LITORALPRESS

Fecha: 20-09-2024
Medio: El Mercurio
Supl. : El Mercurio - Cuerpo B
Tipo: Noticia general
Título: Cómo los instrumentos de imagen de IA están generando nuevos problemas de moderación de contenido

Pág. : 5
Cm2: 811,7
VPE: $ 10.661.888

Tiraje: 126.654
Lectoría: 320.543
Favorabilidad: No Definida

Regulación:

Cómo los instrumentos de imagen de IA están generando nuevos problemas de moderación de contenido

xAI y Google lanzaron productos que han creado controversia sobre los límites éticos y legales.

MEGHAN BOBROWSKY Y MILES KRUPPA
The Wall Street Journal

Uno muestra a Mickey Mouse tomando una cerveza. Otro muestra a Bob Esponja con un atuendo nazi. Un tercero: Donald Trump y Kamala Harris besándose.

Estos elementos visuales están entre las muchas imágenes extrañas y a veces inquietantemente vívidas producidas por los nuevos instrumentos de Google y xAI de Elon Musk que están generando un debate sobre cómo —y si— las compañías tecnológicas pueden controlar la producción de un *software* vanguardista de creación de imágenes con inteligencia artificial (IA).

Los generadores de imágenes con IA están poniendo a prueba los límites de las políticas de las plataformas y la capacidad de las empresas para poner barreras de seguridad eficaces en torno al uso público de esta nueva y poderosa tecnología visual. Los instrumentos tienen el potencial de propagar información falsa durante un ciclo electoral, dicen expertos en moderación de contenido.

Google informó a fines de agosto que permitirá que su chatbot Gemini genere imágenes de personas de nuevo, seis semanas después de que la compañía interrumpió la función luego de la reacción que hubo en línea porque había producido imágenes racialmente diversas de soldados nazi. Google señaló que en un principio va a desplegar la función solo para usuarios en idioma inglés que pagan por una versión *premium* de Gemini.

La capacidad para crear imágenes de personas reales conocidas surgió como uno de los puntos principales de tensión en este nuevo debate sobre la moderación de contenido. Una serie de compañías, entre ellas Google y OpenAI, no permitirá que los instrumentos generen imágenes de personas específicas y reconocibles.

Un generador de imágenes con IA que lanzó hace poco xAI de Elon Musk sí lo hace. La nueva empresa fue criticada por organismos de vigilancia de la industria tecnológica después de que diera a conocer un modelo grande de lenguaje llamado Grok-2 con capacidades de generación de imágenes a principios de agosto.

Musk, quien se ha autocalificado como un absolutista de la libertad de expresión, criticó a Google por las imágenes no históricas que produjo su instrumento de IA.

Su generador de imágenes, el que es activado por un emprendimiento alemán llamado Black Forest Labs y solo está disponible para suscriptores que pagan en la plataforma de redes sociales X, ha producido imágenes de políticos en situaciones comprometedoras o desagradables y otras de personajes protegidos por derechos de autor como Mickey Mouse haciendo cosas ofensivas como saludar a Adolf Hitler.

Las empresas que están detrás de los generadores de imágenes con IA también se enfrentan a amenazas legales por las imágenes que han utilizado para capacitar sus productos. Stability AI y Midjourney son dos de esos emprendimientos de generación de imágenes que han sido demandados por artistas que sostienen que violaron sus derechos.

Stability declinó entregar algún comentario sobre el litigio, y Midjourney no respondió a una solicitud de entregar algún comentario. Un abogado de los artistas tampoco respondió a una solicitud de emitir un comentario.

Una imagen generada por inteligencia artificial que circulaba en X muestra retratos de la vicepresidenta Kamala Harris y del expresidente Donald Trump.

Exclusión voluntaria y filtros

OpenAI —que lanzó su generador de imágenes llamado DALL-E al público en general en 2022— agregó el año pasado para los creadores la posibilidad de excluir sus imágenes de capacitación de futuras versiones del instrumento después de que la compañía recibió amenazas legales. News Corp, dueño de The Wall Street Journal, tiene una asociación de licencia de contenido con OpenAI.

xAI lanzó su generador de imágenes el 13 de agosto. En los días posteriores, usuarios de X inundaron la plataforma con imágenes que, según ellos, fueron generadas por Grok-2, entre ellas unas de Trump y Harris en momentos de intimidad uno con el otro y Mickey Mouse sosteniendo un arma. Al día siguiente, algunas de las imágenes ya no se podían encontrar en la plataforma, pero era posible crearlas haciendo uso de Grok-2.

Musk, xAI, representantes del magnate y Black Forest no respondieron a las solicitudes de entregar algún comentario.

Un día antes de que empezara la Convención Nacional Demócrata en agosto, Trump publicó lo que parecía ser una imagen generada con IA de Harris dando un discurso en Chicago, donde se realizaba la convención, con una bandera roja con la hoz y el martillo como telón de fondo, lo que implicaba que Harris es comunista.

Las políticas de X prohíben que los usuarios compartan contenidos manipulados que pudieran confundir o engañar y hacer daño. La campaña de Trump no respondió a las solicitudes de entregar algún comentario.

Musk, el dueño de la plataforma, hace poco volvió a publicar un aviso de campaña falso manipulado de Harris en el que ella dice que es la "contratación fundamental en cuanto a diversidad". X no respondió a solicitudes de entregar algún comentario.

Es probable que los generadores de imágenes con IA enfrenten los mismos problemas que tienen las redes sociales tradicionales, en que personas utilizan los instrumentos para crear 'deepfakes' (falsificaciones de videos, imágenes o audios) y difundir información falsa, observó Sarah T. Roberts, profesora de la Universidad de California, en Los Ángeles, quien estudia moderación de contenido.

PELIGRO
Las nuevas herramientas de IA tienen el potencial de propagar información falsa durante un ciclo electoral.

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

Otro problema es que las personas aprenden rápido cómo evitar las palabras clave que han sido prohibidas en un esfuerzo por mantener cierto contenido fuera de las plataformas, señaló Pinar Yildirim, profesora de la Universidad de Pensilvania."Empiezan a ser más inteligentes sobre cómo ser capaces de crear ese contenido", dijo.

Crear salvaguardas

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 le pidieran representaciones de personas. Esperaban que las instrucciones protegieran contra los prejuicios comunes en los generadores de imágenes, como la tendencia de producir hombres blancos cuando se le piden imágenes de médicos.

En cambio, los usuarios de X empezaron a publicar pantallazos de imágenes no históricas que se crearon con Gemini. Musk avivó las llamas mientras promovía su propio chatbot.

El director ejecutivo de Google, Sundar Pichai, manifestó que era inaceptable que los resultados de Gemini hubieran ofendido a los usuarios y mostrado prejuicios, y que la compañía "lo repararía a gran escala". La empresa informó luego que había hecho un avance significativo en la generación de imágenes pero que no todas las imágenes que cree Gemini serán perfectas.

"Queremos asegurarnos de que el modelo siga las instrucciones", dijo Sissie Hsiao, vicepresidenta de Google a cargo de la supervisión del chatbot Gemini, en una entrevista reciente. "Este es uno de nuestros principios; este es el Gemini del usuario, así es que estamos a sus órdenes".

Google hace poco irritó a los usuarios en una forma diferente; por ser demasiado permisivo. El mes pasado, el gigante tecnológico lanzó un nuevo teléfono Pixel que permitía a los usuarios producir imágenes generadas con IA del personaje de dibujos animados Bob Esponja usando una suástica.

Un vocero de Google señaló que la compañía estaba "mejorando y perfeccionando continuamente las salvaguardas que hemos puesto en práctica".

Dudas legales

Más allá de los dilemas éticos, persisten las dudas sobre las posibles responsabilidades legales de las compañías.

Además de la demanda colectiva de artistas, Stability AI está enfrentando una demanda de Getty Images, que sostiene que la compañía de IA violó sus derechos para capacitar su modelo. Stability AI declinó hacer algún comentario sobre el litigio. Una vocera de Getty Images comunicó que la compañía no hace comentarios sobre un litigio activo. Getty lanzó su propio generador de imágenes con IA, agregó.

La demanda colectiva apunta a una serie de compañías de IA que han lanzado generadores de imágenes, entre ellas Midjourney, DeviantArt y Runway.

Las causas tanto de los artistas como de Getty Images están pendientes. DeviantArt y Runway no respondieron a las solicitudes de entregar algún comentario.

Estas batallas legales podrían sentar un precedente en cuanto a qué imágenes y datos pueden utilizar las compañías de IA para capacitar sus chatbots, observó Geoffrey Lottenberg, un abogado que se especializa en derechos de propiedad intelectual.

"Vamos a tener claridad sobre esto en algún momento", aseguró Lottenberg.

Artículo traducido del inglés por "El Mercurio".