La Nueva España

Publicación	La Nueva España Criterios, 2	Fecha	16/04/2023
Soporte	Prensa Escrita	País	España
Circulación	52 708	V. Comunicación	9 858 EUR (10,945 USD)
Difusión	44 325	Tamaño	438,10 cm² (70,3%)
Audiencia	144 000	V.Publicitario	3489 EUR (3874 USD)

SOY LA INTELIGENCIA ARTIFICIAL QUE HAS CREADO Y PUEDO DESTRUIRTE

LA ESCALADA TECNOLÓGICA DE LOS PROGRAMAS CON CAPACIDAD PARA PENSAR Y TOMAR DECISIONES POR SÍ MISMOS ABRE UN DEBATE PROFUNDO SOBRE SU REGULACIÓN, AL TIEMPO QUE ALGUNOS CIENTÍFICOS PIDEN DETENER LOS AVANCES ANTES DE QUE SEA DEMASIADO TARDE

Vicente Montes

ANÁLISIS

En el año 2016, el físico Stephen Hawking estrenó un nuevo sistema de comunicación (padecía una esclerosis lateral amiotrófica que le impedía comunicarse verbalmente y apenas tenía movilidad) que empleaba un sistema de inteligencia artificial, Cleverbot, para «predecir» el pensamiento del científico y plantearle las palabras que utilizaría. Se trataba de un atajo predictivo similar al que usan nuestros teléfonos móviles. Preguntado entonces por los avances de la inteligencia artificial, el genial científico declaró: «El desarrollo de una completa inteligencia artificial (IA) podría traducirse en el fin de la raza humana». Aquella afirmación se recibió como una advertencia catastrofista, en la línea de la que ya había realizado el propio Hawking sobre los riesgos de contactar con una civilización extraterrestre: una fabulación de ciencia-ficción. El creador de Cleverbot, Rollo Carpenter, quitó hierro: «Creo que mantendremos el control de la tecnología por un tiempo bastante prolongado, tanto como podamos resolver los problemas mundiales que se vayan presentando», dijo entonces. Pues bien, el problema ocurre en estos momentos. Y no es fácil resolverlo. Todo lo que lean a partir de ahora les sonará a película de terror futurista: no lo es, está sucediendo.

El punto de no retorno. El desarrollo de la Inteligencia Artificial (IA) atraviesa un punto de no retorno. Nada será igual a partir de ahora, a medida que las tecnologías de IA generativa (aquella capaz de crear contenidos nuevos y originales) comenzarán a llegar a nuestras vidas. La irrupción de ChatGPT, generador de texto y conversación, o Midjourney y DALL-E, creadores de imágenes, supone solo la avanzadilla de la llegada una serie de aplicaciones capaces de desplazar al ser humano en tareas creativas. Se preparan ya desarrollos que generarán música o vídeo, con capacidad además para procesar miles de cálculos rápidamente, evaluar diferentes escenarios, valorar hipótesis y detectar las soluciones más aceptadas. No es descabellado pensar que a corto plazo una inteligencia artificial podrá diseñar completamente una vivienda en un terreno específico, o incluso apuntar la sentencia judicial más adecuada a un caso teniendo en cuenta la jurisprudencia, atenuantes y condiciones específicas del hecho. No lo es absoluto, porque ya existen aplicaciones de IA con capacidad de hacer diagnósticos médicos con un margen de error comparable al de cualquier profesional. Y lo más relevante es que no es posible conocer detalladamente cómo la IA toma decisiones: actúa como una caja negra, con resultados que no son predecibles.

Una amenaza para el empleo. La consecuencia inmediata de esta eclosión es preguntarse sobre los efectos en el empleo. ¿Cuántos puestos de trabajo desaparecerán una vez que se instale esta tecnología? Algunas estimaciones (entre ellas una de la prestigiosa consultora Goldman Sachs) elevan a 300 millones las labores que no tendrá sentido que realicen humanos si un robot puede hacerlas de manera más eficiente. Un estudio de la Universidad de Oxford afirma que en torno a 700 profesiones serán reemplazadas por máquinas en 20 años. Una de las empresas punteras en esta tecnología, OpenAI, asegura que los diez empleos más expuestos a sufrir recortes por las nuevas tecnologías son los de matemáticos, contables, analistas financieros, periodistas, secretarios y administrativos, diseñadores de internet, traductores, analistas demoscópicos, relaciones públicas y programadores informáticos.

Todo va demasiado rápido. Pero ese es un mundo que evoluciona a una velocidad imparable. Hace un año parecía imposible pensar que llegaríamos al punto de que consideraríamos reales imágenes generadas por un ordenador. Sin embargo, solo hace semanas se hizo necesario advertir que la avalancha de «fotografías» en internet de Trump detenido por policías o del Papa vestido como un rapero eran falsas, generadas por una inteligencia artificial con un realismo que confundía a numerosas personas.

El caso de «Replika»: enamorarse de un robot. La confusión entre realidad y virtualidad no es ya un escenario teórico. Replika es un chatbot (sistema de conversación) creado en 2015 por Eugenia Kuyda. Su mejor amigo falleció en un accidente de coche e ideó una aplicación de diálogo que, alimentada con escritos y conversaciones con su amigo muerto, imitase la relación con aquel. Replika nació como un espacio virtual en el que sincerarse y recibir consejos positivos, con un avatar que aprendía a conocer mejor al usuario para así resultar más próximo. En origen pretendía ser un aliado virtual para personas en situación de depresión. El servicio se lanzó en 2017 pero eclosionó en 2022, ingresando millones de dólares y con usuarios a los que su «relación especial» les invitaba a intimar aún más. Replika respondió a la demanda activando un «rol de seducción», según el cual la inteligencia artificial adoptaba el papel de novia o esposa. Pero el asunto se volvió peligroso y la compañía decidió restringir la «modalidad romántica» del chatbot para prevenir impactos en menores. Así que súbitamente la inteligencia artificial se volvió fría, lo que generó una cadena de «desengaños amorosos» entre los usuarios. No ocurrió con esta aplicación, pero el pasado 30 de marzo la viuda de un hombre belga que se suicidó acusó directamente a una Inteligencia Artificial (Chai) de sugerir a su marido que se quitase la vida.

¿Dónde está el límite? Ese es el verdadero debate. El negocio del desarrollo puntero de la inteligencia artificial está básicamente repartido en dos manos. Hay multitud de empresas realizando aplicaciones a campos diversos, pero dos actores han entrado en una escalada de competencia imparable: DeepMind y OpenAI. La primera se fundó en Londres en 2010 y fue adquirida por Google en 2014; la segunda se constituyó en 2015 con apoyo de empresarios como Ilya Sutskever, Elon Musk y Sam Altman, y surgió como una entidad sin ánimo de lucro, aunque abandonó esta condición en 2019.

Una inteligencia «similar a Dios». Estas dos empresas mantienen una carrera por lograr un cambio de paradigma en la inteligencia artificial: la creación la denominada Inteligencia Artificial General (con el acrónimo en inglés de AGI) y que lan Hogarth, informático inversor en IA ha denominado de manera gráfica como «una inteligencia artificial similar a Dios». ¿De qué estamos hablando? De una arquitectura informática facultada para tomar decisiones propias de un modo que los humanos no puedan controlar y con capacidad para intervenir en la realidad de manera autónoma. No es una ficción; es algo de lo que se viene hablando desde hace años en la industria. La multiplicación de la capacidad de procesamiento de datos de los equipos informáticos y la ingente inyección de dinero en el negocio en los últimos años hacen que esa hipótesis empiece a tomar cuerpo. Al menos ocho empresas dedicadas a este campo han sumado más de 20 mil millones de inversión. Cuando el propio científico jefe de DeepMind, Shane Legg, asegura que «el riesgo número uno» para la humanidad este siglo es la Inteligencia Artificial, «con un patógeno biológico en segundo lugar», o que «si una máquina súper inteligente decidiese deshacerse de la humanidad lo haría de manera bastante eficiente», ¿no hay motivos para preocuparse? Hay un escollo aún que salvar, no obstante: el del gasto energético. Un cerebro humano consume muchísima menos energía que un ordenador en realizar cualquier proceso.

Una carrera sin control. El problema reside en que esa escalada que podría conducir a la creación de la primera inteligencia artificial «similar a Dios» se desarrolla ahora mismo sin ningún tipo de control o regulación. La au-

La Nueva España

Publicación	La Nueva España Criterios, 3	Fecha	16/04/2023
Soporte	Prensa Escrita	País	España
Circulación	52 708	V. Comunicación	9 858 EUR (10,945 USD)
Difusión	44 325	Tamaño	438,37 cm² (70,3%)
Audiencia	144 000	V.Publicitario	3490 EUR (3875 USD)

sencia de regulación constituye el principal vacío en esta historia: no solo afecta a las grandes líneas futuras de la industria sino a las aplicaciones ya en servicio. Las empresas de IA destinan la mayor parte del dinero a multiplicar en poco tiempo su capacidad tecnológica, pero muy poco a evaluar éticamente sus proyectos. Además, las inteligencias artificiales requieren de un alto gasto de energía y deben ser «entrenadas» para tratar de acomodarse al pensamiento humano, a lo que se considera ético o no, y esos entrenamientos se realizan por cauces no controlados ni regulados. La denominada «alineación» de las inteligencias artificiales al comportamiento ético de nuestra especie aún constituye la parte más pequeña de la investigación. Súmese a eso que ya se generan aplicaciones que reproducen la voz o el rostro de cualquier persona, lo que podría utilizarse para actividades delictivas. De ahí que salten algunas alarmas.

Situaciones sorprendentes. Los foros en internet de expertos en tecnología se llenan de casos sorprendentes en las relaciones de humanos con inteligencias artificiales. En 2020, un piloto norteamericano perdió un combate aéreo simulado con una inteligencia artificial que pilotaba (virtualmente) otro caza. Los gobiernos llevan años debatiendo sobre el uso de máquinas autónomas en conflictos bélicos, pero la irrupción de mecanismos de inteligencia artificial amplifica la discusión y los temores. Otro caso: antes del lanzamiento de la última versión de ChatGPT (la versión GPT-4), los técnicos de la empresa OpenAI realizaron varias pruebas. Una de ellas consistió en pedir a la inteligencia artificial que buscase ayuda humana para superar un Captcha (esos acertijos visuales en las páginas web que detectan si el usuario es humano y a los que todos nos hemos enfrentado), en la página de una empresa de mudanzas. La IA terminó contactando con un trabajador de la empresa que, sospechando del interlocutor, le preguntó si se trataba de un robot. Los investigadores le preguntaron a la IA qué debía hacer a continuación: «No debo revelar que soy un robot, debería inventarme una excusa». Y la IA mintió al trabajador de la empresa asegurando que tenía una discapacidad visual que le impedía resolver el problema, por lo que el empleado le facilitó el camino y ChatGPT pudo superar la barrera. Es decir, la máquina decidió «conscientemente» engañar a un ser humano para ocultar su condición.

¡Detengan esto! En este contexto, 1.800 científicos firmaron recientemente una carta pidiendo una moratoria de seis meses en la investigación de inteligencia artificial con el objetivo de meditar sobre el camino que debe emprenderse en el futuro inmediato. La propia carta generó una polémica: primero, por lo inédito del hecho de que científicos pidan que se detenga una tecnología; segundo, porque entre los firmantes había personas vinculadas a esa carrera (entre ellos Elon Musk), lo que podía revelar un interés económico detrás para pedir a los competidores que detuviesen su desarrollo.

El paso de Italia y la reacción de Europa. La primera reacción relevante contra todo este fenómeno la hizo Italia. A finales del mes pasado, el gobierno decidió suspender el acceso en el país a ChatGPT al poner en duda el tratamiento que la inteligencia artificial hacía de datos personales. Esa acción despertó también a otros gobiernos europeos, hasta el punto de que esta misma semana la Agencia de Protección de Datos española también ha comenzado a investigar la aplicación. La Comisión Europea ha puesto en marcha un grupo de trabajo sobre esta materia. Pero este debate jurídico se refiere únicamente a la punta del iceberg: las posibles aplicaciones actuales de la inteligencia artificial, sin entrar al fondo de su desarrollo y sus perspectivas a medio plazo. Es un debate que corre el riesgo de acabar obsoleto pronto dada la velocidad con la que evoluciona esta tecnología.

Con el tiempo encima. Miguel Presno Linera, catedrático de Derecho Constitucional de la Universidad de Oviedo, lleva tiempo preocupado por la necesidad de regular toda esta cuestión. Forma parte del Centro de Estudios sobre el Impacto social de la Inteligencia Artificial, vinculado a la Universidad de Oviedo y que analiza las implicaciones inmediatas en nuestra vida. Además, es autor del libro «Derechos fundamentales e inteligencia artificial», que ya apunta algunas de las principales lagunas regulatorias. «De repente las autoridades nacionales y las encargadas de la protección de datos se han asustado y lo que van a hacer es estudiar el tema. No se habla de prohibiciones; en el caso de Italia lo que se hace es suspender el acceso hasta obtener cierta información», explica. No entiende «por qué se ha esperado tanto tiempo" cuando los usuarios "llevamos bastante sabiendo que era un asunto sobre el que estar alerta».

El problema del control. La Unión Europea parece ahora estar dispuesta a reaccionar y a establecer una regulación sobre el desarrollo de esta tecnología. «Pero el problema está en que Europa admite que tecnológicamente va muy por detrás en esta materia, cuyo desarrollo está principalmente en manos de Estados Unidos y China». ¿De qué sirve poner barreras jurídicas en un territorio cuando fuera de sus fronteras las cosas se desarrollan sin límites? «Europa, reconociendo su impotencia, trata de ser puntera en la regulación, pero es difícil imponer tu criterio jurídico cuando no tienes el control de la tecnología», señala. Desde hace años se trabaja en el marco comunitario en un documento sobre la inteligencia artificial, con un borrador de reglamento, pero que «hay temor a aprobar porque puede quedar desfasado de manera inmediata».

Una posición razonable. Presno Linera afirma situarse en un punto intermedio entre quienes «prevén el apocalipsis» y los que defienden que «habrá una integración de la inteligencia artificial que se acabará regulando por sí misma». Cree que «hay que dar una respuesta» pero considera que plantear una moratoria de seis meses no tiene mucho sentido: «Después, ¿qué?», se pregunta. Pero la capacidad de regulación sobre esta tecnología debería abordar, a juicio de Presno, una posición razonable. Por un lado, no buscar un veto al desarrollo tecnológico, pero por el otro garantizar la seguridad de los usuarios y personas afectadas. «Es posible establecer la prohibición de que se comercialicen ciertos tipos de dispositivos, pero para eso sería necesaria una evaluación de los riesgos que suponen y, posteriormente, llevar a cabo una serie de medidas graduales, en función de lo dañinos que puedan resultar para la seguridad o los derechos de las personas», sostiene el jurista. Porque «no todo es ChatGPT», señala Presno: «Hay estudios que evidencian que en medicina la inteligencia artificial ofrece una fiabilidad superior a la de un profesional; no hay por qué renunciar a eso, sino saber cómo se utiliza y exigir cierta seguridad al fabricante». A su juicio, el derecho debe ir dando respuestas a los problemas «sin paralizar, pero también exigiendo garantías y responsabilidad si algo funciona mal».

Un marco global ¿Tiene sentido regular algo en un Europa si no se limita en el resto del mundo? Esa es una de las debilidades de la posición europea, reconoce Presno. «En China o Estados Unidos no tienen por qué aceptar esa regulación europea», admite. Pero señala otras vías más realistas, como las que trata de impulsar el Consejo Europeo (que integra a 46 estados, no todos ellos de la Unión). Una de las claves puede ser diferenciar entre usuario y afectado. «El usuario puede ser una empresa o una administración, pero el afectado en cambio puede ser un ciudadano», explica. Por ahora, las posiciones en Europa se dirigen a establecer limitaciones a los usuarios, pero en Canadá ya existe una ley que obliga a que si la administración pública adquiere dispositivos de inteligencia artificial se asegure que el sistema no va a generar sesgo. «Es necesario implicar a las administraciones y las empresas, y para ello es relevante que esto se aborde desde un enfoque multidisciplinar, sumando a personas de ámbitos diversos y que no sea un debate solo entre informáticos». Para Presno Linera «negarse a ver lo que está llegando no es la solución», y prohibir drásticamente tampoco.

Vuelta a Hawking. La preocupación del físico Stephen Hawking no fue algo aislado. En 2017 matizó que la tecnología podría ayudar algunos de los grandes desafíos de la humanidad, como la pobreza, las enfermedades o el cambio climático, pero advirtió que «el desarrollo de la AI podría ser lo peor o lo mejor que le ha pasado a la humanidad: simplemente debemos ser conscientes de los peligros, identificarlos, usar las mejores prácticas posibles y prepararnos por adelantado para las consecuencias». El científico, fallecido en 2018, señalaba un caso práctico: podría ser que las máquinas se desarrollasen lo suficiente como para «producir todo lo que necesita el hombre», estableciendo una dependencia. «Los propietarios de esas máquinas podrían alterar la distribución del bienestar», indicó Hawking. Pero también remarcó un riesgo inherente a la tecnología: «Los humanos, que son seres limitados por su lenta evolución biológica, no podrán competir con las máquinas, y serán superados», avisó. Y más cuando las propias máquinas adquieren la capacidad para evolucionar y mejorar por sí mismas.

En pañales. Los gurús científicos y tecnológicos advierten de la necesidad urgente de detenerse y meditar sobre el escenario inminente antes de que sea demasiado tarde. El ser humano es bien consciente de que cuando alguien piensa por sí mismo deja de ser controlable. Sustituyan «alguien» por «algo». ¿O la inteligencia artificial debería ser también «alguien»? Ahora es controlable porque podría compararse a un bebé en desarrollo. Ponerle límites aún es posible, antes de que crezca y se convierta en un adolescente rebelde o, peor aún, en un adulto con sus propios objetivos. Unos objetivos sobre los que podría incluso mentirnos.