El desafío de los *deepfake* en la esfera política

Francisco José García-Ull

Universidad Europea de Valencia

franciscojose.garcia@universidadeuropea.es

Instituto Mediterráneo de Estudios de Protocolo (IMEP)

fran.garcia@protocoloimep.com

Resumen

Un *deepfake* es un vídeo hiperrealista, manipulado digitalmente, para representar a personas que dicen o hacen cosas que en realidad nunca sucedieron. Estas representaciones sintéticas plantean serias amenazas para la privacidad e incrementan los riesgos derivados de las suplantaciones de identidad. Con la sofisticación de los *deepfake*, resulta cada vez más complicado detectar si las apariciones públicas o declaraciones de personajes influyentes responden a parámetros de realidad o, por el contrario, son resultado de representaciones ficticias. Este estudio tiene como objetivo analizar los distintos métodos existentes para la detección de falsificaciones profundas. Se pretende, además, plantear el debate sobre la aplicación de una tecnología que podría alimentar el actual escenario de desinformación y de agotamiento del pensamiento crítico.

Palabras clave: *Deepfake*, Desinformación, IA, Comunicación Política

Biografía:

Francisco José García Ull es licenciado en Publicidad y Relaciones Públicas y Doctor en Comunicación. Es miembro del Grupo I+D Mediaflows (Universitat de València) y profesor en la Universidad Europea de Valencia y en el Instituto Mediterráneo de Estudios de Protocolo (IMEP). Ha realizado estancias en Estonia, Alemania y Estados Unidos y su investigación gira entorno a la Comunicación y la Tecnología.

Introducción

Las nuevas herramientas basadas en Inteligencia Artificial (IA) permiten la recreación de representaciones audiovisuales realistas originales que simulan la apariencia y el habla de los seres humanos. Estas representaciones sintéticas se conocen como *deepfake* y plantean serias amenazas para la privacidad, en un nuevo escenario en el que se incrementan los riesgos derivados de las suplantaciones de identidad.

Los *deepfake* son el producto de las Redes Generativas Antagónicas (RGA), también conocidas como GAN en inglés. Son una clase de algoritmos de IA que se utilizan en el aprendizaje no supervisado, implementadas por un sistema de dos redes neuronales que

Bibliografía

Das, S., Datta, A., Islam, M., & Amin, M. (2021). Improving DeepFake Detection Using Dynamic Face Augmentation. arXiv preprint arXiv:2102.09603.

Day, C. 2019. "The Future of Misinformation". Computing in Science & Engineering, 21(1): 108–108. https://doi.org/10.1109/MCSE.2018.2874117

Fletcher, J. (2018). "Deepfakes, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance". Theatre Journal, 70(4): 455–471. ProjectMUSE, https://doi.org/10.1353/tj.2018.0097

Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; XU, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. (2014). "Generative Adversarial Networks". arXiv:1406.2661

Guarnera, L., Giudice, O., & Battiato, S. (2020). Deepfake detection by analyzing convolutional traces. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 666-667).

Hasan, H. R., & Salah, K. (2019). "Combating Deepfake Videos Using Blockchain and Smart Contracts". IEEE Access, 7: 41596–41606. https://doi.org/10.1109/ACCESS.2019.2905689

Li, Y., Chang, M. C., Farid, H., & Lyu, S. (1806). In ictu oculi: Exposing AI generated fake face videos by detecting eye blinking. 2018. arXiv preprint arXiv:1806.02877.

Maras, M. H., & Alexandrou, A. (2019). "Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos". International Journal of Evidence & Proof, 23(3): 255–262. https://doi.org/10.1177/1365712718807226

Mittal, T., Bhattacharya, U., Chandra, R., Bera, A., & Manocha, D. (2020, October). Emotions Don't Lie: An Audio-Visual Deepfake Detection Method using Affective Cues. In Proceedings of the 28th ACM international conference on multimedia (pp. 2823-2832).

Neekhara, P., Dolhansky, B., Bitton, J., & Ferrer, C. C. (2021). Adversarial threats to deepfake detection: A practical perspective. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 923-932).

García-Ull, F. J. (2021). «Deepfakes: el próximo reto en la detección de noticias falsas». Anàlisi: Quaderns de Comunicació i Cultura, 64, 103-120. DOI: <https://doi.org/10.5565/rev/analisi.3378>

Giudice, O. Guarnera, A. Paratore, and S. Battiato. 1-D DCT domain analysis for JPEG double compression detection. In Proceeedings of International Conference on Image Analysis and Processing, pages 716–726. Springer, 2019.

Popescu, A. y Farid, H. Exposing digital forgeries in color filter array interpolated images. IEEE Transactions on Signal Processing, 53(10):3948–3959, 2005.

Westerlund, M. (2019). "The Emergence of Deepfake Technology: A Review". Technology Innovation Management Review, 9(11): 39-52. http://doi.org/10.22215/timreview/1282