

Aula Invertida

El "aula invertida" (también conocida como Flipped Classroom o aprendizaje invertido) es un modelo pedagógico que cambia el enfoque tradicional de la enseñanza. En esencia, invierte lo que se hace dentro y fuera del aula.

La IA y su papel en la desinformación.

Realizar la lectura del documento anexo, profundizar en sus temáticas y responder las siguientes preguntas y preparar una participación activa en clase!.

Nota Realizar entrega en la carpeta "Entregas" de esta carpeta del repositorio, en formato .pdf y con la nomenclatura Apellido_Nombre

Contexto

El documento de estudio profundiza en el creciente problema de la desinformación, resaltando su impacto en áreas cruciales como la salud y la política y distinguiendo entre diferentes tipos de información falsa. El documento explora el doble rol de la Inteligencia Artificial (IA), tanto como facilitadora de la creación y difusión de bulos, especialmente a través de modelos generativos, como herramienta esencial para combatirla. Se presenta FacTeR-Check, una herramienta desarrollada por el grupo AIDA de la Universidad Politécnica de Madrid, que utiliza IA para la verificación y monitoreo de desinformación, destacando su enfoque multilingüe y análisis en redes sociales. Finalmente, se reflexiona sobre el futuro de la IA en esta lucha, considerando avances como la IA explicativa y la importancia de la privacidad.

CUESTIONARIO

1. ¿Cuál es la diferencia fundamental, según el texto, entre "misinformation" y "disinformation"?

RTA:

Misinformation Se refiere a información falsa sin intención de causar daño. Es decir, quien la difunde cree que es verdadera y no pretende engañar.

Disinformation, Se refiere a información falsa difundida deliberadamente con la intención de causar perjuicio.

2. Según el Reuters Institute Digital News Report 2023, ¿qué tendencia preocupante se observa en España con respecto al interés por las noticias?

RTA:

Según el **Reuters Institute Digital News Report 2023**, en España se observa una preocupante tendencia a la **caída del interés por las noticias**.

- El porcentaje de personas que se declaran "**muy o extremadamente interesadas**" en las noticias ha caído de aproximadamente **85 % en 2015** a **52 % en 2023** —lo que supone una pérdida de **33 puntos porcentuales**.

- Además, España es uno de los países con los mayores aumentos en **evasión selectiva de noticias**, alcanzando un **37 %** que evita activamente informarse.

3. ¿Cómo se comparan, según los experimentos de Vosoughi, Roy y Aral (2018), la velocidad y facilidad de difusión de noticias falsas frente a las verdaderas?

RTA:

Según **Vosoughi, Roy y Aral (2018)**, las noticias falsas se difunden más rápido y con mayor facilidad que las verdaderas.

- El 1 % de las noticias falsas más difundidas llegó a entre 1.000 y 100.000 personas.
- En cambio, el 1 % de las noticias verdaderas más difundidas rara vez superó las 1.000 personas.

4. ¿Qué ventaja clave ofrecen las redes latentes de difusión sobre los modelos epidemiológicos para el estudio de la desinformación?

RTA:

Permiten identificar quién propaga la información y cómo lo hace, es decir, modelan las interacciones específicas entre usuarios, mostrando la dirección y la intensidad de la influencia entre ellos.

- **Redes latentes de difusión** = permiten trazar e identificar actores e influencias.
- **Modelos epidemiológicos** = permiten detectar el fenómeno, pero no quién lo causa.

5. ¿Qué son los "grandes modelos de lenguaje" y cuál es su principal riesgo en el contexto de la desinformación?

RTA:

Los **grandes modelos de lenguaje** (*large language models*, LLMs) son sistemas de inteligencia artificial entrenados con enormes cantidades de texto para comprender y generar lenguaje natural. Están basados en arquitecturas como *Transformer* y pueden producir texto coherente y convincente en múltiples idiomas y contextos.

RIESGOS

- Su **capacidad para generar contenido falso de forma rápida, convincente y a bajo costo**, incluso sin que el usuario tenga conocimientos técnicos avanzados
- Pueden ser usadas para **automatizar y escalar la desinformación**, haciendo más difícil su detección y control.

6. ¿Cómo facilita la accesibilidad de los modelos de IA la generación de desinformación?

RTA:

1. **Baja barrera técnica:** Esto permite que cualquier persona, incluso sin formación especializada, pueda generar contenido sofisticado y convincente.

2. **Disponibilidad gratuita o de bajo costo:** Muchos modelos son gratuitos o de código abierto, lo que elimina barreras económicas para su uso. Además, ya pueden ejecutarse en equipos domésticos de alta gama, sin necesidad de infraestructura profesional.
3. **Automatización y rapidez:** Los modelos permiten generar grandes volúmenes de contenido falso en segundos, lo que antes requería tiempo y esfuerzo humano. Esto acelera y amplía la capacidad de lanzar campañas de desinformación.
4. **Calidad creíble del contenido:** La IA genera contenido con alto nivel de coherencia lingüística, estilo periodístico o realismo visual, lo que aumenta la credibilidad del bulo y la dificultad para distinguirlo de información legítima.
5. **Anonimato y escala:** Al usarse dentro de plataformas de microencargos o en redes anónimas, estos modelos pueden ser empleados por actores maliciosos de forma encubierta para producir y difundir desinformación a gran escala y con muy poco rastro.

7. ¿Qué son las "cajas negras" en el contexto de la IA explicativa y cuál es el desafío asociado?

RTA:

Cajas negras" se refieren a aquellos modelos de inteligencia artificial (como redes neuronales profundas) cuyo funcionamiento interno es **opaco o difícil de interpretar para los humanos**.

Desafíos asociados

El gran reto es lograr que la IA sea **transparente y comprensible**, especialmente en contextos sensibles como la lucha contra la desinformación. Si no podemos entender cómo y por qué una IA etiqueta un contenido como verdadero o falso:

8. ¿Qué implicaciones tiene el concepto de "Inteligencia Artificial General (AGI)" para la lucha contra la desinformación?

RTA:

Implicaciones positivas:

- **Detección avanzada de desinformación:**
Una AGI podría entender el contexto, las intenciones y las sutilezas del lenguaje humano mejor que los sistemas actuales, permitiéndole identificar bulos complejos, sarcasmos, manipulaciones y campañas coordinadas con mayor precisión.
- **Capacidad de adaptación:**
Podría adaptarse en tiempo real a nuevas estrategias de desinformación y plataformas emergentes, reentrenándose y evolucionando sin intervención humana constante.
- **Educación y alfabetización mediática automatizada:**
AGI podría ofrecer intervenciones personalizadas para educar a los usuarios sobre cómo reconocer y resistir la desinformación, incluso anticipándose a la vulnerabilidad individual.

Implicaciones negativas:

➤ **Generación masiva de desinformación sofisticada**

Si cae en manos maliciosas, una AGI podría crear campañas de desinformación altamente convincentes y personalizadas, a gran escala y de forma automática, mucho más allá de lo que es posible hoy.

➤ **Manipulación psicológica avanzada**

Al entender mejor el comportamiento humano, la AGI podría ser usada para diseñar narrativas o imágenes que manipulen creencias y emociones con precisión quirúrgica.

➤ **Dificultad para regularla y auditarla**

El uso dual (benigno y malicioso) de una AGI plantea enormes desafíos éticos, legales y sociales. Controlar su comportamiento y asegurar la transparencia será extraordinariamente difícil si actúa como una “caja negra” con autonomía total

9. ¿Qué normativas europeas importantes se mencionan en relación con la IA y la privacidad?

RTA:

Las normativas mencionadas —**RGPD y el AI Act**— buscan garantizar que la inteligencia artificial se desarrolle y utilice de forma **ética, transparente y respetuosa con los derechos fundamentales**, especialmente en ámbitos como la **privacidad**, el **consentimiento informado** y la **no discriminación**

1. Reglamento General de Protección de Datos (RGPD / GDPR)

- Es la normativa europea más importante sobre protección de datos personales.
- Establece principios como la minimización de datos, el consentimiento explícito, el derecho al olvido y la transparencia en el tratamiento de información personal.
- Tiene implicaciones directas sobre los sistemas de IA que manejan o procesan datos personales, exigiendo responsabilidad y explicabilidad en su funcionamiento.

2. Propuesta de Reglamento de la UE sobre la Inteligencia Artificial (AI Act)

- Es una normativa pionera en regulación específica para la IA dentro del marco jurídico europeo.
- Clasifica los sistemas de IA por niveles de riesgo (inaceptable, alto, limitado, mínimo).
- Impone requisitos estrictos de seguridad, transparencia y supervisión humana a los sistemas de alto riesgo, como los que pueden influir en elecciones o derechos fundamentales.
- También aborda la necesidad de prevenir sesgos y garantizar la protección de derechos fundamentales frente a la automatización.

10. ¿Cómo garantiza FacTeR-Check el cumplimiento de la normativa de protección de datos al analizar redes sociales? Preguntas de Formato Ensayo

RTA:

El cumplimiento de la normativa de protección de datos en herramientas de análisis automatizado, como FacTeR-Check, es una condición indispensable para su legitimidad, eficacia y aceptación social. FacTeR-Check, diseñada para detectar y contrarrestar la desinformación mediante inteligencia artificial, implementa una serie de medidas técnicas y éticas orientadas a garantizar el respeto a los derechos fundamentales, especialmente en lo que concierne a la privacidad y protección de datos personales en redes sociales. Este compromiso se fundamenta especialmente en el marco normativo del Reglamento General de Protección de Datos (RGPD) de la Unión Europea.

- En primer lugar, FacTeR-Check aplica el principio de **minimización de datos**, recolectando únicamente aquella información que es estrictamente necesaria para la detección de desinformación.
- En segundo lugar, FacTeR-Check incorpora un enfoque de **anonimización y pseudonimización de datos**. Esto significa que los datos personales se transforman de modo que los individuos no puedan ser identificados directa o indirectamente.
- Otra dimensión clave es la implementación de **evaluaciones de impacto sobre la protección de datos** (DPIA, por sus siglas en inglés), particularmente relevante en proyectos que involucran tecnologías automatizadas o a gran escala.
- Además, FacTeR-Check opera bajo un marco de **transparencia y trazabilidad**, es decir, documenta las decisiones que toma el sistema, lo que permite auditar sus procesos y verificar que actúe conforme a criterios legítimos y no discriminatorios.
- Finalmente, FacTeR-Check promueve una **arquitectura de “privacidad por diseño y por defecto”**, tal como lo exige el artículo 25 del RGPD. Desde su concepción, la herramienta se ha estructurado de modo que la privacidad no sea un añadido posterior, sino una condición estructural del sistema, garantizando que las opciones más respetuosas con la privacidad sean las predeterminadas y no dependan del usuario final.

11. Analice las diferentes formas en que la Inteligencia Artificial puede ser utilizada tanto para generar como para combatir la desinformación, basándose en los ejemplos y conceptos presentados en el texto.

RTA:

La inteligencia artificial (IA) representa una de las herramientas más poderosas e influyentes en la era de la información digital. Su impacto en el fenómeno de la desinformación es profundamente ambivalente:

- Ofrece capacidades sin precedentes para identificar y neutralizar información falsa.
- Permite la creación masiva y sofisticada de contenido engañoso.

La inteligencia artificial como generadora de desinformación

La IA se ha convertido en una herramienta efectiva para actores maliciosos que buscan difundir desinformación de forma rápida, persuasiva y a gran escala. Uno de los desarrollos más significativos en este sentido ha sido la aparición de los grandes modelos de lenguaje.

- La IA aplicada a la **generación de imágenes** (como Stable Diffusion o DALL·E) permite crear fotografías falsas extremadamente realistas, que refuerzan narrativas engañosas.

La inteligencia artificial como herramienta para combatir la desinformación

La IA también ofrece respuestas prometedoras. Herramientas como FacTeR-Check, desarrolladas dentro del Proyecto CIVIC, demuestran que la misma tecnología que puede ser usada para engañar también puede emplearse para proteger.

- El uso de **análisis de similitud semántica**, que permite identificar cuándo un nuevo contenido se parece sospechosamente a piezas ya clasificadas como falsas.
- La IA en la lucha contra la desinformación es su capacidad para **identificar patrones de influencia y actores clave** en una red. Mientras que los modelos epidemiológicos permiten detectar flujos anómalos de información.

12. Discuta el papel de la Inteligencia Artificial Explicativa (XAI) en la mejora de la confianza pública en los sistemas de detección de desinformación y en la educación de los usuarios. ¿Cuáles son los principales obstáculos para su desarrollo?

RTA:

La inteligencia artificial (IA) ha demostrado ser una herramienta poderosa **para identificar, rastrear y neutralizar contenidos engañosos**.

1. XAI como herramienta para mejorar la confianza pública

Cumple una función democratizadora del conocimiento. Permite a periodistas, verificadores, académicos e incluso al público general comprender los criterios utilizados por los algoritmos.

2. XAI como herramienta educativa

Las explicaciones que acompañan la detección de desinformación pueden convertirse en oportunidades de alfabetización mediática para los usuarios

3. Principales obstáculos para el desarrollo de XAI

- **Complejidad técnica de los modelos**
Los modelos más potentes para el análisis del lenguaje, como las redes neuronales profundas (deep learning), son también los más opacos. Traducir sus decisiones en explicaciones comprensibles **sin perder precisión** es un reto técnico considerable.
- **Equilibrio entre explicabilidad y rendimiento**
En muchos casos, aumentar la explicabilidad puede significar **reducir la eficacia** del modelo.

Modelos más simples son más fáciles de explicar, pero pueden ser menos precisos en la detección de desinformación compleja.

➤ **Falta de estándares y marcos regulatorios claros**

Aunque existen principios éticos y propuestas como el Reglamento de Inteligencia Artificial de la Unión Europea, **aún no hay un consenso global** sobre qué constituye una explicación adecuada, comprensible y suficiente para el usuario.

13. Compare los modelos epidemiológicos y las redes latentes de difusión como enfoques para estudiar la propagación de la desinformación en las redes sociales. ¿Qué información específica puede obtenerse de cada tipo de modelo?

RTA:

Criterio	Modelos epidemiológicos	Redes latentes de difusión
Nivel de análisis	Macroscópico (poblacional)	Microscópico (usuario a usuario)
Requiere conocer relaciones entre usuarios	No	Sí
Identificación de agentes clave	No	Sí
Datos necesarios	Tiempos de activación	Interacciones y trazabilidad
Explicabilidad de la propagación	Limitada	Alta
Aplicación principal	Predicción de tendencias	Trazabilidad y análisis forense

Tanto los modelos epidemiológicos como las redes latentes de difusión aportan perspectivas complementarias para el análisis de la desinformación. Mientras los primeros son útiles para **detectar y prevenir fenómenos masivos**, los segundos resultan indispensables para **entender la lógica interna de una campaña y actuar con precisión**. En un entorno informativo saturado y vulnerable, como el de las redes sociales actuales, **la combinación de ambos enfoques** puede ofrecer una estrategia robusta: una visión panorámica de las amenazas junto con una capacidad de respuesta quirúrgica para neutralizarlas.

14. Examine la relación entre la accesibilidad de las herramientas de IA generativa y el aumento potencial de la desinformación. ¿Qué estrategias se sugieren para mitigar este riesgo?

RTA:

La accesibilidad de las herramientas de IA generativa ha democratizado tanto la creatividad como la desinformación. Si bien su potencial positivo es enorme, su **mal uso representa una amenaza real para la confianza pública, la estabilidad democrática y la integridad informativa.**

Por ello, la estrategia más efectiva debe combinar **tecnología avanzada, regulación ética, compromiso institucional y educación ciudadana.** La clave no está en prohibir el acceso a estas herramientas, sino en **orientar su uso hacia fines transparentes, responsables y verificables,** asegurando así que la innovación tecnológica no se convierta en una fuente de vulnerabilidad social.

ESTRATEGIAS

- Desarrollo de herramientas de detección automatizada.
- Integración de inteligencia artificial explicativa (XAI)
- Normativas y regulaciones internacionales
- Identificación de contenido generado por IA
- Alfabetización mediática y formación ciudadana

15. Analice las consideraciones éticas y de privacidad asociadas con el uso de la Inteligencia Artificial para combatir la desinformación, haciendo referencia a las normativas europeas mencionadas e identificado si existen normativas en nuestro país similares.

RTA:

El uso de inteligencia artificial para combatir la desinformación debe estar enmarcado en principios éticos sólidos y normativas legales claras, que garanticen la protección de los derechos fundamentales. Mientras que la Unión Europea ha avanzado en este terreno con instrumentos robustos como el RGPD y la propuesta de AI Act, países como Colombia han empezado a construir políticas públicas y marcos legales que reconocen estos desafíos.

En este contexto, la clave no está en frenar la innovación, sino en guiarla con responsabilidad y vigilancia institucional, promoviendo el uso ético y transparente de la IA. Solo así se puede garantizar que el combate a la desinformación no se convierta en una fuente de nuevas vulnerabilidades para la ciudadanía, sino en una herramienta para reforzar la verdad, la confianza pública y la calidad democrática.

El uso de la Inteligencia Artificial (IA) para combatir la desinformación plantea un dilema ético de gran envergadura: ¿cómo se puede proteger a la sociedad del daño que causa la desinformación sin vulnerar los derechos fundamentales de las personas.

Privacidad de los usuarios

La desinformación suele difundirse a través de redes sociales y plataformas digitales donde los usuarios comparten contenidos, a veces públicamente y otras en entornos cerrados o semi-privados (grupos, chats

Transparencia y explicabilidad

La **transparencia algorítmica** y la posibilidad de ofrecer explicaciones comprensibles son fundamentales para legitimar estas tecnologías y proteger a los ciudadanos de decisiones arbitrarias.

No discriminación y equidad

Las herramientas de IA aplicadas a la detección de desinformación no vulneren los principios de equidad, inclusión y no discriminación.

Libertad de expresión y pluralismo

La IA debe operar bajo criterios claros, auditables y con salvaguardias para preservar el derecho a la expresión libre y diversa

Marco normativo europeo

- Reglamento General de Protección de Datos (RGPD / GDPR)
- Ley de Inteligencia Artificial de la UE (AI Act, en propuesta)

Marco normativo en Colombia

- Ley 1581 de 2012 (Régimen General de Protección de Datos Personales)
- Documento CONPES 3975 de 2019 – Política Nacional de Explotación de Datos
- Estrategia Nacional de Inteligencia Artificial (2020)