

Aula Invertida



Mayor (FAC) Daniel Alfonso Torres Jaramillo

Docente:

Dr. Jaider Ospina Navas

Escuela Superior de Guerra "General Rafael Reyes Prieto"

Bogotá D.C., Colombia

2025

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"?**

R/: La diferencia principal entre misinformation y disinformation radica en la intencionalidad.

Misinformation: es información falsa que se difunde sin intención de causar daño; quien la transmite cree que es verdadera.

Disinformation: es información falsa que se genera y difunde deliberadamente con la intención de causar daño, perjudicar, manipular o engañar.

Esta diferencia es esencial para determinar las estrategias de intervención.

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

R/: El informe revela una tendencia alarmante en España: una notable disminución del interés por las noticias.

Es así como se evidencia que en 2015, un 85 % de las personas declaraban tener un interés alto o muy alto en las noticias, cifra que cayó al 51 % en 2023. Además, se ha registrado un aumento significativo en la desconfianza hacia los medios, alcanzando un 40 %, especialmente entre los menores de 45 años.

Esta “tormenta perfecta” crea un entorno muy favorable para la proliferación de desinformación.

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?**

R/: Los experimentos de Vosoughi, Roy y Aral (2018) demostraron que las noticias falsas se difunden más rápido y más ampliamente que las verdaderas.

Específicamente, el 1 % de las noticias falsas más compartidas llegó a un aproximado de entre 1.000 y 100.000 personas, mientras que el 1 % de las verdaderas más compartidas rara vez superó las 1.000 personas.

Este hallazgo ha permitido desarrollar filtros de detección rápida basados en los patrones de propagación.

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

R/: La ventaja clave de las redes latentes de difusión es que a diferencia de los modelos epidemiológicos, las redes latentes de difusión no solo permiten predecir cómo evolucionará la propagación de piezas de información, sino que también permiten conocer quién la propaga y cómo lo hace.

Estas redes permiten mapear influencias directas y construir estructuras de relaciones en la red, facilitando intervenciones más eficaces contra la desinformación.

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

R/: Los "grandes modelos de lenguaje" (Large Language Models) son sistemas basados en arquitecturas de redes neuronales llamada Transformer, capaces de generar textos coherentes y convincentes de manera automática.

Su principal riesgo es que no comprenden la veracidad de lo que producen y pueden ser usados para crear desinformación de forma automatizada y masiva, bajando la barrera técnica para hacerlo. Esto los convierte en herramientas muy peligrosas cuando son mal utilizadas.

Ejemplo: Chat Gpt (Generative Pre-trained Transformer) de OpenAI.

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

R/: La accesibilidad de estos modelos ha aumentado drásticamente: lo que antes requería recursos técnicos y computacionales avanzados, ahora puede ejecutarse desde un portátil doméstico.

Esto significa que cualquier persona con intenciones maliciosas puede crear y difundir desinformación de forma barata y sencilla, sin necesidad de conocimientos técnicos profundos, aumentando así la amenaza global de las campañas de manipulación.

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

R/: Las “cajas negras” son modelos de inteligencia artificial tan complejos que no permiten entender fácilmente cómo llegan a sus conclusiones, ya que están compuestos por millones de parámetros numéricos.

El principal desafío es convertir estos sistemas en explicables y transparentes, para que usuarios y expertos puedan entender, confiar y corregir los sistemas de verificación automática de desinformación.

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

R/: La IA General (AGI) aspira a replicar la inteligencia humana en múltiples tareas y aunque aún no es una realidad plena, los avances actuales ya apuntan hacia sistemas capaces de resolver diversas tareas complejas.

Esto implica que los modelos futuros podrán generar desinformación aún más convincente, masiva y difícil de detectar, aunque también puede ofrecer mejores herramientas para combatirla. Sin embargo, el problema es que el desinformador siempre actúa primero.

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

R/: Se mencionan dos normativas claves: el Reglamento General de Protección de Datos (RGPD), que protege los datos personales en la UE, y el Libro Blanco sobre la IA, que establece el marco para un desarrollo ético y seguro de la IA en Europa.

Ambos documentos subrayan la importancia de equilibrar la innovación con los derechos fundamentales, incluyendo la privacidad, libertad de expresión y no discriminación.

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

R/: FacTeR-Check se asegura de cumplir la normativa de protección de datos mediante el uso exclusivo de datos públicos, como tuits visibles sin restricción. No accede a contenido privado ni realiza perfilado individual sin consentimiento. Además, implementa principios como la explicabilidad y la transparencia, exigidos por el RGPD y otras regulaciones europeas en desarrollo.

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.

R/: La Inteligencia Artificial (IA) cumple un rol dual: por un lado, puede ser usada para crear desinformación mediante modelos generativos como los grandes modelos de lenguaje (ChatGPT), los cuales pueden generar textos falsos convincentes rápidamente y en múltiples idiomas. Además, existen herramientas de manipulación de imágenes (inpainting, deepfakes) y de audio (clonado de voz, manipulación de discursos) que permiten crear contenido multimedia falso con gran realismo.

Por otro lado, la IA también sirve para combatir la desinformación, como lo ejemplifica la herramienta FacTeR-Check, desarrollada por el grupo AIDA. Esta plataforma aplica modelos basados en transformers para detectar similitud semántica, inferencia de lenguaje natural y análisis en redes sociales, contrastando la información con bases de datos verificadas. Así, la IA puede ser tanto una amenaza como una solución, dependiendo de su uso.

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?

R/: La IA explicativa (XAI) es crucial para incrementar la confianza del público y de los expertos en los sistemas de detección de desinformación, al permitir comprender por qué una IA clasifica cierto contenido como falso. También tiene un valor educativo, ya que facilita que los usuarios aprendan a detectar patrones de desinformación por sí mismos. La transparencia y la explicabilidad son elementos que refuerzan la credibilidad y ayudan a depurar sesgos o errores del sistema.

Sin embargo, el principal obstáculo es que la mayoría de los modelos más potentes actuales funcionan como “cajas negras”, debido a su complejidad y la imposibilidad de rastrear exactamente cómo toman decisiones, lo que complica su explicación y validación.

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?

R/: Los modelos epidemiológicos, como el modelo SIR, permiten detectar la existencia de flujos anómalos de información, dividiendo a los usuarios en categorías como susceptibles, infectados y recuperados. Sin embargo, no identifican a los actores específicos ni cómo interactúan entre ellos; es decir, su limitación principal es la anonimidad.

En cambio, las redes latentes de difusión permiten modelar las relaciones entre usuarios, identificando quién influye en quién y con qué intensidad. Así, permiten construir mapas de influencia e identificar nodos clave o propagadores de desinformación, lo que resulta mucho más útil para una intervención dirigida. En resumen, los modelos epidemiológicos son útiles para monitoreo general, mientras que las redes latentes lo son para acción directa.

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?

R/: La accesibilidad de los modelos generativos de IA ha hecho que cualquier persona, incluso sin conocimientos técnicos, pueda generar contenido falso con calidad profesional. Esto se debe a que modelos como GPT o Stable Diffusion ahora pueden ejecutarse en equipos domésticos y están disponibles en repositorios abiertos, lo que reduce costos y barreras de entrada.

Esta facilidad de acceso ha contribuido a un aumento significativo del riesgo de campañas masivas de desinformación. Para mitigar estos riesgos, el texto sugiere desarrollar herramientas ágiles y escalables de detección, como FacTeR-Check, que puedan contrarrestar la generación maliciosa mediante análisis semántico, verificación automatizada y monitorización en redes sociales. También se aboga por mayor regulación y supervisión ética de estas tecnologías.

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.

R/: El documento subraya que cualquier herramienta de IA que procese datos personales para detectar desinformación debe cumplir con principios éticos y legales. En Europa, destacan dos normativas principales: el Reglamento General de Protección de Datos (RGPD) y el Libro Blanco de la IA. Estas regulaciones exigen respeto por derechos como la privacidad, la libertad de expresión y la no discriminación. Además, enfatizan la necesidad de transparencia, explicabilidad y supervisión humana.

En el caso de Colombia, aunque no existe aún un marco legal específico equivalente al RGPD europeo, sí se cuenta con una normativa robusta en materia de protección de datos personales, encabezada por la Ley 1581 de 2012.

Esta legislación establece principios similares a los del RGPD, como la finalidad, libertad, veracidad, transparencia, acceso restringido y seguridad de la información. Además, Colombia cuenta con una autoridad de control, la Superintendencia de Industria y Comercio (SIC), que vigila el cumplimiento de estas normas.

Bibliografía

Mochales, R., & Cuadros, M. (2023). *Luchando contra la desinformación mediante la inteligencia artificial*. Universidad Politécnica de Madrid, Grupo AIDA. Recuperado de https://aida.etsisi.upm.es/docs/Luchando_contra_la_desinformacion.pdf