

Facebook y Twitter usarán inteligencia artificial para limitar el ciberacoso

written by Otros medios | 19/10/2021

Tras un año de pandemia, ya es definitiva la transición de gran parte de las discusiones de la sociedad al entorno digital de las redes sociales. Lo que no ha quedado tan claro es, cómo pueden los propietarios de estas plataformas asegurarles a sus usuarios una buena experiencia con respecto **al ciberacoso o simplemente evitar un ecosistema tóxico**. Su respuesta es la **Inteligencia Artificial**:

Recientemente, Facebook ha sufrido acusaciones de reportes sobre datos extraídos de documentos filtrados los cuales, según alega la empresa de Mark Zuckerberg, crean una narrativa engañosa sobre el progreso de sus tecnologías que ayudan a identificar, ocultar y posiblemente eliminar contenidos que inciten al odio. Justificaban en su [About Facebook](#) blog:

Los informes sugieren que nuestro enfoque para abordar el discurso de odio es mucho más limitado de lo que realmente es, ignorando el hecho de que la prevalencia del discurso de odio se ha reducido al 0.05%, es decir, 5 visitas de cada 10.000... Cuando empezamos a informar de estas métricas en 2016, nuestros sistemas solo detectaron el 23,6% del contenido que eliminamos, mientras que ahora esa cifra asciende al 97%. Y la prevalencia de estos contenidos se ha visto reducida en un 50% en los últimos 3 meses...

Mientras que en Twitter, acaba de llegar parte de los usuarios de IOS y Android una funcionalidad que, según explicaba la cuenta de [Twitter Support en su tweet](#), avisa de que una conversación es violenta, intensa o «acalorada». El mecanismo de advertencia no deja del todo claro qué se

entiende como intenso, pero ya sabemos que este tipo de herramientas nunca son infalibles.

Muchos usuarios denuncian que esta herramienta es contraproducente y que en vez de ocultar conversaciones de este tipo, se debería centrar en usuarios concretos. Ya que la mayoría de conversaciones acaloradas no són más que un par de usuarios que, tras la protección que les otorga el anonimato, causan caos allá donde van. Por otro lado, otros usuarios también denuncian la [limitación de la libertad de expresión](#), pero habrá que ver como avanza la implementación.

Fuente: Holahorro.mx