

# Detección de la desinformación mediante IA

POSVERDAD A DEBATE | CÁTEDRA DE FILOSOFÍA DE LA UGR

Estamos lejos de crear una herramienta unificada que reciba los bulos, los examine y juzgue para calificarlos como mentiras, contenidos engañosos o verdades

---

**JAVIER CANTÓN**

Universidad de Granada

Martes, 27 diciembre 2022, 00:02



Al aumentar el tráfico de información por Internet, también aumenta el número de mentiras y paparruchas que llegan a nuestras vidas, siendo además cada vez más difíciles de detectar: desde caras falsas de personas que no existen para simular cuentas en redes sociales, a esas mismas cuentas falsas que viralizan una mentira que te indigna y compartes sin pensar en redes, o un vídeo falso de un famoso bailando la 'Despechá', de Rosalía. Son solo algunos ejemplos de la aplicación perversa de las técnicas de Inteligencia Artificial (IA) para engañarnos y expandir desinformación.

Pero estas técnicas de IA también pueden utilizarse para luchar contra esos mismos bulos y engaños. La verificación de información es una parte fundamental del periodismo, pero las circunstancias actuales, con unos

ecosistemas informativos infocicados, la han convertido en un área con entidad propia. No es extraño que hayan aparecido organizaciones especializadas en verificar información como Maldita, Newtral o Verificat, conocidos como 'fact-checkers' o verificadores.

La IA está proporcionando a estos verificadores herramientas avanzadas y potentes que se integran en su flujo de trabajo y les ayudan en su importante labor. Pueden usarse para detectar de manera temprana qué tipo de contenidos falsos podrían hacerse virales, determinar si son verificables (lo que es difícil, por ejemplo, en el caso de afirmaciones irónicas y verdades a medias o fuera de contexto), comprobar si ese contenido ya se ha verificado previamente o asistirles en su tarea de recuperar evidencias y pruebas en contra de esa afirmación falsa.

Estamos lejos de crear una herramienta unificada que reciba los bulos, los examine y juzgue para calificarlos como mentiras, contenidos engañosos o verdades. Tampoco debe ser el objetivo final una máquina que evalúe y juzgue si algo es mentira o no. La verdad es mucho más ambigua, imprecisa y llena de matices y escalas de grises. Estas herramientas y técnicas, a disposición también de cualquier persona que desee emplearlas (como, por ejemplo, la búsqueda inversa de imágenes, que permite rastrear si una imagen ya se ha publicado anteriormente) deben ser una especie de Watson para el Holmes moderno que es el verificador de información: una ayuda para leer una realidad que es compleja y ser capaces así de llegar mejor a la verdad de los hechos.