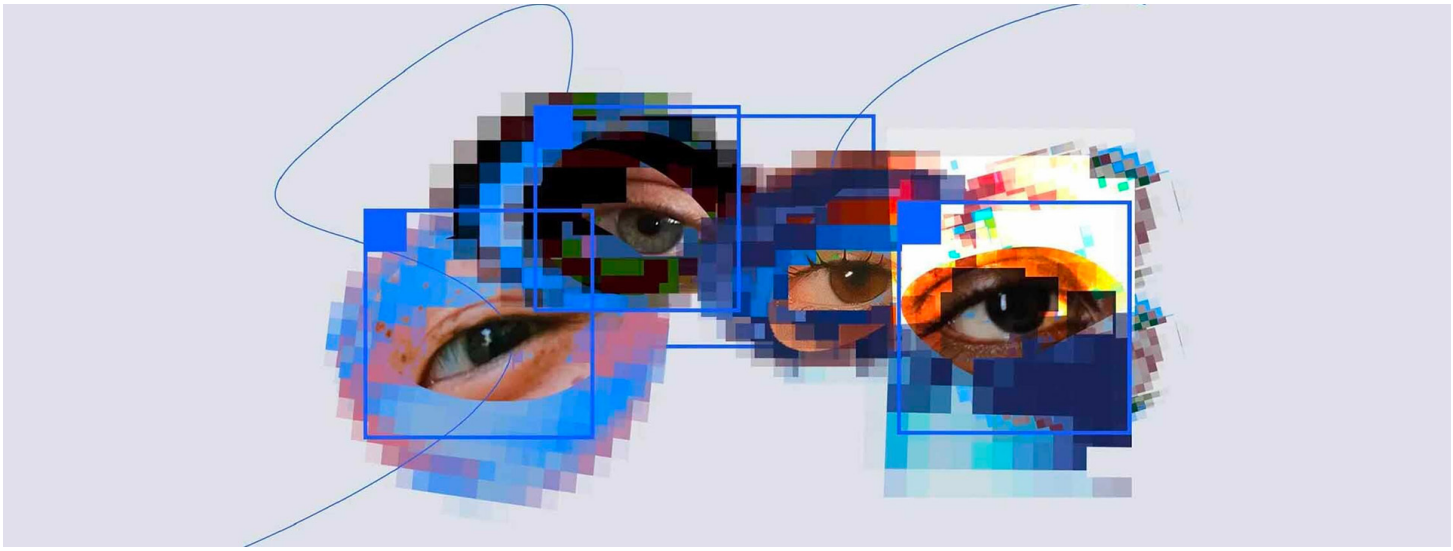
IBM Newsroom

Here's What Policymakers Can Do About Deepfakes, Right Now

By Christina Montgomery, Chief Privacy & Trust Officer, IBM and Joshua New, Senior Fellow, IBM Policy Lab



WASHINGTON, Feb. 28, 2024 /PRNewswire/ -- Deepfakes – realisticAI-generated audio, video, or images that can recreate a person's likeness – are one of the most pressing challenges posed by generative AI, given the potential for bad actors to use it to undermine democracy, exploit artists and performers, and harass and harm everyday people.

What this moment requires is both technical and legal solutions. That's why IBM (NYSE:IBM) signed the Tech Accord to Combat Deceptive Use of AI in 2024 Elections (Munich Tech Accord), pledging to help mitigate the risks of AI being used to deceive the public and undermine elections. It's also why IBM has long advocated for regulations that precisely target harmful applications of technology.

We outline below three key priorities for policymakers to mitigate the harms of deepfakes:

- Protecting elections,
- Protecting creators, and
- Protecting people's privacy

Protecting elections

Democracy depends on a population's ability to participate in free and fair elections. Unfortunately, bad actors can use deepfakes to impersonate public officials and candidates to deceive voters in a variety of ways that would undermine this key principle. For example, deepfakes could mislead voters about where, when, and how they can vote, or falsely portray a candidate making controversial statements or participating in scandalous activity.

Policymakers should prohibit the distribution of materially deceptive deepfake content related to elections. For example, the Protect Elections from Deceptive AI Act, introduced by Senators Klobuchar, Hawley, Coons, and Collins, would curtail the use of AI to generate deceptive content falsely depicting federal candidates in political advertisements with the intent of influencing an election. Other policy approaches could enable candidates targeted by materially deceptive AI-generated content used in

political advertisements or fundraising campaigns to seek damages or remove deceptive content, while preserving protections for free speech.

In the EU, IBM has been supportive of the Digital Services Act, which imposes on large internet platforms certain obligations regarding the moderation of online content. Recent guidelines published by the European Commission have also proposed additional requirements for consumer-facing platforms to mitigate against "systemic risks for electoral processes."

Protecting creators

Musicians, artists, actors, and creators of all kinds use their talents and likeness to help shape culture, inspire, entertain, and make a living. Deepfakes can enable bad actors to exploit creators' likenesses to push deceptive advertising, scam and mislead consumers, improperly reduce a creator's ability to profit from their talents, and more.

Policymakers should hold people accountable for producing unauthorized deepfakes of creator performances and hold platforms accountable if they knowingly disseminate such unauthorized content. Some jurisdictions already have what are known as "likeness laws" that prohibit the unauthorized use of a person's likeness for commercial exploitation, but these can be inconsistent, and few explicitly cover digital replicas or the rights to use a person's likeness after they die. Given these jurisdictional inconsistencies, IBM supports the NO FAKES Act in the U.S., which would create federal protections for individuals whose voices and/or likenesses are generated by third parties without their consent.

Protecting people's privacy

Everyday people are already being harmed by deepfakes in profoundly concerning ways, particularly by bad actors using their likeness to create nonconsensual pornography. This abuse primarily targets women, victims have included minors, and could potentially enable further abuse and extortion by bad actors. Nonconsensual sharing of intimate imagery, also known as revenge porn, is expanding with the use of deepfakes but ultimately is not a new problem. But few existing laws adequately hold bad actors accountable for sharing or threatening to share this material, or necessarily cover AI-generated content.

Policymakers should create strong criminal and civil liability for people that distribute nonconsensual intimate audiovisual content, including AI-generated content, as well as for people that threaten to do so. Penalties should be particularly severe when the victim is a minor. Legislators can act on this recommendation now by supporting and passing the bipartisan Preventing Deepfakes of Intimate Images Actin the U.S. This bill would create liability for individuals who disclose, or threaten to disclose, a nonconsensual intimate digital depiction of someone, including AI-generated content, and allow affected parties to pursue damages. This legislation would create a much-needed federal baseline of accountability that is inconsistently addressed in various revenge porn laws at the state level, providing greater protections for victims and individuals across the United States.

The EU AI Act – which IBM has long-supported – already addresses many of these types of issues, covering deepfakes more generally and imposing transparency requirements that clarify when particular content is not authentic. As policymakers look towards the Act's implementation in the coming months, particular attention should be paid to ensuring individuals are protected from non-consensual intimate audiovisual content.

Conclusion

Solving the problems posed by deepfakes will require thoughtful, whole-of-society approaches leveraging both changes in law

and technology. Technology companies have a responsibility to pursue technical and governance solutions that address AI-generated content, such as those articulated in the Munich Accord, the White House Voluntary AI Commitments, and Canada's Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems

IBM encourages policymakers to seize this opportunity to swiftly target three of the most significant harmful applications of deepfakes to ensure that AI remains a positive force for the global economy and society.

Click here to download the PDF.

Media Contact:
Ashley Bright
brighta@us.ibm.com

SOURCE IBM

Additional assets available online: Photos

https://stage.mediaroom.com/ibmnewsroom/Blog-Heres-What-Policymakers-Can-Do-About-Deepfakes