WILSON MEMO

Wilson Center
Congressional Relations

GANs, Deepfakes, & Synthetic Media

What are generative adversarial networks (GANs)?

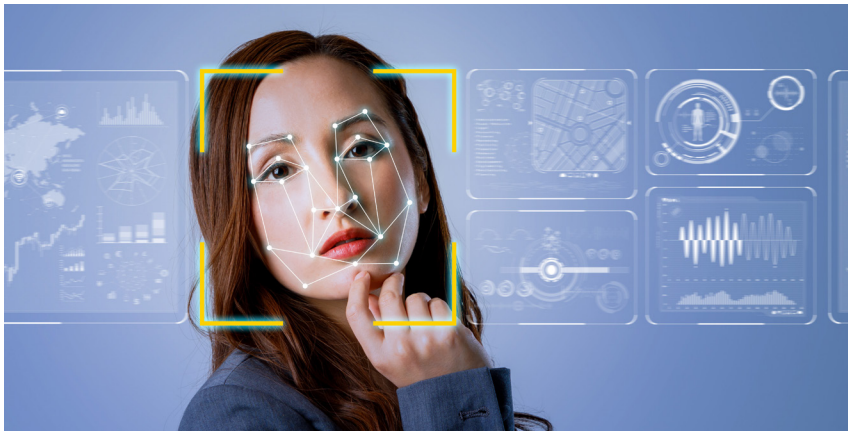
Generative adversarial networks (GANs) are an innovation in machine learning which uses adversarial training to pit two deep learning algorithms against one another. The goal of a GAN is to **create new, synthetic data that is indistinguishable from the real, sample data**. The model architecture of a GAN consists of a generator network and a discriminator network. Generators generate new, synthetic data, while discriminators evaluate this new data for authenticity against the real, sample data.

Deepfakes and Synthetic Media

Deepfakes are created through generative adversarial networks. GANs are **best utilized for re-creating static photos, but can also generate audio, text, drawings, handwritten notes, video, etc**.

Recent advancements in GAN technology have led to **re-created videos impersonating various individuals as well as filters on video conferencing platforms that allow impersonation in real time**. Although these videos are not 100% foolproof, it can be difficult for average Internet users to tell the difference between deepfakes and real content.

GANs were developed for several positive uses such as Internet content moderation, health data and image recognition, short term weather disaster prevention, and much more. However, this technology has proven to be easily weaponized for malicious causes.



Malicious Use of Deepfakes

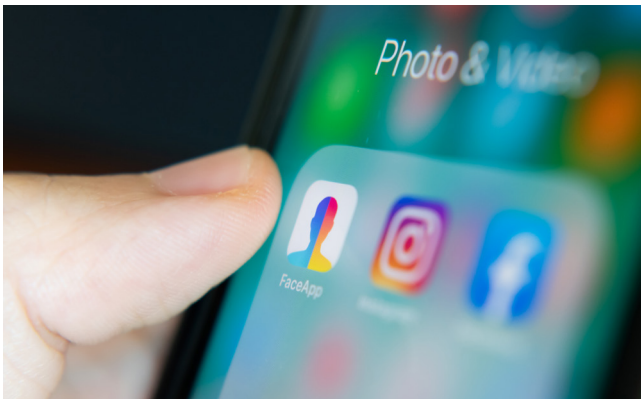
As the amount of fake accounts on social media decreases from heights during the 2016 and 2020 elections, **malicious actors are improving their usage of deepfakes to manipulate public opinion**.

In particular, Russia's motive isn't necessarily to get disinformation to trend virally or through conventional news, but instead takes a rather simple approach by **infiltrating our news sources**. Russia's "Peace Data Operation," for example, created 13 well-groomed accounts using stolen identities tied to specific bank and Bitcoin accounts. Spoofs such as these, despite having many mistakes, lead to real risks and manipulation of public opinion.

WILSON MEMO

Wilson Center
Congressional Relations

GANs, Deepfakes, & Synthetic Media



Malicious Use of Deepfakes (Continued)

Well-resourced states will be able to carry out these steps, but **opportunities for counter-moves have also emerged**. The **more nations act, the more visible their patterns and behaviors are**. The more we can learn about the technologies, the more we can deal with their repercussions.

Overall, however, most malicious use of deepfakes are against women in the form of deepfake pornography.

Quotes from Wilson Scholars

Nina Jankowicz & Melissa K. Griffith

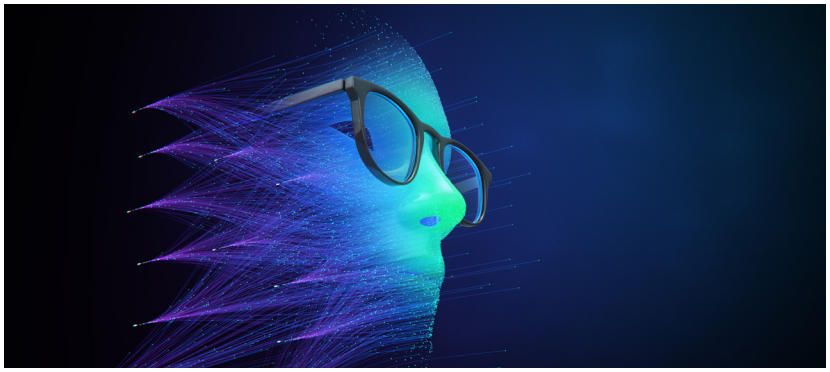
Wilson Center Disinformation Fellow & former Fulbright-Clinton Public Policy Fellow. Learn more about Ms. Jankowicz here.

Wilson Center Public Policy Fellow, Research Fellow at UCLA, & Adjunct Assistant Professor at Georgetown. Learn more about Ms. Griffith here.

Actionable Ideas

Policymakers must **utilize legislative measures to first stamp out deepfake pornography, the largest use of deepfake technology**. Policymakers should also **authorize programs allowing users to install plug-ins in browsers to detect manipulated images**.

Combatting deepfake technology and giving law enforcement the tools to crackdown on it will create the technology necessary to address prospective national security threats. The same technology used to create this problem can be used as a solution to defend against security threats and build resiliency.



Go Deeper! More Resources

Wilson Center
Science and Technology Innovation Program

***Need To Know* Podcast**
High Stakes Deepfakes with Melissa Griffith

Digital Futures Project
Malign Creativity: How Gender, Sex, and Lies are Weaponized Against Women Online