

DEPROFAKE

Defensa Proactiva frente a la Ciberamenaza Desinformativa



Investigación con impacto

En la actual era digital, la variedad de canales de comunicación ha favorecido la emergencia de contenidos falsos y una consecuente falta de credibilidad en la información. DEPROFAKE aborda ese desafío desde una perspectiva social e informática para estudiar sus antecedentes y proponer nuevas formas de respuesta ante este desafío global gracias a herramientas basadas en la Inteligencia Artificial.



Descripción

Este proyecto cubre algunas lagunas que los sistemas e investigaciones actuales no abordan respecto a las **campañas de desinformación y los contenidos falsos** que provocan la manipulación e intoxicación de la opinión pública.

Así, se concentra en:

- **Definición** de los antecedentes y consecuencias de la exposición a contenido falso
- **Propuesta** de mecanismos para su predicción
- **Respuesta** ante los mecanismos que evitan su detección

Aportación de valor

DEPROFAKE contrarresta el problema de los contenidos falsos desde una perspectiva social-in-

formática para aproximarse a los factores concurrentes de las dos disciplinas.

- Examen de los **contenidos falsos más frecuentes**, así como sus **antecedentes** demográficos, económicos, políticos y de alfabetización mediática del contenido fabricado (encuesta panel).
- Investigación de las **consecuencias** de la difusión y la exposición al contenido falso en varios niveles del conocimiento y la participación política.
- Proposición de **nuevos mecanismos de predicción** de la emergencia de este tipo de contenidos falsos gracias a técnicas matemáticas y computacionales.

«DEPROFAKE: Comprender, predecir y frenar los contenidos falsos»

Impacto

El proyecto ha generado **tres herramientas** basadas en Inteligencia Artificial (registradas) que permiten:

- Predecir volúmenes de **noticias falsas**
- Análisis el impacto de **ataques de contaminación**
- Generar **reseñas falsas**.

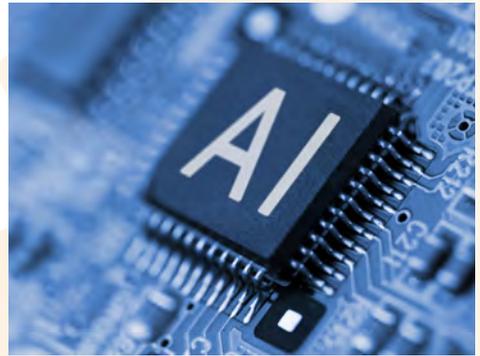
Los resultados del proyecto se han publicado en **revistas y conferencias científicas**, como medio de difusión principal.

Con el fin de transferir el conocimiento a la sociedad y el público general se han llevado a cabo diversas acciones de **diseminación**.



Últimas publicaciones

- 2024. Ibanez-Lissen, Luis; de Fuentes, José María; González-Manzano, Lorena; Anciaux, Nicolas. Continuous Authentication Leveraging Matrix Profile. 19th International Conference on Availability, Reliability and Security (ARES 2024). Springer.
- 2024. Ibanez-Lissen, Luis; González-Manzano, Lorena; de Fuentes, José María; Goyanes, Manuel. Poster: On the feasibility of predicting fake news appearance. Jornadas Nacionales de Investigación en Ciberseguridad (JNIC 2024). JNIC.
- 2024. Ibanez-Lissen, Luis; de Fuentes, José María; González-Manzano, Lorena; Goyanes, Manuel. Use of transfer learning for affordable in-context fake review generation. IEEE Transactions of Big Data.



¿Quieres saber más?

Investigadores principales

Manuel Goyanes

Dpto. Comunicación

José María de Fuentes García

Romero de Tejada

Dpto. Informática

Contacto

mgoyanes@hum.uc3m.es

jfuentes@inf.uc3m.es

Enlaces de interés

[Charla](#) en C1b3rwall (Policía Nacional).

[Mesa redonda](#) "Desinformación y democracia" (Semana de la Ciencia).