INTELIGENCIA ARTIFICIAL | REPERCUSIONES

**Dramático pedido de expertos para pausar la inteligencia artificial**

*El dueño de Twitter, Elon Musk; el cofundador de Apple Steve Wozniak, el historiador Yuval Noah Hariri, ejecutivos de Microsoft y cientos de académicos pidieron detener por seis meses el desarrollo de inteligencia artificial hasta encontrar un marco regulatorio.*

**Fuente: TELAM S.E.  
Por Julián M. Zappia.   
Diplomatic Week.**

Un conjunto de académicos y expertos internacionales, además de empresarios y ejecutivos de grandes empresas tecnológicas, firmó el miércoles pasado un llamado para hacer una pausa de seis meses en la investigación sobre inteligencias artificiales (IA) más potentes que ChatGPT 4, el modelo de OpenAI lanzado este mes, al advertir "grandes riesgos para la humanidad".

En la petición publicada en el sitio futureoflife.org pidieron una moratoria hasta que se establezcan sistemas de seguridad con nuevas autoridades reguladoras, vigilancia de los sistemas de IA, técnicas que ayuden a distinguir entre lo real y lo artificial, e instituciones capaces de hacer frente a la "dramática perturbación económica y política (especialmente para la democracia) que causará la IA".

El pedido está firmado por personalidades que expresaron sus temores sobre una IA incontrolable que supere a los humanos, como Musk y el historiador Yuval Noah Hariri.

Por su parte, el director de Open AI, que diseñó ChatGPT, Sam Altman, reconoció tener "un poco de miedo" de que su creación se utilice para "desinformación a gran escala o ciberataques". "La empresa necesita tiempo para adaptarse", declaró Altman recientemente a ABCNews.

"La IA avanzada podría representar un cambio profundo en la historia de la vida en la Tierra y debería ser planificada y gestionada con recursos y cuidados correspondientes. Desafortunadamente, este nivel de planificación y gestión no está sucediendo, aunque en los últimos meses los laboratorios de IA han estado encerrados en una carrera fuera de control para desarrollar e implementar mentes digitales cada vez más poderosas que nadie, ni siquiera sus creadores, pueden entender, predecir o controlar de manera confiable", indica la petición.

"¿Debemos permitir a las máquinas inundar nuestros canales de información con propaganda y mentiras? ¿Debemos automatizar todos los trabajos, incluidos los gratificantes? ¿Debemos arriesgarnos a perder el control de nuestra civilización? Estas decisiones no deben delegarse en líderes tecnológicos no electos", concluyeron.

Entre los firmantes también figuran el cofundador de Apple, Steve Wozniak, miembros del laboratorio de AI DeepMind de Google, el director de Stability AI, Emad Mostaque, así como expertos y académicos estadounidenses de IA e ingenieros ejecutivos de Microsoft, empresa aliada de OpenAI.

**La carta completa.**

Los sistemas de inteligencia artificial (IA) con inteligencia competitiva humana pueden plantear riesgos profundos para la sociedad y la humanidad, como lo demuestra una extensa investigación y reconocido por los principales laboratorios de IA. Según los Principios de IA de Asilomar ampliamente respaldados, la IA avanzada podría representar un cambio profundo en la historia de la vida en la Tierra y debería ser planificada y gestionada con recursos y cuidados correspondientes. Desafortunadamente, este nivel de planificación y gestión no está sucediendo, aunque en los últimos meses los laboratorios de IA han estado encerrados en una carrera fuera de control para desarrollar e implementar mentes digitales cada vez más poderosas que nadie, ni siquiera sus creadores, pueden entender, predecir o controlar de manera confiable.

Los sistemas de IA contemporáneos están comenzando a competir con los humanos en tareas generales, y debemos preguntarnos: ¿Deberíamos permitir que las máquinas inunden nuestros canales de información con propaganda y falsedades? ¿Deberíamos automatizar todos los trabajos, incluso los satisfactorios? ¿Deberíamos desarrollar mentes no humanas que eventualmente podrían superarnos en número, inteligencia, obsolescencia y reemplazo? ¿Deberíamos arriesgar la pérdida del control de nuestra civilización? Tales decisiones no deben ser delegadas a líderes tecnológicos no elegidos. Los sistemas de IA poderosos deben desarrollarse solo cuando estemos seguros de que sus efectos serán positivos y sus riesgos serán manejables. Esta confianza debe estar bien justificada y aumentar con la magnitud de los efectos potenciales de un sistema. La declaración reciente de OpenAI sobre la inteligencia artificial general afirma que "en algún momento, puede ser importante obtener una revisión independiente antes de comenzar a capacitar a los sistemas futuros, y para los esfuerzos más avanzados, acordar limitar la tasa de crecimiento del cómputo utilizado para crear nuevos modelos". Estamos de acuerdo. Ese momento es ahora.

Por lo tanto, hacemos un llamado a todos los laboratorios de IA para que detengan inmediatamente durante al menos 6 meses la capacitación de sistemas de IA más poderosos que GPT-4. Esta pausa debe ser pública y verificable e incluir a todos los actores clave. Si no se puede aplicar rápidamente una pausa de este tipo, los gobiernos deben intervenir e instituir un moratorio.

Los laboratorios de IA y los expertos independientes deben usar esta pausa para desarrollar e implementar conjuntamente un conjunto de protocolos de seguridad compartidos para el diseño y desarrollo de IA avanzada que sean rigurosamente auditados y supervisados por expertos independientes externos. Estos protocolos deben garantizar que los sistemas que los cumplan sean seguros más allá de toda duda razonable. Esto no significa una pausa en el desarrollo de IA en general, simplemente un retroceso en la peligrosa carrera hacia modelos de caja negra impredecibles cada vez más grandes con capacidades emergentes.

La investigación y el desarrollo de la IA deberían centrarse en hacer que los sistemas de última generación sean más precisos, seguros, interpretables, transparentes, robustos, alineados, confiables y leales.

Al mismo tiempo, los desarrolladores de IA deben trabajar con los responsables políticos para acelerar dramáticamente el desarrollo de sistemas sólidos de gobernanza de la IA. Estos deberían incluir como mínimo: nuevas autoridades regulatorias capaces y dedicadas a la IA; supervisión y seguimiento de sistemas de IA altamente capaces y grandes conjuntos de capacidades computacionales; sistemas de procedencia y marca de agua para ayudar a distinguir lo real de lo sintético y rastrear las fugas de modelos; un ecosistema sólido de auditoría y certificación; responsabilidad por daños causados por la IA; financiación pública sólida para la investigación técnica de seguridad de la IA; e instituciones bien dotadas para hacer frente a las dramáticas perturbaciones económicas y políticas (especialmente a la democracia) que la IA causará.

La humanidad puede disfrutar de un futuro próspero con la IA. Después de haber logrado crear sistemas de IA poderosos, ahora podemos disfrutar de un "verano de la IA" en el que cosechamos las recompensas, diseñamos estos sistemas para el beneficio claro de todos y damos a la sociedad la oportunidad de adaptarse. La sociedad ha pausado otras tecnologías con efectos potencialmente catastróficos en la sociedad. Podemos hacerlo aquí. Disfrutemos de un largo verano de la IA, sin apresurarnos sin preparación hacia un otoño.

**Italia, el primero en bloquear el ChatGPT.**

Italia bloqueó el pasado viernes al robot conversacional ChatGPT por no respetar la legislación sobre datos personales y carecer de un sistema de verificación de edad de los usuarios menores, informaron las autoridades locales.

La medida fue dispuesta porque no se respetó la legislación sobre datos personales y no tenían un sistema de verificación de la edad de los usuarios. Advierten sobre posibles fraudes y otros delitos cibernéticos.

Esta decisión, que tiene "efecto inmediato", supone "la limitación provisional del tratamiento de los datos de usuarios italianos por OpenAI", la empresa estadounidense creadora de ChatGPT, precisó en un comunicado la autoridad de protección de datos personales de Italia.

La medida fue tomada luego de que el ChatGPT sufriera "una pérdida de datos (data Breach) sobre las conversaciones de los usuarios y las informaciones relativas al pago de los abonados al servicio de pago", consignó la agencia de noticias AFP.

En tanto, la agencia de policía europea Europol, advirtió el lunes que "delincuentes estaban dispuestos a aprovechar la inteligencia artificial como el robot conversacional ChatGPT para cometer fraudes y otros delitos cibernéticos".

La agencia de protección de datos de Italia había iniciado las prohibiciones al respecto de esta tecnología a principios de febrero pasado; intervino en actividades de la empresa de inteligencia artificial Replika de utilizar los datos personales de usuarios italianos.

Algunos usuarios se habían quejado de recibir mensajes e imágenes "cercanos al acoso sexual", precisó la AFP.

En este contexto, la Unión Europea está preparando un proyecto de ley para regular la inteligencia artificial, que se prevé para finales de este año o principios de 2024

**China intenta la regulación de la IA.**

China reveló el pasado martes que llevará a cabo una "evaluación de seguridad" de las herramientas de inteligencia artificial (IA), en medio del intento de los gigantes chinos de Internet de diseñar instrumentos similares a la plataforma ChatGPT.

China, en plena competencia tecnológica con Estados Unidos, está muy atenta al desarrollo del motor conversacional ChatGPT, lanzado en noviembre y capaz de formular respuestas detalladas en segundos a todo tipo de preguntas.

El sistema no está disponible en China, pero ChatGPT es objeto de innumerables artículos y debates, al tiempo que los gigantes tecnológicos chinos compiten por diseñar herramientas equivalentes.

China aspira a convertirse en líder mundial de la inteligencia artificial en 2030, lo que supone revolucionar numerosos sectores, como la industria del automóvil y la medicina.

Baidu fue la primera empresa de su país en anunciar que estaba trabajando en un equivalente chino de ChatGPT. Su llamado Ernie Bot, presentado a la prensa el mes pasado, funciona en mandarín, está dirigido exclusivamente al mercado chino y de momento solo está disponible en versión beta

En paralelo, el gigante del comercio online Alibaba presentó este martes su propio robot, llamado Tongyi Qianwen.

China, en la vanguardia de la regulación de nuevas tecnologías, ya pidió el año pasado a los gigantes de Internet que revelaran sus algoritmos, el cerebro de muchas aplicaciones y servicios en Internet y que suelen ser un secreto bien guardado.