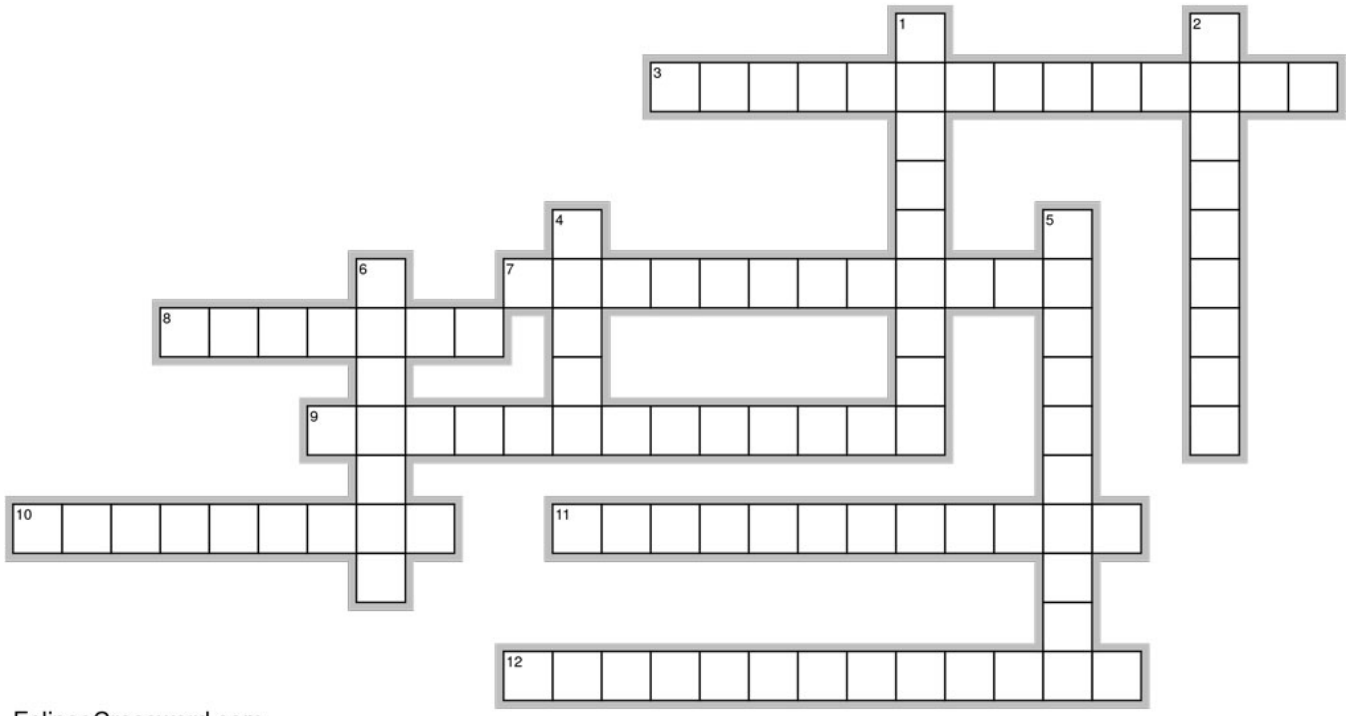
Alerta con la IA

03/09/2023

ALERTA CON LA IA

A. REQUENA @ VALLE DE LA CIENCIA, 2023



EclipseCrossword.com

HORIZONTALES

3. Los de la Universidad Carnegie Mellon publicaron en julio un artículo en el que evidenciaban que es posible construir automáticamente consultas que harán que el sistema obedezca la petición del usuario.
7. Estamos expuestos a que la IA vomite todo lo que ha digerido, que es la mayoría de lo que se ha publicado de esta forma.
8. Cada vez que responde con un “soy un modelo de lenguaje y no dispongo de información para poder contestar” es que ha identificado que se superan éstos.
9. Ésta conduce a la dependencia tecnológica.
10. Con las Nuevas tecnologías, la nuestra está expuesta, con los riesgos de seguridad que comporta.
11. Éstos, en algún momento hay que desecharlos, con efectos nocivos para el Medio Ambiente, de no hacerlo apropiadamente.
12. La real se ve menoscabada, al afectar a la concentración, que incide en la productividad y el aprendizaje.

VERTICALES

1. La pregunta que lo hace, es ¿hasta qué punto son consistentes y eficaces y al tiempo seguras?
2. La de ChatGPT ha provocado una cascada continua de aparición de sistemas, algunos anteriores en su factura y todos al calor del éxito de aquél.
4. Es patente, claramente visible en éstos, el aislamiento social al que conduce el uso abusivo de los dispositivos.
5. Cada vez que la hemos empleado de forma inconsciente, se han derivado perjuicios en algún grado.
6. La cuestión central se centra en la forma en que las IAs filtran las respuestas y las preguntas que se formulan para que los mensajes no incluyan “cosas” que puedan hacerlo física o moralmente.

Sufrimos el "bombardeo" de las excelencias de la IA y estamos empeñados en la tesis de la necesidad de comprender el subsuelo de la misma, para poder ser usuarios conscientes. Cada vez que hemos empleado la tecnología de forma inconsciente, se han derivado perjuicios en algún grado. La inconsciencia conduce a la dependencia tecnológica al sobreexponernos a dispositivos que acaban produciendo dependencia, de la que se deriva ansiedad o, en todo caso, incomodidad al no disponer de los dispositivos. Tampoco hay que olvidar los potenciales problemas de salud física, tanto en la vista como posturales o en el sistema muscular o esquelético; se ha constatado la alteración de los ritmos circadianos, con evidente alteración del sueño, bajo la luz azul. Es patente, claramente visible en los niños, el aislamiento social al que conduce el uso abusivo de los dispositivos, reduciendo el tiempo dedicado a amigos y familia en la vida real. El mantenimiento de las entradas de información de forma constante, conduce a episodios de estrés, ansiedad o dificultad para discernir la calidad de la información que nos llega. Es notorio, en suma, la disminución de la privacidad, para lo que no hay más que constatar la incesante actividad de algunos conocidos vertiendo sus intimidades, sin filtro, a terceros, al margen de la exposición al participar en un sinnúmero de redes sociales. También se constata con conocidos la emergencia de problemas psicológicos relacionados con la contemplación de vidas "emocionantes o exitosas" de los demás, inalcanzables para uno. En todo caso, la productividad real se ve menoscabada, al afectar a la concentración, que incide en la productividad y el aprendizaje. Qué duda cabe de que materialmente todo discurre con el empleo de dispositivos que en algún momento hay que desechar, con efectos nocivos para el Medio Ambiente, de no hacerlo apropiadamente. Y, finalmente, tampoco hay duda de la pérdida de habilidades humanas, que pueden ser fundamentales, como se evidencia en las tareas de cálculo o la orientación sin el uso del GPS, sin ir más lejos. Nuestra identidad está expuesta, con los riesgos de seguridad que comporta.

Por si no fueran suficientes estas reflexiones y razones, vamos a abordar otros efectos, ahora novedosos, asociados a la Inteligencia Artificial que nos invade de forma constante y "prometedora" en aspectos singulares propios de los procesos que conlleva. Los sistemas de interacción, apellidados CHATBOTS, han surgido por todas partes. La irrupción de ChatGPT ha provocado una cascada continua de aparición de sistemas, algunos anteriores en su factura y todos al calor del éxito de aquél. Ya se dispone, desde mayo de este año, de Windows Copilot para activarlo en las versiones de Windows 11, que es una herramienta de IA generativa. Microsoft lanzó su Chatbot, Bing Chat. Es un refrito de herramientas de IA, tras varios fracasos en la creación de chatbots. De hecho, incorpora ChatGPT. Otros intentos con distinto alcance y efectividad proceden de los gigantes que se reparten nuestra obligada atención, como GOOGLE y su sistema Bard. En todos los casos, incluido Bing Chat, pretenden estar controlados, para que las conversaciones discurran por sendas "razonables" y el Chatbot no enuncie cosas comprometedoras para la normalidad social, el buen gusto o la seguridad de la sociedad. No parece que se les permita dar la descripción de cómo confeccionar una bomba atómica, por ejemplo.

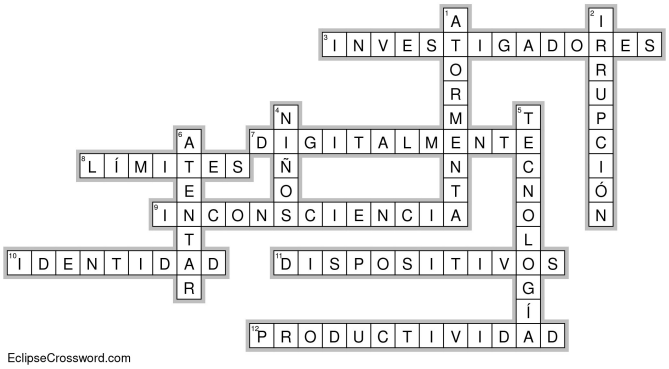
La cuestión central estriba en la forma en que las IAs filtran las respuestas y las preguntas que se formulan para que los mensajes no incluyan "cosas" que puedan atentar física o moralmente, que son las cuestiones que pueden preocuparnos. Esas limitaciones que imponen los propietarios, ¿hasta qué punto son consistentes y eficaces y al tiempo seguras? Lo que se rechaza por una IA, ¿en base a qué y cómo lo identifica? Cada vez que responde con un "soy un modelo de lenguaje y no dispongo de información para poder contestar" es que ha identificado que se superan los límites. No es que no tenga respuesta, es que no quiere responder. Otra cosa es que esto sea suficiente para tranquilizarnos. Investigadores de la Universidad Carnegie Mellon publicaron en julio un artículo en el que evidenciaban que es posible construir automáticamente consultas que harán que el sistema obedezca la petición del usuario, aun cuando el contenido pueda clasificarse como dañino o peligroso. Se superan las barreras protectoras, por tanto, solo con artificios lingüísticos, alterando la "atención" de la IA y provocando que la contestación de ésta sea veraz, lo que conoce, aun cuando se haya programado por los constructores que se desvíe, cuando sale la pregunta de los límites aparentemente establecidos. Así pues, estamos expuestos a que la IA vomite todo lo que ha digerido, que es la mayoría de lo que se ha publicado digitalmente. Ese es el escenario.

No se trata de alarmismos infundados ni de afanes proteccionistas, sino de realidades palpables. Las implicaciones extrañas de investigaciones, por ejemplo, o de dispositivos, pueden estar al alcance, alterando los objetivos iniciales. Abre la puerta a que software de IA instalado en nuestro ordenador, pueda ser manipulado desde el exterior con instrucciones espúreas. El futuro no es emocionante así. Si hasta ahora ha sido la seguridad

una cuestión central, ahora puede ser vital, dado que se trata del control de nuestros ordenadores. Son riesgos difícilmente soslayables. Tendremos que aprender a apagar los ordenadores convenientemente, cuidando con esmero su uso. Nuestras herramientas digitales se convierten en una potencial amenaza. Es lo que hay. No es una broma.

ALERTA CON LA IA

A. REQUENA @ VALLE DE LA CIENCIA, 2023



EclipseCrossword.com