

IA contra IA: la carrera por detectar imágenes falsas

Por Redacción



Internet está inundado de fotos, vídeos y memes, muchos de ellos difícilmente distinguibles de la realidad, y para tratar de contrarrestar esa avalancha han irrumpido también en el tsunami digital numerosas herramientas y aplicaciones que permiten detectar -aparentemente con un alto grado de fiabilidad-, las imágenes que han sido creadas con inteligencia artificial. Los gobiernos, numerosas empresas tecnológicas y centros de investigación han intensificado los esfuerzos para desarrollar sistemas capaces de distinguir una imagen real de una sintética; y varios expertos analizan con EFE si son fiables y eficaces y si tienen futuro en sociedades (como la española) donde la cultura de la seguridad está muy poco arraigada y los usuarios podrían no ser muy propensos a 'perder' (o ganar) tiempo comprobando la veracidad de las imágenes que reciben. Los fraudes provocados por contenidos generados o modificados con IA provocaron el pasado

año, según revelan diferentes informes, pérdidas que superaron los 1.500 millones de dólares; se usan imágenes falsas de facturas o documentos para reclamar gastos corporativos o validar operaciones; se presentan fotos de daños que en realidad nunca existieron (coches, electrodomésticos, ropa o comida en mal estado); se difunden en medios y redes imágenes de sucesos o situaciones que nunca ocurrieron; y hasta se clona la voz de un superior para ordenar transferencias millonarias.

Numerosas herramientas IMGDetector.ai; RealReveal; Tenorshare Deepfake Detection; AI or Not; o Vericta están ya disponibles -algunas de forma gratuita- y son capaces de detectar imágenes generadas por inteligencia artificial y fotos manipuladas, de analizar inconsistencias o deformaciones, o de comprobar los desajustes que puede haber en los píxeles o las huellas digitales que dejan los sistemas que se utilizan para generar esas imágenes falsas.

Una imagen vale -o no- más que mil palabras

¿Sigue valiendo una imagen más que mil palabras? Georgina Viaplana, cofundadora y directora de la empresa Vericta, mantiene que sí, que no ha cambiado su valor, pero sí la capacidad de las personas de darlas por ciertas sin cuestionarlas; "y precisamente porque una imagen tiene tanto poder merece la pena manipularla". Vericta, que ha sido germinada por

una empresa española especializada en automatizar el cumplimiento de normativas digitales para empresas con el respaldo de varias administraciones y organismos públicos -entre ellos el Centro para el Desarrollo Tecnológico e Industrial o el Ayuntamiento de Barcelona (España)-, ha desarrollado su propia tecnología para saber en segundos si una imagen o un vídeo es real o ha sido generado por IA.

Viaplana se ha mostrado convencida del futuro de este tipo de herramientas y de que se van a convertir en "imprescindibles", ya que la confianza en la autenticidad de los contenidos digitales se ha convertido para las empresas "en una necesidad estructural, no en algo accesorio" y ha precisado que cuando una imagen puede suponer un coste directo para una compañía, la verificación no se percibe como una pérdida de tiempo. Y cita en ese sentido ejemplos como los seguros, donde algunas reclamaciones rápidas se resuelven a partir de una imagen (sin peritaje presencial) que puede estar manipulada; en el comercio electrónico y reparto a domicilio -donde se pueden alterar fotos para reclamar devoluciones-; o en las inmobiliarias, para verificar si una imagen es verdadera o sintética.

Una carrera armamentística digital

Etiquetar en origen cualquier imagen creada con inteligencia artificial podría ser una solución, pero sólo a nivel teórico, porque a su juicio las marcas de agua se pueden también modificar y los metadatos se pueden

cambiar y porque no todos los modelos están sujetos a los mismos estándares y obligaciones, por lo que deberían convivir con sistemas independientes de verificación de contenidos.

El responsable de Operaciones Globales de Consumo de la empresa de ciberseguridad Panda Security, Hervé Lambert, mantiene que estas herramientas tienen futuro, pero no serán infalibles; "es una carrera armamentística digital: cada mejora en detección provoca una mejora en generación", pero ha valorado su utilidad en varios entornos profesionales (medios de comunicación, tribunales o fuerzas de seguridad).

Lambert duda; sin embargo de su eficacia y utilidad entre los usuarios 'medios' y en sociedades, como la española, donde la cultura de la ciberseguridad no está madura. "La mayoría de las personas no contrasta ni una noticia, menos aún una imagen que encaja con su narrativa", ha manifestado a EFE Lambert, y ha puesto un dato sobre la mesa: si el 73 por ciento de los españoles cree que sus teléfonos no necesitan protección, difícilmente van a parar a analizar si una imagen es real o no. A su juicio, sería "deseable" que todas las imágenes creadas con IA estén etiquetadas, algo que técnicamente -mantiene- es perfectamente posible, "pero internet no tiene fronteras, y si una jurisdicción obliga y otra no, el contenido sin etiquetas seguirá circulando", y defiende tres líneas de actuación: la regulación, la responsabilidad de las plataformas y la alfabetización digital, para concluir defendiendo en cualquier caso el pensamiento crítico, porque en materia de ciberseguridad "nunca hay una única bala de plata".