VISUAL REVIEW | Vol. 17, No. 4, 2025 | ISSN 2695-9631
International Visual Culture Review / Revista Internacional de Cultura Visual
https://doi.org/10.62161/revvisual.v17.5864

DESINFORMACIÓN DISFRAZADA DE HUMOR: impacto de *Deepfakes* en la confianza de Instagram

FERNANDO FUENTE-ALBA CARIOLA[1]
ffuentealba@ucsc.cl

CLAUDIO TORRES ARAVENA[1]
ctorres@ucsc.cl

[1] Universidad Católica de la Santísima Concepción, Chile

PALABRAS CLAVE

Deepfakes
Inteligencia artificial
Confianza
Información
Instagram
Chile

RESUMEN

Esta investigación analiza el impacto de los deepfakes o videos falsos generados mediante inteligencia artificial en la confianza de Instagram, en Concepción, Chile. Además de analizar los deepfakes se aplicó un focus group y un cuestionario a más de 100 personas. Dentro de los resultados destaca que un 85,2 % de los participantes ha visto al menos un video de deepfakes en Instagram. Cuando lo identifican más de la mitad de ellos siente risa cuando los visualiza. Un 57,6 % admite dudar de su veracidad y un 81.4 % indicó que los deepfakes afectan su confianza en Instagram.

Recibido: 11/ 04 / 2025
Aceptado: 05/ 07 / 2025

1. Introducción

El uso de la inteligencia artificial para crear videos falsos o deepfakes plantea un desafío significativo para la confianza e información en redes sociales. La posibilidad de comprender los efectos potencialmente perjudiciales de este tipo de videos en la sociedad resulta primordial, especialmente si se considera que la difusión no controlada de ellos podría erosionar la confianza pública en instituciones, redes sociales y medios tradicionales (García, 2021). La relevancia del problema de investigación sobre el impacto de estos videos falsos en la confianza en Instagram es significativa dentro de la disciplina de comunicación y periodismo. Los deepfakes, definidos como videos o imágenes falsificadas generadas por inteligencia artificial, tienen el potencial de alterar la confianza pública en las redes sociales y socavar la integridad de la información. El desafío a la veracidad en los medios de comunicación y redes sociales se ve intensificado por el aumento de estas realizaciones, que constituyen una novedosa forma de desinformación capaz de comprometer la confianza en los medios digitales. Este estudio, explora cómo las técnicas avanzadas de manipulación pueden socavar la confianza en plataformas visuales como Instagram, un entorno ya susceptible a la desinformación visual, y cómo ello impacta en la percepción de usuarios de la red.

Como indica Molina (2024), la capacidad de la inteligencia artificial para crear videos falsos realistas, plantea desafíos críticos para la integridad de la información y la confianza pública en los medios, especialmente pues la mayoría de las personas no logra identificar la falsedad de la realización, lo que aumenta su vulnerabilidad a la desinformación (Ballesteros-Aguayo y Ruiz del Olmo, 2024). El término deepfake surgió en 2017 en foros de internet, específicamente en Reddit, cuando un usuario con el nombre de deepfakes comenzó a compartir videos falsificados de celebridades utilizando técnicas de IA, combinando la palabra deep learning (aprendizaje profundo) con fake (falso) (Chesney & Citron, 2019). Este tipo de videos comenzó a llamar la atención por la calidad de la manipulación y la facilidad con la que podían ser producidos. Sin embargo, fue en el ámbito político donde los deepfakes se consolidaron como una amenaza significativa.

Un caso destacado fue el video manipulado de Nancy Pelosi, presidenta de la Cámara de Representantes de Estados Unidos en el 2019. En este video, Pelosi parecía estar balbuceando y actuando de manera incoherente durante un discurso, lo que llevó a una rápida difusión en redes sociales, generando dudas sobre su estado de salud mental (Vaccari & Chadwick, 2020). Aunque este video no fue estrictamente un deepfake (fue manipulado mediante técnicas de edición más simples), su viralización mostró el impacto potencial de las falsificaciones en la política.

Desde entonces, varios deepfakes han captado la atención mundial. Uno de los casos más emblemáticos fue en el 2020, cuando un video falso de Barack Obama fue creado por el cineasta Jordan Peele, en colaboración con BuzzFeed, para mostrar cómo esta tecnología podría ser utilizada para difundir desinformación (Ajder et al., 2019). En el video, Obama parecía advertir sobre los peligros de la desinformación y las noticias falsas, pero el mensaje real fue pronunciado por Peele, utilizando la tecnología deepfake para sincronizar los movimientos de labios y expresiones de Obama con las palabras del cineasta.

Por otra parte, Wardle y Derakhshan (2017) abordan el concepto de desinformación como parte del fenómeno más amplio del desorden informativo. Según los autores, este desorden incluye tres tipos de problemas relacionados con la información: la desinformación (información falsa con la intención de engañar), la malinformación (información verdadera pero utilizada para dañar) y la misinformación (información falsa pero sin la intención de engañar). Los deepfakes pertenecen claramente a la categoría de desinformación, ya que su objetivo es crear falsificaciones audiovisuales convincentes que distorsionen la realidad con intenciones maliciosas. Los autores destacan que este tipo de contenido puede erosionar la confianza en los medios tradicionales, especialmente cuando se propaga de manera masiva a través de redes sociales como Instagram o Facebook, donde las barreras de verificación son limitadas.

Por su parte, Lazer et al. (2018) argumentan que las fake news, en las que se incluyen los deepfakes, representan un serio desafío para la democracia y la sociedad digital. Estos autores subrayan que las plataformas de redes sociales facilitan la difusión rápida de contenido manipulado, lo que agrava la situación.

Las fake news y los deepfakes no solo distorsionan la realidad, sino que también juegan un papel crucial en la polarización política, ya que los usuarios tienden a compartir contenidos que confirmen sus creencias previas, sin verificar su veracidad. De hecho, hay autores como Ibrahim (2021) que realizan un análisis profundo sobre los efectos psicológicos de los deepfakes en la percepción pública.
Sosteniendo que los humanos están predispuestos a confiar en las pruebas visuales y auditivas, lo que hace que los deepfakes sean especialmente peligrosos.

Al manipular las imágenes y los videos, se ataca directamente una de las principales formas en que las personas construyen su comprensión del mundo. A diferencia de otros tipos de desinformación, los deepfakes tienen un impacto emocional mucho más fuerte, ya que engañan no solo el intelecto, sino también los sentidos. Esto no solo complica los esfuerzos para verificar la información, sino que también profundiza la desconfianza en los medios y en las fuentes oficiales.

La regulación y gobernanza deben garantizar un uso ético y responsable de la Inteligencia Artificial, IA, en los medios, considerando las implicaciones prácticas de su implementación. La creación de videos con IA representa un avance tecnológico significativo, pero también plantea serias preocupaciones éticas y sociales. Según Hueso (2022), la falta de regulación y supervisión en este ámbito deja espacio para el uso malintencionado de deepfakes con fines de manipulación y desinformación. La creciente sofisticación de estos videos representa una amenaza significativa para la autenticidad y veracidad de la información en las redes sociales. Según Fletcher y Park (2017), «la confianza en canales de información es un componente esencial para una democracia saludable» (p. 3). La falta de marcos regulatorios adecuados para enfrentar la creación y difusión de estas realizaciones pone en riesgo la seguridad informativa y la confianza en los procesos democráticos . Al proporcionar evidencia empírica y análisis detallados, este estudio puede informar las discusiones sobre políticas públicas y ayudar a desarrollar medidas para mitigar los efectos negativos de los videos falsos (Arteaga et al, 2023).

Entender cómo este tipo de realizaciones audiovisuales afectan la confianza de quienes los ven, es fundamental para desarrollar estrategias efectivas que protejan la integridad de la información y la confianza pública. De hecho, autores como Chesney y Citron (2019) sostienen que «los deepfakes pueden ser utilizados para difundir desinformación a gran escala, lo que podría tener efectos devastadores en la opinión pública y la cohesión social» (p. 1754).

En el contexto chileno y también Latinoamericano, donde las redes sociales juegan un papel crucial en la formación de la opinión pública, el aumento de videos falsos podría exacerbar la desconfianza en las instituciones mediáticas y dificultar la toma de decisiones informadas por parte de la ciudadanía. Además, la capacidad de la inteligencia artificial para crear contenido extremadamente realista plantea importantes cuestiones éticas. Como señala Vaccari y Chadwick (2020), «la manipulación de imágenes y videos mediante IA plantea desafíos únicos para la ética de la información, ya que puede ser difícil, incluso para los expertos, distinguir entre lo real y lo falso» (p. 6). Esta dificultad puede ser aún más pronunciada entre la población general, que puede no estar familiarizada con la existencia o las capacidades de los deepfakes, cuestión que queda clara en esta investigación donde más del 80% de las personas que contestaron el cuestionario, no logran identificar de manera certera si es un video falso o no.

De ahí que sea relevante el objetivo general de esta investigación que es analizar el impacto de los deepfakes generados mediante inteligencia artificial en la confianza de los usuarios de la red social Instagram en la ciudad de Concepción, Chile. Objetivo que será abordado a través de objetivos específicos como identificar y analizar los tipos de deepfakes más populares generados mediante inteligencia artificial en las publicaciones de Instagram en Chile durante el 2024; interpretar la confianza y percepción de usuarios de Instagram frente a la exposición de dichos videos falsos y categorizar las expectativas y experiencias de los usuarios de Instagram de Concepción, Chile, frente a los deepfakes visualizados en Instagram. Chile cuenta con un ecosistema mediático diverso que incluye tanto medios tradicionales como digitales, lo que permite el acceso a una amplia variedad de fuentes de datos para el análisis. Las redes sociales y las plataformas de noticias en línea son entornos ricos en datos donde los deepfakes pueden ser identificados y estudiados. Además, la colaboración con instituciones académicas y organizaciones de medios puede facilitar el acceso a datos y archivos necesarios para la investigación (Hu et al., 2020).

La viabilidad de esta investigación está fundamentada en la disponibilidad de recursos tecnológicos y académicos, el acceso a datos, la aplicabilidad de metodologías adecuadas y el potencial apoyo

institucional. Estas condiciones permiten una investigación significativa sobre el impacto de los deepfakes la veracidad y confianza de los usuarios de la red social Instagram a través de un estudio de caso en la ciudad de Concepción, Chile.

Esta investigación aporta al campo de estudios sobre comunicación digital y medios de redes sociales, dado que el consumo de noticias y la interacción con el contenido en Instagram está en crecimiento. Evaluar el impacto de los deepfakes en este contexto ayuda a generar conocimientos sobre cómo las tecnologías emergentes afectan la confianza pública, un tema central en las investigaciones contemporáneas sobre medios de comunicación.

2. Diseño y método

2.1. Objeto formal de estudio

Esta investigación busca analizar el impacto que tienen los deepfakes en la confianza de los usuarios de la red social Instagram, a través de un estudio de caso situado en la ciudad de Concepción Chile. Este enfoque parte de la premisa de que este tipo de videos realizados por inteligencia artificial no solo desafían la confianza del público en la información compartida en plataformas digitales, sino que también plantean dilemas éticos y transforman las prácticas comunicativas. Como señala Zuboff (2019), los avances en las tecnologías de manipulación mediática, como los deepfakes, exigen una reflexión crítica sobre la responsabilidad ética de los comunicadores, el impacto de estas tecnologías en la integridad de la información y el riesgo de desinformación en los medios. La selección de casos permitirá identificar patrones, percepciones y actitudes frente a contenidos manipulados, generando una comprensión de sus efectos en el ecosistema informativo digital. Este enfoque posibilita no solo abordar el fenómeno desde una perspectiva teórica, sino también contribuir con un análisis empírico que examine las dinámicas particulares entre los usuarios de Instagram y su interacción con los deepfakes, permitiendo una reflexión más profunda sobre la formación en ética comunicacional y la competencia crítica frente a los retos tecnológicos actuales.

2.2. Tipo de investigación

Es una investigación de carácter mixta que integra técnicas cualitativas y cuantitativas (Sierra Bravo, 1992). Esta metodología combina la interpretación profunda de los fenómenos sociales y culturales, con la capacidad de medir y analizar tendencias generales mediante datos cuantitativos. Desde la perspectiva cualitativa, el estudio se orienta a explorar cómo los usuarios seleccionados perciben y reaccionan ante el contenido manipulado mediante deepfakes. Este enfoque permite interpretar las dinámicas sociales subyacentes y contextuales de un fenómeno emergente, proporcionando una visión más rica de las experiencias individuales y grupales. Según Sierra Bravo (1992), los métodos cualitativos son particularmente útiles en investigaciones sobre fenómenos complejos como la manipulación digital en redes sociales. En cuanto al enfoque cuantitativo, se utilizan herramientas como cuestionarios cerrados para recoger datos estructurados sobre la frecuencia de exposición a estos videos falsos, niveles de confianza en la información consumida y datos respecto a experiencias de usuarios de Instagram con estas realizaciones de inteligencia artificial. En el ámbito cuantitativo se sumará un análisis de contenido realizado a los deepfakes más visitados y viralizados en Chile durante el 2024, con la finalidad de clasificarlos y categorizarlos. Este componente permite identificar patrones y relaciones entre variables claves que facilitarán el análisis por categoría. El enfoque mixto asegura una aproximación metodológica que permite explorar en profundidad las percepciones subjetivas de los usuarios de Instagram y, al mismo tiempo, analizar sistemáticamente los efectos de los deepfakes en la confianza de los contenidos visuales en la plataforma. En este marco, se empleará un análisis de casos seleccionados que permitió comprender cómo los usuarios interactúan con contenido manipulado y cómo afecta su confianza en la información. Este método asegura un acercamiento al objeto de estudio, manteniendo la flexibilidad necesaria para capturar las particularidades del fenómeno en cuestión.

2.3. Hipótesis

Los deepfakes generados por inteligencia artificial tienen un impacto negativo en la percepción y confianza respecto de la veracidad de la información en la red social Instagram en usuarios de Concepción, Chile. Esta hipótesis sostiene que la exposición a deepfakes afecta la percepción de autenticidad de la información, lo que influye en la confianza en las redes sociales, particularmente en

situaciones donde la precisión de la información es esencial, como en el ámbito político (Martínez, 2023).

De hecho, la difusión de deepfakes podría socavar la confianza en la veracidad de la información, reflejando las preocupaciones sobre el impacto de la desinformación y la manipulación en la percepción de la realidad (Ballesteros-Aguayo y Ruiz del Olmo, 2024).

2.4. Ámbito de estudio

El presente estudio se centra en analizar el impacto de los deepfakes generados mediante inteligencia artificial en la confianza de usuarios de Concepción, Chile, La investigación aborda específicamente cómo los deepfakes afectan la confianza de la información difundida por la red social Instagram, así como las implicaciones éticas y sociales derivadas de su uso. El universo propuesto en esta investigación responde a las características de un estudio de muestreo intencionado, ya que busca comprender un fenómeno emergente en un contexto local específico. Este tipo de estudios puede enmarcarse dentro de una metodología mixta, que integra enfoques cualitativos y cuantitativos para proporcionar una comprensión más amplia y profunda del fenómeno en estudio (Creswell, 2014).

2.5. Variables de análisis

La investigación considera a personas con diferentes niveles de conocimiento sobre los deepfakes y la manipulación digital, así como una variedad de edades, géneros, intereses y niveles de alfabetización digital. Los criterios de selección son: A)Geográficos : El estudio se enfoca en el Gran Concepción; B)Demográficos : Se incluye a usuarios mayores de edad, ya que son individuos con mayor exposición y comprensión potencial de los deepfakes y su impacto en la confianza en los contenidos digitales y C) Temáticos : El universo está delimitado a usuarios activos de Instagram que interactúan regularmente en esta red social, considerando que este es un espacio clave para la proliferación y el consumo de contenido visual manipulado. La investigación se centró en la ciudad de Concepción Chile y los deepfakes seleccionados para su análisis documental y focus group corresponden a los 8 videos realizados por inteligencia artificial con más visitas y «Me gusta» de la red Instagram que hayan sido difundidos en Chile durante el 2024.

2.6. Técnicas de investigación

Para llevar a cabo esta investigación, se empleará una estrategia metodológica que combina técnicas cuantitativas y cualitativas con el propósito de obtener una comprensión profunda del impacto de los deepfakes en la confianza de los usuarios de la red social Instagram. Para ello, se implementarán diversas técnicas de recolección y análisis.

2.6.1. Técnicas de Recolección de información

El primer instrumento es un cuestionario semiestructurado que se administró en formato digital mediante la plataforma Google Forms. Este cuestionario está diseñado para captar tanto datos cuantitativos como cualitativos, a través de una combinación de preguntas cerradas y abiertas. La estructura del instrumento se organiza en secciones temáticas que buscan indagar en el uso cotidiano de Instagram por parte de los participantes, su exposición a contenidos generados mediante deepfakes y el nivel de confianza que estos manifiestan hacia las publicaciones que encuentran en dicha red social. La muestra estuvo compuesta por 107 personas activas en Instagram, 53 mujeres, 50 hombres y 4 que prefirieron no manifestar preferencia de género, seleccionadas mediante un muestreo no probabilístico por conveniencia, asegurando diversidad en cuanto a edad, género y nivel de interacción con la plataforma. Este enfoque facilitó un análisis descriptivo de tendencias y patrones, mientras que las respuestas cualitativas se integraron al análisis temático para una interpretación más profunda de los fenómenos observados.

El segundo instrumento es un análisis documental que se enfocó en ocho videos de deepfakes seleccionados estratégicamente por su relevancia y circulación en Instagram. Los videos incluyen representaciones de personajes tanto chilenos como extranjeros, asegurando una diversidad de estilos y niveles de manipulación. Para esta etapa, se aplicó una matriz de análisis diseñada específicamente para evaluar aspectos como el estilo técnico (resolución, calidad de la manipulación), el tipo de

alteración (visual, auditiva o mixta) y el propósito percibido de cada contenido (humorístico, informativo, desinformativo o político). Este análisis permitió comprender cómo las características intrínsecas de los videos influyen en la percepción de su autenticidad y en la respuesta emocional de los usuarios. El tercer instrumento corresponde a la realización de un focus group compuesto por siete personas residentes en el área metropolitana del Gran Concepción.

La selección de los integrantes se realizó considerando criterios de heterogeneidad en términos de edad, género y experiencia en el uso de redes sociales. Durante la sesión, a los participantes se les mostró la misma serie de videos de deepfakes seleccionados para el análisis documental. Posteriormente, se les administró un cuestionario diseñado para recoger sus percepciones inmediatas y reflexivas sobre los videos, con énfasis en la autenticidad percibida, el impacto emocional experimentado y las implicancias que identifican en términos de confianza hacia los contenidos de Instagram. La dinámica del focus group incluyó una discusión abierta moderada, que fue grabada, transcrita y analizada cualitativamente mediante técnicas de codificación temática.

3. Trabajo de campo y técnicas de análisis

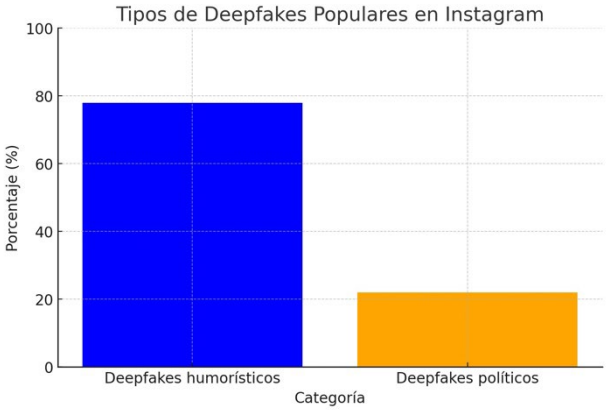
La integración de los datos provenientes de estos tres instrumentos se realizó mediante una triangulación metodológica. Este enfoque permitió contrastar y relacionar los resultados obtenidos a través del cuestionario, el análisis documental y el focus group, maximizando la validez y la fiabilidad de los hallazgos. Los datos cuantitativos obtenidos del cuestionario online fueron analizados mediante técnicas estadísticas descriptivas, permitiendo la identificación de patrones generales en la muestra. Por otro lado, los datos cualitativos, tanto del análisis documental como del focus group, se sometieron a un análisis temático que buscó identificar categorías clave, relaciones entre conceptos y significados emergentes. Esta triangulación de datos permitió no solo validar los hallazgos, sino también ofrecer una interpretación matizada del impacto de los deepfakes en la confianza de los usuarios de Instagram, abordando así la complejidad del fenómeno estudiado desde diferentes perspectivas.

4. Resultados

La triangulación de los datos obtenidos tiene como objetivo integrar los hallazgos provenientes de tres herramientas metodológicas: cuestionario online, focus group y análisis de los ocho videos seleccionados. Esta combinación permite corroborar la hipótesis planteada sobre el impacto negativo de los deepfakes en la percepción de los usuarios del Área Metropolitana del Gran Concepción sobre la veracidad de la información en Instagram. Además, se da respuesta a los objetivos específicos definidos en el estudio a través de diferentes categorías:

Estilo: Aspectos visuales y narrativos de los deepfakes, como calidad técnica y estética. Esta categoría permitió identificar patrones y características comunes en los videos analizados. Respecto a esta categoría , la Figura 1 indica que un 78% de los deepfakes identificados respecto a los patrones de selección, corresponden a estilo humorístico y un 22 % a políticos. Destacando que dentro de los videos humorísticos se centran en figuras conocidas como el actual presidente de los Estados Unidos Donald Trump, el actor Tom Cruise o el futbolista chileno Alexis Sánchez. Mientras que los Deepfakes políticos en figuras como el presidente chileno Gabriel Boric y el presidente argentino Javier Milei.

Figura 1. Tipos de Deepfakes

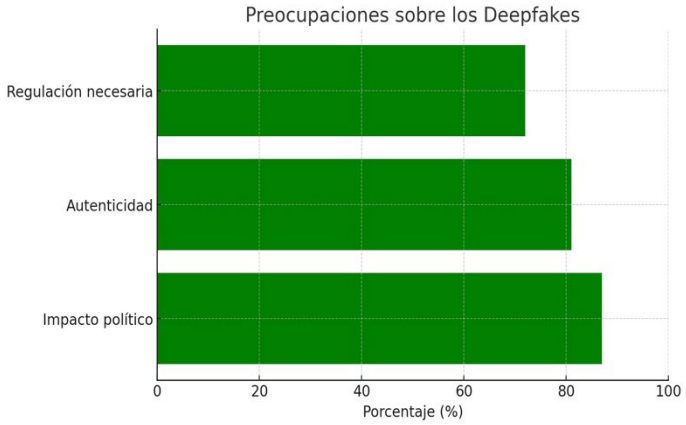


Fuente: elaboración propia, 2024.

Veracidad y Regulación: Analiza las percepciones sobre la autenticidad de los contenidos y las opiniones respecto a la necesidad de normativas en torno a los deepfakes. De hecho, un 100 por ciento de los participantes del Focus Group indica que los videos visionados les hace tomar más conciencia respecto a los riesgos asociados a la veracidad de la información en Instagram. En el mismo instrumento, los videos políticos fueron percibidos como manipuladores por el 92% de los participantes, destacándose un patrón de desconfianza hacia este tipo de contenido, mientras que el otro 8% les genera risa.

Respecto al cuestionario online, Figura 2, el 81% de los encuestados afirma que los deepfakes generan preocupación sobre la autenticidad de la información. Además, un 72% cree que debería haber una regulación estricta sobre su uso.

Figura 2. Alertas sobre Deepfakes



Fuente: elaboración propia, 2024

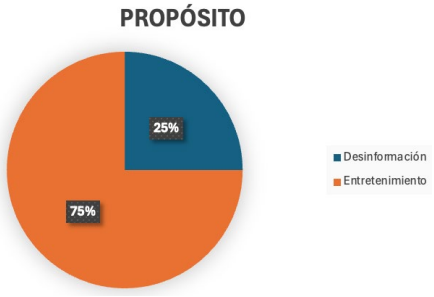
En la misma categoría, pero respecto a si los usuarios logran distinguir un deepfakes de un vídeo real, menos de la mitad, un 45,6%, logra identificarlo siempre, Figura 3.

Figura 3. Identificación de Deepfakes

Fuente: elaboración propia 2024

Manipulación y propósito: Explora las técnicas de alteración digital empleadas en los deepfakes y cómo estas afectan la percepción de los usuarios. Pero también cubre las intenciones que existen detrás de la elaboración de este tipo de videos. De los ocho videos falsos seleccionados con más likes en Instagram Chile un 50% de ellos corresponde a Deepfakes con manipulación de audio y un 50% a videos con manipulación de voz. Mientras que respecto al propósito, Figura 4, un 75 % de los videos tienen un propósito de entretenimiento, mientras que sólo un 25% busca la desinformación.

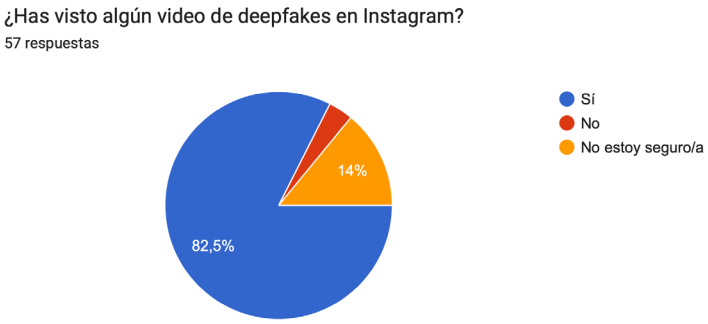
Figura 4. Finalidad de los deepfakes



Fuente: elaboración propia, 2024

Interacción del Usuario: Esta categoría evalúa cómo los usuarios reaccionan e interactúan con los deepfakes en Instagram, considerando métricas como likes, compartidos, y comentarios. Respecto a la interacción un 82,5% ha visto algún video de este tipo en Instagram, Figura 5. Más de la mitad de ellos, un 54,2%, considera que son peligrosos, pero a pesar de ello un 74,6% lo ha compartido en Instagram.

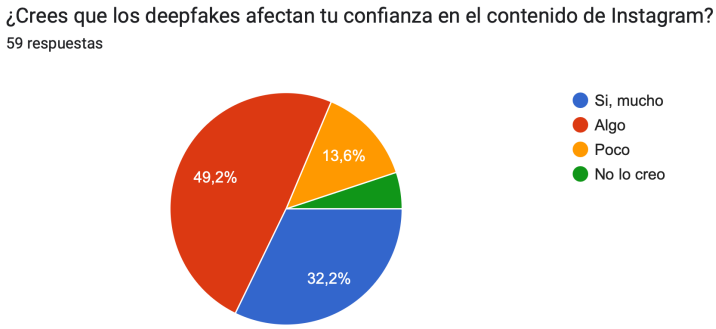
Figura 5 Interacción con deepfakes



Fuente: elaboración propia 2024

Respecto al impacto de la confianza en la red social Instagram, un 81,4% de los usuarios cree que afecta mucho o algo su confianza en la red social, Figura 6. En esta misma categoría y considerando el análisis documental hecho a los Deepfakes, los videos humorísticos generan mayor interacción positiva, likes, pero los videos políticos tuvieron un 89% de comentarios relacionados con la preocupación por la desinformación.

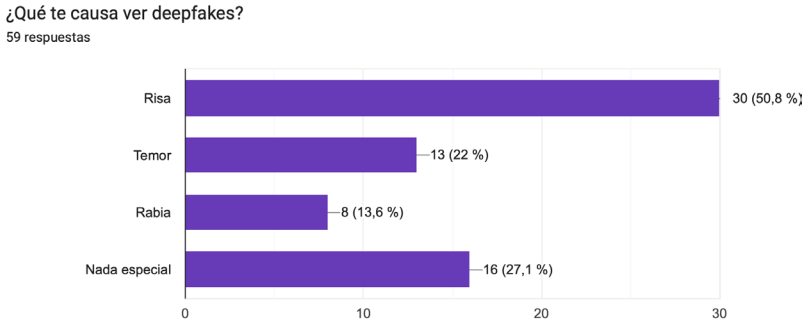
Figura 6. Confianza en Instagram



Fuente: elaboración propia, 2024

En la misma categoría, la risa es el efecto mayoritario de quienes ven este tipo de video en Instagram con un 50,8% de las preferencias, Figura 7. Dichos datos coinciden con otros instrumentos como el focus group, donde los participantes mencionan sentir desconfianza, detallando un temor al uso malintencionado en discursos políticos. El propósito como patrón emergente, identificó que los deepfakes creados con fines humorísticos son aceptados en general, pero hay una línea difusa cuando el contenido puede malinformar o manipular percepciones.

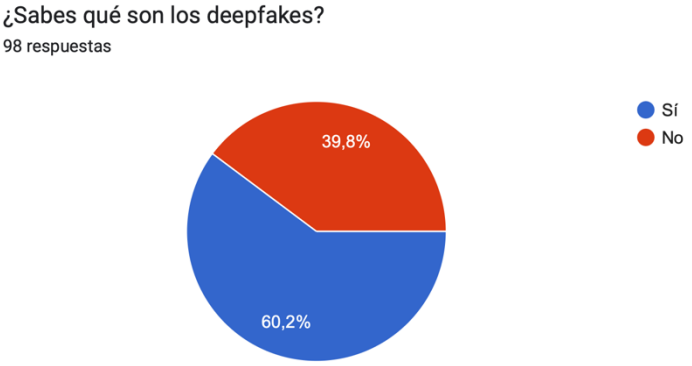
Figura 7. Percepción ante la visualización



Fuente: elaboración propia 2024

Conocimiento de Deepfakes: Examina el nivel de entendimiento que los usuarios tienen sobre esta tecnología y sus implicancias, destacando la relación entre el conocimiento previo y la percepción de confianza. En relación a esta categoría, Figura 8, un 60,2 % de los encuestados saben que son los deepfakes

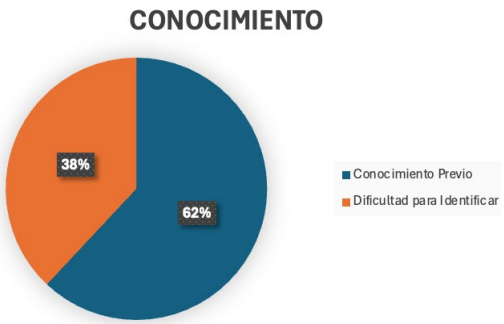
Figura 8. Conceptualización de los Deepfakes



Fuente: elaboración propia 2024

Ya respecto al conocimiento previo de este tipo de videos falsos, Figura 9, un 62% de los encuestados dice tener conocimiento respecto a la presencia y difusión de este tipo de realizaciones hechas por inteligencia artificial, mientras que un 38% de ellos tiene dificultad para identificarlos de contenidos reales. Cifras muy similares a la Figura 8 respecto a si saben qué son los deepfakes.

Figura 9. Conocimiento previo



Fuente: elaboración propia 2024

5. Conclusiones

Los deepfakes o videos falsos representan uno de los mayores desafíos emergentes en el ámbito digital, dada su capacidad para manipular la percepción de la realidad. La creciente sofisticación de estas tecnologías, combinada con su facilidad de creación y dificultad de detección, exige una preparación rigurosa para mitigar su impacto potencial. Resulta preocupante a la luz de los resultados que 8 de cada 10 usuarios de Instagram hayan interactuado con videos Deepfakes y que más de la mitad de ellos, sabiendo que se tratan de videos falsos hechos a través de inteligencia artificial, los hayan compartido sin una verificación real de lo que están distribuyendo con sus redes. Sobre todo, pues este tipo de videos no se autoconstruye, siempre hay un humano tras ellos dirigiendo la herramientas con cierto propósito específico. La mayoría busca entretener y causar una risa en el que los ve, pero en el fondo existe desinformación importante difundida en las redes sociales disfrazada de humor, especialmente en aquellos videos que tienen como protagonistas a líderes políticos. Cuestión que es preocupante cuando se considera que sólo un 26% de los usuarios logra distinguir un deepfake de un video real, entonces la posibilidad de que el video difundido sea «creído» es importante. El análisis de deepfakes en Instagram revela tanto el poder como el peligro relacionado a esta tecnología emergente.

Si bien los deepfakes pueden ser una herramienta creativa poderosa en el campo del entretenimiento y la sátira política, su potencial para difundir desinformación es significativo. En particular, los deepfakes políticos son una amenaza directa para la credibilidad de las figuras públicas y la confianza en las instituciones. La manipulación de la voz y la imagen, puede distorsionar la realidad de una manera tan convincente que incluso los espectadores más críticos podrían ser engañados. A pesar de que una parte significativa de los encuestados es consciente de la existencia de esta tecnología, la mayoría carece de herramientas o conocimientos suficientes para identificar y contrarrestar su impacto. Esto pone en evidencia la necesidad urgente de implementar estrategias educativas, que fomenten tanto el pensamiento crítico como la alfabetización digital. Dichas estrategias deben ser diseñadas para capacitar a los usuarios en la detección de contenidos manipulados y en la evaluación de la autenticidad de las fuentes de información.

Contestando el objetivo general de esta investigación, los resultados indican que los deepfakes generados por inteligencia artificial tienen un impacto negativo en la percepción de los usuarios del de la ciudad de Concepción Chile, en torno a la veracidad de la información presente en Instagram. Este impacto se traduce en una disminución notable de la confianza hacia la plataforma, particularmente en contextos donde la precisión y la autenticidad de la información son críticas.

Los resultados obtenidos a través de las herramientas empleadas en el trabajo de campo destacan el impacto significativo de los deepfakes en la confianza de los usuarios de Instagram. En particular, se identificó que la mayoría de los participantes enfrenta dificultades para detectar contenidos manipulados, lo que los hace vulnerables a la desinformación, especialmente en el ámbito político y en temas relacionados con figuras públicas. Este fenómeno se ve exacerbado por la viralidad inherente a las redes sociales, donde el contenido se comparte rápidamente sin una verificación exhaustiva Dichos usuarios, no sólo confían menos, sino que apelan a regulaciones claras y sanciones estrictas que limiten el uso malintencionado de los deepfakes. En este sentido, las plataformas digitales tienen un rol crucial en la mitigación del impacto negativo de los deepfakes. Asimismo, los datos revelaron que los usuarios con mayor alfabetización digital presentan una mayor capacidad crítica para identificar deepfakes, lo que subraya la importancia de implementar estrategias educativas que promuevan habilidades de análisis y verificación en el entorno digital.

La implementación de sistemas avanzados de detección de contenidos manipulados, combinados con iniciativas de transparencia, que informen a los usuarios sobre la procedencia y autenticidad de los contenidos, puede ser determinante para restaurar la confianza en estas plataformas. Sin embargo, también es crucial la propia conciencia de los usuarios respecto a lo que están difundiendo en redes sociales.

Finalmente, esta investigación subraya la responsabilidad compartida entre usuarios, plataformas tecnológicas y legisladores para enfrentar los desafíos éticos y sociales que plantea esta tecnología emergente. Solo a través de una acción conjunta y coordinada será posible mitigar el impacto negativo de los deepfakes y construir un entorno digital más seguro, confiable y transparente. Además, este

estudio abre la puerta a futuras investigaciones que exploren estrategias específicas para combatir los efectos adversos de los deepfakes en diversos contextos culturales y sociales, consolidando así un enfoque integral frente a este fenómeno.

Referencias

Ajder, H., Patrini, G., Cavalli, F., Cullen, L., & Deeptrace. (2019). The State of Deepfakes: landscape, threats, and impact. In Deeptrace [Report]. https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

Ballesteros, L., y Ruiz del Olmo, F. (2024). Vídeos falsos y desinformación ante la IA: El deepfake como vehículo de la posverdad. Revista De Ciencias De La Comunicación E Información, 29, 1–14. https://doi.org/10.35742/rcci.2024.29.e294

Citron, D. (2018, December 11). Deepfakes and the new disinformation war. Stanford CIS. https://cyberlaw.stanford.edu/publications/deepfakes-and-new-disinformation-war/

Citron, D. K., & Chesney, R. (n.d.). Deepfakes and the new disinformation war. Scholarly Commons at Boston University School of Law. https://scholarship.law.bu.edu/shorter_works/76/

Creswell, J. W. (1995). QUALITATIVE INQUIRY AND RESEARCH DESIGN. In Investigación Cualitativa Y Diseño Investigativo. https://academia.utp.edu.co/seminario-investigacion-II/files/2017/08/INVESTIGACION-CUALITATIVACreswell.pdf

Fletcher, R., & Park, S. (2017). The impact of trust in the news media on online news consumption and participation. Digital Journalism, 5(10), 1281–1299. https://doi.org/10.1080/21670811.2017.1279979

García-Ull, F. J. (2021). Deepfakes: el próximo reto en la detección de notícias falsas. Análisis, 64, 103. https://doi.org/10.5565/rev/analisi.3378

Hu K, Wang S, Lee D, Liu H (2020c) Mining disinformation and fake news: concepts, methods, and recent advancements. In: Disinformation, misinformation, and fake news in social media. Springer, Berlin, pp 1–19 https://doi.org/10.1007/978-3-030-42699-6_1

Hueso, L. (2022). Quién, cómo y qué regular (o no regular) frente a la desinformación. Teoría y realidad constitucional, (49), 199-238. https://dialnet.unirioja.es/servlet/articulo?codigo=8450040

Ibrahim, S. (2024, February 10). Cómo los 'deepfakes' están cambiando nuestra visión de la realidad. https://www.swissinfo.ch/spa/ciencia/cómo-los-ultrafalsos-cambian-nuestra-visión-de-la-realidad/46866008

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. Science, 359(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Martínez, E. (2023). Desinformación y fake news en TikTok: técnicas para su detección y prevención. https://oa.upm.es/76481/

Molina, O. (2024). Dissonance between technological transformation and journalistic ethics: A critical discussion of the impact of artificial intelligence on the media. adComunica Revista Científica De Estrategias Tendencias E Innovación En Comunicación, 91–114. 4 https://doi.org/10.6035/adcomunica.8002

Sierra Bravo, R. (1992). Técnicas de investigación social: teoría y ejercicios (p. 33)

Vaccari, C., & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. Social Media + Society, 6(1). https://doi.org/10.1177/2056305120903408

Wardle, C., y Derakhshan, H. (2017). Desorden de la información: Hacia un marco interdisciplinario para la investigación y la formulación de políticas. Consejo de Europa. https://rm.coe.int/information-disorder-toward-an-interdisciplinary- framework-for-researc/168076277c

Wagner, A. [. (2024, September 13). Desinformación en la era digital. DIGITAL.CSIC. http://hdl.handle.net/10261/367658

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier ofPower (PublicAffairs,Ed.).ilustrada https://www.hbs.edu/faculty/Pages/item.aspx?num=56791