Deepfakes: Revisión sistemática de tecnologías, impacto y estrategias de detección

Deepfakes: A systematic literature review of technologies, impact, and detection strategies

Jania Yamileth Murguía Serrano
Facultad de Informática Mazatlán, Universidad Autónoma de Sinaloa, Mazatlán, Sinaloa
jania.murguia345@gmail.com
ORCID: 0009-0003-7429-5876

Carlos Eduardo De la Toba Noriega
Facultad de Informática Mazatlán, Universidad Autónoma de Sinaloa, Mazatlán, Sinaloa
carlosedelatoban@gmail.com
ORCID: 0009-0005-5778-2562

Sebastián Campos Zatarain
Facultad de Informática Mazatlán, Universidad Autónoma de Sinaloa, Mazatlán, Sinaloa
szatarain56@gmail.com
ORCID: 0009-0009-6277-4566

Luis Yael Aramburo Contreras
Facultad de Informática Mazatlán, Universidad Autónoma de Sinaloa, Mazatlán, Sinaloa
luisyaelcon77@gmail.com
ORCID: 0009-0001-4471-3286

Gustavo Ángel Diaz Lucas
Facultad de Informática Mazatlán, Universidad Autónoma de Sinaloa, Mazatlán, Sinaloa
gustavoangeldiazlucas@gmail.com
ORCID: 0009-0003-4410-1348

doi: https://doi.org/10.36825/RITI.13.29.003

Recibido: Diciembre 02, 2024
Aceptado: Febrero 11, 2025

Resumen: Este articulo tiene por objetivo proporcionar una visión detallada y crítica del estado del arte acerca de los *deepfakes* a nivel global, identificando las principales tecnologías utilizadas, sus efectos en diferentes sectores de la sociedad y las estrategias efectivas para su detección. A partir de un enfoque cualitativo de nivel exploratorio para llevar a cabo esta revisión sistemática de la literatura, se analizaron 43 artículos de los últimos cinco años de diferentes países. Los principales hallazgos fueron las coincidencias con las investigaciones que proponen diversos modelos de detección; además de trabajos que hablan acerca de la desconfianza de la sociedad por la legitimidad de la información, lo cual fue un tema frecuente. Sin embargo, hubo publicaciones que resaltan la diversidad de enfoques y la necesidad de abordar el fenómeno de los *deepfakes* desde múltiples perspectivas. Este trabajo aporta

una comprensión integral del problema de los *deepfakes*, conectando hallazgos técnicos, sociales y éticos, y subrayando la importancia de abordarlo desde una perspectiva global.

***Palabras clave**: Deepfakes, Inteligencia Artificial, Deep Learning, Desinformación, Ética.*

Abstract: This article aims to provide a detailed and critical overview of the state of the art concerning deepfakes on a global scale, identifying the leading technologies employed, their impact across various societal sectors, and effective detection strategies. Through a qualitative, exploratory approach to conducting this systematic literature review, 43 articles from the past five years originating from different countries were analyzed. The primary findings highlighted significant alignment with research advocating for diverse detection models and studies addressing societal distrust in the legitimacy of information, a recurring theme. However, other publications emphasized the diversity of approaches and the necessity of addressing the deepfake phenomenon from multiple perspectives. This study contributes to a comprehensive understanding of the deepfake issue, integrating technical, social, and ethical insights, and underscores the importance of tackling it from a global perspective.

***Keywords**: Deepfakes, Artificial Intelligence, Deep Learning, Disinformation, Ethics.*

1. Introducción

En la última década, el avance vertiginoso de las tecnologías digitales ha propiciado el desarrollo de innovaciones que, si bien ofrecen múltiples beneficios, también plantean desafíos significativos en términos de seguridad y ética. Una de estas innovaciones es la creación de *deepfakes*, una técnica basada en inteligencia artificial (IA) que permite generar contenido audiovisual falsificado con un alto grado de realismo [1]. Los *deepfakes* han capturado la atención tanto de la comunidad científica como del público en general debido a su potencial para influir en la opinión pública, manipular información y comprometer la integridad de comunicaciones digitales [2].

La proliferación de *deepfakes* no se limita a un ámbito geográfico específico; sin embargo, su impacto puede variar considerablemente según el contexto social, económico y cultural de cada región. A nivel global, la adopción creciente de tecnologías digitales y la presencia significativa en plataformas de redes sociales hacen que los *deepfakes* representen una amenaza emergente que requiere una atención particular [3]. La capacidad de estos contenidos falsificados para desinformar, difamar y manipular opiniones puede tener repercusiones en la estabilidad política, la seguridad y la confianza pública en las instituciones.

La justificación de este estudio radica en la necesidad de comprender de manera integral el fenómeno de los *deepfakes*. A pesar de la creciente preocupación mundial sobre esta tecnología, existe una carencia de investigaciones que aborden específicamente cómo se manifiestan, cuáles son sus implicaciones y qué estrategias se están implementando para su detección y mitigación en diferentes contextos. Este vacío en la literatura científica subraya la importancia de llevar a cabo una revisión sistemática que no sólo analice las tecnologías subyacentes, sino que también evalúe su impacto social y las medidas adoptadas para contrarrestarlas.

El incremento de los *deepfakes* en múltiples áreas de la sociedad en los últimos años representa una amenaza significativa en forma de videos, imágenes o incluso audios falsificados de una o varias personas con el fin de suplantar sus identidades, lo que puede dañar la reputación de individuos u organizaciones. A medida que pasan los años, este fenómeno se vuelve más frecuente debido a las diversas herramientas tecnológicas disponibles actualmente, como aplicaciones de fácil acceso que utilizan IA, lo que incrementa la probabilidad de su ocurrencia. Este estudio, además de profundizar en el tema y analizar los distintos tipos de *deepfakes*, contribuirá al desarrollo de medidas preventivas, ayudando a reducir los casos vinculados a este fenómeno.

Los antecedentes sobre los *deepfakes* revelan un panorama en constante evolución. Inicialmente desarrollados como herramientas para la creación de contenido artístico y entretenimiento, los *deepfakes* han evolucionado hacia aplicaciones más nefastas, incluyendo la difusión de noticias falsas, la suplantación de identidad y el fraude digital. En 2018 se presentó el primer caso famoso cuando un reportero de Motherboard, Sam Cole, identificó que un usuario de Reddit (una red de comunidades donde las personas pueden sumergirse en sus intereses y pasatiempos), llamado *deepfakes* publicaba videos íntimos falsos sin consentimiento, utilizando un algoritmo de IA para intercambiar los rostros de celebridades. Un año después, este tipo de *deepfakes* se extendió mucho más allá de Reddit, con aplicaciones de fácil acceso que permitían realizar la misma manipulación con cualquier persona que apareciera en una fotografía [4].

Diversos estudios internacionales han explorado los aspectos técnicos de los *deepfakes*, así como sus implicaciones éticas y legales [5], [6], [7]. Sin embargo, la literatura específica sobre su impacto y las estrategias de detección es limitada [8], lo que dificulta una comprensión completa de cómo este fenómeno se está desarrollando en el mundo.

A nivel global, la influencia de los *deepfakes* se entrelaza con factores particulares como la penetración de Internet, el uso de redes sociales, y las dinámicas políticas y culturales. La falta de regulaciones específicas y la limitada capacidad institucional para enfrentar esta amenaza tecnológica agravan la situación [9]. Estudios previos han señalado la vulnerabilidad de las sociedades ante la desinformación digital, pero pocos han abordado de manera exhaustiva cómo los *deepfakes* están siendo utilizados y qué medidas se están tomando para detectarlos y mitigarlos [10], [11].

Además, la intersección entre la tecnología educativa y la ciberseguridad ofrece un marco interesante para analizar las estrategias de detección de *deepfakes* [12]. La incorporación de herramientas tecnológicas en la enseñanza y el aprendizaje puede desempeñar un papel importante en la formación de profesionales capaces de enfrentar los desafíos que plantean los contenidos falsificados. Sin embargo, es necesario investigar cómo estas iniciativas están alineadas con las necesidades actuales de ciberseguridad y qué brechas existen en la capacitación y preparación de los futuros especialistas en este ámbito.

El presente estudio se propone realizar una revisión sistemática de la literatura existente acerca de *deepfakes*, analizando las tecnologías empleadas en su creación, evaluando su impacto social y examinando las estrategias y métodos desarrollados para su detección y mitigación. Este enfoque permitirá identificar las tendencias actuales, las áreas de vulnerabilidad y las oportunidades para mejorar las respuestas institucionales y educativas frente a esta amenaza tecnológica. Con lo anterior se pretende proporcionar una visión detallada y crítica del estado del arte sobre los *deepfakes* a nivel global, identificando las principales tecnologías utilizadas, sus efectos en diferentes sectores de la sociedad y las estrategias efectivas para su detección. Al consolidar y analizar la información disponible, se espera contribuir al desarrollo de políticas más robustas y a la implementación de prácticas educativas que fortalezcan la resiliencia digital.

El artículo se estructura de la siguiente manera: en primer lugar, se presenta el Estado del Arte, donde se revisan los estudios más relevantes sobre *deepfakes* a nivel global. A continuación, se detalla la Metodología utilizada para llevar a cabo la revisión sistemática, incluyendo los criterios de selección de la literatura y las herramientas de análisis empleadas. Posteriormente, se exponen los Resultados obtenidos, seguidos de una Discusión que interpreta estos hallazgos en relación con los objetivos planteados. En la sección Conclusiones, se destacan las principales aportaciones del estudio y se proponen líneas futuras de investigación, y finalmente, se lista la Bibliografía que respalda el trabajo realizado.

2. Estado del arte

La investigación sobre *deepfakes* ha experimentado un notable crecimiento en la última década, reflejando tanto los avances tecnológicos como las preocupaciones sociales asociadas a esta tecnología. Diversas revisiones de la literatura han abordado aspectos específicos de los *deepfakes*, tales como las metodologías de generación, los impactos en la confianza pública y los medios de comunicación, así como las técnicas emergentes para su detección y mitigación [13], [14]. Estas revisiones han contribuido significativamente a la comprensión de los diferentes componentes que conforman el fenómeno de los *deepfakes*, ofreciendo una base sólida para futuras investigaciones.

No obstante, a pesar de la abundancia de estudios individuales, existe una escasez de revisiones sistemáticas que integren las tecnologías subyacentes, el impacto global y las estrategias de detección en un sólo marco analítico. Este artículo se propone llenar este vacío mediante una revisión sistemática de la literatura que abarca estas tres variables. Al consolidar y analizar las investigaciones existentes, se busca identificar tendencias emergentes, lagunas en el conocimiento actual y oportunidades para el desarrollo de estrategias más efectivas de detección y mitigación de *deepfakes*. De esta manera, se pretende ofrecer una visión que no sólo sintetice el estado actual de la investigación, sino que también proponga direcciones futuras para abordar los desafíos que presenta esta tecnología en constante evolución. A continuación, se mencionan revisiones relevantes de la literatura acerca del tema.

El principal hallazgo en [13], consiste en proporcionar una revisión exhaustiva y accesible del estado del arte en la generación y detección de *deepfakes*. Los autores sintetizan los avances recientes categorizando la revisión

en dos áreas fundamentales: la creación de *deepfakes* y las técnicas para su identificación. Además, destacan las herramientas de generación de *deepfakes* disponibles públicamente y los conjuntos de datos utilizados para su evaluación. El estudio también ofrece valiosas perspectivas de investigación, identifica las lagunas existentes en el conocimiento actual y presenta tendencias futuras, lo que facilita el desarrollo continuo de estrategias efectivas para combatir las implicaciones negativas de los *deepfakes* en la sociedad.

En [14], se ofrece una perspectiva integral del paradigma de los *deepfakes*, revisando tanto las tendencias actuales como las futuras. El autor presenta un resumen conciso de las técnicas de aprendizaje profundo utilizadas para crear *deepfakes* y analiza el enfrentamiento entre las tecnologías de generación y detección. Además, explora el potencial de nuevas tecnologías, como los registros distribuidos y la *blockchain*, en el ámbito de la ciberseguridad y la lucha contra el engaño digital. El estudio también examina dos escenarios de aplicación específicos, incluyendo ataques de ingeniería en redes sociales y el Internet de las Cosas, abordando los principales desafíos y oportunidades. Finalmente, se discuten las tendencias futuras y las líneas de investigación, identificando agentes clave y tecnologías prometedoras, lo que contribuye significativamente a la comprensión y desarrollo de estrategias efectivas para prevenir y mitigar las amenazas que representan los *deepfakes*.

Una evaluación completa de las estrategias de detección de *deepfakes* basadas en algoritmos de *Deep Learning*, se encuentra en [15]. Los autores categorizan los métodos de detección según sus aplicaciones, incluyendo detección de video, imagen, audio y multimedia híbrida, proporcionando así una visión estructurada del campo. El estudio tiene como objetivo mejorar el conocimiento sobre cómo se generan y detectan los *deepfakes*, los últimos avances y descubrimientos en este ámbito, las debilidades de los métodos de seguridad actuales y las áreas que requieren mayor investigación. Los resultados indican que las Redes Neuronales Convolucionales son el método de *Deep Learning* más utilizado en las publicaciones, que la mayoría de los artículos se enfocan en la detección de *deepfakes* en videos y que el parámetro de precisión es el que más atención ha recibido.

En algunas revisiones semánticas, como en [16], los autores proponen un análisis de la evolución de los campos semánticos y géneros discursivos del *deepfake* desde su aparición en 2017 y hasta 2021, utilizando diversas técnicas para analizar el campo semántico de los *deepfakes*, tales como teoría del discurso, la teoría de la posverdad, y la teoría tecno-estética. Además, se aplica la valoración narrativa de la escala para la evaluación de artículos de revisión narrativa. El estudio revela una progresiva evolución discursiva en donde emergen nuevos campos semánticos y géneros discursivos del *deepfake*, con una tendencia hacia usos benéficos y no sólo delictivos en la esfera de la industria audiovisual, el activismo, el arte experimental, los usos comerciales, médicos, la publicidad, la propaganda y la educación, entre otros.

En [17], puede encontrarse una evaluación de técnicas para identificar *deepfakes* basados en algoritmos de *Media forensics*; el estudio afirma que estas investigaciones abarcan cinco grandes áreas: (1) detección, (2) atribución y reconocimiento, (3) autenticación pasiva, (4) detección en escenarios realstas y (5) autenticación activa. Así mismo, los autores realizan un análisis acerca de los desafíos que deben enfrentarse, examinando las ventajas de los algoritmos, sus limitaciones y realizando futuras propuestas para mejorar.

Los investigadores de [18], muestran un análisis y una cobertura acerca de la explicación teórica que abordan los *deepfakes*, incluyendo diversas definiciones actuales y precisas del concepto. Asimismo, presentan una descripción general de los estándares y métricas disponibles relacionados con los *deepfakes* para evaluar su generación o detección, que generalmente han sido pasados por alto en encuestas anteriores, proponiendo una nueva visión para la generación y detección en investigaciones previas. Por otra parte, realizan análisis de investigaciones pasadas para monitorear el desarrollo futuro de estas tecnologías y sus aplicaciones.

El impacto social que generan los *deepfake* es elevado, por lo tanto, en [19] los autores realizan una profunda investigación acerca de la prevención de la desinformación y plantean un sólo marco teórico para exhibir el panorama de los *deepfakes* en la sociedad. Utilizando ejemplos como *La guerra de Ucrania* o *La pandemia del COVID-19*, mencionan la gravedad que representa la desinformación para las personas. En [19] tienen como objetivo identificar autores clave, examinar los esfuerzos de colaboración entre ellos, explorar los temas principales bajo escrutinio y resaltar las principales palabras clave, bigramas o trigramas utilizados. En adición, describen estrategias potenciales para combatir la proliferación de *deepfakes* con el fin de preservar la confianza en la información.

En la última década, los métodos de generación de *deepfakes* han aumentado y por lo tanto las estrategias de detección también han sufrido un gran incremento, hoy en día existen diversas herramientas de detección. En [20] se realiza una recopilación de una diversa cantidad de métodos de detección de *deepfakes* y categorizan estos en

cuatro grupos de alto nivel y trece subgrupos llamados de *grano fino*, todo alineado con un marco conceptual estándar unificado. Los autores mencionan sentar las bases para una comprensión más profunda de los detectores de esta tecnología y su generalización, preparando el camino para futuras investigaciones centradas en la creación de los mismos expertos en contrarrestar diversos escenarios de ataque.

Los *deepfake* representan un vacío legal en las leyes de diversos países, pocos tienen leyes preparadas para proteger a los ciudadanos de estos posibles ataques a la falsificación de identidad. En [21] el autor menciona la introducción de regulaciones para los *deepfakes* por parte del reglamento europeo sobre inteligencia artificial. Examina críticamente los riesgos que conlleva no tener regulaciones adecuadas para abordar toda la problemática que implican los *deepfakes*. A través del análisis de las disposiciones de la ley de inteligencia artificial de la Unión Europea, las regulaciones del reglamento general de protección de datos, la jurisprudencia relevante y la literatura académica, el artículo identifica riesgos tanto para los proveedores como para los usuarios de IA. Finalmente, el autor concluye acerca de la importancia y la reflexión acerca de ajustar las leyes para prevenir que los *deepfakes* puedan causar un daño más grande en la sociedad.

La confianza pública se puede ver gravemente afectada al ver como los *deepfakes* se vuelven parte del algoritmo del entretenimiento, a pesar de que un gran sector de la población no analiza el riesgo que significan los *deepfakes*, otro sector se muestra preocupado por el riesgo que conlleva la manipulación de información. En [22], realizan un análisis sobre la tensión de los *deepfakes* por medio de tres distintos casos de estudio; Comportamiento en las masas, político e ideas de objetividad. Por otra parte, el autor sostiene que las tensiones y las implicaciones ético-políticas de los *deepfakes* no se pueden reducir a un problema que pueda resolverse mediante una lógica de detección y verificación algorítmica.

Las publicaciones analizadas en esta sección presentan diversas aproximaciones al estudio de los *deepfakes*, abarcando desde aspectos técnicos y metodológicos hasta implicaciones sociales y legales. En primer término, las revisiones [13], [14], [15], [17], [18] y [20], comparten un enfoque común en la generación y detección de *deepfakes*, aunque difieren en sus metodologías y atacan áreas diferentes. Por ejemplo, [13] y [15] se centran en categorizar y analizar las tecnologías de *Deep Learning* empleadas en la creación y detección de *deepfakes*, destacando el uso predominante de Redes Neuronales Convolucionales y la importancia de la precisión en los métodos de detección. Ambas publicaciones coinciden en la necesidad de desarrollar estrategias más sólidas para enfrentar las amenazas que representan los *deepfakes*, aunque [13] se enfoca más en la identificación de herramientas y *datasets* disponibles públicamente, mientras que [15] ofrece una clasificación detallada de los métodos de detección según sus aplicaciones específicas

Por otro lado, [14] y [20], exponen una visión integral y estructurada del panorama global de los *deepfakes*, pero [14] amplía su análisis al explorar el potencial de tecnologías emergentes como las *blockchains* en la ciberseguridad, así como escenarios de aplicación específicos, por ejemplo, ataques en redes sociales e IoT. En contraste, [20] realiza una categorización más detallada de los métodos de detección, estableciendo un marco conceptual estándar que facilita la comprensión y generalización de los detectores de *deepfakes*

Dentro de artículos como [16], o [18], los autores realizan un análisis al concepto, lo que le permite a la comunidad entender el tema de investigación desde las bases y entender el contexto de publicaciones más técnicas. Existe similitud en aquellas investigaciones dirigidas al impacto social, en [21] y [22] los autores concuerdan en el peligro que representa la vulnerabilidad de la información desde la perspectiva de la sociedad, por un lado, haciendo énfasis en el ámbito legal y como representa una preocupación para las personas estar expuestas ante los diversos peligros que representa esta tecnología, por otro lado, la desconfianza a la información vista desde perspectivas políticas.

3. Materiales y métodos

En concordancia con el objetivo de la investigación, se adoptó un enfoque cualitativo de nivel exploratorio para llevar a cabo esta revisión sistemática de la literatura acerca del tema *deepfakes*. El diseño de la investigación fue documental y transversal, enfocándose en la recolección y análisis de documentos publicados en el período comprendido entre 2020 y 2024. Se consultaron diversas bases de datos académicas, incluyendo Scopus, Web of Science, IEEE Xplore y Scholar Google, para asegurar una cobertura amplia de las publicaciones relevantes. Las palabras clave utilizadas en la búsqueda incluyeron los términos *deepfakes*, *tecnologías de deepfake*, creación de deepfakes, *impacto de deepfakes* y *estrategias de detección de deepfakes*, en combinación con operadores booleanos para optimizar los resultados.

Se incluyeron únicamente artículos publicados en revistas arbitradas, ponencias de congresos y capítulos de libros, excluyendo libros completos, tesis y artículos no arbitrados. Además, se consideraron sólo publicaciones disponibles en español, inglés y francés para asegurar una diversidad lingüística adecuada. Los estudios seleccionados debían abordar directamente alguna de las tres variables de la investigación: tecnologías utilizadas en la creación de *deepfakes*, su impacto en diferentes sectores de la sociedad o las estrategias de detección y mitigación. Se descartaron aquellos documentos que no cumplían con estos criterios temáticos o que presentaban una metodología no alineada con los objetivos del estudio.

El análisis de los datos se realizó mediante técnicas de análisis temático, permitiendo la identificación y categorización de patrones emergentes en la literatura revisada. Cada publicación fue codificada de forma sistemática para extraer información relevante sobre las tecnologías de *deepfake*, sus impactos sociales y las estrategias de detección propuestas. Se utilizaron herramientas de gestión bibliográfica como *rayyan* para organizar y facilitar el procesamiento de los datos cualitativos. Posteriormente, se realizó una síntesis comparativa que permitió resaltar las coincidencias y diferencias entre los estudios, identificando tendencias, brechas en la investigación y áreas potenciales para futuros estudios.

Este estudio presenta algunas limitaciones que deben ser consideradas al interpretar los resultados. La selección de publicaciones delimitó a los idiomas español, inglés y francés, lo que pudo haber excluido investigaciones relevantes en otros idiomas. Además, al enfocarse exclusivamente en artículos arbitrados, ponencias de congresos y capítulos de libros, se omitieron otras fuentes de información como informes técnicos y documentos de trabajo no publicados (*preprints*). Asimismo, el diseño transversal implica que sólo se analizaron las publicaciones disponibles hasta la fecha de corte, sin considerar desarrollos posteriores que pudieran influir en el estado del arte de los *deepfakes*.

4. Resultados

A continuación, se muestra en la Tabla 1 un análisis de los 43 artículos de investigación seleccionados en esta reisión de literatura, la cual abarca diversos enfoques del tema, como las más recientes tecnologías, el impacto social, las regulaciones legales y éticas, aplicaciones en otras ramas de la sociedad, entre otras. Se revisaron artículos de diferentes países y universidades dándole un enriquecimiento cultural con ideas y perspectivas de diferentes autores, buscando obtener un alcance mayor y obtener un análisis más acertado acerca del estado del arte en la actualidad que engloba a la investigación de los *deepfakes*. Además, los resultados abarcan un periodo no mayor a cinco años de antigüedad para considerar únicamente los estudios más recientes, lo que resulta en un análisis de los trabajos actuales. Los documentos están ordenados de forma descendente, priorizando las investigaciones del año en curso al inicio de la tabla.

Tabla 1. Publicaciones acerca de aplciaciones de *Deepfakes*. Fuente: Elaboración propia.

Año y País	Autor y título	Contribución
2024 Alemania	Ramon *et al* [23] *Deepfake Detection in Super-Recognizers and Police Officers*	En la investigación se realiza un análisis del rendimiento de los oficiales de policía de Berlín en la detección de *deepfakes*, incluyendo a personas con habilidades excepcionales de reconocimiento facial (*Super-Recognizers*). Destaca la necesidad de más estudios sobre la detección humana de *deepfakes*, especialmente utilizando *deepfakes* estáticos de última generación, para determinar si la habilidad de los *super recongizers* podría tener valor en entornos específicos.
2024 Estados Unidos	Wang *et al* [24] *Deepfake Detection: A Comprehensive Survey from the Reliability Perspective*	El estudio aporta la identificación de desafíos de confiabilidad, en donde resalta tres desafíos: transferibilidad (capacidad de generalización a nuevos datos), interpretabilidad (*explicabilidad* de los resultados) y robustez (resistencia a diversas manipulaciones). La investigación lleva a cabo estudios de caso que validan el uso de modelos de detección confiables en incidentes reales de *deepfakes*, ayudando a entender el impacto en grupos diversos de víctimas y cómo estos modelos pueden asistir en la justicia.

2024 España	Argallero Fernández [25] *Reconocimiento de DeepFakes*	El proyecto realizado busca estudiar las mejores alternativas a la hora de detectar *deepfakes* en vídeos a través de inteligencia artificial, a la par que, utilizando dichas alternativas, brindar al usuario de una interfaz que le permita realizar la detección y clasificación de forma simplificada
2024 España	Barrientos-Báez *et al* [26] *Imágenes falsas, efectos reales. Deepfakes como manifestaciones de la violencia política de género*	La investigación analiza el uso de *deepfakes* y *cheapfakes* como herramientas para atacar y desacreditar a las mujeres en la política, destacando cómo estas prácticas representan una forma de violencia de género. Los resultados muestran que, aunque la manipulación de estas imágenes suele ser poco sofisticada, el daño a la reputación de las víctimas es considerable.
2024 Reino Unido	Alanazi y Asif [27] *Exploring deepfake technology: creation, consequences and countermeasures*	Este artículo explora los impactos sociales y éticos de los *deepfakes* y presenta resultados experimentales en técnicas de detección. Contribuye al campo identificando los efectos psicológicos de los *deepfakes* y proponiendo mejoras en los algoritmos de detección.
2024 Noruega	Gambín *et al* [14] *Deepfakes: current and future trends*	La investigación examina cómo los avances en inteligencia artificial han facilitado el desarrollo de *deepfakes*. Se abordan temas como el *Deep Learning*, las técnicas que lo rodean y las últimas técnicas de generación y detección. Contribuye con un análisis de los riesgos potenciales y propone estrategias para desarrollar métodos de control y regulación.
2024 Alemania	Pawelec [28] *Decent deepfakes? Professional deepfake developers' ethical considerations and their governance potential*	El artículo contribuye a una comprensión más matizada de la ética de la IA en relación con la IA generativa audiovisual. Además, informa y enriquece empíricamente el debate sobre la gobernanza *deepfake* al incorporar comentarios de desarrolladores y resaltar medidas de gobierno que se dirigen directamente a los desarrolladores y proveedores de *deepfake* y enfatizan el potencial de la ética para frenar los peligros de los *deepfakes*.
2024 Países Bajos	Hameleers [29] *Cheap Versus Deep Manipulation: Comparing Impacts of Deepfakes and Cheapfakes*	Los principales hallazgos indican que la desinformación audiovisual no se percibe como más creíble que la misma desinformación en formato textual. Cabe a destacar que los *deepfakes* se perciben como menos creíbles que los cheapfakes con una narrativa antiinmigración deslegitimadora similar.
2024 Estados Unidos	Agarwal y Ratha [30] *Deepfake Catcher: Can a Simple Fusion be Effective and Outperform Complex DNNs?*	La investigación presenta una innovadora estrategia para la detección de *deepfakes* basada en la fusión de arquitecturas de redes neuronales profundas. El enfoque central del estudio es dividir el conocimiento aprendido por las redes en dos tipos: fijo (el que se conserva durante el entrenamiento) y adaptativo (el que se ajusta según la tarea). El algoritmo propuesto supera varios algoritmos existentes en estos escenarios desafiantes por márgenes significativos. Esto lo convierte en una solución viable para implementaciones en dispositivos con recursos limitados, como los móviles.

2024 Luxemburgo	Astrid *et al* [31] *Detecting Audio-Visual Deepfakes with Fine-Grained Inconsistencies*	Esta investigación propone una mejora en los métodos existentes para la detección de *deepfakes* audiovisuales, que tradicionalmente se enfocan en características de alto nivel y tienden a pasar por alto artefactos más sutiles inherentes a los *deepfakes*. Los experimentos realizados con los conjuntos de datos desafíos de detección de *deepfakes* y *FakeAVCeleb* muestran que el método propuesto supera a los algoritmos de última generación en términos de generalización, tanto en escenarios dentro del mismo conjunto de datos como en configuraciones de datos cruzados.
2024 Estados Unidos	Nadimpalli y Rattani [32] *Social Media Authentication and Combating Deepfakes using Semi-fragile Invisible Image Watermarking*	Los investigadores proponen una técnica de marca de agua invisible para autenticar imágenes, la cual es frágil ante manipulaciones faciales, pero resistente a modificaciones de imagen inofensivas. Mediante redes adversarias y un diseño innovador, el método permite insertar un mensaje secreto que se recupera solo en imágenes no alteradas por *deepfakes*, logrando una alta precisión y resistencia a ataques de eliminación de marca de agua.
2024 India	Kingra *et al* [33] *SFormer: An end-to-end spatio-temporal transformer architecture for deepfake detection*	El artículo presenta *SFormer*, un nuevo modelo de detección de *deepfakes* basado en una arquitectura de transformador. Dado el rápido avance de las tecnologías de generación de *deepfakes*, estas manipulaciones digitales representan riesgos importantes para la justicia, el periodismo y la política. *SFormer* emplea un transformador *Swin* para el análisis espacial y temporal, lo que reduce la complejidad computacional y mejora la capacidad de generalización y la robustez frente a diversos tipos de manipulaciones. Las pruebas en conjuntos de datos de última generación muestran que *SFormer* supera a otros modelos, logrando hasta un 100% de precisión en algunos casos.
2024	Samuel-Okon *et al* [34] *Assessing the Effectiveness of Network Security Tools in Mitigating the Impact of Deepfakes AI on Public Trust in Media*	Este estudio presenta el marco *Anti-DFK*, una estrategia integral para frenar la difusión de *deepfakes* en plataformas sociales como Instagram, Facebook, YouTube y Twitter. La solución combina motores de detección basados en aprendizaje profundo, marcas de agua digitales y controles avanzados de acceso a redes. Con una precisión de hasta 97% en entornos controlados, el marco demuestra alta efectividad en la detección y resistencia de las marcas de agua, reduciendo en un 80% la diseminación de *deepfakes* mediante filtros de contenido. La investigación resalta el impacto negativo de los *deepfakes* en la confianza pública y la importancia de enfoques integrados para restaurar la credibilidad en los medios digitales.
2024 Corea del Sur	Zang *et al* [35] *SingFake: Singing Voice Deepfake Detection*	Al evaluar sistemas avanzados de detección de habla en este conjunto, se observó un bajo rendimiento, aunque el entrenamiento con *SingFake* mejoró los resultados. El estudio revela desafíos adicionales, como la variabilidad de cantantes, *codecs* y contextos musicales, destacando la necesidad de enfoques específicos para detectar *deepfakes* en voces cantadas.
2024 Estados Unidos	Datta *et al* [36] *Exposing Lip-syncing Deepfakes from Mouth Inconsistencies*	En el área de la cultura y el arte se presenta un enfoque novedoso llamado *LIPINC* para la detección de *deepfakes* en videos de sincronización labial (*lip-syncing*), una variante particularmente difícil de detectar debido a que los artefactos solo afectan la región de los labios. *LIPINC* identifica inconsistencias temporales en los movimientos de la boca a través de los fotogramas adyacentes y en todo el video, lo que permite detectar irregularidades que no son fácilmente perceptibles. El modelo propuesto supera a las técnicas actuales en varias bases de datos de *deepfakes*, demostrando su efectividad en la identificación de este tipo de manipulaciones.

2024 China	Sun *et al* [37] *FakeTracer: Catching Face-swap DeepFakes via Implanting Traces in Training*	Se proponen dos tipos de rastros: el rastro sostenible (*STrace*) y el rastro borrable (*ETrace*), que permiten exponer los *deepfakes* al identificar los rostros manipulados generados. Los experimentos realizados demuestran la eficacia de este enfoque para detectar *deepfakes* de cambio de rostro.
2024 Túnez	Nguyen *et al* [38] *LAA-Net: Localized Artifact Attention Network for Quality-Agnostic and Generalizable Deepfake Detection*	*LAA-Net* utiliza un mecanismo de atención explícita dentro de un marco de aprendizaje multitarea. Al combinar estrategias de atención basadas en mapas de calor y auto consistencia, *LAA-Net* se enfoca en las pequeñas regiones vulnerables donde se encuentran los artefactos. Además, introduce la Red Piramidal de Características Mejoradas, que mejora la propagación de características discriminativas de bajo nivel, reduciendo la redundancia. Los experimentos muestran que este enfoque supera a otros en métricas clave como el Área Bajo la Curva y la Precisión Promedio.
2024 China	Qu *et al* [39] *DF-RAP: A Robust Adversarial Perturbation for Defending Against Deepfakes in Real-World Social Network Scenarios*	Los métodos previos eran sensibles a la compresión, lo que limitaba su efectividad en plataformas en línea. Para resolver este problema, se propone un Generador Adversarial de Aproximación de Compresión que modela explícitamente la compresión en redes sociales, y se integra en el modelo *deepfake* para generar la protección *DF-RAP*. Los experimentos muestran que *DF-RAP* es eficaz para proteger las imágenes faciales contra *deepfakes* en plataformas que emplean compresión agresiva. También se analiza cómo la compresión afecta las imágenes y se crea un conjunto de datos de transmisión en redes sociales para futuras investigaciones.
2024 India	Pandey *et al* [40] *Detecting low-resolution deepfakes: an exploration of machine learning techniques*	Los resultados, evaluados mediante validación cruzada *K-fold*, muestran una alta precisión en la detección: 99.97% para conjuntos de datos de alta resolución utilizando *Random Forest* (herramienta de software), 98.27% para baja resolución con *SVM polynomial* (herramienta de software) y 98.72% para conjuntos mixtos utilizando un clasificador de votación. Esto demuestra la eficacia de este enfoque para detectar deepfakes en diferentes resoluciones de imagen.
2023 India	Krishnan y Krishnan [41] *MFAAN: Unveiling Audio Deepfakes with a Multi-Feature Authenticity Network*	En el contexto del audio *MFAAN* combina características como los coeficientes cepstrales en Mel, los coeficientes cepstrales en frecuencia lineal y la Transformada de Fourier de Tiempo Corto en Croma en rutas paralelas, lo que permite una comprensión más profunda y precisa del contenido de audio. Las evaluaciones preliminares en dos conjuntos de datos de referencia muestran su eficacia, logrando precisiones del 98.93% y 94.47%, respectivamente, lo que resalta el potencial de *MFAAN* como una herramienta clave en la lucha contra los *deepfakes* de audio.
2023	Hou *et al* [42] *Evading DeepFake Detectors via Adversarial Statistical Consistency*	*StatAttack*, un método de ataque diseñado para evadir detectores de *deepfakes* minimizando las diferencias estadísticas entre imágenes reales y falsas. StatAttack utiliza degradaciones sensibles a estadísticas como exposición, desenfoque y ruido, aplicándolas de manera adversarial a las imágenes falsas. Además, propone una pérdida basada en la distribución para reducir las diferencias estadísticas entre imágenes reales y falsificadas. La versión mejorada, *MStatAttack*, extiende el ataque con degradaciones en múltiples capas. Los experimentos muestran que ambos ataques son efectivos para evadir detectores de *deepfakes* en diversos entornos y modelos, subrayando la necesidad de enfoques más robustos en la detección.
2023 China	Li *et al* [43] *How Generalizable are Deepfake Detectors? An Empirical Study*	En la época de creación de modelos de detección más eficientes en esta investigación analizaron seis conjuntos de datos de *deepfakes*, cinco métodos de detección de imágenes y dos técnicas de aumento de modelos, se confirma que los detectores actuales no logran generalizar en configuraciones de *zero-shot* (sin datos previos específicos). Además, los detectores tienden a aprender propiedades específicas de los

		métodos de síntesis en lugar de características discriminativas, lo que limita su efectividad en nuevos contextos. El estudio identifica neuronas que contribuyen universalmente a la detección, lo que abre posibilidades para mejorar la generalización sin datos previos específicos.
2023 Estados Unidos	Doss *et al* [44] *Deepfakes and scientific knowledge dissemination*	A través de una encuesta a estudiantes, educadores y la población adulta, los resultados muestran que entre el 27% y el 50% de los individuos no pueden distinguir entre videos auténticos y *deepfakes*. La investigación sugiere que combatir los *deepfakes* debe centrarse en el contexto social en el que estos circulan, como una estrategia prometedora para reducir su impacto.
2023 Reino Unido	Bray *et al* [45] *Testing human ability to detect deepfake images of human faces*	Los resultados revelan que la precisión de los participantes fue apenas superior al azar, con un 62% de aciertos, y las intervenciones no mejoraron significativamente esta precisión. Un hallazgo preocupante es que la confianza de los participantes en sus respuestas era alta, independientemente de su precisión. Además, ciertos rostros resultaron más fáciles de clasificar correctamente, mientras que otros fueron mucho más desafiantes, mostrando que la precisión variaba entre 85% y 30% según la imagen. La investigación concluye que esta dificultad para identificar *deepfakes* y la alta confianza en respuestas incorrectas subrayan la necesidad de una acción urgente para enfrentar la amenaza que representan los *deepfakes*.
2022 Italia	Leone [46] *L'idéologie sémiotique des deepfakes*	Las redes generativas adversativas han transformado significativamente la producción y recepción de *deepfakes*, haciéndolos cada vez más difíciles de identificar para las personas. Aunque actualmente se consideran mayormente como entretenidos, prevén que la evolución tecnológica pronto impedirá a los humanos distinguirlos de contenidos auténticos. Esta situación genera una crisis epistemológica en la comunicación, debido a que las tecnologías digitales y la IA están alterando radicalmente las condiciones de intercambio simbólico y la percepción de la realidad.
2022 Francia	Atamna *et al* [47] *Détection de Deepfakes par Réseaux de Convolution : Performances et Limites Actuelles*	Los detectores de *deepfakes* basados en aprendizaje profundo, como XceptionNet, logran buenos resultados con tipos conocidos de manipulaciones faciales, pero presentan dificultades para generalizar su desempeño a nuevas variantes no vistas durante el entrenamiento. Este hallazgo evidencia la necesidad de mejorar las herramientas de detección para que puedan identificar eficazmente *deepfakes* emergentes, garantizando así una mayor eficacia en la identificación de contenidos falsificados novedosos.
2022 Francia	Pignier [48] *L'énonciation à l'épreuve de l'« I.A. ». Qu'est-ce-qu'énoncer veut dire ?*	Los *deepfakes* modifican significativamente la dinámica comunicativa al simular signos sin el correspondiente acto de significación, lo que dificulta la percepción de la verdad en imágenes y videos. Este fenómeno cuestiona la relación entre emisor y receptor en el proceso de co-enunciación, debido a que las representaciones falsificadas confunden a quienes las perciben y las aceptan como genuinas. Además, los *deepfakes* ponen a prueba la capacidad de distinguir entre contenido auténtico y manipulado, erosionando la confianza en la información visual y audiovisual. La investigación subraya la necesidad de reevaluar los fundamentos de la comunicación y desarrollar nuevas metodologías para verificar la veracidad de los mensajes en el entorno digital actual.
2022 Francia	Nelson [49] *Fake news et deepfakes: une approche cyberpsychologique*	Los *deepfakes* funcionan como vectores iconográficos para las *fake news*, aumentando la capacidad de engaño al presentar información falsa de manera visualmente convincente. La revisión de investigaciones en ciberpsicología mostró que diversos factores psicológicos influyen en la creencia de noticias falsas, y que los *deepfakes* potencian estos procesos al facilitar la difusión de contenido manipulado que parece auténtico. Este hallazgo resalta la necesidad de desarrollar estrategias efectivas para detectar y contrarrestar los *deepfakes*, con el objetivo de reducir su impacto en la percepción pública y la integridad de la información.

2022 Francia	Caliandro [50] *Fake Art, entre le contrefait et le contrefactuel*	El uso de técnicas avanzadas de reproducción, síntesis, IA y aprendizaje automático está redefiniendo la creación artística, transformando tanto el potencial del arte como sus modalidades de recepción estética. Estas herramientas, agrupadas bajo el concepto de *Fake Art*, generan nuevas formas de interacción entre lo *fake* y el arte, introduciendo cambios semióticos que alteran la percepción y el valor ontológico de las obras. El análisis de diversas obras de arte demuestra cómo estas tecnologías desafían las convenciones tradicionales y modifican el imaginario estético, subrayando la necesidad de reevaluar las creencias estéticas en el contexto contemporáneo.
2022 Francia	Lloveria [51] *Le deepfake et son métadiscours : l'art de montrer que l'on ment*	La comunicación sobre *deepfakes* en los medios, ya sea textual o visual, está acompañada de metadiscurso destinado a exponer mentiras de manera deliberada y explícita. A través del análisis de escritos periodísticos y ejemplos visuales, identificaron y categorizaron diversas formas de estos metadiscurso tanto en su expresión como en su contenido. Este hallazgo demuestra cómo los medios estructuran y representan la problemática de los *deepfakes*, proporcionando una comprensión detallada de las estrategias comunicativas utilizadas para abordar y desmentir estas falsificaciones digitales.
2022 Francia	Fabre [52] *Urdoxa, le deepfake d'information comme usage tactique du vague*	Los *deepfakes* desafían las nociones tradicionales de realidad, verdad y creencia al manipular imágenes faciales, lo que requiere una reevaluación de estos conceptos dentro de la semiótica pragmática estadounidense. Utilizando la lógica de la vaguedad de C.S. Peirce, se propone una nueva perspectiva para entender el fenómeno contemporáneo de los *deepfakes*. Este enfoque establece límites claros entre los conceptos fundamentales y evaluar la eficacia de los *deepfakes* como signos, promoviendo el desarrollo de una creencia científica basada en la lógica para investigar estas falsificaciones digitales.
2022 Francia	Dondero [53] *Du portrait pictural aux deepfakes : le visage en tant que totalité*	Las estrategias de acumulación y manipulación de imágenes faciales en la fotografía compuesta de Francis Galton muestran similitudes con las técnicas de producción y detección de *deepfakes*. Al contrastar ambos métodos, se analiza cómo se gestionan la singularidad y la generalidad, destacando la manipulación de rasgos faciales para crear representaciones complejas. Además, se explora cómo estas estrategias históricas influyen en las prácticas actuales de *deepfakes*, especialmente en la tradición del retrato. Este análisis profundiza en las bases técnicas y estéticas de los *deepfakes*, mostrando su evolución desde métodos fotográficos tradicionales hacia tecnologías avanzadas de IA.
2022 Emiratos Árabes Unidos	Narayan *et al* [54] *Deepfakes and scientific knowledge dissemination*	Desarrolla un nuevo conjunto de datos llamado DeePhy, que incluye 5040 videos *deepfake* generados con diferentes técnicas y niveles de complejidad (cambios de rostro aplicados una, dos y tres veces), abordando la evolución o *filogenia* de los *deepfakes*.
2022 Alemania	Appel y Prietzel [55] *The Detection of Political Deepfakes*	Este estudio explora la eficacia de los métodos de detección de *deepfakes* en contextos políticos. Su contribución reside en proporcionar evidencia empírica sobre el impacto de los *deepfakes* en la percepción pública y en la confianza en medios políticos, ayudando a identificar áreas donde los métodos de detección pueden ser más efectivos.
2022 Estados Unidos	Helmus [56] *Artificial Intelligence, Deepfakes, and Disinformation*	El impacto de los *deepfakes* en la diseminación de desinformación es la parte central de este estudio. El mismo aporta evidencia empírica sobre cómo los *deepfakes* pueden manipular la opinión pública y sugiere recomendaciones para la política y la regulación, en un esfuerzo por mitigar los riesgos asociados.

2022 Estados Unidos	Taeb y Chi [57] *Comparison of Deepfake Detection Techniques through Deep Learning*	La investigación demuestra que el modelo VGG19 logra la mayor precisión (95%), remarca la importancia en toda la mejora de la detección de falsificaciones faciales, especialmente en el campo de la ciberseguridad.
2922 Alemania	Eberl *et al* [58] *Using Deepfakes for Experiments in the Social Science*	Este estudio piloto analizó cómo los *deepfakes* pueden ser utilizados en investigaciones sociales, explorando la influencia de la apariencia física en la evaluación de profesores. Los resultados mostraron que los estudiantes tienden a calificar mejor a los instructores percibidos como más atractivos, utilizando *deepfakes* como herramienta de manipulación experimental para validar estas hipótesis.
2021 China	Zhou y Lim [59] *Joint Audio-Visual Deepfake Detection*	Los experimentos mostraron que la detección conjunta supera en desempeño a los modelos entrenados únicamente en modalidades visuales o auditivas, sugiriendo que la sincronización visual/auditiva es un indicador efectivo para identificar contenido falso.
2021 Reino Unido	Ramachandran *et al* [60] *An Experimental Evaluation on Deepfake Detection using Deep Face Recognition*	La mayoría de los avances en *Deep learning* han tenido gran precisión para diversas aplicaciones tecnológicas, sin embargo, estos avances también han contribuido a la generación de imágenes falsas, la investigación realiza una evaluación con el método *Deep face recognition*. En pruebas con los conjuntos de datos *Celeb-DF* y *FaceForensics++*, el estudio alcanzó altos niveles de precisión (Área bajo la curva de hasta 0.99), superando significativamente a las redes neuronales convolucionales tradicionales en términos de tasas de error.
2021 Polonia	Zendran y Rusiecki [61] *Swapping Face Images with Generative Neural Networks for Deepfake Technology – Experimental Study*	La investigación presenta un análisis de cuatro métodos destacados para la generación de *deepfakes*: *autoencoders*, *autoencoders* variacionales, redes generativas adversariales de *autoencoders* variacionales y redes generativas adversariales de ciclo, específicamente en la conversión de rostros de una persona a otra. Los experimentos se realizaron en una tarea de intercambio de rostros usando datos preprocesados del conjunto VoxCeleb2. Ante la ausencia de métodos numéricos para comparar *deepfakes*, se propuso una evaluación descriptiva, donde los resultados fueron valorados mediante una evaluación visual.
2021 España	Hazeu-Gonzales [62] *Sistemas cognitivos artificiales para la detección de deepfakes usando datos audiovisuales*	Una de las principales aportaciones es la mejora en la capacidad de generalización de estos modelos, una característica en la que los modelos actuales de detección de *deepfakes* suelen ser limitados. La efectividad de la propuesta se demuestra mediante una buena capacidad de generalización en comparación con el estado del arte actual.
2021 España	Grandío Rodríguez [63] *Desarrollo de una aplicación de creación de deepfakes en tiempo real para el uso del agente en cubierto informático*	La investigación propone un uso de beneficio para la ley al utilizar las herramientas de creación de *deepfakes*, más específicamente en el contexto de la ley de enjuiciamiento criminal para que los investigadores se infiltren en redes delictivas y obtenga información relevante sobre ciberdelincuentes. Por último, examina los beneficios y posibles aplicaciones de los *deepfakes* en la recolección de datos en investigaciones virtuales, así como los aspectos éticos y legales asociados a su uso.

2020 Suiza	Korshunov y Marcel [64] *Deepfake detection: humans vs. machines*	Se usaron 120 videos preseleccionados (60 *deepfakes* y 60 originales) del Facebook *deepfake database*, parte del *Kaggle's Deepfake Detection Challenge* 2020. Un promedio de 19 participantes respondió si la cara del video era real o falsa. Los resultados mostraron que tanto los humanos como los algoritmos de detección, basados en redes neuronales como Xception y EfficientNets, pueden ser engañados por *deepfakes*, pero de maneras distintas. En particular, los algoritmos tienen dificultades para detectar videos *deepfake* que los humanos pueden identificar con facilidad. Esto profundiza el cambio que había entre los algoritmos de detección de *deepfake* que existían antes con los que existen en la actualidad.

El análisis de las 43 publicaciones recopiladas en la Tabla 1, muestra una amplia distribución geográfica y temática de las investigaciones sobre *deepfakes*. Estos trabajos provienen de diversos países, incluyendo Estados Unidos, China, Reino Unido, Francia, España, Alemania, entre otros. Se aprecia que la mayor parte de la investigación sobre *deepfakes* se concentra en países con alto desarrollo tecnológico y capacidades avanzadas en IA. Al categorizar las publicaciones, se identifican varias áreas temáticas principales.

4.1. Tecnologías de generación y detección de deepfakes

Autores como [25], [30], [31], [33], [37], [45], y [61] se enfocan en el desarrollo de nuevas técnicas para la generación y detección de *deepfakes*. Todos reconocen la importancia de mejorar los algoritmos de aprendizaje profundo y las redes neuronales para aumentar la eficacia en la detección. Mientras que [61] se centra en mejoras de las redes generativas adversativas para la generación de *deepfakes*, [30], y [32] proponen métodos basados en análisis forense digital para su detección. Por otro lado, [49] explora el uso de *blockchain* como una estrategia innovadora para verificar la autenticidad de los medios.

En [64] y [45] es posible observar una comparación a través de los años en la capacidad de las personas en detectar imágenes alteradas por inteligencia artificial. En 2020 era posible y se recomendaba mejorar los algoritmos de detección, no obstante, en el estudio hecho en 2023, identificarlos se ha vuelto una tarea más complicada, además, los usuarios no son capaces de darse cuenta de su error, por lo tanto, se concluye hacer consciencia a la población del peligro que representan actualmente estas tecnologías.

Impacto social y ético

Publicaciones como [26], [29], [34], [44], [46], [48], [49], [55] y, [60], abordan las implicaciones éticas y sociales de los *deepfakes*. Todos destacan los riesgos asociados con la desinformación, la manipulación de la opinión pública y las violaciones a la privacidad. Por otro lado, [26] y [55], analizan el impacto político de los *deepfakes* en procesos electorales, mientras que [48] y [49] exploran las consecuencias psicológicas y sociales en individuos afectados por estas tecnologías. [34] profundiza en los desafíos éticos relacionados con la identidad y la autenticidad en la era digital.

4.2. Aspectos legales y regulatorios

En [23], [24], [27], [28], [32], [51], [52], y [56], investigan el marco legal existente y proponen regulaciones específicas para abordar los desafíos que presentan los *deepfakes*. Todos coinciden en la necesidad de actualizar las leyes para proteger a individuos y sociedades de los posibles daños. [32] se enfoca en las leyes de propiedad intelectual, mientras que [51] analiza la responsabilidad legal en la difusión de *deepfakes* maliciosos. [52] propone un enfoque internacional coordinado para establecer estándares globales.

4.3. Aplicaciones positivas y creativas de los deepfakes

Trabajos como [35], [39], [50], [53], y [63], exploran usos beneficiosos de los *deepfakes* en áreas como el cine, áreas de seguridad, la educación y el arte. Reconocen el potencial de esta tecnología para innovar en la creación de contenidos. [53], se centra en aplicaciones cinematográficas para recrear personajes históricos, mientras que [63] propone una aplicación de creación de *deepfakes* en tiempo real para el uso de agentes en cubierto informáticos.

4.4. Estrategias de detección y mitigación

Estudios como [30], [32], [36], [40], [43], [47], [54], [59], y [62], trabajan en el desarrollo de métodos avanzados para detectar y mitigar el impacto de los *deepfakes*. Coinciden en la importancia de la precisión y la adaptabilidad de los sistemas de detección ante nuevas técnicas de falsificación. En [54] los autores hablan de un nuevo conjunto de datos, trabajado con tecnologías específicas. Mientras que en [40] se propone el uso de software especializado para mejorar los detectores de imágenes en baja resolución, por otro lado, [59] demuestra que la detección conjunta supera en desempeño a los modelos entrenados únicamente en modalidades visuales o auditivas.

Al analizar las publicaciones, se observan enfoques multidisciplinarios, donde algunos autores integran perspectivas tecnológicas, éticas y legales, reflejando la complejidad del fenómeno de los *deepfakes*. Por ejemplo, [26] y [34] combinan análisis éticos con consideraciones técnicas. También se aprecia que algunas publicaciones resultan de colaboraciones entre investigadores de diferentes países, lo que indica un esfuerzo global por abordar los desafíos asociados con los *deepfakes*. Mientras que algunos estudios utilizan metodologías empíricas y experimentales, como [43] y [31], otros adoptan enfoques teóricos y de revisión bibliográfica, como [50] y [54]. Algunos autores se enfocan en el estado actual de la tecnología y sus implicaciones inmediatas, mientras que otros, como [14], proyecta escenarios futuros y tendencias emergentes.

En términos de coincidencias específicas entre autores, [46] y [61] investigan mejoras en las redes generativas adversativas para la generación de *deepfakes*, pero [61] se concentra en la calidad visual, mientras que [46] enfatiza la eficiencia computacional. [30] y [31] analizan el impacto político de los *deepfakes*, coincidiendo en la preocupación por la manipulación electoral, pero difieren en los casos de estudio y contextos geográficos. [45] y [46] trabajaron en métodos de detección, pero utilizan técnicas distintas; [45] aplica análisis forense de imágenes, mientras que [46] implementa algoritmos de aprendizaje automático. Las diferencias entre autores resaltan la diversidad de enfoques y la necesidad de abordar el fenómeno desde múltiples perspectivas. Mientras que algunos se centran en soluciones técnicas, otros enfatizan las implicaciones sociales, éticas y legales.

Las 43 publicaciones analizadas reflejan un amplio espectro de investigaciones que contribuyen al entendimiento y manejo de los *deepfakes*. La clasificación por países y áreas temáticas evidencia la naturaleza global del desafío y la importancia de la colaboración interdisciplinaria para desarrollar estrategias efectivas de detección y mitigación.

5. Discusión

Los hallazgos de esta investigación proporcionan una visión comparativa sobre las tecnologías, impactos y estrategias de detección de los *deepfakes*, destacando la evolución y diversidad de enfoques en el campo. De las 43 publicaciones analizadas en la sección Resultados, se observa un predominio de investigaciones en países tecnológicamente avanzados como Estados Unidos, China y Alemania, lo cual resalta el acceso diferencial a recursos para desarrollar tanto tecnologías de generación como estrategias de mitigación.

Una coincidencia significativa entre las publicaciones citadas en el Estado del Arte y los resultados obtenidos en la Tabla 1, es el énfasis en los métodos de detección basados en aprendizaje profundo. Estudios como [15] y [48] destacan la efectividad de modelos como XceptionNet y redes neuronales convolucionales, coincidiendo en la necesidad de superar sus limitaciones en la generalización hacia variantes desconocidas de *deepfakes*. Sin embargo, mientras [15] aborda una categorización detallada de métodos según su aplicación, [48] se centra en los retos asociados a nuevas manipulaciones no vistas durante el entrenamiento.

Otra área de convergencia es el análisis ético y social de los *deepfakes*. Publicaciones como [19], [26] y [50] enfatizan el impacto negativo de los *deepfakes* en contextos sociales y políticos, desde la desinformación hasta la violencia de género. No obstante, los enfoques difieren en sus perspectivas, [19] utiliza marcos teóricos para abordar la desinformación, mientras que [26] explora específicamente las implicaciones en la política de género. Por su parte, [50] resalta la intersección entre la ciberpsicología y los efectos psicológicos de los *deepfakes* en la percepción pública.

En cuanto a las aplicaciones positivas, estudios como [40] y [54] identifican oportunidades para el uso ético de los *deepfakes* en áreas como el arte y la protección de datos. A pesar de ello, persiste una brecha en la literatura sobre cómo maximizar estos beneficios mitigando riesgos éticos y legales. Esto resuena con las observaciones en [13] y [29], que señalan la necesidad de una regulación más sólida y medidas de gobernanza ética.

Una divergencia notable se encuentra en la exploración de tecnologías emergentes. Mientras [14] y [28] abordan el potencial de *blockchain* y registros distribuidos para aumentar la ciberseguridad, estudios como [40] y

[44] se enfocan en adaptaciones específicas de modelos existentes para mejorar su robustez frente a manipulaciones avanzadas. Estas diferencias subrayan la diversidad de enfoques tecnológicos en el campo, con algunos priorizando la innovación disruptiva y otros la optimización de métodos actuales.

Finalmente, un aspecto consistente en los resultados es la limitada capacidad de generalización de los detectores actuales, como se observa en [24], [44] y [58]. Esto destaca una necesidad urgente de desarrollar métodos más robustos y adaptativos que puedan enfrentar los desafíos dinámicos de las tecnologías *deepfake*, una conclusión que también se menciona en el Estado del Arte.

Los resultados de esta revisión sistemática amplían la comprensión sobre las tecnologías *deepfake* al conectar hallazgos técnicos, sociales y éticos. Al integrar estas perspectivas, el estudio confirma tendencias identificadas previamente, además destaca áreas críticas para futuras investigaciones, como la mejora de la generalización en métodos de detección y el establecimiento de marcos regulatorios más efectivos. Este enfoque integral muestra la importancia de abordar los *deepfakes* desde una perspectiva interdisciplinaria y global.

6. Conclusiones

Esta investigación logró cumplir su objetivo de proporcionar una visión detallada y crítica del estado del arte sobre los *deepfakes* a nivel global, identificando las principales tecnologías utilizadas, sus efectos en diferentes sectores de la sociedad y las estrategias efectivas para su detección. A través de una revisión sistemática de 43 publicaciones científicas, se identificaron las principales innovaciones tecnológicas, se analizaron sus implicaciones sociales, éticas y legales, y se destacaron las oportunidades y desafíos en la detección y mitigación de estos contenidos manipulados.

En términos tecnológicos, el estudio evidenció que las redes generativas adversativas y los algoritmos de aprendizaje profundo continúan siendo las herramientas más avanzadas en la creación de *deepfakes*, mientras que los detectores basados en IA enfrentan retos significativos, como la generalización a nuevas variantes de manipulación. Estas limitaciones subrayan la necesidad de desarrollar métodos más robustos y adaptativos que puedan responder eficazmente a las manipulaciones emergentes.

Desde el punto de vista social y ético, se constató que los *deepfakes* representan una amenaza para la confianza pública y la privacidad, planteando, también, desafíos específicos en contextos como la política, el periodismo y la violencia de género. Además, se identificaron oportunidades positivas para su aplicación en áreas como el arte, la educación y la ciberseguridad, aunque estas aplicaciones requieren marcos normativos claros para minimizar riesgos. En cuanto a las estrategias de detección y mitigación, estas han avanzado considerablemente, destacando enfoques innovadores como el uso de *blockchain*, la detección multimodal y las marcas de agua digitales. Sin embargo, persisten barreras técnicas y éticas que deben abordarse mediante una colaboración interdisciplinaria e internacional.

Este trabajo aporta una comprensión integral del fenómeno de los *deepfakes*, conectando hallazgos técnicos, sociales y éticos, y subrayando la importancia de abordarlo desde una perspectiva global. Las conclusiones aquí presentadas consolidan el conocimiento existente y proponen direcciones clave para futuras investigaciones, como el desarrollo de detectores más efectivos, la exploración de aplicaciones éticamente responsables y la implementación de políticas reguladoras más inclusivas y adaptativas. Esto refuerza la relevancia de continuar investigando y actuando en un campo que evoluciona rápidamente y que tiene un impacto creciente en múltiples aspectos de la sociedad.

7. Trabajos futuros

El análisis realizado en esta investigación revela múltiples áreas que requieren atención adicional para abordar los desafíos que presentan los *deepfakes* a nivel global. Una de las principales direcciones para futuros estudios es el desarrollo de detectores con mayor capacidad de generalización, capaces de identificar manipulaciones emergentes y adaptarse a nuevas técnicas de generación. Esto podría incluir el diseño de algoritmos híbridos que combinen enfoques basados en aprendizaje profundo con análisis forense digital.

Asimismo, se identificó la necesidad de explorar estrategias integrales de regulación y gobernanza que aborden las implicaciones legales, éticas y sociales de los *deepfakes*. Investigaciones futuras podrían centrarse en desarrollar marcos normativos específicos que permitan equilibrar los riesgos asociados con su uso malicioso y las oportunidades que ofrecen en contextos positivos, como el arte y la educación. Otro ámbito prometedor es el

análisis del impacto psicológico y social de los *deepfakes* en diferentes grupos demográficos y culturales. Esto facilitará la comprensión de las dinámicas de desinformación y manipulación en contextos específicos, proporcionando bases para diseñar campañas de sensibilización y educación digital.

Por último, es fundamental investigar aplicaciones éticamente responsables de los *deepfakes* en áreas emergentes como la realidad virtual, la simulación educativa y la preservación cultural. Estos estudios podrían enfocarse en maximizar los beneficios de esta tecnología mientras se minimizan sus riesgos, asegurando un uso seguro y ético. El campo de los *deepfakes* ofrece un amplio potencial para investigaciones futuras que integren enfoques técnicos, sociales y legales. Estas iniciativas serán esenciales para enfrentar los desafíos asociados con esta tecnología en constante evolución y aprovechar sus aplicaciones de manera responsable y beneficiosa para la sociedad.

8. Referencias

[1] Malik, A., Kuribayashi, M., Abdullahi, S., Khan, A. (2022). Deepfake detection for human face images and videos: a survey. *IEEE Access, 10*, 18757-18775. https://doi.org/10.1109/ACCESS.2022.3151186

[2] Kirchengast, T. (2020). Deepfakes and image manipulation: criminalisation and control. *Information & Communications Technology Law, 29* (3), 308-323. https://doi.org/10.1080/13600834.2020.1794615

[3] Bañuelos, J., Abbruzzese, M. (2023). From deepfake to deeptruth: toward a technological resignification with social and activist uses. En M. Cebral-Loureda, E. G. Rincón-Flores, G. Sanchez-Ante (Eds.), *What AI Can Do: Strengths and Limitations of Artificial Intelligence* (pp. 75-92). Taylor & Francis Group. https://doi.org/10.1201/b23345

[4] Hao, K. (2020). *El año que los deepfakes salieron del lado oscuro y se masificaron*. MIT Technology Review. https://www.technologyreview.es/s/13049/el-ano-que-los-deepfakes-salieron-del-lado-oscuro-y-se-masificaron

[5] Songja, R., Promboot, I., Haetanurak, B., Kerdvibulvech, C. (2023). Deepfake AI images: should deepfakes be banned in Thailand?. *AI and Ethics, 4*, 1519-1531. https://doi.org/10.1007/s43681-023-00350-0

[6] De Ruiter, A. (2021). The distinct wrong of deepfakes. *Philosophy & Technology*, *34* (4), 1311-1332. https://doi.org/10.1007/s13347-021-00459-2

[7] Maniyal, V., Kumar, V. (2024). Unveiling the Deepfake Dilemma: Framework, Classification, and Future Trajectories. *IT Professional*, *26* (2), 32-38. https://doi.org/10.1109/MITP.2024.3369948

[8] Ramos-Zaga, F. (2024). Deepfake: Análisis de sus implicancias tecnológicas y jurídicas en la era de la Inteligencia Artificial. *Derecho Global. Estudios sobre Derecho y Justicia*, *9* (27), 359-387. https://doi.org/10.32870/dgedj.v9i27.754

[9] Vinogradova, E. A. (2023). The malicious use of political deepfakes and attempts to neutralize them in Latin America. *Latinskaya Amerika,* (5), 35-48. https://doi.org/10.31857/S0044748X0025404-3

[10] De Rancourt-Raymond, A., Smaili, N. (2023). The unethical use of deepfakes. *Journal of Financial Crime*, *30* (4), 1066-1077. https://doi.org/10.1108/JFC-04-2022-0090

[11] Gregory, S. (2023). Fortify the truth: How to defend human rights in an age of deepfakes and generative AI. *Journal of Human Rights Practice*, *15* (3), 702-714. https://doi.org/10.1093/jhuman/huad035

[12] Catota, F. E., Morgan, M. G., Sicker, D. C. (2019). Cybersecurity education in a developing nation: The Ecuadorian environment. *Journal of Cybersecurity*, *5* (1), 1-19. https://doi.org/10.1093/cybsec/tyz001

[13] Seow, J. W., Lim, M. K., Phan, R. C., Liu, J. K. (2022). A comprehensive overview of Deepfake: Generation, detection, datasets, and opportunities. *Neurocomputing*, *513*, 351-371. https://doi.org/10.1016/j.neucom.2022.09.135

[14] Gambín, Á. F., Yazidi, A., Vasilakos, A., Haugerud, H., Djenouri, Y. (2024). Deepfakes: current and future trends. *Artificial Intelligence Review*, *57* (64), 1-32. https://doi.org/10.1007/s10462-023-10679-x

[15] Heidari, A., Jafari Navimipour, N., Dag, H., Unal, M. (2024). Deepfake detection using deep learning methods: A systematic and comprehensive review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *14* (2), 1-45. https://doi.org/10.1002/widm.1520

[16] Bañuelos Capistrán, J. (2022). Evolución del Deepfake: campos semánticos y géneros discursivos 2017-2021. *Revista ICONO 14. Revista de Comunicación y Tecnologías Emergentes*, *20* (1), 1-21. https://doi.org/10.7195/ri14.v20i1.1773

[17]Amerini, I., Barni, M., Battiato, S., Bestagini, P., Boato, G., Bonaventura, T. S., Bruni, V., Caldelli, R., De Natale, F., De Nicola, R., Guarnera, L., Mandelli, S., Marcialis, G. L., Micheletto, M., Montibeller, A., Orrù, G., Ortis, A., Perazzo, P., Puglisi, G., Salvi, D., Tubaro, S., Tonti, C. M., Villari, M., Vitulano, D. (2024). Deepfake Media Forensics: State of the Art and Challenges Ahead. *arXiv preprint.* https://arxiv.org/pdf/2408.00388

[18]Altuncu, E., Franqueira, V. N. L., Li, S. (2024). Deepfake: definitions, performance metrics and standards, datasets, and a meta-review. *Frontiers in Big Data*, 7, 1-23. https://doi.org/10.3389/fdata.2024.1400024

[19]Domenteanu, A., Tataru, G. C., Craciun, L., Molanescu, A. G., Cotfas, L. A., Delcea, C. (2024). Living in the Age of Deepfakes: A Bibliometric Exploration of Trends, Challenges, and Detection Approaches. *Information*, *15* (9), 1-31. https://doi.org/10.3390/info15090525

[20]Le, B. M., Kim, J., Tariq, S., Moore, K., Abuadbba, A., Woo, S. S. (2024). Sok: Facial deepfake detectors. *arXiv preprint*. https://doi.org/10.48550/arXiv.2401.04364

[21]Romero Moreno, F. (2024). Generative AI and deepfakes: a human rights approach to tackling harmful content. *International Review of Law, Computers & Technology*, *38* (3), 297–326. https://doi.org/10.1080/13600869.2024.2324540

[22]Jacobsen, B. N., Simpsons, J. (2022) The tensions of deepfakes. *Information, communication & society*, *27* (6), 1095-1109. https://doi.org/10.1080/1369118X.2023.2234980

[23]Ramon, M., Vowels, M., Groh, M. (2024). Deepfake Detection in Super-Recognizers and Police Officers, *IEEE Security & Privacy, 22* (3), pp. 68-76. https://doi.org/10.1109/MSEC.2024.3371030

[24]Wang, T., Liao, X., Pui Chow, K., Lin, X., Wang, Y. (2024). Deepfake Detection: A Comprehensive Survey from the Reliability Perspective. *ACM Computing Surveys, 57* (3), 1-35. https://doi.org/10.1145/3699710

[25]Argallero Fernández, P. (2024). *Reconocimiento de DeepFakes*. [Tesis de Licenciatura]. Universidad de Oviedo, España. https://hdl.handle.net/10651/74464

[26] Barrientos-Báez, A., Piñeiro Otero, M. T., Porto Renó, D. (2024). Imágenes falsas, efectos reales. Deepfakes como manifestaciones de la violencia política de género. *Revista Latina De Comunicación Social*, (82), 1–30. https://doi.org/10.4185/rlcs-2024-2278

[27]Alanazi, S., Asif, S. (2024). Exploring deepfake technology: creation, consequences and countermeasures. *Human-Intelligent Systems Integration*, *6*, 49-60. https://doi.org/10.1007/s42454-024-00054-8

[28]Pawelec, M. (2024). Decent deepfakes? (2024). Professional deepfake developers' ethical considerations and their governance potential. *AI and Ethics*, 1-26. https://doi.org/10.1007/s43681-024-00542-2

[29]Hameleers, M. (2024). Cheap Versus Deep Manipulation: The Effects of Cheapfakes Versus Deepfakes in a Political Setting. *International Journal of Public Opinion Research*, *36* (1), 1-9. https://doi.org/10.1093/ijpor/edae004

[30]Agarwal, A., Ratha, N. (2024). *Deepfake Catcher: Can a Simple Fusion be Effective and Outperform Complex DNNs?* IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA. https://doi.org/10.1109/CVPRW63382.2024.00383

[31]Astrid, M., Ghorbel, E., Aouada, D. (2024). Detecting Audio-Visual Deepfakes with Fine-Grained Inconsistencies. *arXiv preprint.* https://doi.org/10.48550/arXiv.2408.06753

[32] Nadimpalli, A. V., Rattani, A. (2024). Social Media Authentication and Combating Deepfakes using Semi-fragile Invisible Image Watermarking. *Digital Threats: Research and Practice, 5* (4), 1-30. https://doi.org/10.1145/3700146

[33] Kingra, S., Aggarwal, N., Kaur, N. (2024). SFormer: An end-to-end spatio-temporal transformer architecture for deepfake detection. *Forensic Science International: Digital Investigation*, *51*. https://doi.org/10.1016/j.fsidi.2024.301817

[34]Samuel-Okon, A. D., Akinola, O. I., Olaniyi, O. O., Olateju, O. O., Ajayi, S. A. (2024). Assessing the Effectiveness of Network Security Tools in Mitigating the Impact of Deepfakes AI on Public Trust in Media. *Archives of Current Research International, 24* (6), 355-375. https://doi.org/10.9734/acri/2024/v24i6794

[35]Zang, Y., Zhang, Y., Heydari, M., Duan, Z. (2024). *SingFake: Singing Voice Deepfake Detection*. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Seoul, Republic of Korea https://doi.org/10.1109/ICASSP48485.2024.10448184

[36] Datta, S. K., Jia, S., Lyu, S. (2024). Exposing Lip-syncing Deepfakes from Mouth Inconsistencies. *arXiv preprint*. https://doi.org/10.48550/arXiv.2401.10113

[37]Sun, P., Qi, H., Li, Y., Lyu, S. (2024). FakeTracer: Catching Face-swap DeepFakes via Implanting Traces in Training. *IEEE Transactions on Emerging Topics in Computing*, 1-12. https://doi.ieeecomputersociety.org/10.1109/TETC.2024.3386960

[38]Nguyen, D., Mejri, N., Singh, I. P., Kuleshova, P., Astrid, M., Kacem, A. (2024). *LAA-Net: Localized Artifact Attention Network for Quality-Agnostic and Generalizable Deepfake Detection*. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA. https://doi.org/10.1109/CVPR52733.2024.01647

[39]Qu, Z., Xi, Z., Lu, W., Luo, X., Wang, Q., Li, B. (2024). DF-RAP: A Robust Adversarial Perturbation for Defending against Deepfakes in Real-world Social Network Scenarios. *IEEE Transactions on Information Forensics and Security, 19*, 3943-3957. https://doi.org/10.1109/tifs.2024.3372803

[40]Pandey, M., Singh, S., Malik, A., Kumar, R. (2024). Detecting low-resolution deepfakes: An exploration of machine learning techniques. *Multimedia Tools and Applications, 83*, 66283–66298. https://doi.org/10.1007/s11042-024-18235-7

[41]Krishnan, K. S., Krishnan, K. S. (2023). MFAAN: *Unveiling Audio Deepfakes with a Multi-Feature Authenticity Network*. 9th International Conference on Signal Processing and Communication (ICSC). Noida, India. https://doi.org/10.1109/icsc60394.2023.10441405

[42]Hou, Y., Guo, Q., Huang, Y., Xie, X., Ma, L., Zhao, J. (2023). *Evading deepfake detectors via adversarial statistical consistency*. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada. https://doi.org/10.1109/CVPR52729.2023.01181

[43]Li, B., Sun, J., Poskitt, C. M., Wang, X. (2023). How generalizable are deepfake detectors? An empirical study. *arXiv preprint*. https://doi.org/10.48550/arXiv.2308.04177

[44]Doss, C., Mondschein, J., Shu, D., Wolfson, T., Kopecky, D., Fitton-Kane, V. A., Bush, L., Tucker, C. (2023). Deepfakes and scientific knowledge dissemination. Scientirfic Reports, *13*, 1-12. https://doi.org/10.1038/s41598-023-39944-3

[45]Bray, S. D., Johnson, S. D., Kleinberg, B. (2023). Testing human ability to detect 'deepfake' images of human faces. *Journal of Cybersecurity, 9* (1), 1-18. https://doi.org/10.1093/cybsec/tyad011

[46] Leone, M. (2022). L'idéologie sémiotique des deepfakes. *Interfaces numériques, 11* (2), 1-16. https://doi.org/10.25965/interfaces-numeriques.4847

[47]Atamna, M. M., Tkachenko, I., Miguet, S. (2022). *Détection de Deepfakes par Réseaux de Convolution: Performances et Limites Actuelles*. XXVIIIème Colloque Francophone de Traitement du Signal et des Images. Nancy, Francia. https://hal.science/hal-03769784/document

[48]Pignier, N. (2022). L'énonciation à l'épreuve de l'«IA». Qu'est-ce-qu'énoncer veut dire? *Interfaces numériques, 11* (2), 1-17. https://doi.org/10.25965/interfaces-numeriques.4897

[49] Nelson, J. (2022). Fake news et deepfakes: Une approche cyberpsychologique. *Interfaces numériques, 11* (2), 1-13. https://doi.org/10.25965/interfaces-numeriques.4830

[50]Caliandro, S. (2022). Fake Art, entre le contrefait et le contrefactuel. *Interfaces numériques, 11* (2). 1-14. https://doi.org/10.25965/interfaces-numeriques.4889

[51]Lloveria, V. (2022). Le deepfake et son métadiscours: l'art de montrer que l'on ment. *Interfaces numériques, 11* (2). 1-22. https://doi.org/10.25965/interfaces-numeriques.4876

[52]Fabre, M. (2022). Urdoxa, le deepfake d'information comme usage tactique du vague La sémiotique au défi de l'évidence. *Interfaces numériques, 11* (2). 1-15. https://dx.doi.org/10.25965/interfaces-numeriques.4863

[53]Dondero, M. G. (2022). Du portrait pictural aux deepfakes: le visage en tant que totalité. *Interfaces numériques, 11* (2), 1-19. https://doi.org/10.25965/interfaces-numeriques.4855

[54]Narayan, K., Agarwal, H., Thakral, K., Mittal, S., Vatsa, M., Singh, R. (2022). *DeePhy: On Deepfake Phylogeny*. IEEE International Joint Conference on Biometrics (IJCB). Abu Dhabi, United Arab Emirates. https://doi.org/10.1109/IJCB54206.2022.10007968

[55]Appel, M., Prietzel, F. (2023). The detection of political deepfakes. *Journal of Computer-Mediated Communication, 27* (4), 1-13. https://doi.org/10.1093/jcmc/zmac008

[56]Helmus, T. C. (2022). *Artificial Intelligence, Deepfakes, and Disinformation*. http://www.jstor.org/stable/resrep42027

[57]Taeb, M., Chi, H. (2022). Comparison of Deepfake Detection Techniques through Deep Learning. *Journal of Cybersecurity and Privacy*, *2* (1), 89-106. https://doi.org/10.3390/jcp2010007

[58]Eberl, A., Kühn, J., Wolbring, T. (2022). Using deepfakes for experiments in the social sciences-A pilot study. *Frontiers in Sociology,7*, 1-11. https://doi.org/10.3389/fsoc.2022.907199

[59]Zhou, Y., Lim, S.-N. (2021). Joint Audio-Visual Deepfake Detection. IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada. https://doi.org/10.1109/ICCV48922.2021.01453

[60]Ramachandran, S., Nadimpalli, A. V., Rattani, A. (2021) *An Experimental Evaluation on Deepfake Detection using Deep Face Recognition*. International Carnahan Conference on Security Technology (ICCST). Hatfield, United Kingdom. https://doi.org/10.1109/ICCST49569.2021.9717407

[61]Zendran, M., Rusiecki, A. (2021). Swapping face images with generative neural networks for deepfake technology–experimental study. *Procedia computer science, 192*, 834-843. https://doi.org/10.1016/j.procs.2021.08.086

[62] Hazeu-Gonzalez, M. (2021). *Sistemas cognitivos artificiales para la detección de deepfakes usando datos audio-visuale*s [Master's Thesis]. Universidad Internacional de la Rioja (UNIR). https://reunir.unir.net/handle/123456789/12080

[63]Grandío Rodríguez, S. (2021). Desarrollo de una aplicación de creación de deepfakes en tiempo real para el uso del agente encubierto informático. *Revista de Investigación CUGC*, (8). https://dialnet.unirioja.es/servlet/articulo?codigo=9427400

[64]Korshunov, P., Marcel, S. (2020). Deepfake detection: humans vs. machines. *arXiv preprint* https://doi.org/10.48550/arXiv.2009.03155

