

Pornografía *deepfake* en la era de la IA: nuevos desafíos para la educación de género, humanística y tecnológica

Deepfake pornography in the AI era: New challenges for gender, humanistic and technological education

MARÍA LUISA PEREIRA HERNÁNDEZ • VIRGINIA MIRELLA ZATARAIN AVENDAÑO

María Luisa Pereira Hernández.

Universidad Pedagógica del Estado de Sinaloa, México. Trabaja en el área de conocimiento de ciencias sociales, en el campo de formación docente, y su línea de investigación se centra en tecnologías de la información, investigación e innovación educativa. Entre sus más recientes trabajos se contemplan “Percepción y uso de las TICs en docentes de educación secundaria” y “Ciberataques del estudiantado al docente: la nueva realidad”. Es coordinadora editorial de la revista *IPSUMTEC*, integrante del Sistema Sinaloense de Investigadores y Tecnólogos, y miembro de Red Temática Mexicana para el Desarrollo e Incorporación de Tecnología Educativa. Correo electrónico: marialuisa.pereira@upes.edu.mx. ORCID: <https://orcid.org/0000-0002-4748-5397>.

Virginia Mirella Zatarain Avendaño.

Universidad Pedagógica del Estado de Sinaloa, México. Es Licenciada en Psicología por la Universidad Autónoma de Sinaloa, Maestra en Educación campo Intervención Pedagógica y Aprendizaje escolar por la Universidad Pedagógica

Resumen

La proliferación de *deepfakes*, particularmente en forma de pornografía, emerge como una amenaza creciente en la sociedad digital contemporánea, exacerbando la violencia de género y planteando desafíos significativos para la educación y la seguridad en línea. Este estudio cualitativo se centró en investigar el estado actual de la investigación sobre pornografía *deepfake* y su impacto educativo, utilizando el método de análisis de contenido con la técnica de análisis temático. Los resultados revelaron un preocupante aumento en la producción de este tipo de contenido manipulado, con un impacto desproporcionado en las mujeres. Se destacó la necesidad urgente de fortalecer la legislación, promover la alfabetización digital y fomentar el pensamiento crítico en el ámbito educativo para mitigar los efectos negativos. Se concluyó que abordar la ciberviolencia contra las mujeres requiere respuestas integrales y colaborativas, incluyendo el desarrollo de estrategias efectivas y la identificación de áreas de investigación futuras para proteger los derechos humanos en el entorno digital. Este análisis subraya la importancia de avanzar hacia una sociedad digital más segura y ética, en la cual la educación juega un papel crucial en la preparación de individuos para interactuar de manera responsable y crítica con la tecnología emergente.

Palabras clave: Educación de género, educación humanista, educación tecnológica, inteligencia artificial, violencia de género.

Abstract

The proliferation of deepfakes, particularly in the form of pornography, emerges as a growing threat in contemporary digital society, exacerbating gender-based violence and posing significant challenges for online education and safety. This qualitative study focused on investigating the current state of research on deepfake pornography and its educational impact, using the content analysis method with the thematic analysis technique. The results revealed a worrying increase in the production of this type of manipulated content, with a disproportionate impact on women. The urgent need to strengthen legislation, promote digital literacy and encourage critical thinking in the educational field to mitigate the negative effects of deepfakes was highlighted. It was concluded that addressing cyberviolence

del Estado de Sinaloa. Maestra de tiempo completo en la Universidad Pedagógica del Estado de Sinaloa y Miembro del Sistema Sinaloense de Investigadores y Tecnólogos. Correo electrónico: virginia.zatarain@upes.edu.mx. ORCID: <https://orcid.org/0009-0008-8324-2916>.

against women requires comprehensive and collaborative responses, including the development of effective strategies and the identification of future research areas to protect human rights in the digital environment. This analysis highlights the importance of moving towards a safer and more ethical digital society, in which education plays a crucial role in preparing individuals to interact responsibly and critically with emerging technology.

Keywords: Gender education, humanistic education, technological education, artificial intelligence, gender-based violence.

PROBLEMÁTICA

La proliferación creciente de tecnologías de manipulación digital, como los *deepfakes*, plantea desafíos cada vez más complejos en la sociedad. En una comunidad donde la información y los medios digitales son omnipresentes, los *deepfakes* representan una amenaza particularmente preocupante para la educación, debido a sus efectos, como alterar la confianza en la veracidad de los contenidos creados en línea y distorsionar la percepción de la realidad. Este fenómeno no solo afecta a la credibilidad de las fuentes informativas sino que también erosiona la capacidad crítica de los usuarios para discernir entre lo auténtico y lo fabricado. A medida que las herramientas para crear *deepfakes* se vuelven más accesibles, la posibilidad de manipular imágenes y videos para difundir desinformación se incrementa, lo que requiere un enfoque renovado en la educación mediática.

Entre junio y agosto del 2022 el Módulo sobre Ciberacoso –MOCIBA– reveló que el 79.5% de la población de 12 años y más en México usó Internet, y que un 20.8% de esta población experimenta ciberacoso, siendo el 29.3% las mujeres de 20 a 29 años quienes, en el último año, representaron las principales víctimas (INEGI, 2023). Estos datos reflejan que la violencia contra las mujeres es una manifestación arraigada de desigualdad de género y discriminación que persiste y se adapta a las tecnologías digitales. La intersección entre la violencia digital y el ciberacoso se convierte en un espacio crítico que demanda atención, ya que los entornos virtuales pueden amplificar comportamientos abusivos y perpetuar estigmas.

Una de las formas más actuales de violencia de género es la llamada pornografía *deepfake*, que mostró un crecimiento significativo del 464% en su producción entre los años 2022 y 2023 (Bigas, 2023). Este tipo de contenido no solo se reproduce sin el consentimiento de las personas afectadas, sino que también contribuye a la objetivación y deshumanización de las mujeres en el espacio digital. En el 2023, el 98% de los videos de falsa creación eran pornográficos, afectando mayoritariamente al género femenino (Fitzgerald, 2024). Según datos de PornHub, México es el quinto país en consumo de pornografía, y los *deepfakes* no consensuales afectan en un 90% a mujeres (Yessica con “Y” y Desde Mujeres, 2023).

Los datos muestran la relevancia del problema que representa la pornografía en línea y los *deepfakes*, especialmente en lo que respecta a la violencia digital contra las mujeres. Además plantea preocupaciones adicionales sobre la privacidad y la seguridad en línea (Yessica con “Y” y Desde Mujeres, 2023). La gran cantidad de videos falsos de contenido pornográfico que afectan al género femenino subraya la vulnerabilidad específica de este grupo demográfico en el espacio digital. La magnitud de esta problemática pone de relieve la necesidad urgente de abordar la educación sobre la sexualidad y el consentimiento en el entorno digital, así como de fomentar la alfabetización mediática entre los jóvenes.

Surge, entonces, la necesidad de dar respuesta a la interrogante efectuada: ¿Cuál es el estado actual que guardan las investigaciones existentes de pornografía *deepfake* y su impacto en la educación? Esto implica la necesidad de investigar el estado actual del tema para analizar y valorar las tendencias y modelos emergentes en la literatura académica correspondiente, abordar los puntos de convergencia e identificar áreas de oportunidad en el ámbito educativo y posibles direcciones para futuras investigaciones. A medida que la tecnología avanza, también lo deben hacer nuestras estrategias educativas y de sensibilización, asegurando que las futuras generaciones estén equipadas para navegar en un entorno digital cada vez más complejo y desafiante.

DESARROLLO

Los *deepfakes* son videos manipulados digitalmente con un nivel de realismo alto, que muestran a personas realizando acciones que nunca ocurrieron. Utilizan redes neuronales y algoritmos de aprendizaje profundo para intercambiar las caras de personas de manera conveniente (Rössler et al., 2018) y están diseñados para ser compartidos de manera rápida y sencilla en plataformas de redes sociales, propagando teorías conspirativas, rumores y desinformación, intensificando el contexto de posverdad (Jang y Kim, 2018). En este sentido, la población confía, comparte y juzga el contenido, incluso si hay indicios de que podría ser falso, contribuyendo así a la propagación de desinformación y teorías conspirativas en las redes sociales. Esta dinámica se ve potenciada por el diseño de las plataformas, que a menudo priorizan el contenido que genera interacción, sin una revisión crítica de su veracidad.

Actualmente hay aplicaciones disponibles en línea que facilitan la creación de *deepfakes* de manera sencilla y accesible, incluso para usuarios con poca experiencia técnica (García-Ull, 2021). Según el informe *State of deepfakes—2023—* de Home Security Heroes, para producir un video porno manipulado de 60 segundos se necesitan menos de 25 minutos y no tiene costo alguno (aunque también existen opciones de pago) (Bigas, 2023). Esta accesibilidad democratiza la creación de contenido manipulador, aumentando el riesgo de que se utilice con fines maliciosos. Y aunque la creación de imágenes y videos sexuales realizados con inteligencia artificial –IA– aún no está tipificada como delito en el código penal, se argumenta que esta práctica constituye

una forma actual de violencia de género (Bigas, 2023), pues estas manipulaciones tienen repercusiones graves en la vida personal y profesional de las mujeres afectadas, dañando su reputación, seguridad y bienestar emocional. Las víctimas pueden experimentar desde acoso en línea hasta daños en sus relaciones personales y profesionales, lo que subraya la urgencia de abordar este fenómeno.

La proliferación de este tipo de contenido falso o manipulado plantea desafíos significativos para la detección y mitigación de la ciberviolencia contra las mujeres. Dada la naturaleza altamente realista de las imágenes y videos generados por IA, puede resultar difícil para los usuarios discernir entre contenido auténtico y manipulado, lo que aumenta el riesgo de que la desinformación y la difamación se propaguen sin control. Esto también puede llevar a una erosión de la confianza en los medios de comunicación y en las plataformas digitales, donde la veracidad de la información es cada vez más cuestionada.

Ante un fenómeno en constante evolución, es fundamental adoptar enfoques innovadores y adaptativos que aborden tanto las manifestaciones visibles como las raíces profundas de la violencia de género en el ciberespacio, lo cual representa un desafío multifacético que requiere un enfoque integral y colaborativo por parte de actores gubernamentales, empresariales, de organizaciones de la sociedad civil y de la comunidad en general. En este sentido, los estudios examinados proponen cuatro estrategias para enfrentar los *deepfakes*: la implementación de leyes y regulaciones, la promoción de políticas corporativas conscientes, la educación y formación, y el desarrollo de tecnologías anti-*deepfake* como sistemas de detección, así como también autenticación de contenido y medidas de prevención (García-Ull, 2021) y la creación de tecnología que opere bajo elementos jurídicos desde su composición (Niebla, 2019). La regulación es crucial para abordar el contenido ya existente y para prevenir la creación de nuevos *deepfakes*.

Además es fundamental promover la educación digital y la alfabetización mediática entre las mujeres y las niñas, capacitándolas para reconocer, resistir y denunciar la ciberviolencia en todas sus formas. Esto incluye brindar información sobre los riesgos y consecuencias, así como ofrecer herramientas y recursos prácticos para protegerse en línea y buscar ayuda en caso de ser víctimas de violencia digital.

Las investigaciones existentes abordan el fenómeno de los *deepfakes* desde diversas perspectivas geográficas y disciplinarias, destacando su impacto en áreas como la pornografía, la desinformación política, el derecho, la tecnología y la conciencia pública. Se pueden identificar diferentes puntos de convergencia entre las investigaciones revisadas:

1. Amenazas a la privacidad y la integridad personal: varios estudios, incluidos los de Maras y Alexandrou (2018), Maddocks (2020) y Nugraha et al. (2023), destacan cómo los *deepfakes* pueden comprometer la privacidad y la dignidad de las personas al permitir la creación y difusión de contenido explícito sin consentimiento. Los *deepfakes* representan una violación directa a la privacidad

individual, ya que permiten la creación de contenido explícito que puede ser distribuido sin el consentimiento de la persona afectada, dañando la reputación y la imagen de la víctima y repercutiendo de manera emocional y psicológica. La facilidad con que se pueden generar y difundir estos videos crea un entorno donde las personas, especialmente las mujeres, son vulnerables a ataques que pueden arruinar sus vidas personales y profesionales. Este aspecto resalta la necesidad de regulaciones legales y tecnológicas para proteger a las víctimas y prevenir la proliferación de videos pornográficos de contenido falso y dañino.

2. Desafíos para la autenticidad y la evidencia: autores como Mania (2024) y Timmerman et al. (2023) discuten cómo los *deepfakes* plantean desafíos significativos para la autenticidad y la credibilidad de la evidencia visual en contextos legales y forenses. Los *deepfakes* complican la autenticidad de la evidencia visual, un aspecto crítico en contextos legales. La posibilidad de que un video sea manipulado para mostrar eventos o declaraciones que nunca ocurrieron plantea dudas sobre la validez de la evidencia en los tribunales. Esto puede afectar decisiones judiciales así como erosionar la confianza pública en el sistema legal. La creación de tecnologías de autenticación que puedan verificar la originalidad del contenido visual se convierte en una necesidad para mantener la integridad de los procesos legales.
3. Impacto en la política y la desinformación: estudios como el de Maddocks (2020) resaltan el impacto de los *deepfakes* en la esfera política y la desinformación. Estos tienen el potencial de influir negativamente en el ámbito político al crear contenido que puede distorsionar la realidad y manipular la percepción pública. La producción de videos falsos de políticos puede desestabilizar gobiernos y socavar la confianza en las instituciones. Además estos ataques son más frecuentes hacia figuras femeninas y grupos marginados, lo que exacerba las desigualdades existentes. Esto destaca la importancia de abordar las desigualdades subyacentes que hacen que esta tecnología tenga como objetivo de manera desproporcionada a ciertos grupos, como las mujeres.
4. Educación y conciencia pública: autores como Maras y Alexandrou (2018) y Nugraha et al. (2023) enfatizan la importancia de la educación y la conciencia pública para combatir los *deepfakes*. La falta de conciencia sobre el potencial de manipulación de la IA puede hacer que las personas subestimen la amenaza que representa la creación de videos pornográficos de contenido falso y sean más susceptibles a la desinformación. En este estudio se enfatiza en la promoción del pensamiento crítico y la alfabetización digital para ayudar a mitigar tales riesgos.
5. Desafíos legales y regulatorios: investigaciones como las de Karasavva y Noorbhai (2021) y Mania (2024) examinan los desafíos legales y regulatorios

asociados con los *deepfakes*, especialmente en áreas como la pornografía vengativa y la violación de la privacidad. Estos estudios destacan la necesidad de reformas legales específicas y una mayor responsabilidad por parte de las plataformas en línea para abordar el problema. Los marcos legales existentes a menudo son insuficientes para abordar las complejidades que presentan los *deepfakes*, especialmente en contextos de pornografía no consensuada y violaciones de la privacidad.

6. Enfoque en el impacto negativo: todas las investigaciones reconocen el impacto negativo de los *deepfakes*, especialmente en áreas como la privacidad, la política, la desinformación y la victimización de personas (Bañuelos, 2022; De Angeli et al., 2021; Karasavva y Noorbhai, 2021; Kerner y Risse, 2021; Maddocks, 2020; Mania, 2024; Maras y Alexandrou, 2018; Nugraha et al., 2023; Timmerman et al., 2023). Esto resalta la necesidad de un enfoque proactivo y preventivo. El impacto negativo de los *deepfakes* se manifiesta en múltiples dimensiones, desde la violación de la privacidad hasta la erosión de la confianza en la información. Este fenómeno no solo afecta a las víctimas individuales sino que también tiene implicaciones más amplias para la sociedad, creando un clima de desconfianza y temor. Por ello es fundamental adoptar un enfoque proactivo que contemple la educación, la regulación y la concienciación pública.
7. Énfasis en la regulación y la protección legal: varios estudios destacan la necesidad de regulaciones y protecciones legales para abordar los problemas asociados con los *deepfakes*, como la pornografía no consensuada y la difamación (Bañuelos, 2022; Gosse y Burkell, 2020; Karasavva y Noorbhai, 2021; Mania, 2024; Maras, 2018; Nugraha et al., 2023). La regulación de los *deepfakes* es esencial para proteger a las víctimas y prevenir su uso malintencionado. Esto implica el desarrollo de leyes que se ajusten a la naturaleza cambiante de la tecnología y que ofrezcan soluciones efectivas para el manejo de contenido engañoso. La colaboración entre diversos actores, incluyendo legisladores, tecnólogos y defensores de derechos humanos, es vital para establecer un marco legal robusto y efectivo.
8. Objetos de estudio: mientras que algunos estudios se centran en el impacto de los *deepfakes* en la política y la desinformación (Maddocks, 2020), otros se centran en aspectos legales y regulatorios (Gosse y Burkell, 2020), o en la evolución del discurso y los usos beneficiosos o delictivos (Bañuelos, 2022). Los *deepfakes* son objeto de investigación desde diversas perspectivas, lo que permite un análisis multidisciplinario del fenómeno. Algunas investigaciones se centran en su impacto en la política, mientras que otras abordan las implicaciones legales y éticas. Esta diversidad en los enfoques resalta la complejidad del fenómeno y la necesidad de considerar múltiples factores en su análisis.

9. Perspectivas geográficas y culturales: las investigaciones provienen de diferentes países –EE. UU., Canadá, Italia, Indonesia, México–, lo que puede influir en las perspectivas y enfoques adoptados en relación con los *deepfakes* y sus implicaciones. Estas diferencias geográficas pueden resultar en diversas normativas y respuestas sociales que impacten en la eficacia de las medidas adoptadas.
10. Métodos de investigación: algunos estudios utilizan análisis de contenido de medios (Gosse y Burkell, 2020), mientras que otros emplean métodos longitudinales y análisis bibliográficos (Bañuelos, 2022) o enfoques jurídico-normativos (Nugraha et al., 2023). La variedad en los métodos de investigación utilizados en estudios sobre *deepfakes*, como análisis de contenido y enfoques jurídico-normativos, permite una comprensión más rica y completa del fenómeno. Esta diversidad metodológica es crucial para desarrollar estrategias más efectivas y fundamentadas que aborden los retos que presentan los *deepfakes* en diferentes ámbitos.

Los hallazgos revelan aspectos que subrayan la complejidad e impacto del fenómeno *deepfake* en la educación. Primeramente, las amenazas que este genera a la privacidad y la integridad personal provocan un llamado a la atención e implementación de regulaciones legales que protejan a las personas afectadas. La capacidad de crear contenidos falsos que pueden comprometer la reputación y el bienestar emocional de las víctimas es alarmante, especialmente en un contexto educativo donde la vulnerabilidad de los estudiantes puede ser mayor. Esto refiere la necesidad de desarrollar marcos legales que aborden directamente la creación y distribución de *deepfakes*, y de una formación específica que sensibilice a los educadores y estudiantes sobre estos riesgos.

Además, la exposición a *deepfakes* puede tener un impacto negativo en la confianza y la dinámica de las relaciones interpersonales dentro del entorno escolar. Cuando los estudiantes son víctimas de manipulaciones digitales, su autoestima y bienestar psicológico pueden verse gravemente afectados, creando un clima de desconfianza que puede extenderse a la comunidad educativa en general. Por lo tanto, es crucial que las instituciones educativas no solo implementen políticas de protección sino que también integren en su currículo la alfabetización digital, el pensamiento crítico y la educación sobre la privacidad en línea.

En este sentido, la educación debe desempeñar un papel fundamental en equipar a los jóvenes con las herramientas necesarias para identificar y resistir la desinformación. Al fortalecer la comprensión de las implicaciones de los *deepfakes* se fomenta un ambiente donde los estudiantes puedan sentirse seguros y empoderados para actuar ante situaciones de ciberviolencia, convirtiéndose en defensores de su propia integridad y la de sus compañeros. Esto contribuye a crear un entorno escolar más seguro, y a preparar a los jóvenes para enfrentar los desafíos del mundo digital en su vida cotidiana.

METODOLOGÍA

El estudio se basa en un enfoque cualitativo, y el método de análisis de contenido es una técnica de investigación versátil que permite desentrañar significados y patrones en datos textuales y visuales. Este enfoque es especialmente útil en campos en que la comprensión de la experiencia humana y la interpretación de significados son fundamentales. Se centra en identificar y describir patrones, temas y categorías para comprender los significados subyacentes y contextuales, así como explorar cómo se comunican ciertas ideas, valores o sentimientos; la recolección de datos, familiarización con el material, codificación de segmentos de texto, agrupación de códigos en categorías más amplias y la interpretación de estos para extraer conclusiones significativas (Atlas.ti, s.f.; Corral, 2020). Cada una de estas etapas es crucial para asegurar que los datos se analicen de manera sistemática y que las conclusiones sean válidas y relevantes para el contexto del estudio.

Para llevar a cabo el método de análisis de contenido se utilizó la técnica de análisis temático para identificar y analizar patrones o temas dentro de los datos. Este enfoque permite al investigador captar la esencia de las narrativas y las experiencias de los participantes, facilitando una comprensión más profunda de cómo se perciben y se comunican las realidades. El propósito principal fue desentrañar los significados profundos y recurrentes presentes en los textos, permitiendo comprender las percepciones, experiencias y significados compartidos por los participantes. Este proceso es especialmente importante en investigaciones que abordan temas complejos y contemporáneos, como el *deepfake* pornográfico, cuyas implicaciones éticas y sociales son significativas.

El proceso comenzó con la codificación de los datos, en la cual se designaron etiquetas a segmentos específicos del texto que representan ideas o conceptos relevantes. Estos códigos iniciales fueron descriptivos y se aplicaron a frases u oraciones (Vega, 2023). Esta etapa es crítica porque los códigos sirven como puntos de referencia que permiten al investigador organizar los datos de manera efectiva y facilitar el análisis posterior. La codificación inicial también ayuda a captar las diversas dimensiones de los datos y a reconocer las conexiones entre diferentes partes del material analizado.

Una vez codificados los datos, el siguiente paso fue buscar temas recurrentes relacionados al *deepfake* pornográfico y revisar los códigos, agrupándolos en categorías más amplias que reflejaran patrones significativos en los datos. Este proceso permite construir una comprensión más estructurada y coherente del fenómeno estudiado (Acevedo, 2011). Al identificar temas comunes, el investigador puede captar las tendencias y las preocupaciones compartidas entre los participantes, proporcionando así un marco que enriquece el análisis y la interpretación.

Finalmente, el proceso incluyó una etapa de revisión y refinamiento del tema, implicando la subdivisión de temas amplios en subtemas más específicos, la eliminación de temas redundantes o la combinación de temas relacionados para crear categorías

más coherentes. La revisión continua y el refinamiento de los temas aseguran un análisis riguroso. Este enfoque flexible y dinámico permite al investigador adaptarse a nuevos hallazgos y asegurarse de que las interpretaciones sean lo más precisas y representativas posible de los datos recolectados. Al final, este proceso de revisión es esencial para garantizar que el análisis no solo sea exhaustivo, sino también pertinente y significativo en el contexto del estudio.

DISCUSIÓN Y CONCLUSIONES

La ciberviolencia contra las mujeres representa un desafío urgente en el panorama digital contemporáneo, que requiere una respuesta colectiva y coordinada por parte de actores, desde los gobiernos y las empresas de tecnología hasta las organizaciones de la sociedad civil y la comunidad en general. Este fenómeno, que se manifiesta a través de una amplia gama de comportamientos hostiles y abusivos en línea, tiene profundas implicaciones para la seguridad, la dignidad y el bienestar de las mujeres y niñas en todo el mundo.

En este contexto, es fundamental reconocer que la ciberviolencia contra las mujeres no es un problema aislado, sino que está intrínsecamente ligado a estructuras de poder desiguales y normas sociales patriarcales que perpetúan la desigualdad de género y la discriminación sistemática. La normalización y perpetuación de la violencia de género en el ciberespacio reflejan y refuerzan las desigualdades y estereotipos de género arraigados en la sociedad, lo que perpetúa la subordinación y el control de las mujeres en múltiples esferas de la vida. Esto implica que la violencia en línea no solo afecta a las víctimas de manera individual, sino que también alimenta un clima social que desvaloriza y silencia a las mujeres, limitando su participación plena en la sociedad. Este entorno hostil puede llevar a una disminución de la autoestima y la confianza de las mujeres, afectando su capacidad para desenvolverse en espacios profesionales, académicos y sociales.

En este sentido, abordar la ciberviolencia contra las mujeres requiere un enfoque integral que combine medidas legales, políticas, educativas y tecnológicas para prevenir y mitigar sus impactos negativos. Esto incluye fortalecer el marco legal y normativo para proteger los derechos de las mujeres en línea, promover la educación digital y la alfabetización mediática, y fomentar la colaboración entre diferentes sectores y actores para desarrollar respuestas y estrategias de prevención efectivas y sostenibles. La colaboración es esencial para crear un entorno en línea más seguro, donde las mujeres puedan expresarse sin miedo a represalias o violencia. Este enfoque debe incluir la participación activa de las propias mujeres en la formulación de políticas, asegurando que sus voces sean escuchadas y que sus necesidades específicas sean atendidas.

En el caso de la educación, la formación sólida juega un papel crucial en la lucha contra los *deepfakes*. A pesar de la considerable atención mediática y las preocupaciones expresadas por las autoridades, el público aún no ha internalizado completamente la

amenaza que representan las falsificaciones profundas. Recordemos que “calidad no garantiza autenticidad”, y a medida que la tecnología y la desinformación avanzan, será más difícil discernir entre contenidos reales o falsos. Por lo tanto, es fundamental educar sobre las potenciales desventajas de la IA. Para ello, muchos expertos sugieren la inclusión del pensamiento crítico y la alfabetización digital en el currículo escolar, ya que estas habilidades ayudan a los niños, jóvenes y adultos a detectar noticias falsas y a interactuar de manera más segura en línea. La educación no solo debe enfocarse en la detección de contenido engañoso sino también en la promoción de un uso ético de la tecnología, fomentando valores de respeto y responsabilidad hacia los demás en el entorno digital. Esto incluye la comprensión de las consecuencias legales y sociales de difundir información falsa o dañina, así como el impacto emocional que puede tener en las víctimas de ciberviolencia.

En conclusión, este problema representa desafíos importantes para la educación humanista, tecnológica y de género. Desde la educación humanista, se debe enfatizar la importancia de la ética y la moralidad en el uso de la tecnología, fomentar la empatía y la comprensión para prevenir la creación y distribución de pornografía *deepfake*, así como promover la defensa de los derechos humanos, incluyendo el derecho a la privacidad, la dignidad y las responsabilidades personales en el mundo digital. Esta educación es esencial para cultivar una cultura de respeto y responsabilidad en el uso de la tecnología.

Desde la educación tecnológica, es crucial que los estudiantes desarrollen conocimiento y habilidades técnicas en el uso y funcionamiento de las tecnologías de IA y los *deepfakes* para su detección y verificación, así como crear algoritmos avanzados de IA y programas de *software* que identifiquen y eliminen contenidos manipulados. Además la educación tecnológica debe enfatizar la importancia de la ciberseguridad y enseñar a los estudiantes a proteger sus datos personales y su identidad en línea, incluyendo prácticas seguras de navegación, uso de contraseñas seguras y conocimiento sobre la privacidad digital. Esto ayudará a crear un entorno en el que los estudiantes se sientan empoderados para protegerse y proteger a otros en el ciberespacio.

Y desde la educación de género, es importante alzar la lucha contra el acoso y la violencia contra las mujeres, abordando las formas de violencia digital, y equipar a las personas con las habilidades necesarias para identificar, denunciar y combatir tales abusos. La educación en el respeto y la igualdad de género es fundamental para prevenir el uso de *deepfakes* con fines malintencionados, así como proporcionar recursos y apoyo a las víctimas, incluyendo la sensibilización sobre el impacto psicológico y social de dichos abusos, así como ofrecer estrategias para la recuperación y empoderamiento de las mujeres que atraviesan tal problemática. La formación en igualdad de género no solo beneficia a las mujeres sino que también promueve un cambio cultural que puede transformar las dinámicas de poder en la sociedad.

Estas tres áreas (educación de género, humanista y tecnológica) deben trabajar juntas para enfrentar los desafíos que plantea la pornografía *deepfake* creada con IA,

promoviendo en los estudiantes valores éticos, habilidades técnicas y una comprensión profunda de las implicaciones que conlleva el mal uso de las tecnologías. Estos aspectos requieren del compromiso y la preparación de la comunidad docente, que juega un papel crucial en la formación de ciudadanos críticos y responsables.

Es importante destacar la existencia de áreas de investigación que aún no han sido exploradas o, en su caso, resueltas, y que podrían representar futuras líneas de estudio que generen aportes y contribuciones importantes para abordar la problemática:

- a) Desarrollo de tecnologías de detección avanzadas para identificar contenido manipulado,
- b) aspectos legales y éticos asociados a la creación, distribución y consumo de pornografía *deepfake*,
- c) estrategias efectivas para educar al estudiantado sobre el uso adecuado de la IA, mitigar el impacto negativo del mal uso y fomentar el pensamiento crítico para evitar la desinformación creada a través de IA, que permitan al estudiantado tomar decisiones informadas e involucrarse en el desarrollo y regulación de las tecnologías promoviendo su seguridad en el ámbito digital,
- d) estudios que revelen la forma en que el mal uso de la IA y los *deepfakes* afectan las relaciones interpersonales y el desarrollo de estrategias que mitiguen estos efectos negativos y, por último,
- e) el impacto en la salud mental y el bienestar emocional provocado por el uso inadecuado de la IA que vulnera los derechos humanos.

Estos estudios pueden ofrecer *insights* valiosos para diseñar políticas y programas que protejan a las personas, especialmente a las mujeres y niñas, en el entorno digital.

REFERENCIAS

- Acevedo, M. H. (2011). El proceso de codificación en investigación cualitativa. *Contribuciones a las Ciencias Sociales*, (12). <https://www.eumed.net/rev/cccss/12/mha2.htm>
- Atlas.ti (s.f.). *Obtenga una ventaja y potencie su investigación con conocimientos críticos*. <https://atlasti.com/es/investigadores-cientificos>
- Bañuelos C., J. (2022). Evolución del Deepfake: campos semánticos y géneros discursivos (2017-2021). *Icono 14, Revista de Comunicación y Tecnologías Emergentes*, 20(1), 1-21. <https://dialnet.unirioja.es/servlet/articulo?codigo=8560688>
- Bigas F., N. (2023, nov. 2). *'Deepfakes' pornográficos: cuando la IA desnuda tu intimidad y vulnera tus derechos*. Universidad Oberta de Cataluña. <https://www.uoc.edu/es/news/2023/265-deepfakes-pornograficos-cuando-IA-desnuda-tu-intimidad-vulnera-tus-derechos>
- Corral, M. M. (2020, may. 11). *La codificación en la investigación cualitativa*. <https://www.metodos.work/2020/05/11/la-codificacion-en-la-investigacion-cualitativa/#:~:text=codificaci%C3%B3n%20genera%20m%C3%A1s%20familiaridad%20con%20los%20contenidos%20y%20permite%20iniciar>
- De Angeli, A., Falduti, M., Menendez, M., y Tessaris, S. (2021). Reporting revenge porn: A preliminary expert analysis. *CHIItaly '21: Proceedings of the 14th Biannual Conference of the Italian SIGCHI Chapter*, 30, 1-5. <https://doi.org/10.1145/3464385.3464739>

- Fitzgerald, M. (2024, abr. 10). States race to restrict deepfake porn as it becomes easier to create. *Stateline*. <https://stateline.org/2024/04/10/states-race-to-restrict-deepfake-porn-as-it-becomes-easier-to-create/#:~:text=In%202023%2C%20pornography%20made%20up%2098%25%20of%20all%20deepfake%20videos%20online>
- García-Ull, F. J. (2021). *Deepfakes*: el próximo reto en la detección de noticias falsas. *Anàlisi: Quaderns de Comunicació i Cultura*, (64), 103-120. <https://doi.org/10.5565/rev/analisi.3378>
- Gosse, C., y Burkell, J. (2020). Politics and porn: How news media characterizes problems presented by deepfakes. *Critical Studies in Media Communication*, 37(5), 497-511. <https://doi.org/10.1080/15295036.2020.1832697>
- INEGI [Instituto Nacional de Estadística y Geografía] (2023, jul. 13). *Módulo sobre ciberacoso*. *Comunicado de prensa Núm. 404/23*. https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2023/cp404_23.pdf
- Jang, S. M., y Kim, J. K. (2018). Third person effects of fake news: Fake news regulation and media literacy interventions. *Computers in Human Behavior*, 80, 295-302 https://www.researchgate.net/publication/321239042_Third_person_effects_of_fake_news_Fake_news_regulation_and_media_literacy_interventions
- Karasavva, V., y Noorbhai, A. (2021). The real threat of deepfake pornography: A review of Canadian policy. *Cyberpsychology, Behavior, and Social Networking*, 24(3). <https://doi.org/10.1089/cyber.2020.0272>
- Kerner, C., y Risse, M. (2021). Beyond porn and discreditation: Epistemic promises and perils of deepfake technology in digital lifeworlds. *Moral Philosophy and Politics*, 8(1), 81-108. <https://doi.org/10.1515/mopp-2020-0024>
- Maddocks, S. (2020). 'A deepfake porn plot intended to silence me': Exploring continuities between pornographic and 'political' deep fakes. *Porn Studies*, 7(4), 415-423. <https://doi.org/10.1080/23268743.2020.1757499>
- Mania, K. (2024). Legal protection of revenge and deepfake porn victims in the European Union: Findings from a comparative legal study. *Trauma, Violence, & Abuse*, 25(1), 117-129. <https://doi.org/10.1177/15248380221143772>
- Maras, M.-H. y Alexandrou, A. (2018). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *The International Journal of Evidence & Proof*, 23(3), 255-262. <https://doi.org/10.1177/1365712718807226>
- Niebla, J. (2019). Ciberseguridad y derecho: hacia la adopción de un enfoque inteligente para la aplicación de la ley en entornos digitales. *Aequitas Revista cuatrimestral del Poder Judicial del Estado de Sinaloa*, 21(8), 41-57. https://www.academia.edu/45136125/Ciberseguridad_Niebla_Zatarain_2021
- Nugraha, U. A., Kesuma, P. T., y Hidayat, R. M. (2023). Analisis hukum terhadap upaya pencegahan kasus deepfake porn dan pendidikan kesadaran publik di lingkungan digital. *Jurnal Penelitian Hukum*, 7(3). <https://doi.org/10.31004/jptam.v7i3.10815>
- Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., y Nießner, M. (2018). FaceForensics: A large-scale video dataset for forgery detection in human faces. *arXiv*, 1803.09179. <https://doi.org/10.48550/arXiv.1803.09179>
- Timmerman, B., Mehta, P., Deb, P., Gallagher, K., Dolan-Gavitt, B., Garg, S. y Greenstadt, R. (2023). Studying the online deepfake community. *Journal of Online Trust & Safety*, 2(1). <https://doi.org/10.54501/jots.v2i1.126>
- Vega, E. (2023). Tipología de la codificación en el análisis cualitativo. *Medium*. <https://medium.com/@envervega/tipolog%C3%ADa-de-la-codificaci%C3%B3n-en-el-an%C3%A1lisis->

cualitativo-a58a75cbf7f6#:~:text=Codificaci%C3%B3n%20axial%3A%20La%20codificaci%C3%B3n%20axial%20implica%20la%20identificaci%C3%B3n%20de%20relaciones%20y%20conexiones%20entre%20las%20categor%C3%ADas%20o%20c%C3%B3digos

Yessica con “Y” y Desde Mujeres (2023, jul. 31). Deepfake y nuevas formas de violencia digital contra las mujeres. *Zona Docs. Periodismo en Resistencia*. <https://www.zonadocs.mx/2023/07/31/deepfake-y-nuevas-formas-de-violencia-digital-contra-las-mujeres/>

Cómo citar este artículo:

Pereira Hernández, M. L., y Zatarain Avendaño, V. M. (2024). Pornografía deepfake en la era de la IA: nuevos desafíos para la educación de género, humanística y tecnológica. *RECIE. Revista Electrónica Científica de Investigación Educativa*, 8, e2337. <https://doi.org/10.33010/recie.v8i0.2337>



Todos los contenidos de RECIE. *Revista Electrónica Científica de Investigación Educativa* se publican bajo una licencia de Creative Commons Reconocimiento-NoComercial 4.0 Internacional, y pueden ser usados gratuitamente para fines no comerciales, dando los créditos a los autores y a la revista, como lo establece la licencia.
