LT PULSO
LATERCERA

Fecha: 04/09/2024 Audiencia: 271.020 Sección: ACTUALIDAD
Vpe: $7.836.831 Tirada: 76.017 Frecuencia: DIARIO
Vpe pág: $9.570.938 Difusión: 76.017
Vpe portada: $9.570.938 Ocupación: 81,88%

Mickey Mouse fuma: cómo las herramientas de IA para imágenes están generando nuevos problemas de moderación de contenidos

xAI y Google lanzaron productos que han creado controversia sobre los límites éticos y legales. La posibilidad de crear imágenes de personas reales y conocidas ha surgido como uno de los mayores puntos de tensión en este nuevo debate.

Meghan Bobrowsky/Miles Kruppa
THE WALL STREET JOURNAL

Una muestra a Mickey Mouse bebiendo cerveza. Otra muestra a Bob Esponja vestido de nazi. Una tercera: Donald Trump y Kamala Harris besándose.

Estas son algunas de las muchas extrañas y a veces inquietantemente vívidas imágenes generadas por las nuevas herramientas de Google y xAI de Elon Musk, que están provocando un debate sobre si las empresas tecnológicas pueden, y cómo, controlar la producción del software de creación de imágenes de IA (inteligencia artificial) de vanguardia.

Los generadores de imágenes de inteligencia artificial están poniendo a prueba los límites de las políticas de las plataformas y la capacidad de las empresas para colocar barreras eficaces al uso público de esta nueva y potente tecnología visual. Según los expertos en moderación de contenidos, estas herramientas pueden difundir información errónea durante un ciclo electoral.

La semana pasada, Google anunció que volverá a permitir que su chatbot Gemini genere imágenes de personas, seis meses después de que la empresa pusiera en pausa esta función a raíz de las reacciones en línea porque había producido imágenes de soldados nazis de diversas razas. En un principio, Google sólo ofrecerá esta función a los usuarios de habla inglesa que paguen por una versión premium de Gemini.

La posibilidad de crear imágenes de personas reales y conocidas ha surgido como uno de los mayores puntos de tensión en este nuevo debate sobre la moderación de contenidos. Varias compañías, entre ellas Google y OpenAI, no permiten que las herramientas generen imágenes de personas concretas y reconocibles.

Un nuevo generador de imágenes de IA de xAI, la empresa de Elon Musk, sí lo hace. La *startup* fue criticada por los organismos de control de la industria tecnológica tras presentar a principios de agosto un gran modelo lingüístico llamado Grok-2 con capacidad para generar imágenes.

Musk, que se ha descrito a sí mismo como un absolutista de la libertad de expresión, ha criticado a Google por las imágenes ahistóricas que produjo su herramienta de IA.

Su generador de imágenes, impulsado por una *startup* alemana llamada Black Forest Labs y sólo disponible para suscriptores de pago en la plataforma de medios sociales X, ha producido imágenes de políticos en situaciones comprometidas o desagradables y otras de personajes protegidos por derechos de autor, como Mickey Mouse, haciendo cosas ofensivas como saludar a Adolf Hitler.

Las firmas detrás de los generadores de imágenes de IA también se enfrentan a amenazas legales por las imágenes que utilizaron para entrenar sus productos. Stability AI y Midjourney son dos de estas empresas de generación de imágenes que han sido llevadas a los tribunales por artistas que alegan que les han infringido sus derechos.

Stability se rehusó a hacer comentarios sobre el litigio, y Midjourney tampoco hizo comentarios. El abogado de los artistas tampoco habló.

Exclusiones y filtros

OpenAI, que lanzó su generador de imágenes llamado DALL-E al público en general en 2022, añadió el año pasado la po-

sibilidad de que los creadores excluyeran sus imágenes de entrenamiento de futuras versiones de la herramienta, después de que la empresa recibiera amenazas legales. News Corp, propietaria del Wall Street Journal, tiene un acuerdo de licencia de contenidos con OpenAI.

xAI lanzó su generador de imágenes el 13 de agosto. En los días siguientes, los usuarios de X inundaron la plataforma con imágenes que decían haber sido generadas por Grok-2, incluidas las de Trump y Harris intimando y Mickey Mouse sosteniendo una pistola. A partir del miércoles, algunas de las imágenes ya no se podían encontrar en la plataforma, pero era posible crearlas utilizando Grok-2.

xAI, Musk, los representantes de Musk y Black Forest no hicieron comentarios.

Un día antes de que comenzara la Convención Nacional Demócrata en agosto, Trump publicó lo que parecía ser una imagen generada por IA de Harris dando un discurso en Chicago, donde se celebraba la convención, con una bandera roja con una hoz y un martillo colgando de fondo, dando a entender que Harris es comunista.

Las políticas de X prohíben a los usuarios compartir medios manipulados que puedan confundir o engañar a la gente y provocar daños. La campaña de Trump no hizo comentarios.

Musk, el propietario de la plataforma, volvió a publicar recientemente un falso anuncio de campaña manipulado de Harris en el que dice que es la "última contratación de diversidad". X no dio comentarios.

Es probable que los generadores de imágenes por IA se enfrenten a los mismos problemas que tienen las redes sociales tradicionales, con personas que utilizan las herramientas para hacer deepfakes y difundir desinformación, dijo Sarah T. Roberts, profesora de la Universidad de California en Los Ángeles, que estudia la moderación de contenidos.

"Todos esos problemas que han estado presentes en las redes sociales tradicionales están ahí, y en algunos casos son más difíciles de detectar", comentó, añadiendo que los elementos visuales pueden ser a veces más convincentes.

Según Pinar Yildirim, profesor de la Universidad de Pensilvania, otro problema es que la gente aprende rápidamente a eludir las palabras clave prohibidas para mantener determinados contenidos fuera de las plataformas.

"Empiezan a ser más inteligentes a la hora de crear esos contenidos", aseguró.

Creación de salvaguardias

Antes de lanzar una nueva versión de su chatbot Gemini en febrero, los empleados de Google le dieron instrucciones para que creara imágenes diversas cuando se le pidieran representaciones de personas. Esperaban que las instrucciones protegieran contra los sesgos habituales en los generadores de imágenes, como la tendencia a producir hombres blancos cuando se le pedían imágenes de médicos.

En cambio, los usuarios de X empezaron a publicar capturas de pantalla de imágenes ahistóricas creadas con Gemini. Musk avivó el fuego mientras presentaba su propio *chatbot*.

El CEO de Google, Sundar Pichai, señaló que era inaceptable que las salidas de Gemini hubieran ofendido a los usuarios y mostrado sesgo, y que la compañía lo "arreglaría a escala". La empresa comunicó el miércoles que había hecho progresos significativos en la generación de imágenes, pero que no todas las imágenes que Gemini crea serán perfectas.

"Queremos asegurarnos de que el modelo sigue las instrucciones", sostuvo Sissie Hsiao, vicepresidente de Google que supervisa el *chatbot* Gemini, en una entrevista reciente. "Este es uno de nuestros principios: Gemini es del usuario, así que servimos a sus órdenes", explicó.

Hace poco, Google enfadó a los usuarios de otra manera: siendo demasiado permisivo. El mes pasado, el gigante tecnológico lanzó un nuevo teléfono Pixel que permitía a los usuarios producir imágenes generadas por IA del personaje de dibujos animados Bob Esponja con una esvástica.

Un representante de Google dijo que la compañía estaba "mejorando y refinando continuamente las salvaguardas que tenemos en marcha".

Cuestiones jurídicas

Más allá de los dilemas éticos, sigue habiendo dudas sobre las posibles responsabilidades legales de las empresas.

Además de la demanda colectiva de los artistas, Stability AI se enfrenta a una demanda de Getty Images, que alega que la empresa de IA infringió sus derechos para entrenar su modelo.

Stability AI se rehusó a hacer comentarios sobre el litigio. Una representante de Getty Images dijo que la firma no hace comentarios sobre litigios activos. Getty ha lanzado su propio generador de imágenes por IA, añadió.

La demanda colectiva se dirige contra una serie de empresas de IA que han lanzado generadores de imágenes, entre ellas Midjourney, DeviantArt y Runway.

Tanto los casos de artistas como los de Getty Images están pendientes. DeviantArt y Runway no hicieron comentarios.

Según Geoffrey Lottenberg, abogado especializado en derechos de propiedad intelectual, estas batallas legales podrían sentar un precedente sobre las imágenes y los datos que las empresas de IA pueden utilizar para entrenar a sus *chatbots*.

"En algún momento se aclarará", aseguró Lottenberg. WSJ

Traducido del idioma original por PULSO.