

ACTIVIDAD 6

CASO I: “FALSIFICACIÓN DE INFORMACIÓN” (PARTE 2)

Oportunidades curriculares

Tecnología

- **I° Medio - OA 5:** Analizar las formas en que los productos tecnológicos y los entornos evolucionan, caracterizando los diversos factores que influyen en ese cambio.
- **II° Medio - OA 5:** Evaluar críticamente cómo las innovaciones tecnológicas actuales afectan a la sociedad y el ambiente, considerando criterios éticos, económicos, ambientales y sociales.
- **II° Medio - OA 6:** Proyectar escenarios de posibles impactos positivos o negativos de las innovaciones tecnológicas actuales en ámbitos personales, sociales, ambientales, legales, económicos u otros.

Historia

- **II° Medio - OA 22:** Analizar el concepto de Derechos Humanos, considerando características como su universalidad, indivisibilidad, interdependencia e imprescriptibilidad y la institucionalidad creada para resguardarlos a nivel nacional e internacional, reconociendo, en este contexto, los derechos vinculados a los grupos de especial protección.

Visión general

En esta actividad, las y los estudiantes experimentan una adaptación del test de Turing, teniendo que diferenciar contenidos reales de los elaborados por sistemas de inteligencia artificial. Después de revisar un caso cotidiano aplicado a redes sociales, utilizan la plataforma *Detect Fakes* para identificar videos que han sido manipulados a través de IA para distorsionar caras o intercambiar

rasgos faciales. Finalmente, conocen y reflexionan en torno a un caso real de fraude, para volver a responder las preguntas de la guía de la actividad anterior.

Objetivos

- Experimentar con herramientas que permiten crear contenido nuevo a partir de datos preexistentes.
- Reflexionar sobre las implicancias de sistemas avanzados de IA que permiten imitar la realidad con un alto nivel de verosimilitud.

Agenda (60 minutos)

- Introducción (15 min):
 - Conocimientos previos: filtro de Snapchat
 - El test de Turing
- Desarrollo (30 min):
 - Experimentación con plataforma Detect Fakes
 - Caso real: engaño con deepfake
- Cierre (15 min):
 - Reflexión sobre el fenómeno del engaño
 - Comparación de respuestas de la actividad anterior

Materiales para esta actividad

- Guía docente
- Presentación de diapositivas
- Celular con Internet con Snapchat y sesión iniciada
- Computadores con Internet
- Hoja del estudiante (misma que para parte I)

Palabras clave

Test de Turing, IA creativa, deepfake, contenido falso, realidad.

Acceso a presentación de diapositivas

https://docs.google.com/presentation/d/1JRiACuJSVsiDyWZIXMF4zP_a4B6z2rufgnom1lI3lo/edit?usp=sharing

GUÍA DOCENTE

En la primera parte de esta actividad, las y los estudiantes reflexionaron acerca de las implicancias éticas de estos sistemas de IA. El objetivo de esta segunda parte será que puedan conocer nuevos elementos y experimentar el fenómeno del engaño por sus propios medios.

Introducción (15 minutos)

Esta parte introductoria tiene el objetivo de activar los conocimientos previos de las y los estudiantes mediante la revisión de aplicaciones cotidianas de la IA, llamando de paso la atención en torno a lo vinculados/as que están, quieran o no, con estos sistemas.

—¿Cuántos/as de ustedes tienen redes sociales como Snapchat, TikTok o Instagram?

Seguramente una amplia mayoría contesta afirmativamente.

—¿Alguien ha utilizado el filtro de intercambio de caras?

Pídale a un par de parejas que quieran hacer la prueba. Para esto, indíqueles que vayan al buscador de filtros de una de estas aplicaciones y pongan “FaceSwap”. Puede reforzar las instrucciones con fuentes externas¹⁸. Verá cómo sus estudiantes ríen y comentan los resultados del intercambio facial.

—¿Sabían que este filtro usa inteligencia artificial?

Comente a las y los estudiantes que éste es otro ejemplo de cómo este tipo de sistemas de IA generativa se están utilizando con fines recreativos, enfatizando que tecnologías muy similares también pueden usarse con objetivos delictivos. Esta reflexión inicial es importante para el análisis que viene a continuación.

IA Creativa e Inteligencia

Como mencionamos en la actividad anterior, las tecnologías basadas en **IA generativa** son a veces llamadas **IA creativa**, pues su uso permite la creación de dibujos, pinturas, imágenes y canciones. Hasta hace solo unos años, el adjetivo “creativo” era exclusivamente

¹⁸<https://www.infobae.com/america/tecno/2020/04/23/face-swap-como-usar-el-filtro-de-snapchat-para-intercambiar-rostros>

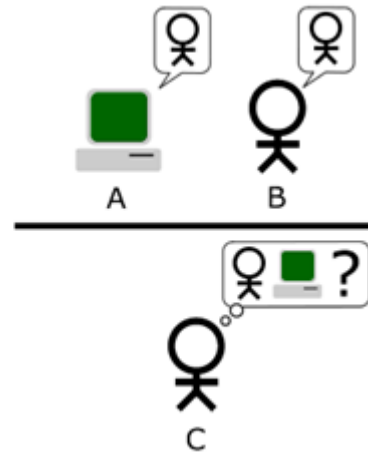
utilizado para describir a los seres humanos. Más aún, **la creatividad artística es aún considerada una de las principales características de la inteligencia humana**. Sin embargo, los resultados de la IA creativa nos hacen preguntarnos si la creatividad es aún una cualidad exclusiva de los seres humanos. Más aún, ¿son los sistemas basados en IA “inteligentes”?

Esta última pregunta es tan antigua como la informática. **Alan Turing**, padre de la computación moderna, se pregunta en su artículo de 1950 “Máquinas computacionales e inteligencia” la pregunta: **¿pueden las máquinas pensar?** Al iniciar su artículo, Turing indica que las definiciones de “inteligencia” y “capacidad de pensamiento” son difíciles y polémicas. En vez de intentar definir estos términos, Turing propone una prueba llamada “El juego de la imitación” (The imitation game), en el cual un interrogador aislado en una sala se comunica a través de mensajes escritos con personas y con máquinas sobre las cuales se espera evaluar inteligencia. El interrogador, que no sabe quiénes de sus interlocutores son máquinas o personas, puede hacer todas las preguntas que estime necesarias. Al final del juego, el interrogador debe identificar quiénes de sus interlocutores son máquinas o personas. Si en varias rondas del juego, con distintos interrogadores, la máquina y las personas interrogadas no logran diferencias sustantivas en ser identificadas, entonces la máquina es considerada “inteligente”, o lo suficientemente “inteligente” como para suplantar a una persona¹⁹.

¹⁹ Puede apoyarse con este video de TED-Ed, activando los subtítulos en español:
<https://www.youtube.com/watch?v=3wLqsRLvV-c>



Alan Turing a los 16 años²⁰.



Prueba de Turing²¹. El interrogador (C) desconoce quién es el ser humano (B) y quién la máquina (A).

Esta “prueba de inteligencia” es actualmente conocida como el “test de Turing”.

La IA creativa ha tenido éxito en *imitar* la inteligencia humana en el sentido de la creatividad artística. Las “pruebas de Turing creativas” han sido utilizadas para evaluar la “creatividad” de las máquinas, pidiéndoles a personas intentar distinguir entre imágenes generadas por IA creativa o por seres humanos. Los resultados de estos tests son mixtos, aunque en algunas aplicaciones son muy positivos en favor de los algoritmos de inteligencia artificial [1].

—¿Les gustaría probar una adaptación del test de Turing? ¿Podrán identificar qué es real y qué es elaborado por máquinas?

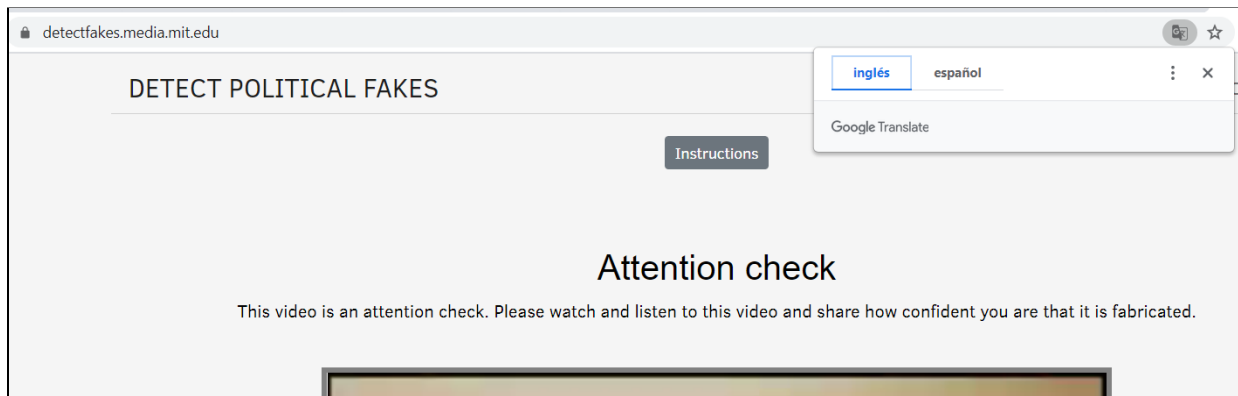
Desarrollo (30 minutos)

Para esta actividad, diríjase a la sala de computación o Enlaces para que sus estudiantes tengan acceso a un computador con Internet. Una vez allí, pídale ingresar al sitio con el navegador Google Chrome: <https://detectfakes.media.mit.edu>. Indíqueles que activen la traducción automática desde la esquina superior derecha de la pantalla (como se indica en

²⁰ Imagen de dominio público vía Wikimedia Commons

²¹ Imagen de dominio público vía Wikimedia Commons. Autor: [Bilby](#)

la imagen). Note que la plataforma solicita el consentimiento informado de los padres o apoderados, dado que recolecta los datos de las respuestas con fines investigativos²². Considere un tiempo para informar esta situación antes de realizar la actividad.



Una vez sus estudiantes hayan ingresado a la página, explíqueles que la actividad se trata de identificar si los videos que se muestran son reales o falsos, es decir, si están alterados o no por un algoritmo de IA (contenido falso o *fake*). El sitio define un *deepfake* como aquel contenido en el que “una cara ha sido manipulada por IA para distorsionarla parcialmente o intercambiar rasgos faciales”. Para hacer la prueba, tienen que ver los videos e indicar en cada caso qué tan seguros/as están de que se trata de un *deepfake* y presionar “enviar”.

Una vez enviada la respuesta, el sistema retroalimentará con la predicción de un sistema automatizado y, posteriormente, con una respuesta taxativa. Para ver otro video, es necesario pulsar “próximo”. La plataforma irá entregando indicaciones y explicaciones mientras se usa, como el número de preguntas que se ha respondido y la advertencia sobre el uso de distractores.

Deje que sus estudiantes revisen entre 5 y 10 videos, de modo que puedan experimentar el fenómeno del engaño. Para ello, deben usar la hoja del estudiante de la clase anterior y completar el cuadro con sus respuestas. En lo posible, mantenga los mismos grupos.

—¿Cómo les fue con el test? ¿Cuántas veces fallaron en identificar el contenido falso?

²² <https://detectfakes.media.mit.edu/informed-consent>

Una vez que hayan comentado distintas experiencias, presente el siguiente caso:

—**¿Recuerdan que dijimos que estos sistemas estaban relacionados a la entretención, como en el caso de Instagram, TikTok o Snapchat? Pues bien, la IA creativa también puede ser usada para usos maliciosos.**

Presente el titular de esta noticia “Los deepfakes de audio se usan para realizar estafas telefónicas”. Ingrese al link a continuación. Haga una lectura general y apóyese con la presentación de diapositivas.

<https://www.lavanguardia.com/tecnologia/20190910/47263675640/deepfake-audio-voz-suplantacion-estafa-telefonica-llamada-inteligencia-artificial-seguridad.html>

Si cuentan con tiempo, recomendamos revisar también los titulares y el contenido principal de los siguientes contenidos noticiosos:

- “Los videos virales del falso Tom Cruise en TikTok con tecnología deepfake: ¿cuál es el peligro?” El objetivo de este artículo es reflexionar sobre la proliferación y los riesgos de generar perfiles falsos de figuras públicas.

<https://www.infobae.com/america/tecno/2021/03/01/los-videos-virales-del-falso-tom-cruise-en-tiktok-con-tecnologia-deepfake-cual-es-el-peligro/>

- “ZAO, la app móvil china que a través de deepfake te convierte en DiCaprio en segundos”. El objetivo de este video es graficar la facilidad con que se pueden desarrollar hoy videos con suplantación de identidad, incluso sin la necesidad de software o conocimiento especializado.

https://www.youtube.com/watch?v=VnUvGeLK_8w

Cierre (15 minutos)

Para finalizar esta actividad de dos partes, realice la siguiente dinámica de cierre y consolidación.

—¿Qué se siente ser engañado/a por una máquina? ¿Qué pasa cuando este engaño ya no es un simple juego o entretenimiento, sino para robar o generar un testimonio o evidencia falsos?

Entréguales un espacio para contestar estas preguntas.

—Observen nuevamente sus respuestas de la actividad anterior. En ellas podrán encontrar las implicancias que ustedes indicaron que podía tener el uso de este tipo de sistemas de IA. ¿Qué cambió? ¿Qué efectos creen que podría tener el uso malicioso de estas tecnologías, ahora que han experimentado el fenómeno del “engaño”? Agreguen la respuesta a esta pregunta en la última parte de la guía anterior.

Bibliografía

[1] <http://bregman.dartmouth.edu/turingtests/node/72>