

CANADIAN GLOBAL AFFAIRS INSTITUTE
INSTITUT CANADIEN DES AFFAIRES MONDIALES

The Uses and Abuses of Deepfake Technology

by Abby MacDonald

February 2022

POLICY PERSPECTIVE

THE USES AND ABUSES OF DEEPFAKE TECHNOLOGY

by Abby MacDonald

CGAI Fellow
February 2022



CANADIAN GLOBAL AFFAIRS INSTITUTE
INSTITUT CANADIEN DES AFFAIRES MONDIALES

Prepared for the Canadian Global Affairs Institute
1800, 150 – 9th Avenue S.W., Calgary, AB T2P 3H9
www.cgai.ca

©2022 Canadian Global Affairs Institute
ISBN: 978-1-77397-231-2

Black-and-white photos of people taken 100 years ago can be brought to life. There are companies that can clone your own voice for you, and apps that allow you to become a celebrity in a movie scene. You can watch news broadcasts in which a computer-generated image of a person is on-screen instead of an actual person. All these scenarios are possible due to deepfake technology, which is essentially artificial intelligence capable of creating realistic but false videos, photos and audio of people. Not all deepfakes are so harmless, however, as the technology can, and has been, used to commit fraud, sexually harass women, exacerbate tensions and cause violence. With the increasing dependence on the internet for news as well as the speed of online communication, deepfakes will pose challenges to national security and public safety, to individuals, especially women, and governance of cyber-security. The technology is improving exponentially and policy-makers must act on this now, developing both a national and multilateral strategy to combat deepfakes. One of the most challenging issues democracies face today is the decreasing trust in institutions and the democratic process fuelled by disinformation; it is critical that democracies balance control over technology with fundamental freedoms of expression.[1]

What is a Deepfake?

Two terms are commonly used for altered videos and photos circulating online: deepfakes or cheap fakes. A cheap fake is essentially fake content that has been altered using readily available and easy-to-use technology, requiring very little expertise or technical skills.[2] For instance, there was a viral video of U.S. House Speaker Nancy Pelosi in 2019 in which her speech was slow and slurred, making it look like she was intoxicated. This was done simply by slowing the video by 75 per cent, which could be done by using iMovie or any basic video editing program.[3] Typically, cheap fakes are not particularly convincing due to the low level of sophistication, but they can still have an impact, as demonstrated by the reactions to this video. Deepfakes, on the other hand, require far more sophisticated technologies and the use of artificial intelligence (AI) to "realistically depict real people doing or saying things they never actually did."[4] Deepfakes are far more convincing, and increasingly difficult to detect, usually requiring the use of AI to find inconsistencies.

Deepfake technology relies on two important breakthroughs in machine learning and artificial intelligence. The first is a neural network. The more information these algorithms are exposed to, the more accurately they can repeat it back. The second is generative adversarial networks (GANs), which essentially combine two neural networks together and make them compete against

[1] Thomas Paterson and Lauren Hanley, "Political Warfare in the Digital Age: Cyber Subversion, Information Operations and 'Deep Fakes," *Australian Journal of International Affairs* 74, no. 4 (2020): 439–54, doi:10.1080/10357718.2020.1734772, 442.
[2] Britt Paris and Joan Donovan, "Deep Fakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence," Report, *Data & Society*, 2.
[3] Ibid., 30.
[4] Paterson and Hanley, doi:10.1080/10357718.2020.1734772, 448.

one another to produce a better final product.[5] GANs were created by Ian Goodfellow, a former Google researcher in the fields of AI and machine learning. To create a deepfake, the first algorithm, called the generator, is given a set of original data which it uses to produce the fake. Then, both the original and fake data are downloaded to the second algorithm, called a discriminator, which attempts to distinguish which is fake. If it is successful, the generator then learns how the discriminator was able to determine the fake and subsequently corrects itself the next time, thus improving with each attempt.[6]

A realistic deepfake carefully created and released at a certain time could sway the results of a democratic election, incite violence against certain groups or exacerbate political and social divides.[7] The potentially dangerous effects of deepfakes have begun to emerge, with the technology having been used to commit fraud, fool people into connecting online and discredit individuals.

National Security and the Democratic Process

Online disinformation has exacerbated security incidents in democratic countries, including violent protests which have led to health and safety risks, property damage, injuries and even death. The prospect of deepfakes as a cause of instability is not difficult to imagine, nor far off. We have not seen many deepfakes during recent elections, with altered videos of politicians and other influential people mostly being of cheap-fake quality. However, deepfakes are still improving, at a rate far faster than detection techniques. Efforts underway in the private sector include initiatives such as Facebook, Microsoft, Amazon and the Partnership on AI developing a "Deepfake Detection Challenge" with datasets of videos and photos with a competition and prize money to develop better detection methods.[8] The results have not been especially impressive, with the winner of a competition in 2020 only being able to detect 82 per cent of the competition's dataset, with that result dropping to 65 per cent on new, unseen deepfakes.[9] Microsoft announced a new detection tool in 2020; however, it admitted that the tool would quickly become obsolete without frequent updating.[10] Clearly, in order to prepare for coming elections there must be a co-ordinated effort to improve detection techniques, and this research needs to be both consistently used and constantly maintained if it is to be effective.

---

[5] Konstantin A. Pantserev, "The Malicious Use of AI-Based Deepfake Technology as the New Threat to Psychological Security and Political Stability," in *Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity*, Jahankhani, Hamid, Stefan Kendzierskyj, Nishan Chelvachandran and Jaime Ibarra, eds., (Cham, Switzerland: Springer, 2020) 39–40.

[6] Ibid.

[7] Robert Chesney and Danielle K. Citron, "Disinformation on Steroids: The Threat of Deep Fakes," Council on Foreign Relations, October 16, 2018, https://www.cfr.org/report/deep-fake-disinformation-steroids.

[8] William A. Galston, "Is Seeing Still Believing? The Deepfake Challenge to Truth in Politics," Brookings, May 06, 2020. Accessed November 18, 2020. https://www.brookings.edu/research/is-seeing-still-believing-the-deepfake-challenge-to-truth-in-politics/.

[9] Will Knight, "Deepfakes aren't Very Good. Nor are the Tools to Detect Them," *Wired*, June 12, 2020, https://www.wired.com/story/deepfakes-not-very-good-nor-tools-detect/.

[10] Leo Kelion, "Deepfake Detection Tool Unveiled by Microsoft," BBC News, September 01, 2020, https://www.bbc.com/news/technology-53984114.

However, authenticating content will not always be possible, and even if something that has been altered is marked as such, this will not necessarily change people's responses. The fact is, misinformation doesn't need much help from deepfakes to be effective, especially as people will choose to believe whatever reinforces their existing beliefs. With the speed and reach of the internet and rapidly eroding trust in intuitions, misinformation in all forms is a reality that can only be mitigated by awareness and transparency.

Individuals

The impact of deepfakes has been felt mostly by individuals at this point, largely females. This is because, according to a survey done in 2019 of 15,000 deepfake videos online, about 96 per cent of them are pornographic, with 99 per cent of those portrayed being women.[11] This phenomenon is probably due in part to the first viral deepfake videos that emerged on Reddit in 2017, where the user "Deepfakes" photoshopped the faces of famous actresses onto the bodies of women in real pornographic videos.[12] Since then, this has been used as a tactic to shame and discredit women, such as an Indian investigative journalist who criticized the government, a competitive mother who wanted her daughter's gymnastics teammates eliminated and the creators of a "nudifying" filter that removes the clothing from a person in a photo – which is designed to work only on women. Furthermore, the pandemic has brought about an increase in technology-facilitated gender-based violence (TFGBV).[13]

Canada's commitments to a feminist foreign policy and to end sexual and gender-based violence, abuse and harassment in digital contexts make this an especially significant problem to address. While the national security threat deepfakes pose is real, the impact on people is already happening and must become a more urgent issue than it currently is. The use of deepfakes against women and other individuals could lead to serious physical and emotional harm, not only regarding content, but people's responses to it. For example, inflammatory videos have led to the increasing use of doxing, which is a dangerous trend involving posting someone's personal information, such as phone number, email, home address, workplace or other private material online for anyone to see.[14] Furthermore, even with the timely detection and removal of such content, it is essentially impossible to ensure that it does not remain in cyber-space. By the time the content is addressed, it has likely already been downloaded, copied and shared with many others.[15]

[11] Suzy Dunn, "Women, Not Politicians, are Targeted Most Often by Deepfake Videos," Centre for International Governance Innovation, March 03, 2021, https://www.cigionline.org/articles/women-not-politicians-are-targeted-most-often-deepfake-videos/.
[12] Pantserev, 42.
[13] Suzy Dunn, "Technology-Facilitated Gender-Based Violence: An Overview," Supporting a Safer Internet Paper No. 1, December 7, 2020, 1-30, Center for International Governance Innovation, 2, https://www.cigionline.org/sites/default/files/documents/SaferInternet_Paper%20no%201.pdf.
[14] Ibid.
[15] Raina Davis, Chris Wiggins and Joan Donovan, "Rep," *Tech Fact Sheets for Policymakers,* Amritha Jayanti, ed.,. (Cambridge, MA, 2020) 10.

Deepfakes used to blackmail, discredit or humiliate pose challenges not only to those in public positions, but to any vulnerable group, especially those who are already marginalized.[16] For instance, video evidence was used last year to prove the guilt of police officers in cases of violence against unarmed civilians, largely Black men, and the use of body cameras on police officers has become a solution to cases on the unjust use of force.[17] However, the emergence of sophisticated deepfakes brings audio and visual evidence into question. It will be important for law enforcement to begin preparing for the need to thoroughly verify such evidence, and this can pose challenges to the credibility of those with less social power. For example, deepfakes introduce something called the "liar's dividend." As people grow increasingly aware of the capabilities of deepfake technology, it's easier to believe the denials of a person who claims they are simply being framed.[18]

Deepfakes for Good?

The use of deepfake technology is not always malicious, and the stakeholders who are interested in using this technology for research must be included in policy-making. For instance, the medical and entertainment industries have interests in the use and distribution of this technology. In the medical field, this technology is used for research, such as generating realistic MRI images for training, and can be used in some cases to help those with illnesses regain control over their voices.[19] Some of the earliest uses of this technology were by Disney to convincingly recreate original Star Wars characters; given Disney's interests in using this for creative purposes, it fought hard against legislation in New York state that attempted to completely ban the technology.[20] Businesses can also use deepfakes to create training videos in multiple languages, as some have already begun to explore.

This technology can also simply be used for creative purposes. For example, the popular app Face Swap allows the user to put their face on the body of a celebrity in an action movie. Those with the skills and equipment may choose to entertain on social media – for example, Chris Ume, behind the viral Tom Cruise Tik Tok deepfakes, who chooses to create for fun and to educate. In fact, there are plenty of examples of deepfakes that were created specifically to spread awareness of this technology. In 2018, Buzzfeed teamed up with actor and comedian Jordan Peele to create a deepfake video of then-president Barack Obama disparaging Donald Trump, which was intended to encourage people to be critical of the information they consumed online.[21] It is important to consider all the ways in which this technology supports creativity, art, education and critical research when making policy.

---

[16] Paris and Donovan, 8.
[17] Riana Pfefferkorn, "The Threat Posed by Deepfakes to Marginalized Communities," Brookings, April 21, 2021, https://www.brookings.edu/techstream/the-threat-posed-by-deepfakes-to-marginalized-communities/.
[18] Danielle K. Citron and Robert Chesney, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," 107, *California Law Review*, 1753/85? (2019), https://scholarship.law.bu.edu/faculty_scholarship/640.
[19] Arthur Nelson and James A. Lewis, "Trust Your Eyes? Deepfakes Policy Brief," Center for Strategic and International Studies, October 2019, 3.
[20] Ibid.
[21] Paris and Donovan, 38.

Responses to deepfakes must be flexible and dynamic. A soft-law framework should be the approach in any context, as the technology is moving far too quickly for traditional methods to follow it.[22] Creating standards, guidelines for use and best practices, while not necessarily enforceable by law, provides a way to keep up with the varying and quickly changing uses of the technology so that some of the negative effects can be mitigated and controlled, but is not left completely behind by drawn-out and outdated processes; another reason this approach is ideal is the wide range of possible uses for this technology.[23] Some important issues to consider include both ethics and ownership. For example, deepfakes have been created of those who are dead, and this leads to questions over the rights to the video, and other ethical considerations.[24]

Recommendations

Canada is in a unique position to lead the initiative on countering deepfakes. Within Canada, some of the most cutting-edge AI research is being conducted by government in partnership with a number of other domestic and foreign actors. Furthermore, Canada is a member and leader in many related multilateral initiatives, such as the Paris Call for Trust and Security in Cyberspace, NATO Cooperative Cyber Defence Centre of Excellence, or the Global Partnership on Artificial Intelligence. Canada also belongs to the Organisation internationale de la Francophonie, positioning it well to engage with francophone African countries which have more fragile democracies that could be destabilized by this technology. It can use these forums to co-ordinate with global and domestic actors to improve deepfake policy in different areas.

It seems the only way for verification technologies to be effective is if they are standardized and used by a wide variety of relevant news organizations and social media companies.[25] However, adopting an industry-wide standard is not so simple. Co-ordinating and implementing such a standard is only the beginning; after that, the constant need for updating and the use of increasingly innovative methods would be necessary to keep pace with the perpetual changes and improvements across a variety of techniques.[26] Furthermore, the creators of deepfake content could test themselves against the standard and constantly seek to evade it. Governments should facilitate the adoption and use of such technology by businesses, keeping these challenges in mind. Ideally, these collaborations should include small organizations that lack the resources of the larger corporations and tech companies.

The effects of disinformation cannot be eliminated, but with increasing awareness of deepfakes and the consistent use of verification tools across major companies and outlets, it will at least be clear more quickly if content has been manipulated or not. Furthermore, the development of

[22] John Villasenor, "Soft Law as a Complement to AI Regulation," Brookings, July 31, 2020. Accessed November 19, 2020. https://www.brookings.edu/research/soft-law-as-a-complement-to-ai-regulation/.
[23] Ibid.
[24] Bernd Debusmann, Jr., "'Deepfake is the Future of Content Creation,'" BBC News, March 8, 2021, https://www.bbc.com/news/business-56278411.
[25] Tim Huang, "Deepfakes: A Grounded Threat Assessment," July 2020, Center for Security and Emerging Technology, https://cset.georgetown.edu/wp-content/uploads/CSET-Deepfakes-Report.pdf., 14.
[26] Ibid.,18.

verification tools for personal use would also be a worthwhile option to explore by private companies or possibly media organizations. Considering the number of apps that can be used to create deepfakes, perhaps apps could help users determine the authenticity of video or audio content, giving individuals the option to fact-check their own information on their own time.[27]

Heightened awareness and the designation of specific, trusted news sources around election periods will also be necessary, as during times of transition it can be expected that disinformation will be employed by both foreign and domestic actors. To mitigate this risk as much as possible, it is important to ensure trustworthy information, especially during transitional periods such as elections.[28] To do this, Canada should support independent journalism and media organizations and ensure that they have the tools and training to effectively communicate this trusted content. Perhaps they could run awareness campaign around deepfakes through departments such as the Canadian Center for Cyber Security, which can design awareness products and mandatory training materials that could be used by journalists, and also be accessible to the wider public.

While legal mechanisms will have a difficult time keeping up with the pace of deepfake innovation, further steps must be taken to better protect victims of harassment or sexual exploitation alongside existing criminal law, which can be applied to such cases.[29] Canada should encourage other governments and organizations involved in forums such as the Paris Call, which seeks safety and trust online, to work on better responses and solutions to the issue of TFGBV, perhaps conducting a gender impact assessment on laws applicable to cyber-security and TFGBV, identifying gaps and weaknesses to be improved upon.

Finally, industries that seek the use of deepfake technologies should be consulted about their research to better understand the implications of their work and raise relevant legal and ethical questions to govern the landscape of deepfakes and allow for innovation. A group like the Global Partnership on Artificial Intelligence could be an appropriate forum for this topic, which brings together governments and expertise to responsibly research and implement AI. Consultations could lead to establishing best practices on a multinational level for the ethical development and regulation of these technologies, encouraging both innovation and accountability.

Conclusion

Given the pace of change in artificial intelligence and the growing influence of disinformation around the world, including in Canada, it is imperative that policy-makers act now and start to prepare for a reality in which there is a constant need to verify audio and visual information.

[27] Ibid., 26.
[28] Hannah Smith and Katherine Mansted, "Weaponised Deep Fakes – National Security and Democracy," April 29, 2020, https://www.aspi.org.au/report/weaponised-deep-fakes.
[29] Jane Bailey and Carissima Mathen, "Technology-facilitated Violence against Women & Girls: Assessing the Canadian Criminal Law Response," *Canadian Bar Review* 97 (3): 664–96, 666.

Canada is committed to technological innovation, maintaining trust and safety online and tackling disinformation, eliminating violence against women and girls and protecting other marginalized groups; it also is committed to multilateralism. Cyber-security issues are global problems that require strong international co-operation. Canada should take advantage of its global reach and exceptional position on the topic to begin addressing this issue and meeting its policy goals.

About the Author

Abby MacDonald *received her Master's in International Affairs, where she specialized in security and defence policy, in 2021. Before that, she earned her B.A. in International Relations from Western University in 2019. Her research interests include cybersecurity policy, the impact of technology on conflict, artificial intelligence, and conflict economics. Abby has worked as a research assistant studying national economic security and geoeconomics, and has worked in information security policy and information management.*

Canadian Global Affairs Institute

The Canadian Global Affairs Institute focuses on the entire range of Canada's international relations in all its forms including (in partnership with the University of Calgary's School of Public Policy), trade investment and international capacity building. Successor to the Canadian Defence and Foreign Affairs Institute (CDFAI, which was established in 2001), the Institute works to inform Canadians about the importance of having a respected and influential voice in those parts of the globe where Canada has significant interests due to trade and investment, origins of Canada's population, geographic security (and especially security of North America in conjunction with the United States), social development, or the peace and freedom of allied nations. The Institute aims to demonstrate to Canadians the importance of comprehensive foreign, defence and trade policies which both express our values and represent our interests.

The Institute was created to bridge the gap between what Canadians need to know about Canadian international activities and what they do know. Historically Canadians have tended to look abroad out of a search for markets because Canada depends heavily on foreign trade. In the modern post-Cold War world, however, global security and stability have become the bedrocks of global commerce and the free movement of people, goods and ideas across international boundaries. Canada has striven to open the world since the 1930s and was a driving factor behind the adoption of the main structures which underpin globalization such as the International Monetary Fund, the World Bank, the World Trade Organization and emerging free trade networks connecting dozens of international economies. The Canadian Global Affairs Institute recognizes Canada's contribution to a globalized world and aims to inform Canadians about Canada's role in that process and the connection between globalization and security.

In all its activities the Institute is a charitable, non-partisan, non-advocacy organization that provides a platform for a variety of viewpoints. It is supported financially by the contributions of individuals, foundations, and corporations. Conclusions or opinions expressed in Institute publications and programs are those of the author(s) and do not necessarily reflect the views of Institute staff, fellows, directors, advisors or any individuals or organizations that provide financial support to, or collaborate with, the Institute.