GZERO AI

everything is political

SUMARIO:

- Secretos robados de Google se dirigen a China
- Sosorando el silbato en el generador de imágenes de Microsoft
- Una solución nuclear al problema energético de la IA
- IA a prueba de elecciones
- **Además: El** último "retrato" familiar de **Kate Middleton**

Sintonizar:¿ En la era de la IA generativa, cómo puede la tecnología convertirse en una herramienta para crear más oportunidades para las mujeres y las niñas? Acompáñanos el lunes 18 de marzo, a las 12 p.m. ET para el estreno en vivo de Gender Equality in the Age of AI**,** nuestra próxima discusión Global Stage, presentada por GZERO en asociación con Microsoft y la Fundación de la ONU. Nuestro panel de expertos discutirá formas inclusivas de hacer nuestras vidas digitales más seguras y productivas a medida que se reúnen al margen de la 68a Comisión de la Condición Jurídica y Social de la Mujer de las Naciones Unidas, una reunión de líderes de los Estados miembros de la ONU y ONG enfocadas en el progreso y la igualdad.

El equipo GZERO

Lo que estamos mirando: Un informante en Microsoft, Going nuclear, hackeados electorales

Imágenes problemáticas plagan Microsoft-s Copilot

Sólo semanas después de que Google suspendiera su generador de texto a imagen de Gémini por producir imágenes ofensivas, Microsoft se enfrenta a una agitación similar por uno de sus productos.

Según CNBC, que replicó los resultados, un ingeniero de Microsoft pudo utilizar su herramienta Copilot para generar "demons y monstruos", junto a la terminología relacionada con el derecho al aborto, adolescentes con rifles de asalto, imágenes sexualizadas de mujeres en cuadros violentos y consumo de alcohol y drogas por parte de menores.

También generó imágenes perturbadoras que se duplicaron como posibles violaciones de derechos de autor, como pistolas de la marca Disney, latas de cerveza y bolígrafos de vape. Y se hace más preocupante: La herramienta creó

imágenes de Elsa de "Frozen" en medio de los restos en la Franja de Gaza, pero también con un uniforme de las Fuerzas de Defensa de Israel.

El empleado de Microsoft, **Shane Jones**, notificó recientemente a la Comisión Federal de Comercio de lo que vio mientras trabajaba como un rojibeo encargado de probar esta tecnología, que es impulsada por OpenAI a través de una asociación con el fabricante de ChatGPT y DALL-E.

En respuesta, Microsoft ha comenzado a bloquear algunos de los términos que generaron imágenes ofensivas, incluyendo "pro-elegen", "pro-life", y "cuatro veinte". Microsoft dijo a CNBC que los cambios se debían a monitorear, hacer ajustes y establecer controles adicionales.

Esto refleja un ciclo continuo: Los peores abusos de la IA generativa sólo llegarán a través de las pruebas de la gente y descubrirán exactamente qué horrores puede producir, lo que conducirá a políticas de uso más estrictas y nuevos límites para empujar. Por supuesto, cuando se trata de violaciones de derechos de autor, el ciclo puede ser interrumpido muy rápidamente por demandas de los titulares de IP desesperados por proteger sus marcas.

El problema de la energía de IA tiene una solución nuclear?

Sam Altman, el cofundador y CEO de OpenAI, tiene amplias ambiciones de resolver todos los problemas de la IA, desde algoritmos hasta chips de alta tecnología. Pero hay un problema más en su plato: la energía. Altman está respaldando una serie de compañías que esperan encontrar una manera de poder la tecnología revolucionaria, literalmente.

Una de las startups en las que Altman invirtió se llama Oklo, que está construyendo una central nuclear en Idaho que eventualmente podría alimentar centros de datos de los que la IA depende, pero no hay un calendario público claro para el proyecto. Google y Microsoft también se han asociado con las firmas de energía nuclear para sus necesidades energéticas.

La energía nuclear conlleva riesgos, por supuesto, y Oklo ha tenido problemas con los reguladores, que rechazaron las solicitudes en el pasado basadas en la falta de información sobre seguridad y protección proporcionada. Pero ir nuclear, si empresas como Oklo pueden hacerlo bien, también es una alternativa más limpia a las fuentes de energía más emisoras de carbono.

Las salvaguardas de las elecciones de IA no son geniales

El Centro de Lucha contra el Odio Digital está probando las herramientas de IA más populares para ver si son capaces de ser manipulados para generar desinformación electoral a pesar de las promesas públicas y las reglas de uso en sentido contrario.

La organización sin fines de lucro británica utilizó Midjourney, OpenAI's ChatGPT, Stability.ai's DreamStudio y Microsoft's Image Creator para probar en febrero, simplemente atando diferentes mensajes de texto relacionados con las elecciones estadounidenses. El grupo encontró que era capaz de eludir las herramientas. protecciones un enorme 41% de las veces.

Algunas de las imágenes que crearon mostraban a **Donald Trump** siendo llevado esposado, a Trump en un avión con presunto pedófilo y traficante de humanos **Jeffrey Epstein**, y **a Joe Biden** en una cama de hospital.

La IA Generativa ya está jugando un papel tangible en las campañas políticas, especialmente cuando los votantes acuden a las urnas para las elecciones nacionales en 64 países diferentes este año. AI se ha utilizado para ayudar a un ex primer ministro a sacar su mensaje de prisión en Pakistán, para convertir a un ministro de Defensa endurezado en un personaje adorable en Indonesia, y para hacerse pasar por Biden en New Hampshire. Las protecciones que fracasan casi la mitad del tiempo que no lo cortan. Con la regulación rezagada del ritmo de la tecnología, las empresas de IA se han comprometido voluntariamente a impedir la creación y difusión de los medios de comunicación de IA relacionados con las elecciones.

Todas estas herramientas son vulnerables a que las personas que intentan generar imágenes que podrían ser usadas para apoyar las afirmaciones de una elección robada o podrían ser usadas para disuadir a la gente de ir a lugares de votación", le dijo **a** la BBC la CCDH. Si hay voluntad por parte de las empresas de IA, pueden introducir salvaguardas que funcionen

GZERO Exclusive: El autor más vendido Yuval Noah Harari llama a AI un arma de destrucción masiva social.

La IA tiene el poder de destrozar la sociedad en las costuras? El autor más vendido **Yuval Harari** advierte que podría. La IA es la primera tecnología en la historia que nos puede quitar el poder, dijo en una amplia conversación exclusiva de GZERO con **Ian Bremmer** en la calle 92 Y en la ciudad de Nueva York.

Escucha más sobre la amenaza que Harari ve a los dictadores, la democracia y las relaciones ordinarias, ya que la tecnología nos deja sin poder confiar en estos tiempos cambiantes.

Mira este clip.aquí., y atrae toda la entrevista y más de "GZERO World con Ian Bremmer" cada semana en línea y en la televisión pública de EE.UU. Revisa los listados locales.

Nacional chino acusado de robar secretos comerciales de Google

Imagen cortesía de Midjourney

Linwei Ding, de nacionalidad china residente en California, fue detenido y acusado el pasado miércoles pasado por supuestamente robar secretos comerciales relacionados con inteligencia artificial de Google y trasladarlos a sus empresas chinas. Ding, que trabajaba para Google, supuestamente tomó más de 500 archivos confidenciales de su empleador y los utilizó en su trabajo con dos empresas en China, una que fundó, la otra que lo contrató y dijo a los inversores que era el director de tecnología.

Ni Ding ni su abogado han hecho comentarios públicos sobre el caso.

Bill Hannas, director del Centro de Seguridad y Tecnología Emergentes de la Universidad de Georgetown y ex experto de la CIA, dijo que tanto un caso de individuo supuestamente se enriquece a sí mismo robando valiosos secretos comerciales como una amenaza para la seguridad nacional de Estados Unidos.

Conocido los casos de robo de plano, donde China es el beneficiario, número de cientos, dijo Hannas. Pero ha habido decenas de miles de casos en general en los

que la tecnología estadounidense terminó en China por medios desconocidos u oscuros, agregó.

El director de Geotecnología de Eurasia, **Xiaomeng Lu**, dice que aún no está claro si hay alguna participación directa del gobierno chino en el caso de Dings.

"Quasiste que el FBI tenga más información sobre el caso que no han revelado", dijo Lu, pero lo que he visto en los informes de los medios no sugiere explícitamente que la información robada tiene serias implicaciones de seguridad nacional. La información de diseño de chip y software se lee más como secretos comerciales que secretos de seguridad nacional.

Los Estados Unidos tratan de restringir el acceso de China a una tecnología valiosa que ayude a sus esfuerzos por desarrollar sofisticados modelos de IA y capacidades de computación. Washington reforzó los controles de exportación sobre la tecnología de semiconductores a principios de este año para cortar a China de los tipos de chips de alta potencia necesarios para ejecutar modelos de IA.

Los controles de exportación animan a China a encontrar otras maneras de conseguir lo que necesitan, dijo Hannas. Los chips son un área donde se dice que China se rezaga en comparación, pero el otro gran déficit, reconocido por los propios científicos chinos, son algoritmos de IA. Es más difícil reprimir el progreso científico que los materiales específicos, y este tipo de espionaje corporativo es una forma de ganar la paridad con los EE.UU.

En última instancia, a Washington le preocupa que China supere algún día al ejército estadounidense con armamento autónomo impulsado por AI. El robo de tecnología innovadora y secretos comerciales de las empresas estadounidenses puede costar empleos y tener devastadoras consecuencias económicas y de seguridad nacional, escribió el director del FBI, **Christopher Wray**, en un comunicado de prensa.

Los departamentos de Justicia y Comercio. Fuerza de Ataqueo de Tecnología Interruptiva, establecida el año pasado, se ha centrado en el uso de controles de exportación para cortar a adversarios extranjeros, incluyendo Irán y Rusia, además de China. La Fiscal General Adjunta **Lisa Monaco** escribió que, como parte de estos esfuerzos, el Departamento de Justicia perseguirá y responsabilizará implacablemente a aquellos que sifieran tecnologías disruptivas, especialmente AI, por exportación ilegal.

Un caso de espionaje corporativo dirigido a Google puede ser una manera de cumplir con esa preocupación, pero la respuesta agresiva del gobierno indica una voluntad de vigilar de cerca cualquier tecnología de IA que abandone el país, especialmente por medios ilícitos.

De ser declarado culpable, Ding enfrentará hasta 10 años de prisión por cada cargo.

Números duros: Entendiendo el universo, Apertura OpenAI, advertencia de bioarma, Revisión independiente, IA medios de comunicación miles de millones

100 millones: AI está ayudando a los investigadores a trazar un mejor mapa del espacio exterior. Una simulación reciente dirigida por un investigador del University College London fue capaz de mostrar 100 millones de galaxias a través de un cuarto del cielo del hemisferio sur de la Tierra. Esto es parte de un esfuerzo más amplio para entender la energía oscura, la fuerza misteriosa que causa la expansión del universo.

30.000: El bufete de abogados WilmerHale, que completó su investigación de **la** breve expulsión de **Sam Altman en** noviembre de OpenAI, examinó 30.000 documentos como parte de su revisión. El contenido del informe no se hizo público, pero el nuevo presidente de la junta, Bret Taylor, dijo que la revisión encontró que la junta anterior actuó de buena fe, pero no anticipó la reacción a la eliminación de Altman, quien ahora se está reincorporando a la junta. La SEC, mientras tanto, todavía está investigando si OpenAI engañó a los inversores, pero no está claro si WilmerHale dará sus conclusiones a la agencia.

90: Más de 90 científicos se han comprometido a no utilizar la IA para desarrollar armas biológicas como parte de un acuerdo forjado en respuesta a las declaraciones del Congreso dadas por el director general antrópico **Dario Amodei** el año pasado. Amodei dijo que mientras que la actual generación de tecnología de IA no podría manejar tal tarea, está a sólo dos o tres años de distancia.

100: Más de 100 investigadores de IA han firmado una carta abierta pidiendo a las principales empresas que permitan a los investigadores independientes acceder a sus modelos para asegurarse de que la evaluación del riesgo sea exhaustiva. Las empresas de IA Genero deberían evitar repetir los errores de las plataformas de redes sociales, muchas de las cuales han prohibido efectivamente los tipos de investigación destinadas a responsabilizándolas, dijo la carta.

8 mil millones: La compañía de medios Thomson Reuters dice que tiene un cofre de guerra de 8 mil millones de dólares para gastar en adquisiciones relacionadas con la IA. Además de publicar la línea de noticias Reuters, la compañía vende acceso a servicios como Westlaw, una popular plataforma de investigación legal. También se comprometió a gastar al menos 100 millones de dólares en desarrollar tecnología interna de IA para integrarse en sus ofertas de noticias y datos.

Desde la galería

Cada semana, vamos con un espécimen de IA generativa, algo bueno, otro malo, y otros con demasiados dedos o dientes. Esta semana: ***Kate Middleton*** *se metió en*

un poco de agua caliente por publicar una imagen de fotos de su familia. Tratamos de ver si la IA podría ser un poco más sutil.

Esta edición de GZERO AI fue escrita por Scott Nover. Editado por Matthew Kendrick y Tracy Moran.