

Inteligencia Artificial y la desinformación

La desinformación y la mala información están muy extendidas. Entre 2013 y 2019, se llevaron a cabo 96 campañas de influencia doméstico y extranjero en las redes sociales, con el objetivo de influir en los resultados políticos en varios países con la difusión de información falsa.¹ El 20 de febrero de 2020, el “World Health Organization” (WHO) declaró una epidemia global definida como un “tsunami de información—algunos precisos, algunos no—que se esparcen con la epidemia”.² La infomedia se intensificó debido a un afluencia de “deepfakes”—imágenes, videos o contenidos de audios realistas fabricados utilizando tecnologías de inteligencia artificial y aprendizaje automático “profundo”, de ahí sale el nombre. Mientras que los “deepfakes” comparten similitudes con editores de fotos o filtros, son fundamentalmente diferentes debido por su dependencia en algoritmos de aprendizaje profundo para crear medios sintéticos hiperrealista prácticamente indistinguibles del contenido auténtico. A pesar que los “deepfakes” pueden ser inofensivos cuando en uso para entretenimiento, como hacer una versión de “Back to the future” protagonizada por los actores de “Spider-man” de 2017, también se puede usar para promover la desinformación, discursos de odio, y polarización de sociedades.

La llegada de los “deepfakes” coincide con la expansión de las redes sociales y el acceso al internet, amplificando la propagación de mala información (información falsa sin la intención de causar daño e información falsa con intención de causar daño, respectivamente). El número de usuarios de internet globalmente ha aumentado de 416 millones en el año 2000 a 4,7 billones solo 20 años después.³ Con el reciente aumento de “deepfakes” mala información, y desinformación se han vuelto cada vez más indistinguibles de la información real.

En un nivel personal, la inteligencia artificial puede ser perjudicial porque la tecnología “deepfake” permite que las personas sin calificaciones puedan crear pornografía sin consentimiento usado para avergonzar, acosar, y extorsión. La firma de inteligencia artificial Diptrace informó que 96% del contenido “deepfake” que circula en el internet es pornografía con caras de mujeres famosas pegadas a estrellas de porno. Un profesor de leyes en la Universidad de Boston Daniell Citron argumenta que “la tecnología deep face se está utilizando para dañar a las mujeres”.⁴

Como la WHO ha explicado, la proliferación de “deepfakes” combinado con el rápido crecimiento de las redes sociales y la digitalización hizo la propagación de información falsa durante una pandemia global mucho más peligrosa. Noticias falsas alegando que el alcohol, calor excesivo, o frío excesivo puede proteger contra el virus convenció al Internet a pesar de que no hubo evidencia científica que apoyara esas afirmaciones.⁵ Temprano durante la pandemia, el Centro para el Control de Enfermedades informó de un estudio que el 39% de respondedores

¹ Martin, Diego A, et al. “Introducing the Online Political Influence Efforts Dataset.” *Journal of Peace Research*, 10 Nov. 2022, p. 002234332210928, <https://doi.org/10.1177/00223433221092815>. Accessed 20 Jan. 2023.

² Martin, Diego A, et al. “Introducing the Online Political Influence Efforts Dataset.” *Journal of Peace Research*, 10 Nov. 2022, p. 002234332210928, <https://doi.org/10.1177/00223433221092815>. Accessed 20 Jan. 2023.

³ Murphy, Julia, et al. “Internet.” *Our World in Data*, 2023, ourworldindata.org/internet.

⁴ Sample, Ian. “What Are Deepfakes – and How Can You Spot Them?” *The Guardian*, 13 Jan. 2020, www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them.

⁵ Group, Internet. “Deepfakes and the Healthcare Industry during COVID-19.” *GlobalSign*, 2 Apr. 2024, www.globalsign.com/en/blog/in/how-deepfakes-could-further-burden-healthcare-industry-during-covid-19. Accessed 2 Jan. 2025.

participaron en prácticas peligrosas en relación a la exposición a productos del hogar desinfectantes, incluyendo “lavando productos alimenticios con lejía, aplicando productos de limpieza directamente sobre la piel, y intencionalmente respirando e ingiriendo desinfectantes para prevenir el contagio del COVID-19.”⁶

La politización de “deepfakes”—como un video de Barack Obama llamando a Donald Trump un “idiota completo”⁷ o un video de la política demócrata Nancy Pelosi ralentizado para que su discurso parezca equivocado—ilustra lo más preocupante de las implicaciones de la inteligencia artificial y la información falsa.⁸ Un estudio comparó la propagación de las noticias falsas y las noticias verdaderas, encontró que las noticias falsas en Twitter se esparcen un 70% más rápido que las noticias verdaderas, con noticias políticas falsas siendo las más virales.⁹

Los “deepfakes” de la inteligencia artificial también proponen grandes amenazas a la seguridad internacional. En Marzo del 2022, poco después de la invasión Rusa a Ucrania, propagandistas rusos fabricaron un video “deepfake” del presidente de Ucrania Volodymyr Zelenskyy urgiendo a su militar que se rinda. El “deepfake: rápidamente agarro mucha atención en las redes sociales y en las noticias, causando confusión y tensión en relación al crisis.”¹⁰ La administración de Zelenskyy rápidamente negó su autenticidad; sin embargo, este ejemplo de la utilización de “deepfakes” como arma en conflictos sugiere que esto no sera la ultima vez que los “deepfakes” amenacen la seguridad internacional si no se toman medidas para prevenir futuro problemas relacionados con los “deepfakes”. A pesar de las peligrosas implicaciones de los “deepfakes” de la inteligencia artificial, investigadores en la Universidad Northwestern sugieren qué, con un código estricto para los gobiernos, los “deepfakes” podrían usarse para contrarrestar y desestabilizar a grupos terroristas u otras organizaciones maliciosas.

La apariencia de los “deepfakes” de la inteligencia artificial impide la comprensión de sus efectos dañinos en la sociedad y ha dado lugar a un gran número de investigadores que sugieren modos de protección contra estas amenazas. El “Disinformation Disruption Framework” desarrollado por el Deep Trust Alliance describe posibles soluciones a las noticias falsas y otros problemas causados por la inteligencia artificial. Se sugiere cuatro estrategias principales para combatir la crisis de la desinformación: desarrollo tecnológico (por ejemplo, tecnología de detección), regulación pública (por ejemplo, legislación), regulaciones privadas (por ejemplo, condiciones de servicio), y educación pública.¹¹ Empresas como Microsoft y Facebook han implementado herramientas de inteligencia artificial para detectar “deepfakes” con una precisión cada vez más segura—algunas herramientas logran una tasa de detección del 80-90%.¹² La ley de prohibición de noticias falsas de 2018, aprobada por el 115vo congreso de

⁶ Nelson, Taylor, et al. “The Danger of Misinformation in the COVID-19 Crisis.” *Missouri Medicine*, vol. 117, no. 6, Nov. 2020, p. 510, [pmc.ncbi.nlm.nih.gov/articles/PMC7721433/](https://pubmed.ncbi.nlm.nih.gov/articles/PMC7721433/).

⁷ Sample, Ian. “What Are Deepfakes – and How Can You Spot Them?” *The Guardian*, 13 Jan. 2020, www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them.

⁸ Vosoughi, Soroush, et al. *FALSE NEWS IS BIG NEWS*. 2017.

⁹ Vosoughi, Soroush, et al. *FALSE NEWS IS BIG NEWS*. 2017.

¹⁰ Buffett Brief. *The Rise of Artificial Intelligence and Deepfakes*. July 2023.

¹¹ Sasse, Ben. *Malicious Deep Fake Prohibition Act of 2018*. 21 Dec. 2018.

¹² Sasse, Ben. *Malicious Deep Fake Prohibition Act of 2018*. 21 Dec. 2018.

Sasse, Ben. *Malicious Deep Fake Prohibition Act of 2018*. 21 Dec. 2018.

los Estados Unidos, propuso penales criminal para las personas que creen o difundan “deepfakes” con el intento de dañar a otras personas, como difundir contenido político falso o pornografía sin consentimiento.¹³