 2024 election collides with AI boom", *Reuters*, May 30, 2023, https://www.reuters.com/world/us/deepfaking-it-americas-2024-election-collides-with-ai-boom-2023-05-30/; Zeke Miller, "Rubio Campaign Fires Back at Cruz Over Photoshopped Image", *Time*, Feb. 18, 2016, https://time.com/4229092/marco-rubio-ted-cruz-photoshop/. While running for re-election in 2019, Houston's mayor said a critical ad ran by a fellow candidate broke a Texas law that bans certain misleading political deepfakes. Ivory Hecker, "Mayor Turner calls for criminal investigation of Tony Buzbee's attack ad", *Fox 26 Houston*, Oct. 17, 2019, https://www.fox26houston.com/news/mayor-turner-calls-for-criminal-investigation-of-tony-buzbees-attack-ad.

[39] James Bickerton, "Deepfakes Could Destroy the 2024 Election", *Newsweek*, Mar. 24, 2023, https://www.newsweek.com/deepfakes-could-destroy-2024-election-1790037.

[40] Sam Gregory, "Tracing trust: Why we must build authenticity infrastructure that works for all", *Witness*, May 2020, https://blog.witness.org/2020/05/authenticity-infrastructure/.

[41] Tiffany Hsu, "As Deepfakes Flourish, Countries Struggle With Response", *The New York Times*, Jan. 22, 2023, https://www.nytimes.com/2023/01/22/business/media/deepfake-regulation-difficulty.html.

[42] *See* Content Authenticity Initiative, https://contentauthenticity.org/.

prohibiting the distribution of materials without such standards) would be deeply problematic, because it would suppress the posting and sharing of lawful images whose creators lacked the resources or awareness to use a provenance tool, who face safety risks if their work can be traced back to them, or who simply do not want to do so.

The challenges of regulating deepfakes does not mean policymakers must sit idle. To the contrary, there are concrete steps Congress can take to increase transparency and accountability in the design, development and use of generative AI tools, as well as appropriations provisions, oversight of relevant federal agencies, and steps such as hearings, convenings, and/or the creation of a Commission to highlight best practices and novel innovations to address potential harms.

1) **Mandating transparency & disclosures of AI risks.** Several legislative proposals introduced last Congress seek to increase the accountable design and transparency of AI systems, including the Algorithmic Accountability Act, and the algorithmic impact assessment provision of the bipartisan American Data Privacy & Protection Act. These measures were drafted before the wide-scale public release of generative AI systems, but their principles lay an important foundation for future work.

   As a starting point, Congress could require the developers of AI systems that can be used in high-risk settings to disclose how their tools are developed and designed, to test them using frameworks based on principles such as those set out in the Blueprint for an AI Bill of Rights and the NIST AI Risk Management Framework, and to share the analysis of those tests with an outside regulator (with some version made available for the public and for independent researchers, balancing concerns about the potential privacy and safety aspects of such disclosures). Such steps would increase transparency and support meaningful public dialogue about how tools are developed and governed. They would also normalize the principle that companies designing and deploying AI tools *must* analyze and document how they work, identify potential risks, and disclose the steps they have taken to mitigate those risks. Such legislation would establish an essential baseline, and need not foreclose potential legislation on minimum design and safety standards, the specific regulation of highly capable foundation models, or further steps to address other high-risk AI uses.

2) **Examining how existing criminal and civil laws map onto harms created by new tools, and filling gaps.** In some instances, the appropriate framework to address harms created by generative AI (and other AI systems) may be litigation under existing laws. For example, people who use AI to perpetrate scams could be prosecuted for fraud, extortion, or harassment; face investigation by the Federal Trade Commission for unfair and deceptive trade practices or the Federal Elections Commission for violating campaign laws; or face civil litigation for claims such as fraud, intentional infliction of emotional distress, harassment, defamation and intellectual property violations. Congress should monitor whether these existing legal frameworks adequately address emerging harms.[43]

[43] Four federal agencies recently announced their efforts to enforce existing laws to protect the American public from AI-related harms. Other agencies should take similar steps, and Congressional committees of relevant jurisdiction can support these efforts to understand how existing

In assessing liability, courts will have to tackle the complex question of whether and when developers of generative AI tools bear legal liability for the content those tools produce. Courts will have to consider whether the content generated by an AI tool is properly considered to be the speech of the user who prompted its creation, or something partially or wholly created or developed by the AI tool itself. This will likely differ depending on the fact pattern: for example, whether a user inputted specific prompts aiming to generate the content that gave rise to litigation, such as soliciting a list of crimes committed by a private individual and publishing that list with reckless disregard for whether the information was true, or whether the AI tool was the source of the content giving rise to litigation, , such as making up dangerously incorrect medical advice in response to a query. In addition to statutes and case law regarding intermediary liability protections, courts will need to consider a range of common law principles from across civil and criminal law, including standards for aiding and abetting liability, and questions of knowledge and intent for both the user and the developer (and, if different, the deployer) of the tool. Companies will need to point to content policies and technical safeguards they have in place to mitigate foreseeable misuses and other harms.

As courts grapple with these and other complex issues, Congress can shine a light and drive public discourse — and then act as appropriate to fill in the gaps. This could include hearings and reports by Congressional committees in their areas of jurisdiction, commissioning reports by the GAO or federal agencies, or, more formally, the creation of an expert Commission to advance such work.[44]

3) **Advancing best practices for responsible design and governance of generative AI systems**. There is an urgent need for companies developing generative AI systems to develop robust safety processes and other governance measures, as many of their CEOs have themselves publicly declared.[45] This can include steps ranging from well-developed content policies and technical safeguards that limit the creation of certain high-risk content or uses of the technology;[46] robust pre- and post-release testing to identify and address bias and potential harms; improved interfaces, labeling and product descriptions to better educate

---

laws map onto novel fact patterns. *See* Joint Statement on Enforcement Efforts Against Discrimination and Bias in Automated Systems, Apr. 25, 2023, https://www.ftc.gov/system/files/ftc_gov/pdf/EEOC-CRT-FTC-CFPB-AI-Joint-Statement%28final%29.pdf.

[44] *See, e.g.,* Deepfake Task Force Act, S.2559, 117th Cong. (2021-2022); American Data Privacy & Protection Act of 2022 (ADPPA), H.R. 8152, 117th Cong. (2021-2022). These proposals both focus on creating a task force (or in the case of the ADPPA, mandating annual reporting by the Commerce Department) on the uses and harms of deepfakes and advancements in deepfake detection technology. But a Commission could also be charged with reporting on and assessing existing legal frameworks for addressing and seeking redress for other harms.

[45] *See e.g.,* Sam Altman, Oversight of A.I.: Rules for Artificial Intelligence Hearing before the U.S. Senate Committee on the Judiciary Subcommittee on Privacy, Technology, & the Law, 118th Cong. (2023), https://www.judiciary.senate.gov/committee-activity/hearings/oversight-of-ai-rules-for-artificial-intelligence; Sundar Pichai, "Why Google thinks we need to regulate AI", *Financial Times*, Jan. 20, 2020, https://www.ft.com/content/3467659a-386d-11ea-ac3c-f68c10993b04 (CEO of Google stating that "there is no question in [his] mind that artificial intelligence needs to be regulated"); Brad Smith, "Meeting the AI moment: advancing the future through responsible AI", Microsoft, Feb. 2, 2023, https://blogs.microsoft.com/on-the-issues/2023/02/02/responsible-ai-chatgpt-artificial-intelligence/ (Vice Chair & President of Microsoft calling for effective AI regulations that "center on the highest risk applications and be outcomes-focused and durable").

[46] For example, OpenAI claims that its image generator DALL-E cannot create images of public figures, and that it restricts any "scaled" usage of its products for political purposes, such as the use of its AI to send out mass personalized emails to constituents. Reporters testing these claims have found significant exceptions and workarounds. Robust, well-tested and publicly disclosed content policies form an important aspect of safety testing. Alexandra Ulmer and Anna Tong, "Deepfaking it: America's 2024 election collides with AI boom", *Reuters*, May 30, 2023, https://www.reuters.com/world/us/deepfaking-it-americas-2024-election-collides-with-ai-boom-2023-05-30/.

users about the systems' limitations and risks of inaccurate results;[47] safeguarding systems against security threats, and more.

Governments in different countries are pressing companies on what these steps should look like.[48] Whether or not these steps are ripe for legislation, Congress can play a role in driving forward these efforts – and, most critically, ensuring they are not taking place behind closed doors with only companies in attendance, but instead with meaningful participation from civil society and independent sources of expertise.

4) **Scaling agencies' capacity to address deepfakes and boost authentic sources of information.** It has long been said that the best remedy to combat undesirable speech is counterspeech[49] – but in our cacophonous information ecosystem, it takes work for counterspeech to be effective. There are steps policymakers can take to mature the systems that can help individuals better understand content authenticity and identify reliable sources of information. As one step, the government could increase funding and other efforts to support the development of technologies that assist in deepfake detection.[50] Policymakers could also support and foster awareness of voluntary efforts to authenticate content, funding research projects through the National Science Foundation and other programs, or raising awareness of private sector efforts to encourage the quick development of such work.[51]

   Critically, Congress and the Administration should significantly ramp up efforts to equip key institutions so they can identify and debunk manipulated content that threatens national security, financial markets, election administration, public health and similar priority areas. The bipartisan Deepfake Task Force Act proposed last Congress provides a good bipartisan foundation from which to start. That measure directed the creation of a task force comprised of government and non-government experts to "investigate the feasibility of, and obstacles to, developing and deploying standards and technologies for determining digital content provenance", and created "a formal mechanism for interagency coordination and information sharing to facilitate the creation and implementation of a national strategy to address the growing threats posed by digital content forgeries."[52]

---

[47] Michal Luria, "Your ChatGPT Relationship Status Shouldn't Be Complicated", *WIRED*, Apr. 11, 2023, https://www.wired.com/story/chatgpt-social-roles-psychology/.

[48] Ryan Browne, "With ChatGPT hype swirling, UK government urges regulators to come up with rules for A.I.", *CNBC*, Mar. 29, 2023, https://www.cnbc.com/2023/03/29/with-chatgpt-hype-swirling-uk-government-urges-regulators-to-come-up-with-rules-for-ai.html; Ryan Browne, "Europe takes aim at ChatGPT with what might soon be the West's first A.I. law. Here's what it means", *CNBC*, May 15, 2023, https://www.cnbc.com/2023/05/15/eu-ai-act-europe-takes-aim-at-chatgpt-with-landmark-regulation.html. In the U.S., the White House issued the AI Bill of Rights in October 2022 and the National Institute of Standards and Technology (NIST) followed in January 2023 with an AI Risk Management Framework, and officials have spoken about ways in which these map onto the risks posed by generative AI. *See* Blueprint for an AI Bill of Rights, https://www.whitehouse.gov/ostp/ai-bill-of-rights/; National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

[49] *Whitney v. Cal.*, 274 U.S. 357, 377 (1927) (Brandeis, J., concurring) ("If there be time to expose through discussion the falsehood and fallacies, to avert the evil by the processes of education, the remedy to be applied is more speech, not enforced silence.").

[50] *See, e.g.,* IOGAN Act, Pub. L. No. 116-258 (2020), directing the National Science Foundation and the National Institute of Standards and Technology (NIST) to support research on generative adversarial networks. The proposed American Data Privacy & Protection Act of 2022 would have required the Secretary of Commerce to publish an annual report on common sources of digital content forgeries, an assessment of the uses, applications and harms of digital content forgeries, and an analysis of the methods and standards available for detection and counter-measures such as labeling. American Data Privacy & Protection Act of 2022, H.R. 8152, Section 305, 117th Cong. (2021-2022).

[51] Shirin Ghaffary, "What will stop AI from flooding the internet with fake images?, *Vox*, Jun. 3, 2023, https://www.vox.com/technology/23746060/ai-generative-fake-images-photoshop-google-microsoft-adobe.

[52] Section 5709 of the National Defense Authorization Act of 2020 also took steps to improve government agency awareness and competency to address deepfakes. It directed the Director of National Intelligence to produce a report on the technological capabilities of foreign actors with

Capacity-building efforts could also include funding training, resources and using oversight pressure to ensure public institutions take steps to best earn public trust when they speak out. To give one simple example, research by my organization, the Center for Democracy & Technology, revealed that only in 1 in 4 official election websites uses the trusted ".gov" domain managed by DHS, while other election officials use ".com" domains that can be easily spoofed. The result is to undermine the role of such websites as a source for people to access trusted information about the administration of elections. Funding, education and oversight could help election officials address this simple vulnerability.

Conclusion

The examples of face recognition and misleading information about elections show two very different ways in which AI is already impacting Americans' human rights and the structure of our democracy. Critically, these examples show that there are concrete steps policymakers can take, today, to address the potential harms that can arise from certain uses of AI. As commentators around the world assess the existential threats posed by AI systems, it is important to remember that existential threats can also include threats to the fabric of society: undermining individual rights, equality and economic mobility, and an informed public discourse that is the bedrock of a functioning democracy. On many of these issues, there are steps that technology companies, regulatory agencies and Congress can take right now to address and reduce AI-driven harms. Thank you for the opportunity to share these thoughts today.

respect to "machine-manipulated media, machine generated text, generative adversarial networks, and related machine-learning technologies", and analysis of the counter-technologies that have been or could be developed and deployed to address such uses, among other factors. National Defense Authorization Act of 2020, Pub. L. No. 116-92 (2019).