

Instagram avisará qué imágenes publicadas tienen inteligencia artificial

03/04/2023



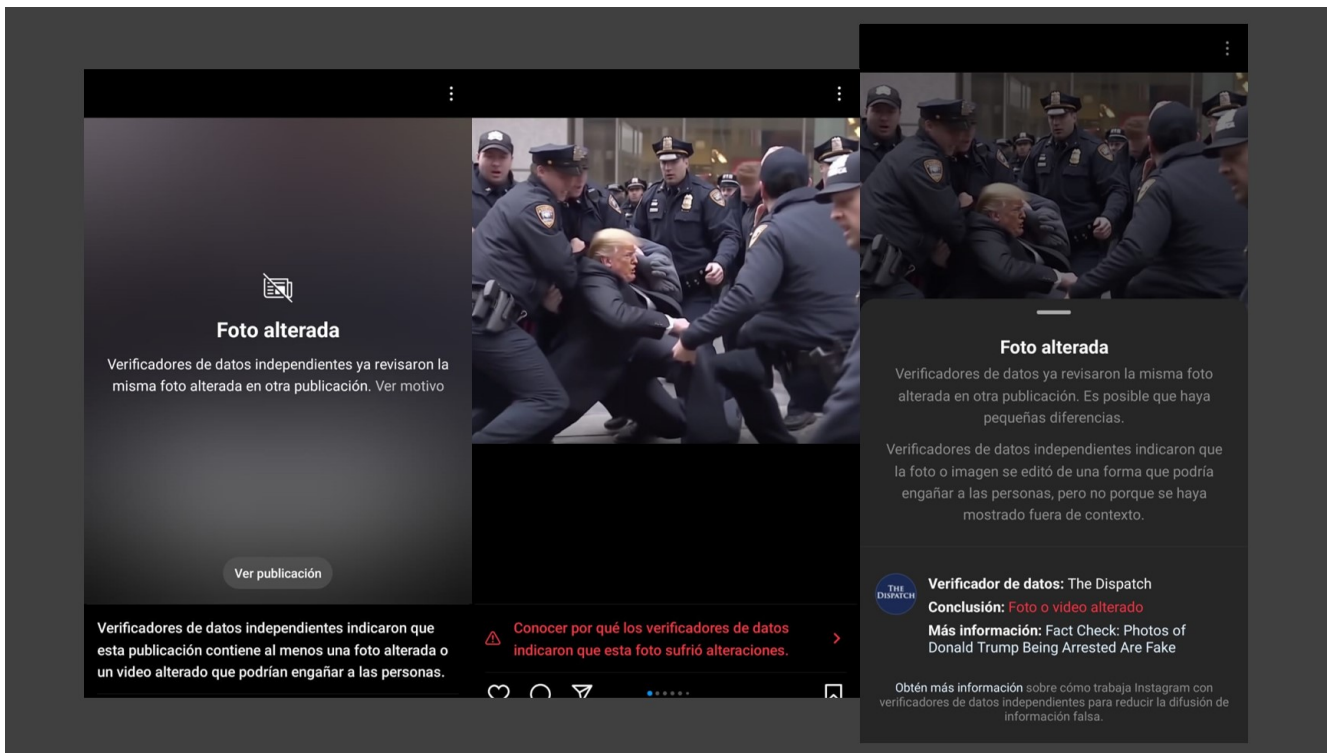
La generación de imágenes con **inteligencia artificial** presenta en la actualidad un gran nivel de realismo que ha generado confusión en algunas personas. Fotografías como las del **Papa Francisco** usando un abrigo Balenciaga o el supuesto arresto del expresidente de Estados Unidos, **Donald Trump**, que fueron producidas usando la tecnología de **Midjourney**, tienen un aspecto tan realista que a simple vista pudieron haber sido consideradas como verdaderas.

Debido a que el uso de esta **tecnología** puede provocar olas masivas de desinformación, tal y como indicó Sam Altman, CEO de OpenAI, la compañía creadora de **ChatGPT**, las redes sociales como **Instagram** han generado recursos que permitirán a las personas identificar cuando están viendo contenido que fue creado por **inteligencia artificial**.

Según la red social, “si los **verificadores** de información externos detectan información falsa, contenido alterado o contenido al que le falta contexto en Instagram, hacemos que sea más difícil de encontrar”.

Entre los métodos usados se encuentra la reducción de la efectividad de los hashtags, además de evitar que aparezcan en la sección **Explorar** e **historias**.

También, se ha desarrollado un sistema de **etiquetas** en el que se puede identificar casos de información falsa y los usuarios son quienes deciden si exponerse o no a este tipo de contenido, además de compartirlo en otras secciones de **Instagram**.



Verificación de imágenes generadas con inteligencia artificial en Instagram. (Infobae)

Las publicaciones que contengan esta característica podrán ver un filtro de difuminado sobre la imagen con el mensaje “Foto alterada. **Verificadores** de datos independientes ya revisaron la misma foto alterada en otra **publicación**.”

Si se comparte el contenido con la etiqueta de advertencia, la

misma aparecerá en las plataformas en las que se comparta. El **texto** también incluirá un enlace directo a la calificación del verificador de la información que la emitió. Por ejemplo, la fotografía de Donald Trump siendo arrestado recibió la calificación de “**Foto o video alterado**”.

Esta calificación se mostrará en una ventana emergente en la que se indicará el nombre de la organización de **verificación** de datos y se brindará información adicional en un texto como “la foto o imagen se editó de una forma que podría engañar a las personas, pero no porque se haya mostrado fuera de contexto”.

Se incluirán enlaces que redirijan a los usuarios a **artículos** de otras fuentes de **confianza** que desacreditan las afirmaciones de la publicación verificada.

Qué pasa con las cuentas que comparten desinformación

Según Instagram, es posible que una persona pueda compartir una **publicación** que contenga datos que promuevan la desinformación. Sin embargo, si este comportamiento se vuelve recurrente puede suponer sanciones a los usuarios, que serán empleadas de forma gradual para restringir el impacto de las **imágenes** compartidas.



Publicar imágenes que generan desinformación puede ser sancionado con la reducción del alcance de la cuenta en Instagram. (REUTERS/Dado Ruvic)

Por ejemplo, si se comparte este tipo de contenido es posible que se inhabiliten las menciones o etiquetas de estas cuentas por parte de otros usuarios. De igual forma, cada persona que intente seguir a este perfil verá una advertencia de contenido que indicará que la cuenta ha compartido información falsa previamente.

Limitar la visibilidad de las publicaciones de estas cuentas también es una forma que tiene el sistema de **Instagram** para evitar que la desinformación se difunda a más personas. Como último recurso, es posible que se ejecute la desactivación de una cuenta como producto de la difusión recurrente de desinformación.

En este tipo de situaciones, los usuarios son denunciados por incumplimiento de las **Normas Comunitarias** y aún cuando es posible recurrir a otras instancias para prevenir la eliminación de la cuenta, el comportamiento recurrente será

tomado en cuenta para tomar una decisión al respecto. Esto no solo incluye a imágenes generadas por **inteligencia artificial**, sino también a todas las publicaciones que sean señaladas por parte de otros usuarios como contenedoras de desinformación.

Fuente: Infobae