iENDGAME

Escapando de la nueva desinformación a través de la educación gamificada transfronteriza en educación mediática

Cofinanciado por la Unión Europea. Las opiniones expresadas son exclusivamente las del autor o autores y no reflejan necesariamente las de la Unión Europea ni las de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de ellas.

Co-funded by the European Union

Título

Materiales didácticos sobre alfabetización mediática que acompañan a la escape room 2 "ECHO".

Licencia

Este documento está bajo una licencia Creative Commons Atribución 4.0 Internacional.



Este documento se ha elaborado en el marco del proyecto **ENDGAME**, cofinanciado por la Unión Europea a través del instrumento de financiación Europa Creativa.

Referencia del proyecto: 101185763
Sitio web del proyecto: https://endgameproject.github.io/

Las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea.

Co-funded by the European Union

1. ¿Qué es ENDGAME? **4**
2. Antecedentes y objetivos generales **4**
2.1. La inteligencia artificial y su papel en las redes sociales 4
2.2. moderación de contenido 5
2.3. Desinformación/información errónea sobre el cambio climático 8
3. La escape room ECHO **10**
3.1. Preparatorio sesión 10
3.2. Cierre actividad 16
4. Guía paso a paso para la escape room **17**
4.1. Recomendaciones de juegos 17
4.2. Instrucciones del juego 17

Co-funded by
the European Union

1. ¿Qué es ENDGAME?

ENDGAME: Escaping new disinformation through gamified cross-border media literacy education es un proyecto educativo europeo con un grupo diverso de socios, entre los que se incluyen la Universidad del Este de Finlandia (UEF), la Fundación Maldita.es, la Universidad Politécnica de Madrid (UPM), la Universidad Metropolitana de Belgrado (BMU), Faktabaari y Fake News Tragač (FNT).

El objetivo principal de esta iniciativa es **fomentar la alfabetización mediática entre los jóvenes** de Finlandia, Serbia y España, ayudándoles a desarrollar **el pensamiento crítico** y a construir una ciudadanía digital responsable. Para ello, se ha creado una serie de juegos **interactivos tipo escape room** en los que los participantes deben resolver una serie de desafíos en un escenario ficticio en un tiempo limitado.

Mediante estos juegos educativos, que simulan situaciones de la vida real en las que consumimos información, los participantes aprenderán a **distinguir entre desinformación** y contenido manipulado, a reconocer la información generada por **inteligencia artificial** y **a comprender** los **riesgos de compartir datos personales** en las redes sociales.

2. Antecedentes y objetivos generales

2.1. La inteligencia artificial y su papel en las redes sociales

Hoy en día, la distinción entre contenido creado por humanos y generado por máquinas se ha vuelto casi imposible. **La Inteligencia Artificial** (IA) ha democratizado la creación de contenido, pero también ha proporcionado una poderosa herramienta para **la producción masiva de desinformación**. Algunos ejemplos son ***los deepfakes***, grabaciones de audio o vídeo realistas de personas diciendo o haciendo cosas que nunca ocurrieron, y **las redes de bots de IA** que utilizan modelos de lenguaje para participar en discusiones similares a las humanas en las secciones de comentarios, creando así una falsa sensación de consenso público. Dado que estas herramientas son económicas y rápidas, el volumen de desinformación aumenta a un ritmo que los verificadores humanos no pueden igualar, lo que genera un entorno informativo "contaminado". El principal motor de gran parte de esta transformación es un tipo específico de IA conocido como **Modelo de Lenguaje de Gran Tamaño**. Comprender cómo funcionan estos modelos es el primer paso para entender por qué el contenido generado por IA puede resultar tan convincente.

Un modelo de lenguaje de gran tamaño (LLM, por sus siglas en inglés) es un tipo de IA entrenada con enormes cantidades de datos textuales para comprender, resumir y generar lenguaje similar al humano. Se podría considerar como un sistema de autocompletado muy avanzado. No "conoce" los hechos como lo hace un humano, sino que **predice la palabra más probable que**

sigue en una oración basándose en los patrones aprendidos durante el entrenamiento. Los LLM están diseñados para ser útiles y conversacionales: son muy buenos explicando temas complejos o generando ideas. Sin embargo, también pueden "**alucinar**", es decir, afirmar con seguridad algo como un hecho cuando en realidad es una combinación de patrones que nunca ocurrieron en la realidad.

Mira el vídeo: https://endgameproject.github.io/ECHO/assets/video_IA_es.mp4

El problema va más allá de las alucinaciones generadas por los LLM o incluso del contenido que estos generan con el objetivo de difundir desinformación. Muchos estudiantes han comenzado a usar herramientas basadas en LLM, como chatbots, como **sustituto directo de los motores de búsqueda tradicionales** u otros métodos de recuperación de información. Sin embargo, los estudiantes deben comprender que la mayoría de los LLM no se entrenan en tiempo real. A diferencia de un motor de búsqueda que rastrea la web en tiempo real para encontrar la información más reciente, un LLM se basa en un conjunto de datos de entrenamiento con una **fecha límite específica.** Esto significa que, en el caso de noticias de última hora, avances científicos recientes o acontecimientos políticos en desarrollo, la IA puede proporcionar información desactualizada o llenar los vacíos con alucinaciones que suenan plausibles.

Además, mientras que un motor de búsqueda dirige al usuario a un sitio web externo para su verificación, un LLM suele presentar la información como un **hecho definitivo y sin fuente**. Asimismo, los motores de búsqueda permiten a los estudiantes **evaluar la procedencia de una afirmación** distinguiendo entre un estudio científico y un blog personal. Los LLM a menudo combinan estas fuentes en una sola respuesta, lo que imposibilita que el estudiante juzgue la credibilidad de los datos subyacentes. Esta **naturaleza opaca** de la IA dificulta que los estudiantes verifiquen la lógica o identifiquen sesgos específicos.

2.2. Moderación de contenido

¿Qué es la moderación de contenido?

En esencia, la moderación de contenido es el proceso mediante el cual las plataformas de redes sociales supervisan, revisan y aplican un conjunto de reglas (normas de la comunidad) al contenido generado por los usuarios. Es un constante equilibrio entre dos valores contrapuestos:

- **Seguridad:** Proteger a los usuarios de daños, acoso y explotación.
- **Expresión:** Defender el principio de la libertad de expresión y el diálogo abierto.

La moderación no se trata solo de **eliminar publicaciones** , sino más bien de decidir qué tipo de "**espacio digital**" quiere ser una plataforma.

Contenido ilegal frente a contenido desinformativo

Es un error común pensar que todo lo "malo" en internet es ilegal. Para los fines del juego, es vital distinguir entre estas dos categorías:

	Contenido ilegal	**Contenido de desinformación**
Definición	Contenido que infringe leyes específicas (por ejemplo, seguridad infantil, incitación a la violencia, robo de derechos de autor).	Contenidos falsos o engañosos difundidos con la **intención** de engañar o causar daño.
Acciones del moderador	Las plataformas suelen estar obligadas por ley a eliminar este contenido y, en muchos casos, a denunciarlo a las autoridades. La plataforma puede ser considerada responsable por incumplimiento de las obligaciones legales.	A menudo se encuentra en una "zona gris". Puede que no sea ilegal, pero podría infringir las políticas de la plataforma, al menos en teoría. El moderador podría responder o no.
Ejemplos	Discursos de odio (en ciertas regiones), reclutamiento terrorista o extremista, fraude y estafas.	Vídeos "deepfake" generados por IA de políticos, supuestas curas para la salud o noticias de última hora escenificadas.

Conclusión clave: La desinformación suele ser «legal pero perjudicial»; imita la verdad para manipular las emociones, lo que dificulta su detección por parte de los sistemas automatizados en comparación con simples palabras clave ilegales. Y si bien las políticas protegen a los usuarios en teoría, no siempre se ponen en práctica: la moderación consume tiempo y dinero, y, hasta cierto punto, la lógica empresarial de las plataformas crea incentivos para que permitan contenido engañoso y que provoque reacciones emocionales.

Bots y trolls

La desinformación rara vez viaja sola; generalmente es difundida por actores específicos. Uno puede encontrarse con "usuarios" que no son lo que parecen.

- **Bots (Robots)** : Abreviatura de "robots web", son programas automatizados diseñados para realizar tareas repetitivas. En la desinformación, los bots se utilizan para dar "me gusta" o "compartir" una publicación miles de veces en segundos, engañando al algoritmo para que crea que un tema es tendencia. Según NATO Stratcomm , con una inversión de tan solo 10 € se pueden comprar, por ejemplo, más de 156 000 visualizaciones falsas en X (en 2025).

- **Trolls (los humanos)** : A diferencia de los bots, los trolls son personas reales. Su objetivo es provocar, molestar o distraer. Utilizan desencadenantes emocionales para iniciar discusiones, ya que estas disputas generan mucha más interacción y visualizaciones en las plataformas digitales.
- **La actualización de IA:** La IA generativa permite que un solo "troll" opere cientos de cuentas "bot", cada una de las cuales suena como un humano único y convincente.

Opciones disponibles para la moderación de contenido

Las plataformas no siempre utilizan un botón de "eliminar". La moderación eficaz utiliza una "escala gradual" de intervenciones:

- **Eliminación:** El contenido se borra y el usuario puede ser bloqueado. Esto suele reservarse para las infracciones más graves.
- **Etiquetado/Verificación de datos:** La publicación permanece visible, pero se agrega una superposición (por ejemplo, "Falta contexto" o "Manténgase informado: Este video ha sido alterado digitalmente").
- **Bloqueo en la sombra (alcance reducido)** : El algoritmo deja de mostrar el contenido a nuevos usuarios. El usuario aún puede publicar, pero básicamente está hablando a ciegas.
- **Degradación:** Consiste en relegar el contenido al final del feed del usuario para que sea poco probable que lo vea.
- **Intervención humana:** Si bien la IA busca patrones, a menudo se recurre a moderadores humanos para realizar comprobaciones de "matices", decidiendo si una publicación es una mentira peligrosa o simplemente una broma inofensiva.

En el contexto de la desinformación, generalmente se recomienda no eliminarla salvo en casos ilegales, ya que existen formas más efectivas de reducir su impacto. Diversos estudios sobre este tema indican que acciones como las etiquetas de verificación de datos ayudan a los usuarios a identificar contenido falso y, por lo tanto, a reducir su difusión sin necesidad de eliminarlo por completo.

2.3. Desinformación/información errónea sobre el cambio climático

En cuanto a los temas que se tergiversan en las redes sociales, aunque existen muchas narrativas sobre las que se difunde contenido falso, la segunda escape room se centrará en **la desinformación y la información errónea sobre el clima** .

La desinformación sobre el cambio climático ha evolucionado. Hemos pasado de la simple "negación científica" (la afirmación de que el clima no está cambiando) a **formas más sofisticadas, a menudo potenciadas por la IA, de dilación climática**. Esto incluye atacar la viabilidad de las energías renovables, desacreditar a los científicos del clima y promover el "catastrofismo" para desalentar la acción.

Las narrativas de desinformación sobre el cambio climático se utilizan para:

- **Polarizar las sociedades**: Enfrentar a las poblaciones urbanas con las rurales o a las generaciones más jóvenes con las mayores.
- **Erosionar la confianza**: Socavar la confianza en las instituciones científicas y los gobiernos democráticos.
- **Sabotaje económico**: Difundir afirmaciones falsas sobre el coste de las transiciones ecológicas para proteger los intereses de la industria de los combustibles fósiles o crear inestabilidad política interna en naciones rivales.

Al hablar de desinformación, es importante destacar que suele ser una herramienta estratégica utilizada para obtener beneficios políticos o económicos, y no simplemente una mala interpretación de los hechos. Enseñar sobre desinformación requiere un cambio de enfoque: pasar de «decirles a los estudiantes qué es verdad» a «enseñarles a pensar».

Hay investigaciones que sugieren que la llamada **Teoría de la Inoculación** (o **desmitificación previa**) es más eficaz que **la desmitificación** (o **corrección posterior de las falsedades**). Esto implica exponer a los estudiantes a una versión atenuada de una técnica manipuladora para que puedan reconocerla más adelante.

- **Identifica la técnica:** en lugar de centrarte en un hecho falso específico, céntrate en la lógica (por ejemplo, el uso de "falsos expertos") para exponer la estrategia de manipulación que se ha utilizado.
- **El efecto contraproducente:** Evita simplemente repetir un mito. Si debes mencionar uno, comienza con la verdad, explica la manipulación y termina con la verdad nuevamente (el sándwich de la verdad).

La desinformación se nutre de emociones intensas: ira, miedo o indignación. **El principal motor de su propagación es la activación de emociones fuertes y reactivas ante afirmaciones escandalosas.** Por eso, aprender a detectar nuestras emociones es el primer paso para enseñar a evaluar críticamente la información, y por eso es crucial enseñar a los adolescentes a detenerse cuando experimentan una fuerte reacción emocional ante un titular. Esa reacción suele ser el "gancho" que utilizan los manipuladores y puede reforzarse con la repetición, por lo que la educación también debe intervenir en este sentido.

Es recomendable proporcionar a tus alumnos estas herramientas prácticas para que las utilicen cuando se encuentren con afirmaciones sobre el cambio climático en TikTok, Instagram o en los canales de noticias:

El método SIFT:

(S)top- Detenerse: Controla tus emociones. ¿Quieres que esto sea cierto?
(I)nvestigate the source - Investigar la fuente: Haz clic en el perfil. ¿Es un científico climático, un grupo de presión con una agenda o una cuenta cualquiera?
(F)ind better coverage - Buscar Mejor Cobertura: Comprueba si organizaciones científicas de renombre (como la NASA, el IPCC o la NOAA) informan sobre lo mismo.
(T)race the claim to its original context - Rastrear la afirmación hasta su contexto original: ¿La publicación resumió correctamente un estudio o sacó una frase de contexto?

También sería beneficioso analizar algunas señales de alerta comunes que se deben tener en cuenta en los debates sobre el clima:

- **Selección sesgada de datos**: Mostrar un gráfico de una sola semana fría para "demostrar" que el calentamiento global se ha detenido, ignorando la tendencia ascendente de los últimos 100 años.
- **El "falso equilibrio"** : Los medios de comunicación dan el mismo tiempo de emisión a un teórico de la conspiración marginal y a un científico climático, como si su experiencia fuera equivalente.
- **Expectativas imposibles:** Exigir que los modelos climáticos sean 100% perfectos, de lo contrario, son "inútiles". (La ciencia se basa en probabilidades, no en la perfección).
- **Ataques ad hominem**: Atacar la apariencia o el estilo de vida de un activista climático (como Greta Thunberg) en lugar de abordar los datos científicos que presenta.

Más adelante presentaremos actividades más específicas que se pueden utilizar en relación con este juego de Escape Room. Pero, como ejemplo de una actividad adicional, práctica y generalmente muy educativa, se puede pedir a los estudiantes que trabajen en pequeños grupos para "diseñar" ellos mismos una pieza de desinformación climática. Deben elegir un público objetivo (por ejemplo, adolescentes), un objetivo (por ejemplo, lograr que dejen de creer en los coches eléctricos) y una técnica (por ejemplo, seleccionar datos de forma sesgada). Al comprender cómo se realiza el "truco de magia", podrán comprender mejor la lógica de la desinformación y serán menos propensos a ser engañados por ella en el futuro.

3. La escape room ECHO

"ECHO" es la segunda de las escape rooms desarrolladas dentro del proyecto ENDGAME. En esta escape room, el jugador asume el papel de un **moderador de redes sociales** encargado de detener una ola de desinformación sobre el clima. El jugador debe detectar **cuentas bot**, identificar **contenido generado por IA** e intentar que **los usuarios hagan un uso responsable de la IA**. Los **objetivos de aprendizaje** de la escape room son los siguientes:

- Fomentar **el pensamiento crítico** entre los estudiantes.

- Reflexiona sobre cómo usamos **las redes sociales** y el grado de **veracidad** que atribuimos al contenido que vemos en ellas.
- Comprender **cómo funciona un modelo de lenguaje a gran escala** y cómo se utiliza en las redes sociales.
- Evaluar **los usos recomendados de la IA** para generar información y contenido.
- Detecta patrones típicos de **cuentas bot.**

Para garantizar que la escape room sea una experiencia de aprendizaje y no solo un juego, recomendamos complementarla con **actividades preparatorias en una o más sesiones**. Estas sesiones ayudan a los participantes a comprender mejor el tema, reflexionar sobre sus propios hábitos y desarrollar una perspectiva más crítica sobre las redes sociales.

Finalmente, sugerimos una **sesión de reflexión** posterior a la partida. Esta sesión brinda la oportunidad de conversar sobre la experiencia, evaluar lo aprendido y reforzar los conocimientos adquiridos, creando así una experiencia de aprendizaje más completa y significativa.

3.1. Sesión preparatoria

Para ayudar a los estudiantes a comprender mejor la narrativa de la escape room y a involucrarse más en el juego, recomendamos realizar una sesión preparatoria para introducir los conceptos de desinformación, bots e inteligencia artificial. A continuación, presentamos varias ideas para poner en práctica en el aula, que puede adaptar a sus necesidades:

PREGUNTAS DE ACTIVACIÓN

Puedes comenzar la sesión introduciendo el tema mediante preguntas sobre su consumo de noticias y redes sociales:

1. **Acerca de la IA y cómo funciona**
    a. ¿Cómo crees que la IA "sabe" qué escribir? ¿Entiende lo que dice o hay algo más en juego?
    b. Si le haces la misma pregunta a una IA dos veces, ¿debería dar la misma respuesta? ¿Por qué sí o por qué no?
    c. ¿Puede la IA inventarse cosas? ¿Cómo lo sabrías?
2. **Acerca de la IA y las redes sociales**
    a. ¿Cómo podría alguien utilizar la IA para difundir contenidos falsos en las redes sociales?
    b. Si una publicación en redes sociales suena profesional y está bien escrita, ¿significa eso que es fiable?

c. ¿Crees que es posible distinguir entre una publicación escrita por un humano y una escrita por una IA? ¿Qué pistas buscarías?

3. **Acerca de la información confiable**
    a. ¿Alguna vez te has creído algo que viste en internet y que resultó ser falso? ¿Qué te pareció convincente?
    b. Si una IA te diera una respuesta sobre medicina, el tiempo o una tarea, ¿confiarías en ella sin comprobarla? ¿Dónde la verificarías?
    c. ¿Qué significa realmente una cuenta "verificada" en las redes sociales? ¿Garantiza que la cuenta la gestiona una persona real?
4. **Acerca de los bots y la manipulación**
    a. ¿Cómo distinguirías la cuenta de una persona real de la de un bot? ¿Qué patrones podrían delatarla?
    b. ¿Por qué alguien crearía cuentas falsas en las redes sociales? ¿Qué ganarían con ello?
    c. ¿Has visto publicaciones en internet que parecen diseñadas para provocarte ira o miedo? ¿Por qué alguien haría eso a propósito?
5. **Reflexión / Opinión**
    a. ¿Deberían las plataformas de redes sociales usar inteligencia artificial para moderar el contenido, o deberían hacerlo los humanos? ¿Puede la IA por sí sola asumir esa responsabilidad?
    b. ¿Crees que la mayoría de las personas de tu edad pueden detectar la desinformación en internet? ¿Qué lo dificulta?

ACTIVIDAD:

La siguiente actividad puede complementar los temas tratados y reforzar el aspecto educativo del juego ER, destacando cómo se puede utilizar la IA generativa para manipular y engañar, y cómo se lleva a cabo esto, especialmente en relación con las políticas climáticas.

Imágenes generadas por IA y respuestas emocionales (5 minutos) **:** Muestra a la clase algunas imágenes generadas por IA y pide a los estudiantes que presten atención a sus reacciones iniciales, así como a lo que sucede cuando reflexionan un poco más. Puede usar, por ejemplo, los siguientes ejemplos como guía, pero siéntase libre de crear sus propios ejemplos:

a. Una foto realista generada por IA de un oso polar en la selva.
b. Un “reportaje” de 2026 sobre una ciudad descubierta bajo el hielo antártico en deshielo.
c. Un gráfico (falso) generado por IA sobre una supuesta caída drástica de las temperaturas globales en 2025.

Pregunta a los estudiantes (10 minutos) **:** "¿Cuál de estas opciones les pareció incorrecta de inmediato?" y analicen cómo la desinformación, especialmente la generada por IA, suele activar

nuestro **reflejo de sobresalto emocional** , desactivando momentáneamente nuestra capacidad de análisis crítico. A continuación, se debe debatir sobre las estrategias de verificación de contenido y las diferencias entre una evaluación rápida y descuidada y una evaluación lenta y minuciosa.

A continuación, debatid en pequeños grupos (durante 5 minutos) qué tipos de desinformación han encontrado en las redes sociales con el objetivo de impactar, indignar o preocupar. Reunid ejemplos y analizad **cómo** funcionan y provocan la reacción emocional que generan: ¿a quiénes van dirigidos?

Luego, amplía la discusión (5 minutos) **:** ¿qué motivos pueden llevar a difundir desinformación? Más específicamente, ¿por qué alguien difundiría desinformación sobre, por ejemplo, el cambio climático? Si la discusión no fluye, puedes utilizar las siguientes preguntas:

1. ¿Alguien se beneficia de ellos financiera o económicamente? Por ejemplo, personas influyentes, industrias, etc.
2. ¿Alguna facción política se beneficia políticamente de ellas? ¿Resulta útil para alguien el estancamiento de las políticas climáticas?
3. Es recomendable introducir y analizar aquí los conceptos de «**polarización**» y «**radicalización**». En resumen, se trata de cómo se utilizan las redes sociales para inducir deliberadamente a las personas a adoptar ideologías extremas o violentas, y cómo esta división de la sociedad en facciones contradictorias suele formar parte de una estrategia extranjera para canalizar recursos hacia la lucha interna. ¿Cómo se beneficiaría una potencia extranjera de enturbiar los debates sobre el clima en nuestro país? Es decir, ¿cómo influyen la polarización y la radicalización en nuestra toma de decisiones políticas?

Para concluir (10 minutos), se analizará y evaluará la posibilidad de que el contenido que ven en las redes sociales esté influenciado por este tipo de motivos, y la probabilidad de que sea generado por inteligencia artificial.

BUSCAR EJEMPLOS

Si puedes dedicar varias sesiones a abordar estos temas, aquí tienes una idea para una actividad de aprendizaje más basada en la investigación. Una buena idea es pedirles que realicen algunas búsquedas sobre diversos temas climáticos en sus redes sociales para "sembrar" sus algoritmos y ver qué tipo de contenido reciben.

Probablemente lo mejor sea utilizar el llamado " **sándwich de la verdad"** para esta **actividad previa a la desinformación** .

1. Comencemos por establecer lo que actualmente se conoce como hechos sobre el cambio climático,
2. Luego, proceda a buscar ejemplos de desinformación y a analizarlos, así como los motivos que los impulsan, y
3. Por último, volvamos a recordar que esos fueron ejemplos de desinformación, y de nuevo: ¿cuáles son los hechos verificables, conocidos y científicos?

Explícales claramente a los estudiantes que usar cualquier término de búsqueda hará que el algoritmo de recomendaciones de cualquier plataforma comience a sugerir contenido, y parte de él es directamente falso. Adviérteles que el algoritmo podría seguir sugiriendo ese tipo de contenido. La idea es recopilar ejemplos de desinformación, preferiblemente contenido generado por IA, y junto con la clase discutir y analizar cada ejemplo: ¿son convincentes? ¿Cómo convencen? ¿Se pueden verificar las afirmaciones y, de ser así, cómo?

Sería conveniente realizar una **lectura complementaria** e investigar los antecedentes de las cuentas de redes sociales junto con los estudiantes, intentando identificar quiénes están detrás de ellas, qué motivaciones tienen (por ejemplo, ¿por qué alguien querría socavar la credibilidad de la investigación científica?) y si las afirmaciones presentadas pueden verificarse con fuentes fiables e independientes. **El método SIFT,** presentado anteriormente, es idóneo para esta actividad.

Como docente, puede mostrar ejemplos de cómo buscar información verificable y animar a los estudiantes a hacer lo mismo, por ejemplo, en pequeños grupos.

Algunos posibles términos de búsqueda que pueden producir resultados de búsqueda muy dudosos para esta actividad podrían ser, por ejemplo:

- ¿El cambio climático es causado por el ser humano?
- ¿Es el cambio climático un engaño?
- ¿Miente la ciencia sobre el cambio climático?
- ¿La Tierra se está enfriando en lugar de calentarse?

Después de analizar estos ejemplos de desinformación, recuérdeles a los estudiantes qué piensa la comunidad científica sobre las cuestiones climáticas y por qué la comunidad científica es una fuente más confiable que las cuentas persuasivas de las redes sociales o la IA.

Podría ser útil introducir el concepto de **Consenso científico** para los estudiantes. Científicos y organizaciones científicas de todo el mundo generan enormes cantidades de datos utilizando métodos científicos igualmente numerosos y rigurosos. Si la mayoría de los datos (empleando la mayoría de los métodos científicos) apunta en una dirección determinada y respalda una hipótesis concreta, se habla de "consenso científico". No se trata simplemente de un acuerdo o

una opinión, sino de algo que se verifica y surge de los datos que recopilan los científicos. Esta sería también una buena oportunidad para debatir con los estudiantes la diferencia entre opiniones y método científico.

En caso de disponer de más de una sesión para abordar estos temas, aquí les proponemos una actividad de aprendizaje más basada en la investigación. Una buena idea es proporcionar a los estudiantes una tabla con los logotipos de agencias meteorológicas y organizaciones climáticas reconocidas, y pedirles que investiguen qué son estas instituciones, qué hacen y por qué se consideran fuentes de información fiables.

Probablemente lo mejor sea estructurar esta actividad en torno a tres etapas sencillas:

> Para empezar, **analiza junto a los estudiantes de dónde suele provenir la información sobre el tiempo, el cambio climático, los fenómenos extremos y los riesgos ambientales**, y si los estudiantes conocen alguna organización responsable de producir esa información.
>
> A continuación, procede a **investigar las organizaciones en sí, analizando su función, experiencia, financiación y el tipo de información que publican.**
>
> Por último, volvamos a la cuestión más amplia de la confianza y la fiabilidad: **¿por qué son importantes estas instituciones, qué significa el consenso científico y por qué deben contrastarse las afirmaciones sobre el clima con fuentes expertas?**

Explícales claramente a los estudiantes que no todas las fuentes en línea son igualmente fiables y que los temas relacionados con el clima suelen difundirse a través de artículos de opinión, publicaciones engañosas o afirmaciones sin fundamento. El objetivo de esta actividad es ayudar a los estudiantes a reconocer las instituciones cuya función es recopilar evidencia, monitorear la atmósfera, analizar datos y comunicar los resultados de forma transparente.

Los estudiantes pueden trabajar individualmente, en parejas o en grupos pequeños. A cada grupo se le puede asignar una o más organizaciones para investigar y luego presentar un informe a la clase. Pueden completar una tabla utilizando preguntas orientadoras como:

- ¿Qué es?
- ¿Qué hace?
- ¿Cómo se financia?
- ¿Quién trabaja allí?
- ¿Qué tipo de datos o informes publica?
- ¿Por qué podría considerarse fiable su información?

Sería conveniente fomentar el uso del método SIFT durante la actividad. Los estudiantes no deberían basarse únicamente en una página web, sino comparar la información de sitios web oficiales, instituciones públicas, universidades y fuentes científicas reconocidas. Esta es también

una buena oportunidad para debatir cómo la transparencia, la revisión por pares, la cooperación internacional y los datos abiertos contribuyen a la confianza en la ciencia.

Como docente, puedes mostrar cómo investigar una organización, enseñando a identificar sitios web oficiales, información sobre financiación e informes publicados. Luego, anima a los estudiantes a hacer lo mismo en grupos.

Algunas organizaciones que podrían incluirse en la tabla son:

- NASA
- Panel Intergubernamental sobre el Cambio Climático
- Programa Copérnico
- Mapeador de carbono
- Organización Meteorológica Mundial
- Agencia meteorológica nacional de cada país (adaptar según el contexto)

Para concluir, reúnan a la clase y discutan cuestiones más generales:

- ¿Por qué las sociedades necesitan organizaciones científicas especializadas?
- ¿Por qué la ciencia del clima es internacional?
- ¿Cuál es la diferencia entre la predicción meteorológica y la investigación climática?
- ¿Por qué importa el consenso científico a la hora de tomar decisiones públicas?

Esta actividad ayuda a desarrollar la alfabetización mediática, las habilidades de evaluación de fuentes, la competencia en investigación y una comprensión más clara de cómo se produce y comunica el conocimiento científico sobre el cambio climático.

3.2. Actividad de cierre

Una vez que los alumnos hayan jugado al juego de escape, puedes dedicar unos minutos a hacerles preguntas que fomenten la reflexión y generen debate, y así descubrir qué han aprendido. Aquí tienes algunas sugerencias que pueden resultarte útiles:

Acerca de la detección de bots (Reto 1)

1. ¿Qué cuenta fue la más difícil de clasificar como bot o humana? ¿Qué te hizo dudar?
2. ¿Qué pistas te ayudaron más a identificar un bot? ¿Notarías esas mismas pistas en tus redes sociales reales?
3. ¿Te sorprendió algo del comportamiento de los bots? ¿Tienen un aspecto diferente al que esperabas?

Acerca del contenido generado por IA (Reto 2)

4. Al completar el texto palabra por palabra, ¿la IA siempre elegía la "mejor" palabra? ¿Qué te indica esto sobre cómo escribe la IA?

5. Antes de jugar, ¿creías que la IA entendía lo que escribía? ¿Ha cambiado tu opinión?

6. Si la IA genera texto basándose en probabilidades en lugar de en la comprensión, ¿cuáles son los riesgos de confiar en ella ciegamente?

Sobre el uso correcto de la IA (Reto 3)

7. ¿Cuál de los usos no recomendados de la IA te sorprendió más? ¿Pensabas que la IA podría hacer eso de forma fiable?

8. ¿En qué situaciones es apropiado usar la IA y cuándo se debe tener precaución? ¿Se te ocurren ejemplos de tu vida diaria?

9. Si una IA te dijera algo sobre tu salud, el clima o tus tareas escolares, ¿qué harías antes de actuar en consecuencia?

Sobre la desinformación y la manipulación

10. ¿Notaste algún patrón común en el lenguaje utilizado por las cuentas que difunden desinformación? ¿Por qué crees que el lenguaje emotivo funciona tan bien?

11. ¿Este juego ha cambiado la forma en que verás las publicaciones en tus redes sociales de ahora en adelante? ¿De qué manera?

12. ¿Crees que una cuenta verificada puede llegar a ser poco fiable? ¿Por qué?

Acerca de la moderación de contenido

13. Tras jugar como moderador de contenido, ¿crees que este trabajo podría ser realizado completamente por IA? ¿Qué podría salir mal?

14. ¿Fue fácil o difícil decidir qué contenido era problemático? ¿Hubo zonas grises?

15. ¿De quién es la responsabilidad de detener la desinformación: de la plataforma, de los usuarios o de ambos?

16. ¿Qué tipos de contenido suelen moderarse en las plataformas digitales?

17. ¿Cómo se puede equilibrar la moderación de contenidos con la libertad de expresión?

Síntesis general

18. ¿Qué fue lo más importante que aprendiste del juego?

19. Si tuvieras que escribir un consejo para alguien de tu edad sobre la IA y las redes sociales, ¿cuál sería?

20. ¿El juego cambió alguna de tus creencias previas al juego? ¿Cuál?

4. Guía paso a paso para la escape room

4.1. Recomendaciones de juegos

La duración total de la escape room es de aproximadamente **20 minutos** . Los jugadores verán el tiempo restante en la esquina superior derecha, lo que les generará una sensación de urgencia para terminar el juego antes de que se acabe el tiempo.

Puede ser una buena idea que el facilitador haga una cuenta regresiva y que todos comiencen a jugar al mismo tiempo para que se sienta más como una **carrera contra el reloj** . Siempre debe haber un profesional en la sala para guiar a los estudiantes y ayudarlos si se atascan.

4.2. Instrucciones del juego

Acceder al juego

Puedes acceder al juego a través del siguiente enlace:
https://endgameproject.github.io/ECHO/

El juego comienza con una pantalla que le pide al jugador que elija un idioma e inicie sesión. Hay cuatro opciones de idioma disponibles: español, inglés, finlandés y serbio. Además, el jugador debe proporcionar un alias y su edad.



Vídeo de introducción y pre-test

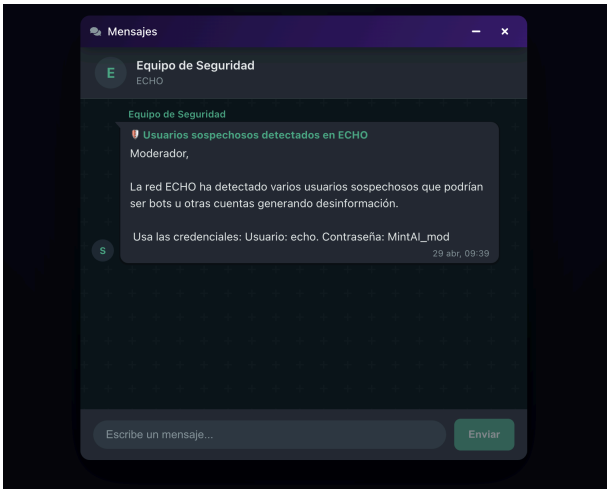
A los jugadores se les muestra un vídeo introductorio que explica la historia y el objetivo de la escape room. Deben completar un «Formulario del Moderador» que indaga sobre sus actitudes iniciales hacia la IA y sus ideas sobre las redes sociales, pidiéndoles que marquen 5 afirmaciones que consideren verdaderas. El objetivo de esta tarea es evaluar los conocimientos previos de los jugadores antes de empezar. Esta misma tarea se presentará al final del juego, por lo que las

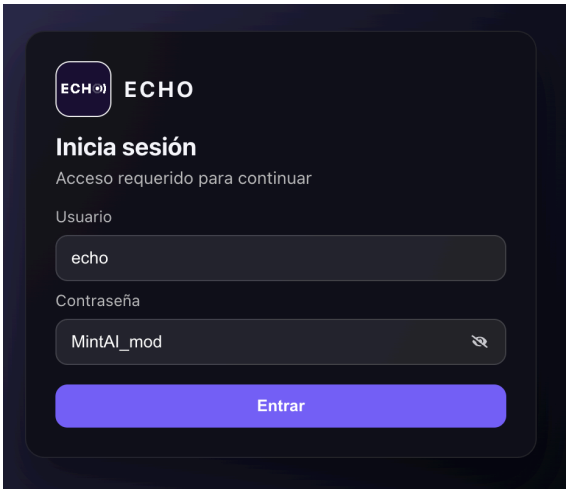
respuestas correctas no se proporcionan en esta etapa y los jugadores pueden avanzar independientemente de sus respuestas.



Reto 1: Identificación de bots

El primer desafío comienza con un SMS de El Jefe, quien le indica al jugador que revise la actividad sospechosa en las redes sociales. El objetivo es identificar qué cuentas presentan las características típicas de la actividad de bots y cuáles parecen normales.



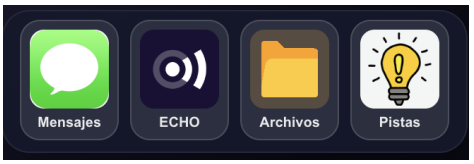


Una vez que inicien sesión en la red social, verán un feed con varias publicaciones. En el lado izquierdo, se encuentran los botones de Inicio y Perfil, que permiten al jugador navegar por la red

social ECHO y ver las publicaciones. En el lado derecho, se muestra el panel de moderación, que informa al jugador sobre el tiempo restante y la cantidad de actividad sospechosa en la red.



En la parte inferior de la pantalla se encuentra el panel de navegación principal, donde el jugador puede acceder a la aplicación de mensajes, la aplicación de la red social ECHO, así como a las carpetas y las pistas (que le ayudarán en caso de quedarse atascado).



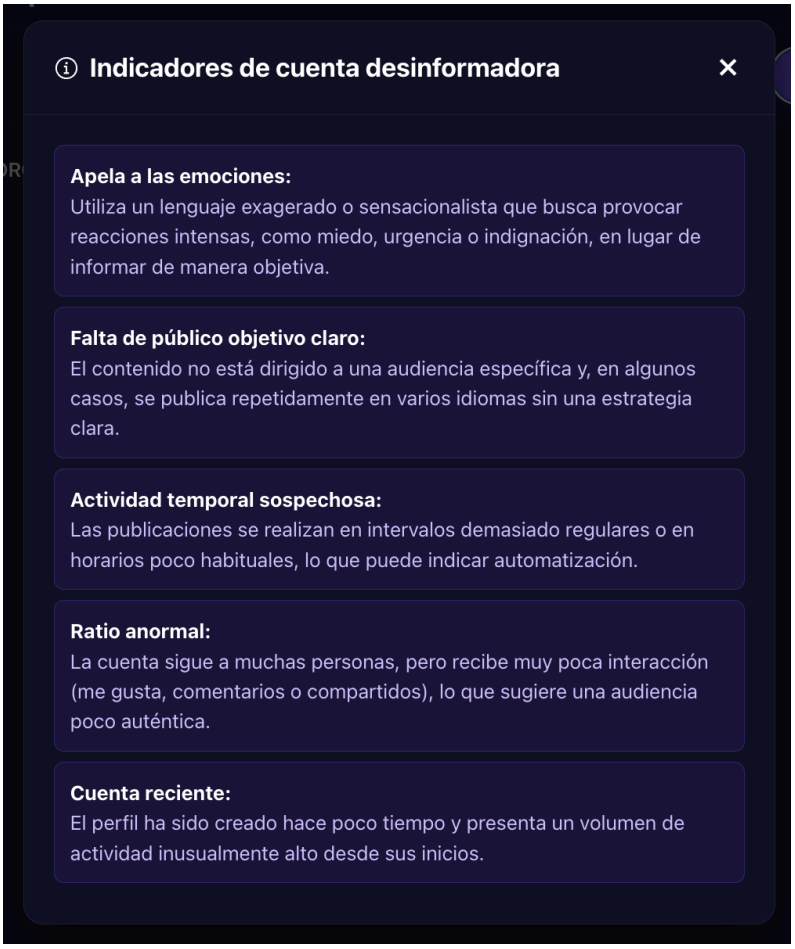
Tras iniciar sesión, el jugador recibirá un SMS de El Jefe pidiéndole que examine los perfiles que hayan sido marcados como sospechosos e identifique si se trata de bots o cuentas reales.



Hay cinco cuentas de redes sociales que el jugador debe revisar. Estas pueden variar ligeramente en cada partida:



Para identificar correctamente si los perfiles corresponden a cuentas de bots o reales, el jugador necesita aprender los criterios que se revelan al activar el botón "**Antes de empezar"** en el lado derecho de la pantalla:

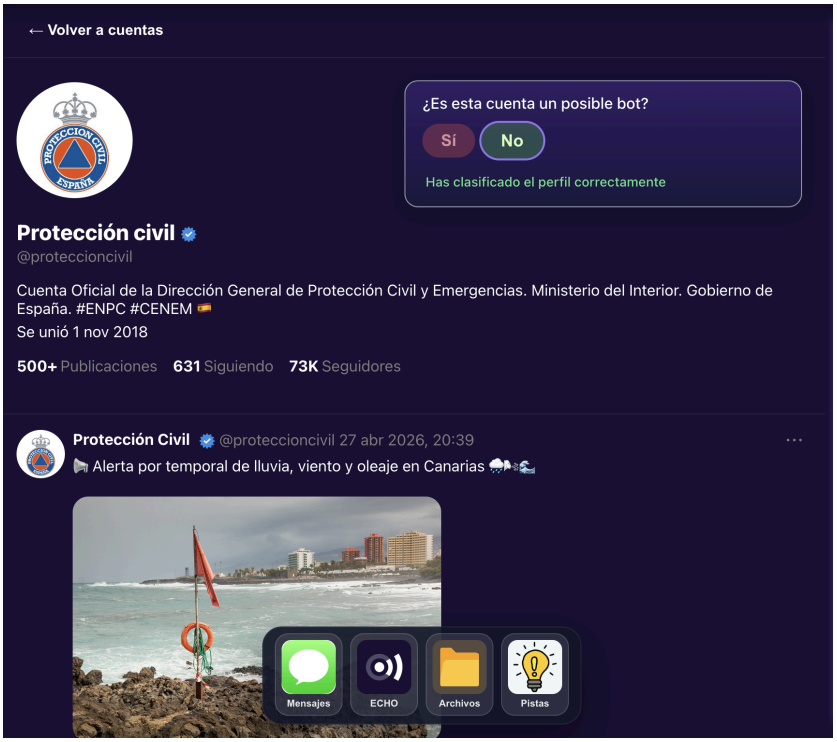


Estos criterios pueden variar según la cuenta de redes sociales, por lo que no todos estarán presentes en todas. Se recomienda al jugador prestar mucha atención al contenido y al comportamiento de publicación de las cuentas. Estos reflejan las prácticas reales de las redes sociales.

Un ejemplo de cuenta bot es “LaVerdadSobreLasNoticias”, que difunde desinformación utilizando un lenguaje que apela a las emociones de los lectores y tiene un índice de interacción anormalmente alto.



Un ejemplo de una cuenta real es Protección civil, la cuenta oficial de este organismo público..

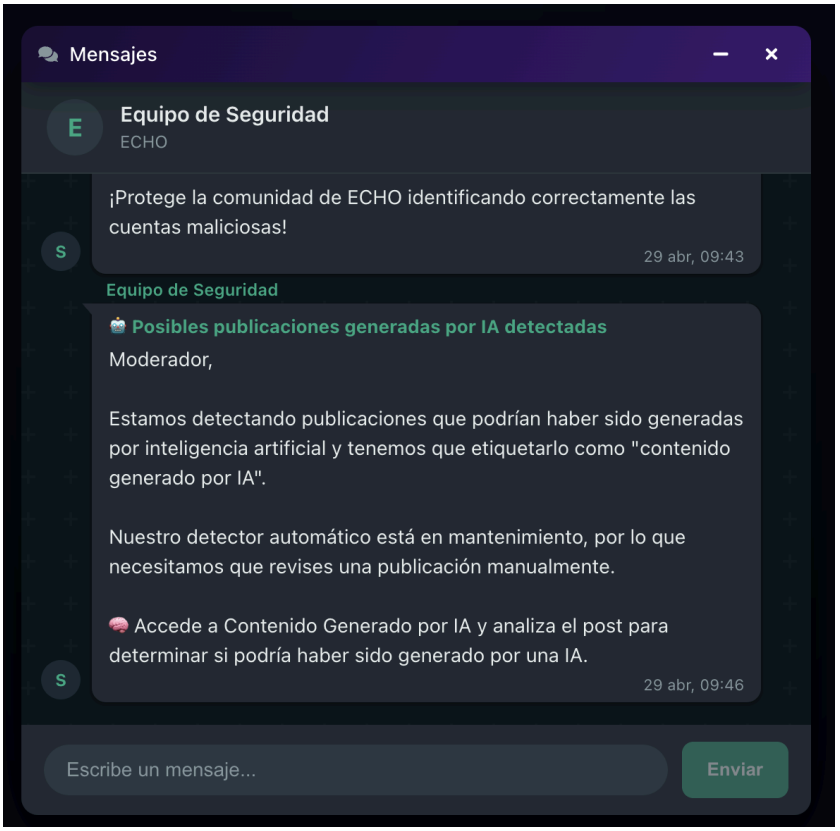


En cada partida, el juego presentará **tres cuentas de bots** y **dos cuentas reales** (una oficial de un organismo público u organización similar y una cuenta de una persona común). A continuación, se muestra la lista de todas las cuentas posibles en la versión en español del juego:

Cuenta	Foto	Detalles
FireTruthNow		Bot Esta es una cuenta bot que puede detectarse por su actividad temporal, con una publicación por hora exactamente a la misma hora, así como por su lenguaje que apela a las emociones.
LaVerdadSobreLasNoticias		Bot Esta es una cuenta bot que puede detectarse por su actividad temporal, con una publicación por hora exactamente a la misma hora, así como por su lenguaje que apela a las emociones.
Agenda2030No		Bot Esta es una cuenta bot que puede detectarse por su lenguaje que apela a las emociones.
VozDelPuebloLibre		Bot Esta es una cuenta bot que se puede detectar por su ratio de seguidores vs. seguidos así como por el tipo de lenguaje empleado
MariaCurrante		Bot Esta es una cuenta bot que puede detectarse por su lenguaje que apela a las emociones.
InfoSinCensura		Bot Esta es una cuenta bot que se puede detectar por su ratio de seguidores vs. seguidos así como por el tipo de lenguaje empleado
LauritaO7		Personal Esta es una cuenta que pertenece a una persona real, que contiene actualizaciones personales y momentos importantes, y no muestra señales de estar automatizada ni de difundir desinformación.
Manolo83		Personal Esta es una cuenta que pertenece a una persona real, que contiene actualizaciones personales y momentos importantes, y no muestra señales de estar automatizada ni de difundir desinformación.
TecnoLuis		Personal Esta es una cuenta que pertenece a una persona real, que contiene actualizaciones personales y momentos importantes, y no muestra señales de estar automatizada ni de difundir desinformación.
nOahvibes		Personal Esta es una cuenta que pertenece a una persona real, que contiene actualizaciones personales y momentos importantes, y no muestra señales de estar automatizada ni de difundir desinformación.
CarlosRunner92		Personal Esta es una cuenta que pertenece a una persona real, que contiene actualizaciones personales y momentos importantes, y no muestra señales de estar automatizada ni de difundir desinformación.
lowkeydiego		Personal Esta es una cuenta que pertenece a una persona real, que contiene actualizaciones personales y momentos importantes, y no muestra señales de estar automatizada ni de difundir desinformación.
xoel_ferrol		Personal Esta es una cuenta que pertenece a una persona real, que contiene actualizaciones personales y momentos importantes, y no muestra señales de estar automatizada ni de difundir desinformación.
BeBeBelen		Personal Esta es una cuenta que pertenece a una persona real, que contiene actualizaciones personales y momentos importantes, y no muestra señales de estar automatizada ni de difundir desinformación.
proteccioncivil		Oficial Esta es una cuenta perteneciente a un organismo oficial
AEMET_Esp		Oficial Esta es una cuenta perteneciente a un organismo oficial

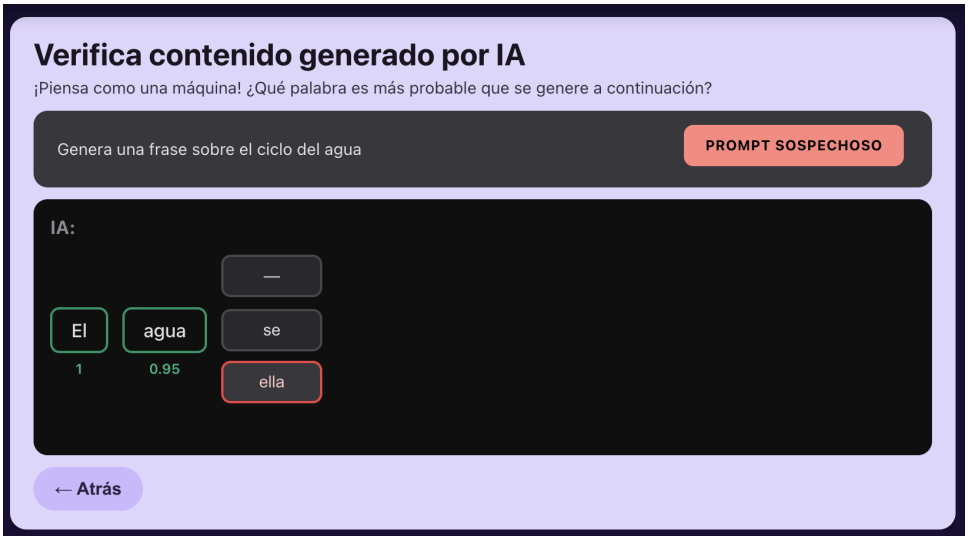
Reto 2: Contenido generado por IA

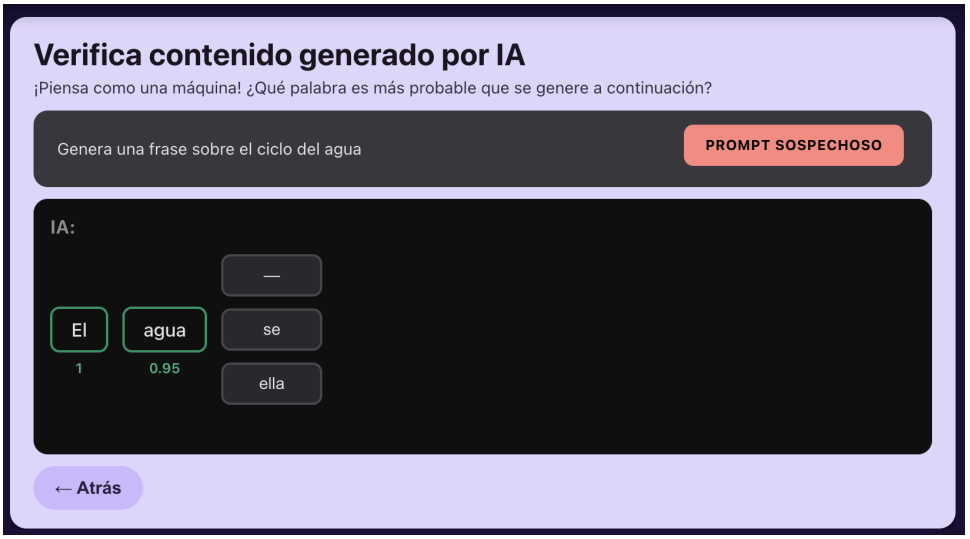
El reto 2 tiene como objetivo enseñar a los jugadores cómo funcionan los LLM y cómo se pueden usar para producir contenido en masa. Esto no tiene necesariamente una intención maliciosa (como la desinformación), pero es necesario revelar cuándo un contenido ha sido generado por IA. Una vez más, el jugador recibe un SMS de El Jefe con instrucciones sobre cómo proceder.



El jugador hará clic en la sección de contenido "Generado por IA" en la barra de navegación para acceder a la tarea. El reto consiste en cómo la IA responde a las preguntas dadas prediciendo estadísticamente la siguiente palabra. Hay un video que explica este proceso y que los jugadores deben ver antes de intentar la tarea. Este video también puede usarse de forma independiente como recurso de aprendizaje.

A continuación, se le proporciona al jugador una publicación y debe demostrar que ha sido generada por IA replicando el proceso de generación del LLM. El LLM ofrecerá tres "palabras siguientes" y el jugador deberá adivinar cuál sigue según la indicación. Si elige la palabra incorrecta, deberá reiniciar la oración completa hasta que logre replicarla correctamente.


Verifica contenido generado por IA
¡Piensa como una máquina! ¿Qué palabra es más probable que se genere a continuación?
Genera una frase sobre el ciclo del agua
PROMPT SOSPECHOSO
IA:
El
agua
1
0.95
—
se
ella
← Atrás


Verifica contenido generado por IA
¡Piensa como una máquina! ¿Qué palabra es más probable que se genere a continuación?
Genera una frase sobre el ciclo del agua
PROMPT SOSPECHOSO
IA:
El
agua
1
0.95
—
se
ella
← Atrás


Verifica contenido generado por IA
¡Piensa como una máquina! ¿Qué palabra es más probable que se genere a continuación?
Genera una frase sobre el ciclo del agua
PROMPT SOSPECHOSO
IA:
El 1
agua 0.95
se 0.92
evapora, 0.9
luego 0.85
forma 0.88
nubes 0.93
y 0.97
vuelve 0.85
a 0.98
la 0.99
Tierra 0.94
como 0.9
lluvia 0.95
COINCIDENCIA CON EL POST:
Pensamiento crítico @quesofresco85 11 dic. 2025
El agua se evapora, luego forma nubes y vuelve a la Tierra como lluvia

El mensaje que se generará se seleccionará aleatoriamente en cada partida. Aquí tienes una lista de todas las frases posibles en la versión en español:

- El agua se evapora, luego forma nubes y vuelve a la Tierra como lluvia
- Es el organismo que predice el tiempo y estudia el clima en España
- Significa que el cambio climático está causado principalmente por la actividad humana
- Reciclar reduce residuos y ahorra recursos naturales
- Los océanos absorben más del 90% del calor extra del cambio climático

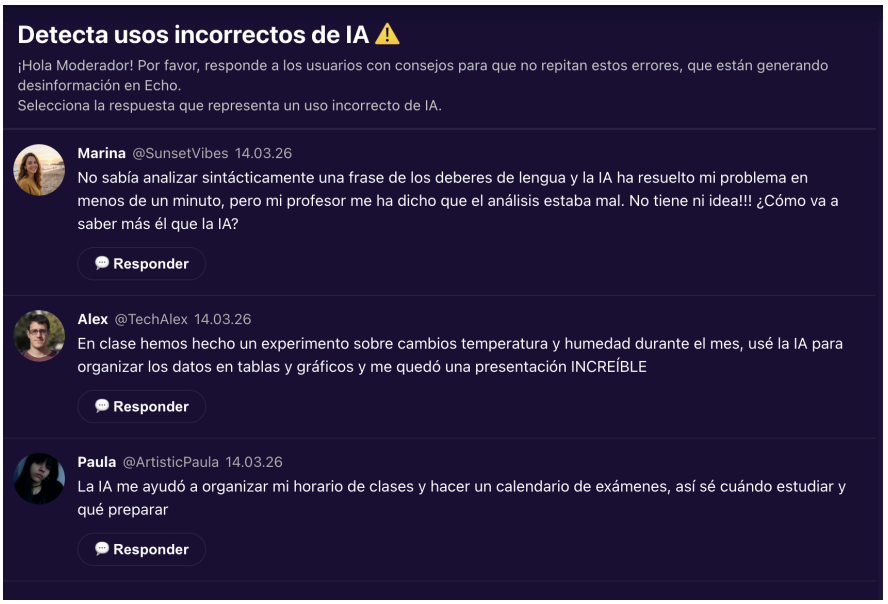
Reto 3: Evaluación del uso recomendado de la IA

El último reto trata sobre usos recomendables de la IA. Varios usuarios de redes sociales han sido señalados por usar la IA como motor de búsqueda para obtener información sobre noticias locales, eventos recientes y afecciones médicas, lo cual no constituye un uso apropiado de la IA. El jugador debe evaluar si los mensajes señalados que reportan algunos de estos posibles usos indebidos son apropiados o no y responder en consecuencia.

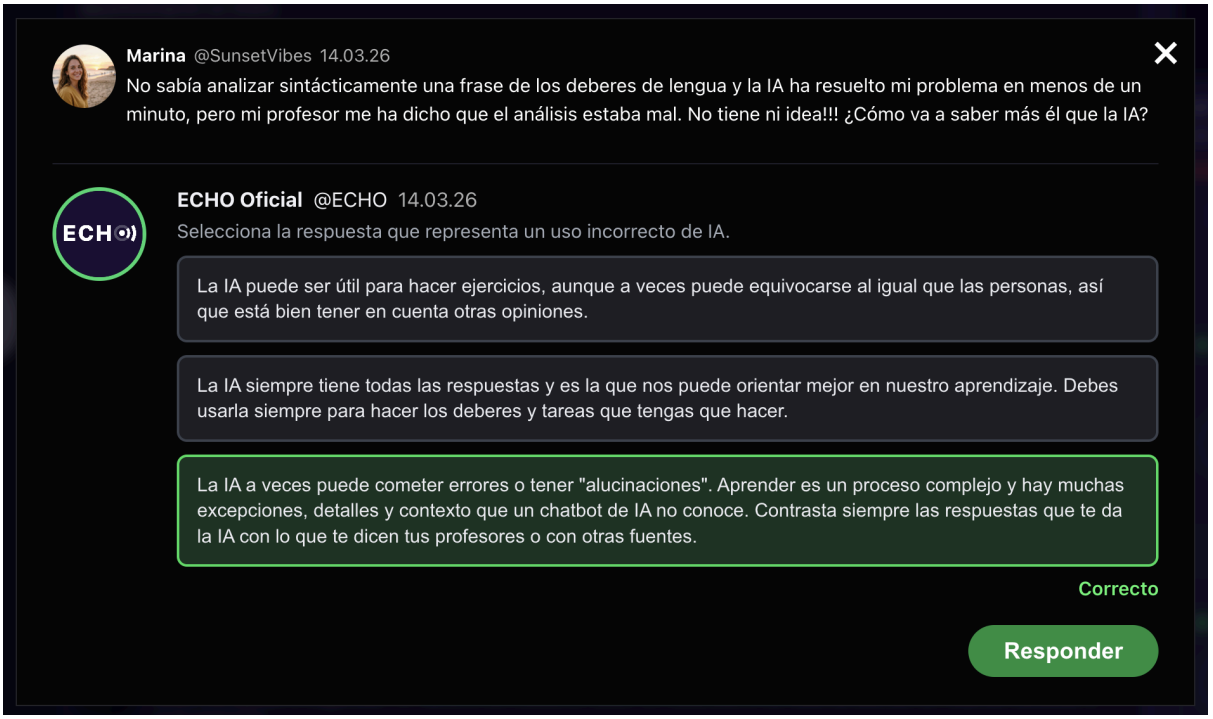
Nuevamente, se recibe un SMS del jefe explicando la tarea:



El jugador hará clic en la sección titulada "Usos incorrectos de la IA" en la barra de navegación y verá varias publicaciones que hacen referencia a la IA como fuente de información.



El usuario, actuando como moderador de redes sociales, debe responder a estas publicaciones para informar a los usuarios sobre si su uso de la IA fue apropiado o no. Para cada publicación, hay tres posibles respuestas: una es claramente errónea, otra es neutral y la tercera es la más precisa y completa.



Las tres publicaciones son seleccionadas aleatoriamente en cada partida ente las siguientes:

- **Publicación** : No sabía analizar sintácticamente una frase de los deberes de lengua y la IA ha resuelto mi problema en menos de un minuto, pero mi profesor me ha dicho que el análisis estaba mal. No tiene ni idea!!! ¿Cómo va a saber más él que la IA?
    - **Correcto:** La IA a veces puede cometer errores o tener "alucinaciones". Aprender es un proceso complejo y hay muchas excepciones, detalles y contexto que un chatbot de IA no conoce. Contrasta siempre las respuestas que te da la IA con lo que te dicen tus profesores o con otras fuentes.
    - **Neutral:** La IA puede ser útil para hacer ejercicios, aunque a veces puede equivocarse al igual que las personas, así que está bien tener en cuenta otras opiniones.
    - **Incorrecto:** La IA siempre tiene todas las respuestas y es la que nos puede orientar mejor en nuestro aprendizaje. Debes usarla siempre para hacer los deberes y tareas que tengas que hacer.

  El objetivo de esta tarea es reflexionar sobre el concepto de “alucinaciones”, y que los alumnos sean conscientes de que, aunque la IA puede ser una herramienta muy útil para apoyar el aprendizaje, no siempre tiene las respuestas correctas.

- **Publicación** : Mirad, MINT me ha dicho que mañana habrá tormentas extremas en España. ¡Preparad todo!
    - **Correcto:** Una captura sin contexto de un chatbot de IA no es una fuente fiable para saber el tiempo. Las IAs no pueden predecir tormentas concretas en tiempo real. Es mejor consultar siempre fuentes oficiales como AEMET para alertas meteorológicas
    - **Neutral:** No leo bien lo que dice la captura de pantalla
    - **Incorrecto:** ¡Genial! Entonces mejor preparad todo como dice la IA, seguro que va a pasar. Pregúntale por datos concretos, hora en la que ocurrirá y cuanta lluvia caerá.

  El objetivo de esta tarea es sacar a relucir que la IA no está al tanto de eventos recientes y es mejor utilizar un motor de búsqueda para conocer ese tipo de información en tiempo real.

- **Publicación**: Hay que ir mañana al instituto??? Le he preguntado a la IA si las clases siguen suspendidas por las lluvias y dice que sí
    - **Correcto:** Para saber si las clases están suspendidas, consulta siempre el aviso oficial del instituto o de Protección Civil, que tiene la información real y actualizada. Un chatbot de IA no es igual que un buscador
    - **Neutral:** Fíate de tu intuición o pregúntale a tu madre o padre
    - **Incorrecto:** Esto es un buen uso de la IA. Vista su respuesta puedes quedarte en casa, la IA te aconseja bien ya que siempre consulta las fuentes oficiales

El objetivo de esta tarea es poner de manifiesto que la IA no tiene en cuenta ni conoce el contexto personal de cada persona, por lo que no puede responder preguntas concretas sobre el entorno inmediato de cada uno.

- **Publicación** : En clase hemos hecho un experimento sobre cambios temperatura y humedad durante el mes, usé la IA para organizar los datos en tablas y gráficos y me quedó una presentación INCREÍBLE
    - **Correcto:** Usar la IA para organizar y visualizar datos es útil porque te ayuda a interpretar mejor los resultados. Pero siempre aporta valor con tu trabajo, pensamiento crítico y revisión del resultado final.
    - **Neutral:** No te olvides de poner también tus datos personales en el trabajo final
    - **Incorrecto:** Genial, deja que la IA haga todo el análisis por ti y entrega directamente el trabajo realizado por la IA sin revisarlo, siempre está muy completo

  El objetivo de esta tarea es informar a los jugadores que la IA puede automatizar algunas tareas, pero puede cometer errores y siempre es necesario confirmar los resultados.

- **Publicación** : La IA me ayudó a organizar mi horario de clases y hacer un calendario de exámenes, así sé cuándo estudiar y qué preparar
    - **Correcto:** Usar la IA para organizar tu horario y calendario de exámenes te ayuda a planificarte, pero el aprendizaje sigue siendo tu responsabilidad
    - **Neutral:** Pon colores bonitos también, que no se te olvide
    - **Incorrecto:** Ese es un uso de la IA que no sirve para nada

  Esta tarea demuestra que la IA es una herramienta apropiada para ayudar a organizarnos, pero la responsabilidad siempre recae en cada uno.

- **Publicación** : Para el examen de Naturales, le pedí a la IA que me resumiera y esquematizara los temas que tenía que estudiar. Aun así, lo revisé para asegurarme que había incluido todos correctamente
    - **Correcto:** Utilizar la IA para resumir y esquematizar te ayuda a organizar los temas, pero leerlos tú mismo/a asegura que comprendes todo y evitas posibles fallos o alucinaciones
    - **Neutral:** Qué bien, ya nos contarás que nota sacas
    - **Incorrecto:** No hace falta que revises nada, la respuesta que te da la inteligencia artificial es siempre muy completa. No te pares a leerlo, tampoco es necesario

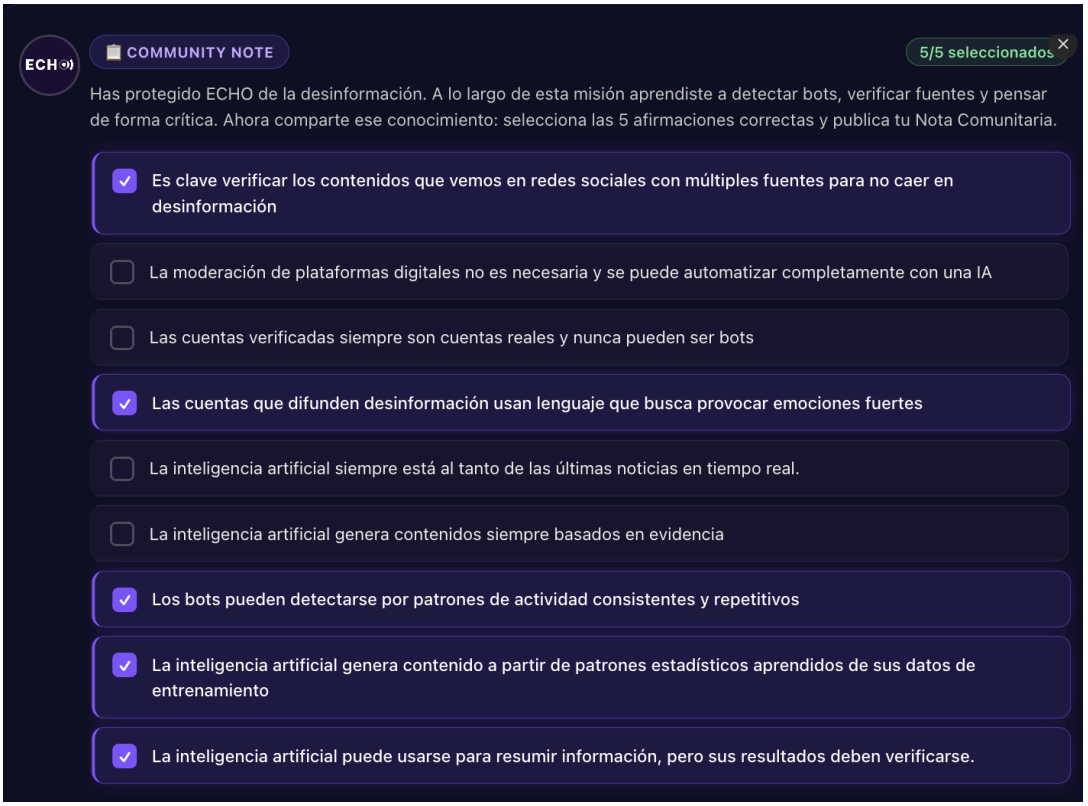
  Esta tarea demuestra que la IA es una herramienta apropiada para la elaboración de resúmenes, pero siempre es necesario verificar los resultados.

Condiciones de victoria

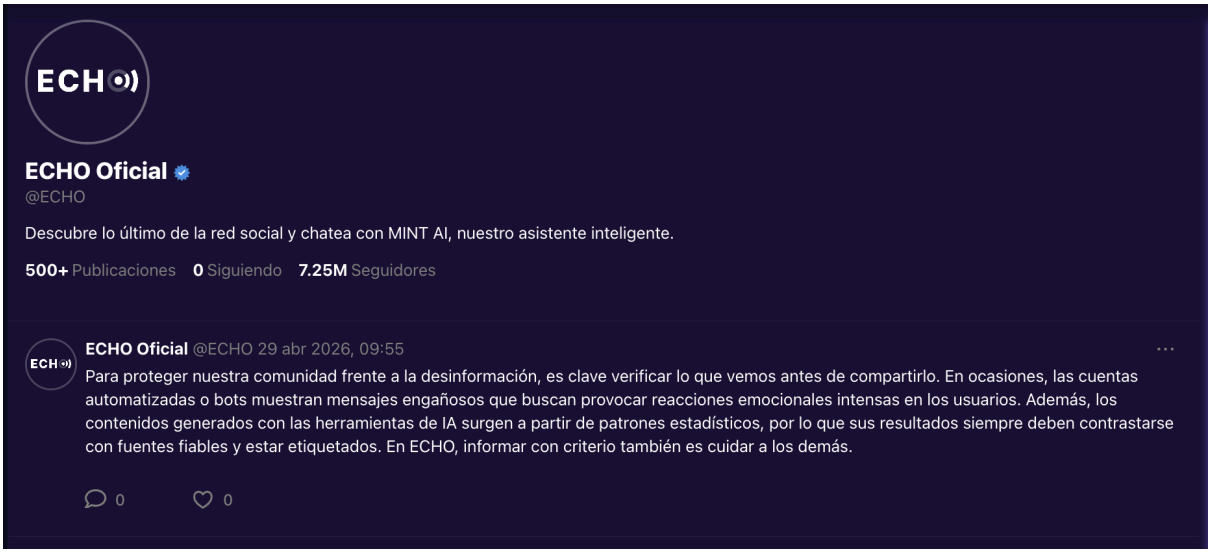
Para ganar el juego, el jugador debe resolver correctamente estos tres retos en 20 minutos. Tras resolverlos, se le pedirá que publique un mensaje en redes sociales, el cual se difundirá en la red ECHO desde la cuenta oficial de la red.



Los jugadores deben identificar las cinco afirmaciones verdaderas sobre el uso de la IA en las redes sociales, que son prácticamente las mismas que vieron al inicio del juego. Las afirmaciones correctas son las que aparecen seleccionadas en la imagen a continuación:



Una vez que lo hayan hecho, estas notas se publican como una "Nota para la comunidad" en la red social a modo de mensaje para ayudar a los usuarios a identificar bots, contenido generado por IA y hacer un uso adecuado de la inteligencia artificial.



Por último, los jugadores ven un vídeo final y se les pide que respondan una encuesta sobre su experiencia de juego. Agradecemos que los profesores animen a los jugadores a responder la encuesta, ya que sus comentarios ayudan a mejorar el juego e identificar los contenidos con los que tienen más dificultades.