



Revista Electrónica de Psicología Iztacala



Universidad Nacional Autónoma de México

Vol. 28 No. 4

Diciembre de 2025

<https://doi.org/10.22402/REPI.2025.28.04.5>

RIESGOS Y ALCANCES EN EL USO DE DEEPPFAKES DESDE LA PERSPECTIVA DE ESTUDIANTES DE PSICOLOGÍA

Rodrigo Daniel Medrano Figueroa¹, Jose Manuel Meza Cano²

Facultad de Estudios Superiores Iztacala

Universidad Nacional Autónoma de México

RESUMEN

El presente artículo describe los riesgos y alcances percibidos ante el uso y aplicación de deepfakes por parte de estudiantes de la carrera en Psicología a distancia de la Universidad Nacional Autónoma de México. Se trató de un estudio con diseño no experimental, transversal, con un alcance descriptivo. La muestra la conformaron 10 estudiantes, con edades desde los 18 hasta los 54 años, siendo 9 mujeres y 1 hombre. Se aplicó una entrevista semiestructurada con 15 reactivos de preguntas abiertas referentes a áreas correspondientes al uso y aplicación de deepfakes. Los principales hallazgos mostraron una percepción negativa, ya que una de las categorías con mayor número de frecuencia mencionaba los riesgos y peligros que trae el uso y desarrollo de deepfakes. Sin embargo, durante la entrevista se buscó informar a los participantes sobre la funcionalidad y distintos usos que se le da a esta tecnología, evitando dar información tendenciosa que afectará al objetivo. Aun así, esta percepción se mantuvo debido a un primer acercamiento con deepfakes, pues solo 1 de los 10 participantes tenía conocimiento sobre la existencia de esta tecnología. Por último, este trabajo demostró que es indispensable informar sobre la existencia de este tipo de tecnologías a las personas, para que reflexionen sobre sus alcances. Sobre todo, porque el uso constante en redes sociales hará que sea habitual la

¹ Profesor de Asignatura de la carrera de Psicología SUAyED, FES Iztacala.

Correo electrónico: rodrigo.medrano@iztacala.unam.mx ORCID: <https://orcid.org/0009-0008-0560-2058>

² Profesor Titular "A" Tiempo Completo de la carrera de Psicología SUAyED, FES Iztacala.

Correo electrónico: manuel.meza@iztacala.unam.mx ORCID: <https://orcid.org/0000-0002-9504-7906>

exposición a deepfakes y, como se ha expuesto en este trabajo, la ignorancia y desinformación son aspectos ideales para que este tipo de contenidos logren engañar a los usuarios.

Palabras clave: Deep fakes, alcances, riesgos, inteligencia artificial, manipulación, desinformación, engaño.

RISKS AND IMPACTS IN THE USE OF DEEPFAKES FROM THE PERSPECTIVE OF PSYCHOLOGY STUDENTS

ABSTRACT

This article describes the risks and implications of the use and application of deepfakes by students of the Psychology distance education at National Autonomous University of Mexico. This was a non-experimental, cross-country study with a descriptive scope. The sample consisted of 10 students, aged between 18 and 54 years, 9 women and 1 man. We applied for a semi-structured interview with 15 items of open-ended questions referring to areas corresponding to the use and application of deepfakes. The main results showed a negative perception, because one of the most frequent categories referred to the risks and dangers involved in the use and development of deepfakes. However, during the interview we informed the participants about the functionality and different uses of this technology, avoiding giving us a tendency to give us some information that would affect the target. Even so, this perception was maintained because of a first approach to deepfakes, since only 1 of the 10 participants was aware of the existence of this technology. Finally, this study showed that it is essential to inform people about the existence of this type of technology, to make them think about its implications. Especially because the frequent use of social networks will make it common to be exposed to deepfakes and, as this study has shown, ignorance and misinformation are ideal aspects for this type of content to be able to deceive users.

Keywords: Deep fakes, scopes, risks, artificial intelligence, manipulation, disinformation, deception.

La difusión de la información ha sido una actividad natural del humano desde tiempos remotos, con ello ha evolucionado la forma en que se comunica y difunde el conocimiento. Sin embargo, surgen hechos o aspectos que modifican la estructura original, como lo son las actividades de manipulación o falseamiento de la información. Un ejemplo de ello es la creación de “fake news”, las cuales

consisten en generar contenido que desinforme a un público objetivo y generen polémica en un sector o en general a la sociedad. Esto es posible gracias a la rapidez con la que una noticia o hecho se puede viralizar gracias a las distintas plataformas digitales o medios de comunicación existentes actualmente (Floridi, 2018). Aunque la evolución de estos medios de comunicación se ha desarrollado con el fin de mejorar la rapidez y eficacia; la manipulación y falseamiento de la información se han colocado como una práctica cada vez más común. Si bien las noticias falsas se presentan comúnmente en forma de texto, actualmente, también se encuentran en imágenes o vídeo, por lo que han llegado a ser un problema para que la sociedad pueda informarse correctamente sobre hechos o acontecimientos específicos (Francis, 2021).

Recientemente se ha ido más allá de las manipulaciones en formatos convencionales, como la manipulación de imágenes o videos, a partir de que se han generado productos a través de inteligencia artificial, mejor conocidos como “deepfakes”. Estas consisten en la creación o modificación de contenido audiovisual engañoso o falso mediante la manipulación e integración de imágenes, sonidos y vídeos, logrando crear contenidos realistas a partir de fotografías o montajes, la edición de audio por medio de inteligencia artificial y la personalización de videos, lo que puede incluir modificación de rostros, de voz, del movimiento de los labios o de las posturas (Hazan, 2020).

Cabe resaltar que el alcance y difusión que han tenido este tipo de prácticas se debe a la facilidad para generarlos. Si bien, el crear contenido que realmente engañe al público tiene un coste elevado, ya que suponen la participación de expertos de diferentes áreas, como montadores de vídeo o especialistas en animación, también hay aplicaciones que permiten generar deepfakes a un coste reducido (Qian et al., 2022). Por esta razón, este tipo de recursos implican un riesgo potencial en la creación de noticias falsas, ya que al ser contenidos generados por inteligencia artificial (la cual es constantemente actualizada y mejorada) aportan a

la creación de deepfakes cada vez más realistas y de fácil acceso para el usuario promedio (Cerdán y Padilla, 2019).

REVISIÓN TEÓRICA

Técnicas de manipulación de imágenes, vídeos y audio

Los deepfakes se han convertido en un elemento cada vez más detallado y realista, es cierto que existen características o rasgos que ayudan a diferenciar de un elemento real a uno manipulado, pero las técnicas de manipulación de contenido son perfeccionadas en cuanto al detalle y profundidad que estas reciben (Kietzmann et al., 2020). Con ello, los recientes avances tecnológicos han abierto paso a la creación de materiales que en ocasiones resultan ser altamente realistas gracias al avance en materia de intercambios de rostros, voces y morfología del cuerpo humano, que, dependiendo de la técnica, dejan pocos rastros de manipulación (Chawla, 2019). Para ello existen distintos métodos o técnicas que dependiendo de la experiencia, conocimiento o recursos tanto económicos como digitales que se posean, se logrará crear un producto realista o lo suficientemente cercano para generar contenido falso (Al-khazraji et al., 2023).

De acuerdo con la literatura, los deepfakes se dividen principalmente en dos tipos:

- Deepfaces: consisten en generar imágenes realistas o convincentes, aunque resultan de un contenido totalmente falso al crearse desde cero. Por medio de aplicaciones de inteligencia artificial, se manipulan y generan nuevas imágenes o vídeos a partir de otros contenidos, y se reemplaza a la persona que originalmente aparece en ellos (Godulla et al., 2021).

Dentro de esta categoría se encuentra una de las principales técnicas utilizadas para manipular imágenes o vídeos, denominada “morphing” (Cerdán y Padilla, 2019). Esta consiste en manipular y adjuntar patrones similares entre dos fotografías para transformar una imagen en otra. Su funcionalidad se encuentra en generar una imagen que tenga su propio movimiento considerando una estructura

que va de un punto A hacia un punto B. Dicho en otras palabras, esta técnica sirve para crear una ilusión de metamorfosis entre varias formas, generando una serie de imágenes que crearán el producto final que puede tomarse como falso (Pavis, 2021). De esta forma será posible generar distintos contenidos, como el cambiar una cara por otra, integrando el rostro de una persona en otra, o la manipulación de expresiones faciales para distorsionar la percepción de la identidad de una persona. Otra de las técnicas existentes dentro de esta categoría es el “warping”, la cual permite modificar digitalmente la estructura de una imagen, esto con propósitos creativos para corregir sus alteraciones. Esta técnica tiene diferentes aplicaciones, como generar caricaturas a través de la exageración de rasgos personales (Güera y Delp, 2018). Un uso común que se le da es el de corregir imágenes, que se acostumbran a utilizar para mejorar la estética de las fotos de las redes sociales.

- Deepvoices: contenido generado a través de la manipulación de voz de una persona a través del funcionamiento de texto estructurado por audio, haciendo que esta voz diga o pronuncie frases que se escuchan como voces reales (Gao, 2022). Al igual que la manipulación de imágenes, en función de los datos que se proporcionen a las aplicaciones que generan estos contenidos, la IA será capaz de procesar, automatizar y organizar estos algoritmos que estructuran los patrones de voz que se buscan moldear, imitar y, por lo tanto, falsificar (López, 2022). Actualmente este recurso se ha vuelto popular en el sector del entretenimiento al crear canciones con voces de distintos artistas, logrando modificar una canción original a un producto totalmente falso. Aunque, de igual forma, ha funcionado para crear estafas o robos por medio de llamadas telefónicas o grabaciones de audio (Biurrun, 2021).

Usos y funcionalidad de los deepfakes

Como se revisó con anterioridad, el deepfake es entendido como todo producto o recurso creado por aplicaciones de inteligencia artificial que combinan, reemplazan o manipulan imágenes y videos para crear contenidos falsos que parecen auténticos (Maras y Alexandrou, 2019). Con ello, podemos destacar que la

funcionalidad que puede alcanzar un deepfake depende del propósito de creación y el alcance que se le busque dar. A continuación, se presentan algunos escenarios en donde los deepfakes han tomado un alcance mediático, ya que su desarrollo va desde la aplicación de recursos con usos provechosos, hasta aquellos que plantean preocupaciones éticas y de seguridad debido a su potencial para el uso malintencionado:

- Deepfakes de entretenimiento: Este campo de aplicación podría considerarse uno de los más relevantes debido a las posibilidades económicas derivadas de los productos creados con este fin. Por ejemplo, industrias como la cinematográfica hacen uso de tecnologías avanzadas en deepfake para mejorar efectos especiales, rejuvenecer a actores o recrear digitalmente a personajes fallecidos.

Otro campo de aplicación con este fin se puede ver en plataformas como YouTube donde existen canales dedicados a compartir deepfakes que en su mayoría son protagonizados por celebridades o actores de cine. Lo que se realiza es básicamente la intervención de videos con intercambio de rostros, cuerpos y voces con una narrativa irónica, satírica, crítica o paródica (Ajder, 2019). Aunque en este campo no nos encontramos necesariamente con la creación de materiales malintencionados, depende del conocimiento previo del usuario que consume estos materiales, ya que para alguien que conoce sobre la existencia de deepfakes encontrará entretenimiento en ellos, pero, para aquellos usuarios que desconocen sobre la existencia de estas tecnologías, podría significar un encuentro con desinformación y generación de noticias falsas, especialmente cuando se trata de materiales compartidos en redes sociales sin algún contexto previo, por lo que la interpretación de estos queda a libertad del usuario (González, 2019).

- Deepfakes comerciales: Al igual que el escenario anterior, los deepfakes también tienen aplicaciones comerciales que involucran al campo del marketing y la publicidad. Ejemplo de ello es la creación de las llamadas “influencers virtuales”, que son personajes creados digitalmente o por inteligencia artificial, que, aunque no

tengan presencia física, pueden tener una relevancia significativa en la cultura popular y medios de comunicación (Rico, 2021). Dicha relevancia trae consigo la asociación con marcas para promocionar productos o servicios a un coste menor.

- Deepfakes políticos: En este campo encontramos vídeos o audios de índole política que son alterados a fin de crear contenidos falsos o engañosos, estos pueden incluir discursos falsos, entrevistas alteradas o incluso la aparición de figuras políticas en eventos falsos (Barari et al., 2021). Por lo anterior, es que este campo es considerado como uno de los de mayor impacto mediático, ya que representan numerosas consecuencias éticas, morales y legales debido a la inmediatez de la publicación, la viralización y el tiempo que duran en la red. Estos factores pueden originar distintas problemáticas ya que pueden ser utilizados como una forma de interferencia en campañas electorales, para crear conflictos entre distintos simpatizantes políticos o generar contenido que desinforme y propague noticias falsas (Paris & Donovan, 2019).

- Deepfakes pornográficos: Estos materiales suponen la manipulación por medio de técnicas de inteligencia artificial de material pornográfico, a fin de superponer digitalmente los rostros de personas en dicho contenido. Su creación se centra en mujeres, artistas, actores y en general cualquier celebridad (Cerdán y Padilla, 2019). En todo caso, la suma de estos materiales supone un riesgo inminente para las personas que aparecen en dichos contenidos ya que existe el riesgo de sufrir acoso y chantaje por aquellos que manipulan el material o bien por los usuarios que lo consumen. Ejemplo del avance del deepfake en este campo es la aplicación “DeepNude” que permite desnudar a una mujer a partir de un retrato. Esto es posible gracias a una base de datos con más de 10,000 fotos de mujeres desnudas tomadas de internet. Tomando en cuenta esta base de datos y el análisis previo que realiza a la imagen que se desea manipular es como se añade un cuerpo desnudo al rostro del retrato (Cole, 2019).

Riesgos potenciales en el uso de deepfakes

El fenómeno de los deepfakes y en general las noticias falsas han hecho evidente el problema de desinformación que enfrenta la sociedad actual (Pennycook & Rand, 2021). Tal como se revisó con anterioridad, estos tienen distintos usos y funcionalidades dependiendo del propósito de su creación. Sin embargo, para que un deepfake tenga relevancia y sobre todo genere cierta influencia en su público objetivo, necesita de la difusión y audiencia correcta, ya que el problema no radica en la creación de estos materiales (al menos en aquellos que no implican contenido sensible o restringido), sino en la audiencia que los consume, puesto que muchas veces no se plantea si la información que se está consumiendo o difundiendo es verídica (Mirsky & Lee, 2021).

Tal como lo demuestra la encuesta “La infodemia y su impacto en la vida digital” realizada por Kaspersky, el 70% de los usuarios que participaron no tenían conocimiento sobre la existencia de los deepfakes. Simplemente el 72% de los participantes mexicanos que conformaron la muestra mencionaron no saber qué era un deepfake o cómo reconocer un contenido de este tipo (Kaspersky, 2021). Lo anterior resulta alarmante, considerando que dichos contenidos son una amenaza potencial para la sociedad porque abren la posibilidad de difundir información errónea que puede influir en la opinión pública, además de permitir violar la privacidad y el consentimiento mediante el uso de datos personales sin permiso. Todo esto posibilita el acto de dañar la reputación y la credibilidad al suplantar o difamar a personas, organizaciones o marcas, creando no solo riesgos de seguridad al enfrentarse a problemáticas como el robo de identidad, el fraude o los ataques cibernéticos, sino que los deepfake también pueden afectar la confianza de distintos ambientes digitales gracias a la poca o nula verificación de la autenticidad y las fuentes de información (Francis, 2021).

En relación con lo anterior, se comprende que la preocupación centrada en este tipo de recursos sea cada vez mayor, tomando en cuenta que en internet la

información está al alcance de cualquier persona en cualquier momento, la intención de estos contenidos puede tener alcances sin más limitación que la imaginación. No solo porque la creación de deepfakes es cada vez más realista, sino porque se crean rápidamente y se hacen a bajo costo con softwares disponibles gratuitamente (Al-khazraji et al., 2023). De este modo, ya no es necesario contar con conocimiento especializado para su creación, solo basta con descargar las herramientas informáticas necesarias, utilizar datos públicos, y crear contenidos falsificados cada vez más convincentes (Brown, 2020).

OBJETIVO

Este estudio de investigación tuvo como objetivos describir el riesgo percibido ante el uso y aplicación de deepfakes y analizar los alcances que percibe una muestra de estudiantes de nivel universitario. En este contexto, el trabajo busca responder a las preguntas sobre ¿cómo perciben los estudiantes universitarios los ejemplos de deepfakes? ¿Qué impacto genera en los estudiantes el uso de estos contenidos?

MÉTODO

Se trató de un estudio con diseño no experimental, transversal, con un alcance descriptivo, ya que se pretende explicar las representaciones subjetivas que surgieron de los participantes sobre el fenómeno de los deepfake.

Muestra

Se trató de una muestra no-probabilística, de sujetos voluntarios, ya que se publicó una invitación por medio de redes sociales a alumnos inscritos en los semestres lectivos de la licenciatura en psicología del Sistema de Universidad Abierta y Educación a Distancia de la Facultad de Estudios Superiores Iztacala, de la Universidad Nacional Autónoma de México. Participaron 10 estudiantes, con edades desde los 18 hasta los 54 años, siendo 9 mujeres y 1 hombre. 5 de ellos

estaban inscritos en noveno semestre, 1 de segundo y 4 de primero. Por último, los participantes radican en distintos estados de la república, donde 7 residían en el Estado de México, los 3 restantes residen en Yucatán, Puebla y Nuevo León.

Instrumento

Se realizó una entrevista semiestructurada con 15 reactivos de preguntas abiertas referentes a áreas correspondientes al uso y aplicación de deepfakes. Se convocó a los alumnos de manera individual y se realizó cada entrevista a través de videoconferencia. Cada entrevista abarcó temas como el desarrollo, uso, alcances y riesgos de los deepfakes. Fue grabada, y se les informó a los participantes sobre la confidencialidad de los datos del estudio, ellos aceptaron un consentimiento informado. Las preguntas realizadas se muestran en el anexo 1.

Procedimiento

Por medio de la publicación realizada en redes sociales se contactó a través de correo electrónico a los alumnos interesados en participar en la aplicación de una entrevista. En este primer correo se expuso brevemente el objetivo y naturaleza de la entrevista a realizar. Para aquellos alumnos que respondieron al primer correo se les propuso un horario en el cual se realizó la entrevista a través de la plataforma Zoom, la cual fue grabada, previo consentimiento del participante. Cada entrevista tuvo una duración aproximada de 25 minutos.

La estructura de la entrevista fue la siguiente: Al inicio de la entrevista se mostró a los participantes cuatro imágenes producto de deepfakes, donde se les solicitó analizar dichos contenidos. Posterior al análisis de las imágenes, se procedió a aplicar los reactivos 1, 2 y 3 con el fin de medir el conocimiento inicial sobre IA y deepfakes. A partir del reactivo 4 se retomaron las imágenes para que los participantes identificaran si eran imágenes reales o falsas. Después de dicho ejercicio, se les mostró a los participantes una presentación que detalló la función y

aplicación que tenían los deepfakes. Ya con un conocimiento previo sobre dichos contenidos, se continuó con la aplicación de los reactivos restantes.

A continuación, se muestran los 4 estímulos empleados durante la entrevista.

Figura 1. Deepfake de una escena de la película “volver al futuro” donde se reemplaza el rostro del actor Christopher Lloyd por el de Robert Downey Jr.



Fuente: (Lloyd,2020).

Figura 2. Deepfake de una escena de la película “James Bond” donde se reemplaza el rostro del actor Sean Connery por el de Burt Reynolds



Fuente: (Strauss, 2020).

Figura 3. Deepfakes del actor Keanu Reeves, dichos contenidos son parte de una cuenta de TikTok titulada “@unreal_keanu”



Fuente: (Theres, 2022).

Figura 4. Deepfake creados por el artista de efectos visuales y de inteligencia artificial Chris Umé con la ayuda de un suplente de Cruise, el actor Miles Fisher



Fuente: (Hill, 2021).

Realizadas las entrevistas se procedió a la transcripción de la información, posteriormente se utilizó el programa Taguette (<https://app.taguette.org/>) para la categorización de la información y conteo de las frecuencias de cada categoría. Para crear las categorías se tomaron en cuenta las preguntas diseñadas en la entrevista, algunas preguntas se conjuntaron en una sola categoría debido a la similitud de las respuestas. Solo las preguntas 4 y 15 no fueron incluidas dentro de las categorías, esto porque las respuestas obtenidas permiten representarlas de forma individual. Una vez que se tuvo una primera versión de las categorías, se reunieron dos jueces para revisar que fueran excluyentes y que las definiciones fueran pertinentes, así como algunos ejemplos. Posteriormente se realizó el análisis de la totalidad de las entrevistas.

RESULTADOS

Los resultados obtenidos muestran en primer lugar, las respuestas de cada participante a la pregunta “Respecto a los materiales que acabas de revisar,

¿Cuáles de ellos consideras que son reales?”, el número de aciertos se muestran en la Tabla 1.

Tabla 1. **Respuestas correctas por participante**

Imagen	Respuestas Correctas (Promedio)	Porcentaje de Aciertos (de 10)
Imagen 2	7	70%
Imagen 1	5	50%
Imagen 3	5	50%
Imagen 4	1	10%
Imagen 5	1	10%
Total	19	38%

De acuerdo con la tabla 1, se puede apreciar en qué imágenes los participantes acertaron respecto a los materiales que consideraban como falsos, para ello, se asignó el valor 1 por cada imagen acertada y el valor 0 para aquellas donde el resultado fue erróneo. Considerando el promedio de los participantes, se tuvo una media de 2 respuestas correctas frente a los 4 materiales que se mostraron. De las cuatro imágenes, la imagen 2 fue la que obtuvo mayor cantidad de respuestas correctas, es decir, las personas identificaron que fue creada por IA.

Para el análisis de la información de las entrevistas se usó la aplicación Taguette, en ella se importaron los documentos que correspondieron a las 10 transcripciones de las entrevistas, las cuales contabilizaron un total de 35 páginas. Esto permitió cumplir el propósito de realizar un análisis a través de la selección de los segmentos más importantes. A partir de este análisis se obtuvieron las categorías y sus frecuencias, las cuales son presentadas en la tabla 2.

Tabla 2. **Frecuencia de categorías**

Categorías	Descripción	Frecuencia de segmentos

Peligros en el uso y aplicación de deepfakes	Descripción sobre las desventajas y riesgos encontrados con el uso y existencia de deepfakes	32
Alcances en la sociedad	Impacto que pueden generar los deepfakes en la sociedad	22
Engaño con deepfakes	Los alumnos consideran que han sido víctimas de engaño por un deepfake	20
Evolución de los deepfakes	Descripción sobre la evolución que tendrán los deepfakes en el futuro	19
Percepción sobre deepfakes	Pensamientos que genera el conocimiento adquirido sobre deepfakes	19
Conocimiento sobre IA	Pensamientos de los alumnos sobre la inteligencia artificial	17
Aplicaciones de deepfakes	Descripción sobre las mejores formas de aplicar un deepfake	12
Usos y aplicaciones de la IA	Conocimiento sobre los distintos campos de aplicación de la IA	10
Conocimiento sobre deepfakes	Información del participante sobre deepfakes antes del estudio	10

Como muestra la tabla 2, la categoría con mayor número de frecuencias fue la asociada a los peligros percibidos en el uso y aplicación de deepfakes con 32 frecuencias, seguido de la categoría “alcances en la sociedad” con 22 y “engaño con deepfakes” con 20 frecuencias. A continuación, se muestran ejemplos de cada una de las categorías empleadas:

1. Peligros en el uso y aplicación de deepfakes

Esta categoría tuvo 32 segmentos, lo cual describe las desventajas y riesgos que la muestra percibió con el uso y existencia de deepfakes. Un ejemplo de ello sería el siguiente:

A: “Estos materiales confunden la realidad y son un peligro, manipulan, para que la gente crea sobre cosas que no son reales”.

B: “Es una tecnología peligrosa, incluso ahora ya no me siento segura con el hecho de que mi imagen esté en redes sociales, ya que no es información privada y cualquiera podría hacer uso de ella”.

Cómo puede notarse, estos ejemplos hacen referencia a los comentarios y opiniones que la muestra tuvo respecto a los riesgos que conlleva la existencia de estos contenidos. Siendo la categoría con mayor número de frecuencia, nos da un indicio sobre las creencias que giraron en torno a los deepfakes.

2. Alcances en la sociedad

Esta categoría contó con 22 segmentos, lo cual describe las creencias que la muestra presentó respecto al impacto que los deepfakes podrían generar en la sociedad. Un ejemplo de ello sería el siguiente:

A: “los deepfakes pueden impactar de forma negativa, ya que hay muchas poblaciones vulnerables que no analizan la información”.

B: "Pueden causar pensamientos negativos hacia ciertos sectores dependiendo del propósito, como puede ser manchar la imagen pública o generar polémica".

Esta categoría nos muestra que la percepción que hubo en torno hacia el impacto que puede generar en la sociedad el uso y desarrollo de deepfakes es negativo, ya que el riesgo no solo está presente en su desarrollo, si no en la influencia que estos pueden tener.

3. Engaño con deepfakes

Esta categoría contó con 20 segmentos, la cual indica si los participantes han sido víctimas de engaño por un deepfake. Un ejemplo de ello sería el siguiente:

A: "Una vez en un sitio de citas tuve contacto con una persona que al final falsificaba su imagen".

B: "Recuerdo que llegué a ver un video que pensé que era real y después descubrí que no era así".

Estos ejemplos muestran que los participantes al menos una vez han caído en el engaño de estos contenidos falsos, demostrando que la existencia y presencia de los deepfakes será cada vez más común en ambientes virtuales.

4. Evolución de los deepfakes

Para esta categoría se tuvieron 20 segmentos, la cual muestra el progreso que tendrán los deepfakes en un futuro. Un ejemplo de ello sería el siguiente:

A: "Creo que estos contenidos serán muy realistas en el futuro. Pues actualmente ya es muy difícil distinguir una imagen de una manipulada. Entonces creo que va a llegar el momento donde sea imposible distinguir estos materiales".

B: “Hay que tomar en cuenta que estos contenidos van mejorando, en un futuro serán cada vez más engañosos”.

Estos ejemplos muestran cómo se percibe a esta tecnología como una herramienta que en esencia solo sirve para generar engaño, ya que se puede deducir que existe una preocupación donde los deepfakes evolucionen al grado de ser indistinguibles de la realidad.

5. Percepción sobre deepfakes

En este caso se tuvieron 19 segmentos, la cual describe los pensamientos que genera el conocimiento adquirido de deepfakes. Un ejemplo de ello sería el siguiente:

A: “Me deja una sensación desagradable, pues ahora veo que es muy fácil manipular no solamente las emociones o los pensamientos de una persona, sino de las masas, de gran número de personas a través de distintas plataformas como son las redes sociales”.

B: “Me genera preocupación, entiendo que ya existía este tipo de contenidos, pero no pensé que ya pudieran lucir tan realistas También me genera inseguridad, pienso que cualquiera podría hacer uso de mi imagen”.

Los siguientes ejemplos muestran las primeras impresiones respecto al conocimiento sobre funcionamiento y aplicación de deepfakes. Debe señalarse que durante la entrevista se les informó a los participantes sobre la existencia de estos contenidos, sus características, funcionalidad y distintos campos de usos. Sin embargo, resulta importante destacar que aun con ese repertorio de conocimiento, los deepfakes generaron opiniones negativas.

6. Conocimiento sobre IA

Con 17 segmentos, esta categoría describe el conocimiento inicial que tuvieron los participantes sobre inteligencia artificial. Un ejemplo de ello sería el siguiente:

A: “Es una herramienta útil para tener acceso más rápido y eficaz a la información”.

B: “Entiendo que es una tecnología que permite crear sistemas o aplicaciones que puedan realizar distintas tareas”.

Tomando en cuenta que la Inteligencia Artificial hace referencia a una disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico (Floridi, 2018), los ejemplos nos muestran que existen diversas opiniones sobre lo que representa la IA para la mayoría, pues sólo uno de los participantes dio una definición que podría considerarse como acertada, tal como lo muestra el siguiente ejemplo:

A: “Es una herramienta tecnológica que se alimenta de millones de datos que hay en internet originados por personas. Estos datos se pueden utilizar para crear nuevo contenido ya sea imágenes, texto, y otros usos”.

7. Aplicaciones de deepfakes

Para esta categoría se obtuvieron 12 segmentos, la cual describe las mejores formas de aplicar un deepfake. Un ejemplo de ello sería el siguiente:

A: “Yo creo que sí pueden tener aplicaciones positivas desde un nivel creativo, siempre y cuando no tenga un uso malintencionado, sino que sea usado para innovar”.

B: “Creo que en la industria del entretenimiento como lo que se hace en películas”.

Esta categoría buscó adjuntar todos los datos que hicieran mención sobre usos positivos para aplicar o desarrollar un deepfake, al ser una de las categorías

con menor frecuencia, podemos deducir que la muestra no percibió a estos contenidos como seguros para su aplicación y uso.

8. Usos y aplicaciones de la IA

Para esta categoría se obtuvieron 10 segmentos, la cual describe el conocimiento sobre los distintos campos de aplicación de la IA. Un ejemplo de ello sería el siguiente:

A: “Solo alcance a utilizar aplicaciones como Chat GPT la cual te da ideas sobre cómo redactar textos”.

Ejemplos como el anterior detallan la mayoría de las respuestas proporcionadas por los participantes, más allá de las distintas aplicaciones que tiene la inteligencia artificial, la muestra sólo logró identificar la aplicación de chatbot “Chat GPT”.

9. Conocimiento sobre deepfakes

Por último, en esta categoría se obtuvieron 10 segmentos, la cual describe el conocimiento previo sobre la existencia de deepfakes. Un ejemplo de ello sería el siguiente:

A: “No, no conozco el término”.

La afirmación anterior ilustra las respuestas de la mayoría de los participantes, pues de los 10 participantes solo 1 tenía conocimiento sobre la existencia de dichos contenidos.

Por último, la tabla 3 ilustra los resultados obtenidos por medio de la pregunta 15, la cual buscaba englobar en una palabra las emociones o pensamientos que le generó a la muestra el conocimiento adquirido sobre deepfakes.

Tabla 3. **Frecuencia de palabras**

Palabra	Frecuencia
Miedo	4
Incertidumbre	1
Precaución	1
Preocupación	1
Protección	1
Riesgo	1
Sorpresa	1

Como puede verse en la tabla 3, de los 10 participantes, 4 de ellos coincidieron con el término “miedo”, el resto de la muestra proporcionó distintas palabras que podrían entrar en una misma categoría referida al miedo o la percepción de un riesgo ante esta tecnología.

DISCUSIÓN

Se aplicó una entrevista semiestructurada a 10 participantes pertenecientes a la carrera de Psicología Sistema de Universidad Abierta y Educación a Distancia de la Facultad de Estudios Superior Iztacala, Universidad Nacional Autónoma de México. Para describir el impacto que generó a la muestra la existencia y el uso de deepfakes, no solo se presentó el uso malintencionado que se le da a estos contenidos, pues parte del objetivo incluía instruir sobre el desarrollo y distintas aplicaciones que tiene esta tecnología, sobre todo porque tal como indican los resultados, solo uno de los diez participantes tenía conocimiento sobre la existencia del término deepfakes. Sin embargo, resulta notorio que la muestra tuvo una percepción negativa frente a esta tecnología, lo podemos deducir con el hecho de que las categorías con mayor número de frecuencia mencionan los riesgos y peligros que trae el uso y desarrollo de deepfakes.

En ese sentido, podría interpretarse que la muestra tuvo una percepción negativa debido a este primer acercamiento sobre la existencia de estos contenidos,

y aunque podría ser cierto gracias a la poca información con la que contaban los participantes y en general, la dificultad que presentaron para definir a la inteligencia artificial o sus distintos campos de aplicación, es importante resaltar que los deepfakes representan grandes peligros, como lo es la manipulación de información y con ello la creación de fake news y desinformación, las cuales por sí solas ya representan un riesgo real (Brown, 2020). Sin embargo, una de las mayores amenazas son los riesgos que afectan a la intimidad. De acuerdo con Al-khazraji et al. (2023), las mujeres son las principales víctimas de esta tecnología, ya que los deepfakes pornográficos son los principales contenidos que se desarrollan actualmente, y con ello, generan distintas consecuencias como la violencia virtual, daños a la reputación y acoso. Esto ya ha sucedido en México, específicamente en el Instituto Politécnico Nacional, dado que un estudiante creó miles de imágenes de compañeras desnudas empleando IA (Rodríguez, 2023). Aunado a ello, cualquier mujer puede ser víctima de los deepfakes, además que es casi imposible defenderse de ellos. Esto, gracias a la rapidez con la que viaja la información y la posibilidad de obtener imágenes de cualquier persona por medio de redes sociales.

Evidentemente, que la muestra haya percibido como peligrosa la existencia de esta tecnología no es algo aislado, a final de cuentas, nos encontramos inmersos en una era digital que está cambiando nuestra forma de interactuar y en general de habituarnos a una vida digital (Pavis, 2021). Por lo tanto, otros de los puntos de mayor importancia es el hecho de que la muestra desconociera sobre la existencia de los deepfakes. Si bien, conforme asimilaban la información se dieron cuenta que ya habían interactuado o visto algún contenido considerado como deepfake, no hay que negar que esta tecnología está en constante avance, pues actualmente ya hay contenidos que lucen realistas, uno de ellos fue la imagen 4 ocupada durante la entrevista, que al retomar el análisis de la tabla 1, podemos percatarnos que solo 1 persona de los 10 participantes pudo identificar a esta imagen como falsa.

Si tomamos en cuenta que en la actualidad ya se habla de una sobreestimulación generada por el uso constante de redes sociales, no estaría

alejado de la realidad el pensar que los deepfakes tienen una ventaja en este contexto (Armaza, 2023). Esto porque al no analizar la información que consumimos puede generar un riesgo latente al ser víctimas de fake news, tal y como los participantes detallaron dentro de la categoría “engaño con deepfakes”, donde gran parte de ellos pudieron identificar que el contenido que estaban revisando era falso gracias a que la descripción del mismo lo especificaba o porque otra persona les había hecho mención de esto, pero no hay que dejar de lado aquellos contenidos malintencionados que claramente no van a especificar su veracidad y buscarán el mayor número de víctimas. Ejemplo de ello son los deepfakes políticos, que de acuerdo con Lavanda (2022), se han visto inmersos en distintas problemáticas gracias a la capacidad que tienen para generar polémica, y no solo eso, puesto que se busca difundir estos materiales en aquellos sectores o poblaciones vulnerables que son más propensos a caer en engaños, tal y como lo mencionan algunos participantes, donde coinciden en que adultos mayores o adolescentes son las víctimas perfectas para manipular por medio de estos contenidos.

CONCLUSIONES

El objetivo del trabajo fue describir los riesgos y alcances percibidos ante el uso y aplicación de deepfakes de estudiantes de Psicología. Para ello, durante la entrevista se buscó informar a los participantes sobre la existencia, funcionalidad y distintos usos que se le da a esta tecnología. Sin embargo, aunque el objetivo inicial buscaba indagar sobre la percepción que los participantes tenían respecto a dicha tecnología, resultó necesario tomar la decisión de instruir con información precisa para que el criterio que la muestra participante formará a lo largo de la entrevista aportará mejores resultados, esto, porque de haber generado información desde la ignorancia o el poco conocimiento sobre el tema no aportaría significancia al trabajo.

Entre las posibles limitaciones del estudio, resaltó el no considerar participantes que ya tuvieran conocimiento sobre la existencia de deepfakes, esto para crear una comparativa sobre distintos puntos de vista frente aquellos que desconocen en su totalidad sobre estos contenidos. Pero, incluso el único

participante que tenía conocimiento sobre deepfakes, compartió puntos de vista similares a sus demás compañeros, por lo que resultó conveniente el no tener criterios de inclusión o exclusión para participar, ya que esta pequeña muestra podría representar lo que percibe un gran sector como sería el sistema educativo y el nivel académico al que pertenecen. Para ello, resulta indispensable considerar a una mayor muestra que represente resultados más robustos, esto con la creación de un instrumento o encuesta que tome en cuenta los aspectos revisados en este estudio para poder ser aplicado a una mayor cantidad de personas y emplear estadística descriptiva e inferencial.

Finalmente, este trabajo demostró que es indispensable informar sobre la existencia de este tipo de tecnologías a las personas, para que reflexionen sobre sus alcances. Sobre todo, porque el uso constante en redes sociales hará que sea habitual la exposición a deepfakes y, como se ha expuesto en este trabajo, la ignorancia y desinformación son aspectos ideales para que este tipo de contenidos logren engañar a los usuarios. Por ello, aprender a identificar contenidos falsos y demostrar los beneficios de analizar la información que se consume puede combatir los efectos negativos que ha traído el uso de esta tecnología, sobre todo dado que uno de los mayores riesgos reside en la facilidad para crear deepfakes con facilidad, ya que con una aplicación en el teléfono móvil cualquier persona puede crear un vídeo falso pero creíble (Mirsky & Lee, 2021). Aunado a ello, hay que tomar en cuenta que ya es difícil detener la viralización de noticias falsas por la red, al tomar en cuenta que pueden crearse también en un formato en vídeo, las herramientas cognitivas para cuestionar lo que se puede revisar a simple vista, serán cada vez más necesarias.

Para futuros estudios, se recomienda tomar una muestra de aquellos profesionales o usuarios que hagan uso de esta tecnología. Si bien, sería difícil incluir a aquellos que le dan un uso malintencionado, considerar a aquellos que se dedican al desarrollo de aplicaciones o de estos contenidos, lo cual podría generar un panorama más amplio sobre el verdadero propósito o expectativa sobre los

deepfakes. Esto, debido a que uno de los usos principales de esta tecnología se ha visto en producciones de entretenimiento como lo son las industrias cinematográficas, que han hecho uso de deepfakes para crear contenidos realistas. Esto permitiría esclarecer el porqué de la existencia de esta tecnología y cómo su propósito inicial no necesariamente corresponde al uso que se le da actualmente para desinformar.

Referencias Bibliográficas

- Ajder, H. (2019). *The State of Deepfakes: Landscape, Threats, and Impact*. Amsterdam: Deepttrace.
- Al-khazraji, S., Saleh, H., Khalid, A., & Mishkhal, I. (2023). Impact of Deepfake Technology on Social Media: Detection, Misinformation and Societal Implications. *The Eurasia Proceedings of Science Technology Engineering and Mathematics*, 23, 429-441. DOI: 10.55549/epstem.1371792
- Armaza, J. (2023). The risk of excessive use of social networks by latin american students. *SciELO Preprints*. <https://doi.org/10.1590/SciELOPreprints.5241>
- Bañuelos, J. (2020). Deepfake: la imagen en tiempos de la posverdad. *RPC*, (1), 51–61. <https://doi.org/10.21555/rpc.v0i1.2315>
- Barari, S., Lucas, C., & Munger, K. (2021). Political deepfakes areas credible as other fake media and (sometimes) realmedia. *OSF preprints*. <https://doi.org/10.31219/osf.io/cdfh3>
- Biurrun, A. (2021). Deep Voice: tecnología que falsifica la voz y con la que han robado 35 millones de dólares. *La razón*. <https://www.larazon.es/tecnologia/20211016/ivt7yazbmzb7lmxknhf2hz2fma.html>
- Brown, I. (2020). "Deepfakes and the Weaponization of Disinformation". *Virginia Journal of Law and Technology*, 23(1).
- Cerdán, V. y Padilla, G. (2019). Historia del fake audiovisual: deepfake y la mujer en un imaginario falsificado y perverso. *Historia y comunicación social*, 24(2), 505-520. <https://doi.org/10.5209/hics.66293>
- Chawla, R. (2019). Deepfakes: How a pervert shook the world. *International Journal of Advance Research and Development*, 4(6), 4-8. <http://doi.org/10.22215/timreview/1282>

- Cole, S. (2019). Esta terrorífica app crea un nude de cualquier mujer con un simple clic. Vice. <https://bit.ly/2li3WbS>
- DFCD (2020). Deepfake Detection Challenge. <https://deepfakedetectionchallenge.ai/>
- Francis, C. (2021). Popular discourse around deepfakes and the interdisciplinary challenge of fake video distribution. *Cyberpsychology, Behavior, and Social Networking*, 24(3):159– 63.
- Floridi, L. (2018). Artificial Intelligence, Deepfakes and a Future of Ectypes. *Philos. Technol*, 31, 317–321. DOI: <https://doi.org/10.1007/s13347-018-0325-3>
- Gao, Y. (2022). *Audio deepfake detection based on differences in human and machine generated speech*. (Doctoral dissertation). Carnegie Mellon University.
- Godulla, A., Hoffmann, C., & Seibert, D. (2021). Dealing with deepfakes, an interdisciplinary examination of the state of research and implications for communication studies. *SCM Studies in Communication and Media*, 10(1), 72-96.
- González, M. (2019). “Fake news: desinformación en la era de la sociedad de la información”. Ámbitos. *Revista internacional de comunicación*, 45, 29-52. <https://doi.org/10.12795/ambitos.2019.i45.03>
- Güera, D., y Delp, E. (2018). *Deepfake video detection using recurrent neural networks*. In the 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 1-6). IEEE. <https://doi.org/10.1109/AVSS.2018.8639163>
- Hazan, S. (2020). *Deep fake and cultural truth custodians of cultural heritage in the age of a digital reproduction*. In the International Conference on Human-Computer Interaction, 65–80.
- Hill, C. (2021). Tom Cruise deepfakes that are almost too real. [Imagen]. <https://www.indy100.com/showbiz/tom-cruise-deepfake-videos-tiktok-b1810325>
- Kietzmann, J., Mills, A., & Plangger, K. (2020). Deepfakes: perspectives on the future “reality” of advertising and branding. *International Journal of Advertising*, 1–13.
- Kaspersky. (2021). La infodemia y su impacto en la vida digital. kaspersky daily. <https://latam.kaspersky.com/blog/investigacion-la-infodemia-y-su-impacto-en-la-vida-digital>

- Lavanda, M. (2022). Deepfake: Cuando la inteligencia artificial amenaza el Derecho y la Democracia. *Lawgic Tec - Revista de Derecho y Tecnología*, 2, 84 - 95.
- López, J. (2022). *Análisis y desarrollo de modelos de generación de voz mediante Deep Voice*. Proyecto Fin de Carrera / Trabajo Fin de Grado, E.T.S. de Ingenieros Informáticos (UPM), Madrid, España.
- Lloyd, B. (2020). This deepfake of Robert Downey Jr. in 'Back To The Future' is incredible. [Imagen]. <https://entertainment.ie/movies/movie-news/deepfake-robert-downey-jr-doc-brown-441444/>
- Mirsky, Y. & Lee, W. (2021). The creation and detection of deepfakes: A survey. *ACM Computing Surveys. CSUR*, 54(1), 1–41.
- Rico, T. (2021). *¿El futuro del marketing digital? Tendencias y características de los influencers artificiales en Instagram*. Trabajo Final de Grado. Universitat Politècnica de València.
- Rodríguez, A. (2023). *Acusado un universitario de alterar con inteligencia artificial miles de imágenes de alumnas para venderlas como pornografía*. El país.
- Paris, B. & Donovan, J. (2019). Deepfakes and Cheap Fakes. United States of America: Data & Society.
- Pavis, M. (2021). Rebalancing our regulatory response to Deepfakes with performers' rights. *Convergence*, 27(4), 974-998.
- Pennycook, G., & Rand, D. (2021). The psychology of fake news. *Trends in cognitive sciences*, 25(5), 388-402.
- Qian, S., Shen, C., & Zhan, J. (2022). Fighting cheapfakes: using a digital media literacy intervention to motivate reverse search of out-of-context visual misinformation. *Journal of Computer-Mediated Communication*. 28(1), 1–12 <https://doi.org/10.1093/jcmc/zmac024>
- Strauss, P. (2020). Burt Reynolds Is James Bond. [Imagen]. <https://theawesomer.com/burt-reynolds-is-james-bond/567581/>
- Theres, S. (2022). Keanu Reeves deepfake. [Imagen]. <https://techstory.in/keep-an-eye-out-for-this-keanu-reeves-deepfake-on-tiktok-that-has-fooled-millions-of-users-already/>