



Anti-
RUMQUR

MODULO 3
IDENTIFICANDO
CONTENTIDO IA



BY SA



**Cofinanciado por
la Unión Europea**

Financiado por la Unión Europea. No obstante, los puntos de vista y opiniones expresados son exclusivamente los del autor o autores y no reflejan necesariamente los de la Unión Europea o la Agencia Nacional. Ni la Unión Europea ni la autoridad que concede la subvención pueden ser consideradas responsables de las mismas.

Anti-
RUMQUR

stiftung
digitale
chancen



CITIZENS
IN POWER

challedu
industry | games | education

Módulo 3: Identificando contenido IA

El auge de los contenidos generados por IA ha dificultado la lucha contra la información falsa y la desinformación. La IA puede crear vídeos, imágenes y textos falsos de gran realismo a gran velocidad, lo que dificulta el trabajo de los verificadores de hechos. Esto significa que la información falsa puede difundirse rápidamente y parecer muy convincente, lo que dificulta saber qué es verdad y qué no. Como resultado, el contenido engañoso puede convertirse fácilmente en viral, dañando la confianza del público y creando confusión.

En el estado actual de la IA, todavía es posible identificar los contenidos generados por ella, ya sea prestando atención a ciertos defectos o haciendo uso de herramientas útiles. Este módulo explorará tanto los retos de la IA como las opciones que tienen los usuarios para afrontarlos.

3.1 Actividad 1: Input – Introducción al texto generado por IA

Tipo: Input

Duración: 10 minutos

Objetivos de aprendizaje: Los participantes comprenderán mejor la IA generativa, los retos que plantea para la difusión de la desinformación y la detección de textos generados por IA..

Materiales/ recursos: Cuadernos y bolígrafos para los participantes. Guión de entrada para el facilitador. Pantalla grande o proyector y un ordenador portátil conectado a él para el facilitador..

Opcional: El facilitador puede preparar una PPT ajustada a la edad y los conocimientos previos del grupo.

The facilitator uses the following information for input.

Introducción:

La IA puede generar imágenes engañosamente realistas de hechos que nunca ocurrieron, voces generadas artificialmente y textos sofisticados sin autores humanos. Cada vez es más difícil saber qué es real y qué es falso. Esto pone de relieve una verdad general sobre la tecnología: puede ser tanto una bendición como una maldición, dependiendo de cómo se utilice y con qué fin.

Aunque gran parte de los contenidos generados por la IA en Internet hoy en día son el resultado de la creatividad de los usuarios, la IA también puede emplearse para engañar y difundir desinformación.

La IA generativa es un tipo de inteligencia artificial que puede crear nuevos contenidos, como textos, imágenes, música o vídeos, aprendiendo de los datos existentes. Utiliza programas informáticos inteligentes llamados redes neuronales para crear cosas que parecen o suenan como si hubieran sido hechas por humanos. Algunos ejemplos son chatbots como ChatGPT que escriben textos, herramientas de IA que crean arte y programas que crean vídeos falsos.



Texto generado por IA:

El texto generado por IA se crea utilizando un tipo de inteligencia artificial llamada modelos lingüísticos. Estos modelos se entrenan con grandes cantidades de datos de libros, artículos y páginas web. Al aprender patrones y estructuras de los datos, la IA puede generar nuevos textos que suenan como si los hubiera escrito un ser humano. Cuando se le da una indicación o un tema, la IA predice las siguientes palabras y frases basándose en lo que ha aprendido.

El texto generado por la IA supone un reto para la desinformación, ya que puede crear artículos o publicaciones en redes sociales realistas y convincentes. Esto facilita la creación deliberada de contenidos engañosos. El contenido puede producirse en grandes volúmenes y adaptarse para dirigirse a audiencias específicas.

También es importante recordar que los contenidos generados por IA no tienen por qué ser ciertos. Uno de los motivos es que los propios modelos lingüísticos que hay detrás de los chatbots no tienen sentido de la realidad ni idea de lo que es verdad y lo que es mentira. Debemos ser conscientes de que una IA -por muy auténticas que puedan parecer sus respuestas- no tiene conocimiento del mundo que nos rodea ni percepción humana. Por eso, aunque una IA pueda calcular qué palabra es más probable que siga a otra, la propia IA no puede captar el significado del texto resultante. La probabilidad de que el texto sea fiel a los hechos depende de la frecuencia con que una secuencia de palabras aparezca en los datos de entrenamiento. **Recognizing AI generated text:**

Cada vez resulta más difícil identificar un texto generado por IA, especialmente para el usuario medio. Sin embargo, algunas herramientas en línea pueden ayudar a identificar contenidos generados por IA.

Por ejemplo:

GPTZero
<https://gptzero.me/>

AI Detector by Content at Scale
<https://contentatscale.ai/ai-content-detector/>

Enseá uno de estas herramientas y explica cómo funciona:

Basta con copiar el texto en el cuadro de texto o copiar una URL en el campo URL.

Recibirá una estimación de la probabilidad de que este contenido haya sido generado por personas..

Sin embargo, es importante saber que esta estimación no es totalmente fiable. Es sólo una predicción y no es 100% exacta. Puede ayudarte a hacerte una idea de si un texto ha sido creado por la IA, pero nunca debes confiar ciegamente en ella. Estas herramientas también funcionan mejor en inglés que en otros idiomas.

Participant Instructions: Escucha lo que dice el facilitador, toma notas y haz preguntas si algo no está claro.

Posibles alteraciones: Dependiendo de la edad, los conocimientos y la formación de los participantes, se puede ajustar la aportación. En general, al tratarse de una actividad de tipo conferencia, el animador puede llevarla a cabo de la forma que mejor se adapte al grupo. No obstante, debe transmitirse toda la información necesaria para que las actividades siguientes sean viables.

Reflexión: Pregunte a los participantes si tienen alguna duda.

Fuentes bibliográficas:

<https://www.digital-mobil-im-alter.de/wissen/kuenstliche-intelligenz/ki-generierte-inhalte> (en Alemán)

<https://www.makeuseof.com/accurate-ai-text-detectors/>

<https://research.ibm.com/blog/what-is-generative-AI>

-



3.2 Actividad 2: testeando los detectores de texto AI.

Tipo: Actividad en grupo + debate

Duración: 10 minutos

Objetivos de aprendizaje: Los participantes sabrán cómo utilizar las herramientas de detección de texto AI y serán más conscientes de sus limitaciones.

Materiales / recursos:

Un dispositivo por grupo con una cuenta ChatGPT (u otro generador de texto AI)

Ordenador portátil y pantalla grande / proyector para el facilitador

Instrucciones para el facilitador

Divida a los participantes en grupos de 2-3 personas. Cada grupo debe tener un dispositivo con una cuenta ChatGPT.

Al cabo de 6-7 minutos, pide a los participantes que terminen y comenten sus conclusiones con las siguientes preguntas informativas.

Instrucciones para los participantes:

- Inicia sesión en ChatGPT. Crea un artículo sobre un tema de tu elección. Puedes elegir deliberadamente crear un artículo falso, por ejemplo sobre Chemtrails.
- Copia el texto en una de las herramientas del Detector de IA. ¿Con qué precisión reconoce el texto generado por la IA?
- Intenta ajustar el texto original con preguntas como «Reformula el texto para que suene menos como un texto generado por IA» o prueba con diferentes idiomas con preguntas como «Traduce este texto al inglés».
- Copia el texto ajustado en el Detector de IA. ¿Cómo ha cambiado la clasificación?

Posibles alteraciones:

Si los participantes no tienen cuentas de ChatGPT o no quieren crear una, esta actividad también se puede hacer con todo el grupo. Los participantes indican al moderador, que está compartiendo su pantalla, qué mensajes deben escribir. A continuación, el moderador copia el texto en un detector de texto de IA y comenta el resultado con el grupo.

Reflexión: ¿Qué ha observado? ¿Hasta qué punto fueron fiables los detectores en tu prueba? ¿Qué te muestra esto sobre los retos de reconocer contenidos generados por IA?

Fuentes bibliográficas:

<https://www.makeuseof.com/accurate-ai-text-detectors/>

3.3 Actividad 3: Entrada y cuestionario - Imágenes generadas por IA

Tipo: Input y Quiz

Duración: 10 minutos

Objetivos de aprendizaje: Los participantes comprenderán mejor las imágenes generadas por IA y cómo detectarlas.

Materiales/recursos:

- Cuadernos y bolígrafos para los participantes.
- Guión de entrada para el facilitador.
- Pantalla grande o proyector y un ordenador portátil conectado a ella para el facilitador.
- Anexo 3: ImágenesEither show the posts on a large screen (projector/TV) via a laptop and PDF/PPT file.

O preferiblemente Mentimeter Quiz (PC + Proyector/TV, un smartphone / tablet por participante)

1. Crea una encuesta de respuesta múltiple en www.mentimeter.com (o una plataforma similar).
2. Prepara cinco diapositivas y añade las cinco imágenes de ejemplo del Anexo 4, todas con la pregunta «¿Crees que esta imagen es real o generada por IA?».
3. Añade una imagen por diapositiva (puede que no se muestre de forma ideal en la diapositiva debido al formato, pero los participantes la verán correctamente en sus dispositivos). Añade las posibles respuestas «Real» y generada por IA» y marca la casilla de la opción correcta
4. Desactiva «Más puntos por respuestas correctas rápidas».

Instrucciones para el facilitador

Presente brevemente el tema de las imágenes generadas por IA y muestre la lista de control en una pantalla grande.

Reconocimiento de imágenes y vídeos generados por IA

Los contenidos generados por la IA, como imágenes y vídeos, se crean mediante redes neuronales avanzadas entrenadas en amplios conjuntos de datos. Estas redes aprenden a generar nuevos contenidos reconociendo patrones en los datos que reciben. A partir de una descripción o una idea, pueden producir imágenes o vídeos de gran realismo. Sin embargo, a pesar de su creciente sofisticación, todavía hay formas de identificar los contenidos generados por IA buscando pistas específicas.

Indicadores clave para identificar videos generados por IA

- **Anatomía y extremidades:** La IA suele tener problemas para representar con precisión la anatomía humana. En las imágenes, busca manos con demasiados dedos o extremidades con formas extrañas. En los vídeos, busca movimientos corporales poco naturales, ya que los deepfakes pueden producir movimientos mecánicos o repetitivos.
- **Texto y símbolos:** Las imágenes generadas por IA pueden manejar mal el texto, dando lugar a palabras mal escritas, letras borrosas o símbolos extraños. Del mismo modo, en los vídeos, cualquier texto en pantalla que aparezca borroso puede ser un indicio.
- **Cabello y piel:** El pelo humano en las imágenes generadas por IA puede parecer trazos de pintura en lugar de fluir de forma natural. En los vídeos, la piel puede parecer demasiado suave o demasiado arrugada, y las incoherencias entre la piel y otros rasgos como el pelo y los ojos pueden ser indicadores de manipulación.

- **Simetría y coherencia:** La IA puede tener problemas con la simetría. Tanto en imágenes como en vídeos, un lado de la cara puede parecer diferente del otro. Los problemas de coherencia también son una pista; si tienes varias imágenes o fotogramas del mismo sujeto, las incoherencias entre ellos pueden sugerir la generación de IA.
- **Texturas y patrones:** Renderizar texturas, patrones y tejidos realistas es un reto para la IA. Pequeñas discrepancias en la apariencia de las superficies pueden indicar que una imagen ha sido generada por IA. En los vídeos, busca texturas poco naturales en objetos o fondos que parezcan fuera de lugar.
- **Geometría y ángulos:** Comprueba cómo encajan los objetos en la escena. En las imágenes, los ángulos extraños o los elementos que no se alinean de forma natural pueden ser un signo de generación de IA. En los vídeos, busca incoherencias en las relaciones espaciales y las interacciones entre objeto.
- **Rasgos faciales y movimientos:** Los deepfakes suelen implicar transformaciones faciales. Busca texturas de piel poco realistas, indicadores de edad que no coincidan entre los distintos rasgos faciales y sombras o reflejos poco naturales. Preste atención a los patrones de parpadeo, ya que los deepfakes pueden parpadear demasiado o demasiado poco. Examine también los movimientos de los labios para ver si coinciden con las palabras pronunciadas de forma natural.
- **Detalles y refinamientos:** Los contenidos generados por IA pueden a veces pasar por alto detalles más sutiles. En las imágenes, esto puede ser evidente en la forma en que la luz interactúa con las superficies, produciendo luces o sombras poco realistas. En los vídeos, presta atención a los reflejos en las lentes y a pequeños rasgos como lunares o vello facial, que pueden no parecer tan naturales.

Now move on to the quiz:

Cuando utilices una encuesta menti, comparte el código con el grupo. Muestra el modo «Presentar» de mentimeter.com en la pantalla grande. Pide a los participantes que seleccionen una respuesta.

Cuando todos hayan votado, los resultados aparecerán en la pantalla principal.

Si no utilizas una encuesta mentimeter, muestra los mensajes uno a uno en una pantalla grande y deja que los participantes voten si creen que una imagen es real o generada por IA, por ejemplo, levantando la mano.

Después de cada imagen, comenta brevemente cuáles son los indicadores de las imágenes generadas por IA.

Aquí está la solución al cuestionario y algunos puntos a tener en cuenta:

1: El Canciller alemán Olaf Scholz : Rea

- El canciller alemán Olaf Scholz sufrió un accidente mientras corría por un bosque en septiembre de 2023. Se cayó y se lesionó la cara. Para cubrirse el ojo herido, se puso un parche y publicó esta foto en su página X
- Fuente: <https://x.com/Bundeskanzler/status/1698627434508063228>
- Crédito de la foto: Steffen Kugler/Gobierno Federa

2: Donald Trump: Generado por IA

- Esta imagen fue creada por partidarios de Donald Trump
- Verificación: <https://www.bbc.com/news/world-us-canada-68440150>
- Indicadores de que es fake:
 - o A la gente le faltan dedos
 - o El espacio entre dos dedos de la mano izquierda de Trump parece poco realista
 - o El texto del sombrero de la izquierda está borroso y es ilegible
 - o Textura de la piel: demasiado brillante

3: Katy Perry: Generado por IA

- Esta imagen circuló durante la Gala MET 2024. Sin embargo, Katy Perry ni siquiera estaba presente
- Verificación: <https://www.politifact.com/factchecks/2024/may/07/facebook-posts/ai-generated-images-of-katy-perry-at-2024-met-gala/>
- Indicadores de que es fake
 - En el fondo, los rostros, las manos, los brazos y los cuerpos de los fotógrafos no están claramente separados entre sí ni del entorno, sino que fluyen unos dentro de otros.

4: Giorgia Meloni: Real

- Esta foto fue tomada el 9 de febrero de 2023 en el Palacio del Elíseo
- Fuente: <https://www.faz.net/aktuell/politik/ausland/darum-aergert-sich-giorgia-meloni-ueber-emmanuel-macron-18672739.html>
- Crédito de la foto: EPA

5: Tom Cruise: Fake

- Verificación: <https://www.snopes.com/fact-check/tom-cruise-stunt-doubles/>
- Indicadores de que es fake
 - A la mano del supuesto doble de Tom Cruise en el centro parecen faltarle las uñas de los dedos
 - La textura de la piel es demasiado suave, brillante y parecida a la cera

Después del cuestionario, dar más consejos y aportaciones sobre el reconocimiento de los contenidos generados por IA.

Técnicas avanzadas de detección

Software de detección basado en IA: Se están desarrollando varias herramientas para ayudar a detectar contenidos generados por IA. Estos programas analizan incoherencias en imágenes y vídeos que podrían no ser visibles para el ojo humano. Pueden ser especialmente eficaces en la identificación de deepfakes examinando detalles minúsculos y comparándolos con patrones conocidos de manipulación.

Para imágenes e.g.:

Hive Detection Tool:

<https://hivemoderation.com/ai-generated-content-detection>

Para videos, e.g.:

[Deepware Scanner](#) or [DeepFake-o-meter](#).

- **Análisis humano** A pesar de los avances en el software de detección de IA, la intuición y el análisis humanos siguen siendo cruciales.
- **Sentido común:** Lo mismo se aplica a toda la información de Internet. Haz una pausa, mantente escéptico y crítico y cuestiona lo que ves y oyes. ¿Qué motivos puede haber para poner algo así en boca de ciertas personas? ¿Hasta qué punto es realista el escenario?

Instrucciones para los participantes

Abre menti.com en tus dispositivos portátiles. Introduce el código que te dé el facilitador. Mira la imagen que aparece en tu dispositivo y elige si crees que es real o generada por IA

Posibles alteraciones Los animadores también pueden utilizar sus propios ejemplos para el cuestionario, si prefieren utilizar imágenes diferentes.

Reflexión: ¿Te ha resultado fácil o difícil? ¿Qué es lo que más te ha gustado de esta actividad?

Fuentes bibliográficas:

<https://www.theguardian.com/technology/2024/apr/08/how-to-tell-if-an-image-is-ai-generated>

3.4 Actividad 4. Batalla de enunciados. ¿Quién genera la imagen IA más realista?

Tipo: Juego / concurso de creatividad

Duración: 15 minutos

Objetivos de aprendizaje: Los participantes comprenderán mejor cómo se crean las imágenes de IA y cómo los enunciados que se dan influyen en el resultado.

Materiales/ recursos

- Cuadernos y bolígrafos para los participantes.
- Guión de entrada para el facilitador.
- Pantalla grande o proyector y un ordenador portátil conectado a él para el facilitador..

Instrucciones para el facilitador:

Divida el grupo en dos equipos. Utilizando el software gratuito Copilot Designer de Microsoft (o un software de pago, como Dall-E o Midjourney), competirán entre sí para generar la imagen de IA más realista sobre un tema determinado.

Jugarán tres rondas. En cada ronda, pueden elegir o dibujar un tema de la siguiente lista o inventar su propio tema. A continuación, cada grupo dispone de 3 a 5 minutos para crear una indicación detallada para obtener una imagen realista. Ten cuidado de que los equipos no oigan ni vean las indicaciones del otro equipo. Una vez introducido el texto, el software creará 4 imágenes similares, entre las que los equipos elegirán su favorita. Cuando se acabe el tiempo, ambos grupos presentan los resultados. Todo el grupo puede ver las imágenes y decidir cuál es la ganadora (dependiendo de qué imagen parezca más realista). Después de tres rondas, habrá un equipo ganador que haya sido capaz de crear las indicaciones más realistas.

Lista de temas:

- una escena de una acción de protesta de activistas climáticos
- un anuncio de perfume con una modelo en la playa
- Un político en campaña electoral, hablando ante una gran multitud (tenga en cuenta que muchos generadores de IA no recrean personalidades reales, por ejemplo Donald Trump, para protegerse contra la desinformación)
- Un incendio devastador
- Paparazzi persiguiendo a famosos

Consejos para crear un buen enunciado (pueden presentarse con antelación)

- Describe con la mayor precisión posible lo que quieres ver en la foto (sujeto en primer plano, fondo, detalles, qué está dispuesto cómo...) Indica el estilo de la imagen: fotorrealista, pintura, blanco y negro, etc. Context: Provide background information, data or context for the precise creation of content

- Palabras clave: Nombra las palabras clave o frases importantes que deben incluirse.
- Ejemplos: Proporciona ejemplos del estilo, la estructura o el contenido deseados
- Utiliza efectos de iluminación y estados de ánimo para dar a tu imagen un estilo determinado, por ejemplo, iluminación cinematográfica, iluminación navideña, o descripciones como sombrío, alegre, brillante, borroso, etc.

Importante: actualmente las indicaciones funcionan aún mejor en inglés

Reflexión: En un debate posterior, pueden fijarse en las distintas indicaciones utilizadas e intentar

detectar errores en las imágenes y señales de que han sido generadas por IA.

Al igual que el juego Get Bad News, esta actividad cambia de perspectiva. Al ponerse en el papel de crear una imagen engañosamente real con IA, aprenden lo rápido y fácil que es hoy en día crear imágenes de aspecto realista con las indicaciones adecuadas. Practican el reconocimiento de errores en sus propias imágenes que son indicativos del producto de la IA. Se trata de un enfoque entretenido y lúdico para sensibilizar a los más jóvenes para que observen con detenimiento las imágenes de Internet..

Posibles alteraciones: Los facilitadores también pueden utilizar sus propios ejemplos de vídeos deepfake, si prefieren utilizar otros diferentes. Los animadores también pueden utilizar software de pago, como Midjourney o Dall-E. En lugar de jugar en dos equipos, también puede haber sólo dos personas compitiendo. En este caso, en cada ronda cambian las personas que compiten. Además de los temas propuestos, también es posible elegir otros temas o ideas para las imágenes de IA.:

