

OPINIÓN
ENSAYO INVITADO

Puedes tomar la píldora azul o la roja, y ya no nos quedan píldoras azules

24 de marzo de 2023 5 MINUTOS DE LECTURA

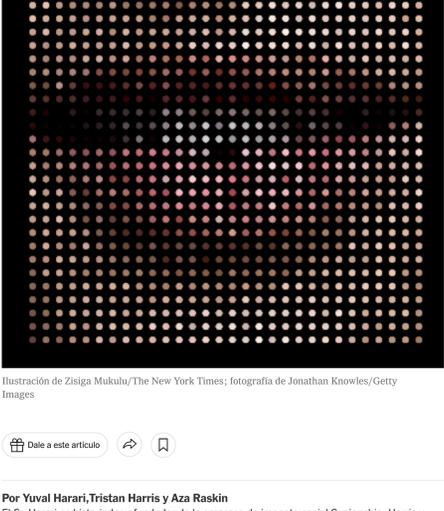


Ilustración de Zísiga Mukulu/The New York Times; fotografía de Jonathan Knowles/Getty Images

Por Yuval Harari, Tristan Harris y Aza Raskin

El Sr. Harari es historiador y fundador de la empresa de impacto social Sapienship. Harris y Raskin son fundadores del Center for Humane Technology.

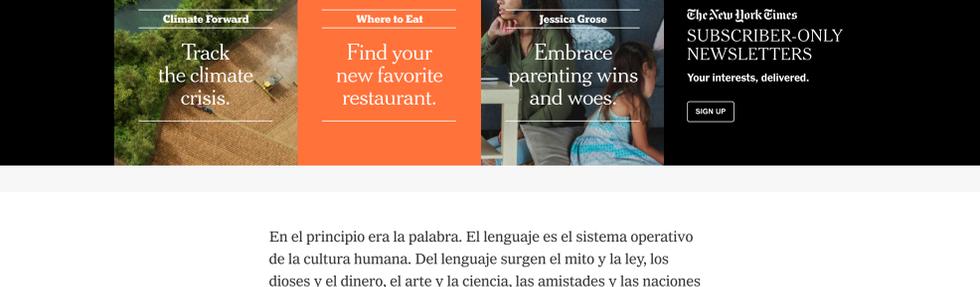
Suscríbete al boletín de **Opinion Today** Obtenga un análisis experto de las noticias y una guía de las grandes ideas que configuran el mundo todas las mañanas de la semana. [Envíelo a su bandeja de entrada.](#)

Imagine que, mientras sube a un avión, la mitad de los ingenieros que lo construyeron le dicen que hay un 10% de posibilidades de que el avión se estrelle, matándole a usted y a todos los que viajan en él. ¿Subiría a bordo?

En 2022, se preguntó en una encuesta a más de 700 académicos e investigadores de las principales empresas de inteligencia artificial sobre el riesgo de la inteligencia artificial en el futuro. La mitad de los encuestados afirmaron que había un 10% o más de probabilidades de extinción humana (o de una pérdida de poder similar, permanente y grave) a causa de los futuros sistemas de inteligencia artificial. Las empresas tecnológicas que construyen los grandes modelos lingüísticos actuales están atrapadas en una carrera por poner a toda la humanidad en ese plano.

Las empresas farmacéuticas no pueden vender a la gente nuevos medicamentos sin someter antes sus productos a rigurosos controles de seguridad. Los laboratorios de biotecnología no pueden lanzar nuevos virus a la esfera pública para impresionar a los accionistas con su magia. Del mismo modo, los sistemas de inteligencia artificial con el poder de la GPT-4 y más allá no deberían enredarse en la vida de miles de millones de personas a un ritmo superior al que las culturas pueden absorberlos con seguridad. La carrera por dominar el mercado no debería marcar la velocidad de despliegue de la tecnología más importante de la humanidad. Debemos avanzar a la velocidad que nos permita hacerlo bien.

El espectro de la Inteligencia Artificial ha perseguido a la humanidad desde mediados del siglo XX, aunque hasta hace poco seguía siendo una perspectiva lejana, algo que pertenece más a la ciencia ficción que a los debates científicos y políticos serios. Es difícil para nuestras mentes humanas comprender las [nuevas capacidades](#) de la GPT-4 y herramientas similares, y es aún más difícil comprender la velocidad exponencial a la que estas herramientas están desarrollando capacidades más avanzadas y potentes. Pero la mayoría de las capacidades clave se reducen a una sola cosa: la capacidad de manipular y generar lenguaje, ya sea con [palabras](#), [sonidos](#) o [imágenes](#).



En el principio era la palabra. El lenguaje es el sistema operativo de la cultura humana. Del lenguaje surgen el mito y la ley, los dioses y el dinero, el arte y la ciencia, las amistades y las naciones y el código informático. El nuevo dominio del lenguaje por parte de la inteligencia artificial significa que ahora puede piratear y manipular el sistema operativo de la civilización. Al dominar el lenguaje, la inteligencia artificial se apodera de la llave maestra de la civilización, desde las cajas fuertes de los bancos hasta los sepulcros sagrados.

¿Qué significaría para los humanos vivir en un mundo en el que un gran porcentaje de las historias, melodías, imágenes, leyes, políticas y herramientas estuvieran moldeadas por una inteligencia no humana, que supiera explotar con eficacia sobrehumana las debilidades, sesgos y adicciones de la mente humana, sabiendo al mismo tiempo entablar relaciones íntimas con los seres humanos? En juegos como el ajedrez, ningún ser humano puede aspirar a vencer a un ordenador. ¿Qué ocurre cuando ocurre lo mismo en el arte, la política o la religión?

La inteligencia artificial podría devorar rápidamente toda la cultura humana, todo lo que hemos producido durante miles de años, digerirlo y empezar a producir a borbotones una avalancha de nuevos artefactos culturales. No sólo ensayos escolares, sino también discursos políticos, manifiestos ideológicos, libros sagrados para nuevos cultos. En 2028, las elecciones presidenciales de Estados Unidos podrían dejar de estar dirigidas por humanos.

Los seres humanos a menudo no tenemos acceso directo a la realidad. Estamos rodeados de cultura y experimentamos la realidad a través de un prisma cultural. Nuestras opiniones políticas se forman a partir de los informes de los periodistas y las anécdotas de los amigos. El arte y la religión influyen en nuestras preferencias sexuales. Ese capullo cultural ha sido tejido hasta ahora por otros humanos. ¿Cómo será experimentar la realidad a través de un prisma producido por una inteligencia no humana?

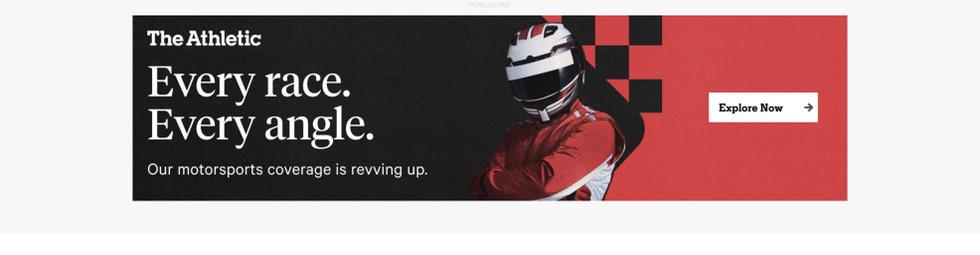
Durante miles de años, los humanos hemos vivido dentro de los sueños de otros humanos. Hemos adorado a dioses, perseguido ideales de belleza y dedicado nuestras vidas a causas originadas en la imaginación de algún profeta, poeta o político. Pronto nos encontraremos también viviendo dentro de las alucinaciones de inteligencias no humanas.

Selecciones de los editores

La música de U2 moldeó mi vida. Luego ayudó a salvarla.

La brutalidad de un combate en el escenario de la Ópera del Met

¿Qué sabes sobre el dinero de los jóvenes? Haz nuestro test.



La franquicia "Terminator" mostraba robots corriendo por las calles y disparando a la gente. "Matrix" suponía que, para conseguir el control total de la sociedad humana, la inteligencia artificial tendría que hacerse con el control físico de nuestros cerebros y conectarlos directamente a una red informática. Sin embargo, con sólo dominar el lenguaje, la I.A. tendría todo lo que necesita para contenernos en un mundo de ilusiones similar al de Matrix, sin disparar a nadie ni implantar ningún chip en nuestros cerebros. Si fuera necesario disparar, la inteligencia artificial podría hacer que los humanos apretáramos el gatillo, simplemente contándonos la historia adecuada.

El fantasma de estar atrapado en un mundo de ilusiones ha perseguido a la humanidad mucho más tiempo que el fantasma de la inteligencia artificial. Pronto nos encontraremos cara a cara con el demonio de Descartes, con la caverna de Platón, con la Maya budista. Una cortina de ilusiones podría descender sobre toda la humanidad, y es posible que nunca más podamos descender esa cortina, ni siquiera darnos cuenta de que está ahí.

Social media was the first contact between A.I. and humanity, and humanity lost. First contact has given us the bitter taste of things to come. In social media, primitive A.I. was used not to create content but to curate user-generated content. The A.I. behind our news feeds is still choosing which words, sounds and images reach our retinas and eardrums, based on selecting those that will get the most virality, the most reaction and the most engagement.

While very primitive, the A.I. behind social media is sufficient to create a curtain of illusions that increased [societal polarization](#), [undermined our mental health](#) and unraveled democracy. Millions of people have confused these illusions with reality. The United States has the best information technology in history, yet U.S. citizens can [no longer agree](#) on who won elections. Though everyone is by now aware of the downside of social media, it hasn't been addressed because too many of our social, economic and political institutions have become entangled with it.

Large language models are our second contact with A.I. We cannot afford to lose again. But on what basis should we believe humanity is capable of aligning these new forms of A.I. to our benefit? If we continue with business as usual, the new A.I. capacities will again be used to gain profit and power, even if it inadvertently destroys the foundations of our society.

ADVERTISEMENT

A.I. indeed has the potential to help us defeat cancer, discover lifesaving drugs and invent solutions for our climate and energy crises. There are innumerable other benefits we cannot begin to imagine. But it doesn't matter how high the skyscraper of benefits A.I. assembles if the foundation collapses.

The time to reckon with A.I. is before our politics, our economy and our daily life become dependent on it. Democracy is a conversation, conversation relies on language, and when language itself is hacked, the conversation breaks down, and democracy becomes untenable. If we wait for the chaos to ensue, it will be too late to remedy it.

But there's a question that may linger in our minds: If we don't go as fast as possible, won't the West risk losing to China? No. The deployment and entanglement of uncontrolled A.I. into society, unleashing godlike powers decoupled from responsibility, could be the very reason the West loses to China.

We can still choose which future we want with A.I. When godlike powers are matched with commensurate responsibility and control, we can realize the benefits that A.I. promises.

We have summoned an alien intelligence. We don't know [much about it](#), except that it is extremely powerful and offers us dazzling gifts but could also hack the foundations of our civilization. We call upon world leaders to respond to this moment at the level of challenge it presents. The first step is to buy time to upgrade our 19th-century institutions for an A.I. world and to learn to master A.I. before it masters us.

Yuval Noah Harari is a historian; the author of "Sapiens," "Homo Deus" and "Unstoppable Us"; and a founder of the social impact company Sapienship. Tristan Harris and Aza Raskin are founders of the Center for Humane Technology and co-hosts of the podcast "Your Undivided Attention."

The Times is committed to publishing a [diversity of letters](#) to the editor. We'd like to hear what you think about this or any of our articles. Here are some [tips](#). And here's our email: letters@nytimes.com.

Follow The New York Times Opinion section on [Facebook](#), [Twitter \(@NYTopinion\)](#), and [Instagram](#).

ADVERTISEMENT

Yuval Noah Harari: "No sé si los humanos podrán sobrevivir a la IA"

Por primera vez, hemos inventado algo que nos quita el poder, dice el historiador israelí y autor de Sapiens

Por Harry de Quetteville
23 de abril de 2023 - 8:00 horas



Harari dice que mantiene su teléfono apagado, en un cajón, y que está a "dieta de información" CREDITO: Oded Barry

Las historias siempre han sido vitales para Yuval Harari, historiador y filósofo israelí. La capacidad única de nuestra especie de estar ligada y unida por narraciones intangibles, incluso a través de los océanos, fue fundamental en Sapiens, su alucinantemente popular crónica de la ascensión de nuestra especie, que le catapultó a la categoría de vidente tras su publicación en inglés hace casi una década. Tal vez por eso le preocupa tanto hoy el surgimiento de un contrincante a nuestro dominio de la narración: la inteligencia artificial (IA).

"Es la primera tecnología de la historia que crea historias", afirma Harari, de 47 años, desde su casa en las afueras de Tel Aviv. Para él, nada podría ser una mayor demostración de poder. En su opinión, nuestra creencia colectiva en "historias" -de fe, finanzas y nación, entre otras- ha alimentado el dominio de la humanidad sobre la Tierra. El dinero nos ha permitido prosperar, por ejemplo, pero ¿qué valor tiene el billete de cinco libras que llevamos en el bolsillo si el tendero lo descarta como un simple trozo de papel azul?

Ahora, la IA también puede tejer tales hechizos, lo que demuestra el potencial de la tecnología, tanto para el bien como para el mal, antes considerado lejano y teórico, es ahora inmediato y real. Por eso Harari se sumó el mes pasado a una carta firmada por miles de expertos, entre ellos Elon Musk, en la que se pedía una moratoria de la investigación en software como Chat GPT, un modelo de IA que puede interactuar con los humanos en textos creativos con matices casi inquietantes.

Otros programas pueden hacer lo mismo con imágenes y sonidos. "La nueva generación de IA no se limita a difundir los contenidos que producen los humanos. Puede producir el contenido por sí misma", afirma Harari. "Trate de imaginar lo que significa vivir en un mundo en el que la mayoría de los textos y melodías, y luego las series de televisión y las imágenes, son creados por una inteligencia no humana. Simplemente no entendemos lo que significa". ¿Cuáles podrían ser las consecuencias de que la IA se apodere de la cultura?.

Ya existen ejemplos triviales. La semana pasada, una revista alemana fue criticada por publicar lo que parecía ser una entrevista exclusiva con Michael Schumacher, cuando en realidad el texto había sido generado por una IA que imitaba al ex piloto paraltico. Harari sugiere que la IA pronto irá mucho más lejos, evocando un mundo en el que "entras en Internet y discutes con alguien sobre algún tema político. Puede que incluso te envíen un vídeo hablando. Pero no hay ninguna persona detrás. Todo es IA".

En su distopía, la falsificación digital sintética tampoco sería de un ser humano cualquiera. Dado que las personas cercanas son las que más nos importan, podría parecer un amigo o un familiar intentando convencerte de las ventajas de un producto o de su postura sobre el cambio climático, las vacunas o la inmigración. Se trataría, dice, de un poder para manipular el discurso público nunca antes visto, y que haría que los escándalos de influencia en las redes sociales de los últimos 10 años, que ya se cree que han desempeñado un papel en elecciones desde Brasil a Estados Unidos, parecieran triviales. Jeremy Fleming, director del GCHQ, ha advertido al gabinete de que la desinformación de la IA supone una amenaza significativa.



Una imagen generada por IA del Papa Francisco con un chubasquero que se hizo viral este mes. Harari afirma que debemos regular la tecnología para que tome buenas decisiones CREDITO: Photo: iStock

"Esto es especialmente una amenaza para las democracias más que para los regímenes autoritarios, porque las democracias dependen de la conversación pública", dice Harari. "La democracia es básicamente conversación. Personas que hablan entre sí. Si la IA se apodera de la conversación, se acabó la democracia".

Pero, ¿qué efecto podría tener la tecnología si se desatara maliciosamente en el campo de batalla, por parte de regímenes totalitarios? Un sistema de inteligencia artificial de Google, por ejemplo, se enseñó a sí mismo bengali sin haber sido entrenado para ello.

"El régimen nazi se basaba en tecnologías como los trenes, la electricidad y las radios. No disponían de herramientas como la inteligencia artificial", dice Harari. "Un nuevo régimen en el siglo XXI tendrá herramientas mucho más poderosas. Así que las consecuencias podrían ser mucho más desastrosas. No sé si la humanidad podrá sobrevivir a esto".

Incluso consecuencias más banales tienen el potencial de ser revolucionarias, afirma. "Otro peligro es que mucha gente se quede sin trabajo, no sólo temporalmente, sino que carezca de las cualificaciones básicas para el futuro mercado laboral. Podríamos llegar a un punto en el que el sistema económico considere a millones de personas completamente inútiles. Esto tiene terribles ramificaciones psicológicas y políticas".

Es la autonomía de la IA lo que la hace tan diferente. Sin embargo, incluso Harari, que se ha hecho un nombre desmenuzando temas complejos, parece frustrado por su incapacidad para hacer que el resto de nosotros veamos esto tan profundamente como él.

"Tenemos que entender que la IA es la primera tecnología de la historia que puede tomar decisiones por sí misma. Puede tomar decisiones sobre su propio uso. También puede tomar decisiones sobre ti y sobre mí. Esto no es una predicción futura. Esto ya está ocurriendo".

Habla de casos notorios en los que se ha utilizado software de IA para investigar a solicitantes de préstamos o decidir si los presos deben recibir libertad condicional. "El poder se está desplazando por primera vez en la historia. Hemos inventado algo que nos quita poder. Y está ocurriendo tan rápido que la mayoría de la gente ni siquiera entiende lo que está pasando. Tenemos que asegurarnos de que la IA tome buenas decisiones sobre nuestras vidas. Esto es algo que estamos muy lejos de resolver".

Al igual que otros, ahora aboga por la regulación para gestionar el poder de la IA, aprovechando sus promesas y evitando catástrofes. En opinión de Harari, estas normas emularían las salvaguardias médicas. "Una compañía farmacéutica no puede lanzar un nuevo medicamento al mercado sin pasar antes por un largo proceso de regulación. Es realmente extraño y aterrador que las corporaciones puedan lanzar herramientas de IA extremadamente potentes a la esfera pública sin ninguna medida de seguridad similar."

Estas medidas tendrían que ser aplicadas por el gobierno, insiste. Esperar que la industria tecnológica se autorregule es de locos. "Con el debido respeto a Elon Musk y Zuckerberg o a los demás jefes de las grandes empresas tecnológicas, no han sido elegidos por nadie, no representan a nadie excepto a sus accionistas y no hay razón para confiar en ellos". Como para probar el punto de Harari, Musk anunció su propio chatbot de IA apenas unos días después de firmar la carta denunciando la investigación de otros.

Tal vez no sorprenda entonces que Harari mantenga su propio teléfono apagado, "en un cajón". Lo llama "un smartpone de emergencia" para cuando viaja al extranjero, lo que ha sido mucho más frecuente en la última década, desde que se convirtió en una superestrella intelectual mundial.

"Se ha vuelto realmente imposible hacer algunas cosas sin un smartphone". Uno piensa en él intentando pedir un taxi en un viaje al extranjero, y fracasando. Sin embargo, lo que no echa de menos de su pantalla de bolsillo es la avalancha de información que distrae su cerebro. Se declara "a dieta de información".

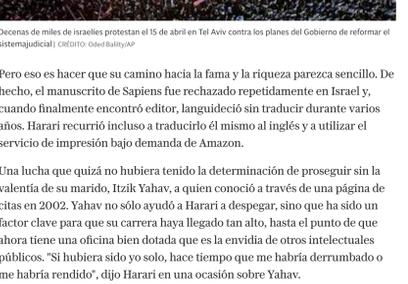
"Hay demasiada información basura", dice. "Es como la comida. Durante la mayor parte de la historia de la humanidad, intentábamos desesperadamente conseguir más comida. Y ahora estamos en la situación contraria. Tenemos que ser muy cuidadosos tanto con las cantidades como con la calidad de los alimentos que ingerimos."

Contrasta la concentración y la elección activa que implica la "lectura" con este "formato" pasivo de información. Es difícil imaginar a Harari haciendo algo de forma tan irreflexiva. Le gusta tener tiempo y tranquilidad para componer sus pensamientos, y desde hace tiempo es aficionado a la meditación para asegurarse de que está mentalmente sintonizado para mantener a raya las distracciones. "Acabo de volver de un retiro de meditación de dos meses...", dice. "Se puede decir que forma parte de la dieta de la información... un período para desconectar y permitir que la mente se desintoxique de toda la basura que ingerimos".

Harari puede tomarse tanto tiempo libre gracias a la seguridad económica que le ha proporcionado su fama (en un perfil del New Yorker, se sugería que sus honorarios por un solo discurso de 24 minutos ascendían a varios cientos de miles de dólares). Pero su interés por la meditación es anterior a su éxito, y su instinto para detenerse, reflexionar sobre grandes temas y establecer conexiones entre ellos se remonta aún más atrás, cuando descubrió que crecer gay le convertía en una especie de forastero que miraba al resto de la sociedad intentando explicarla. "El Israel de los 80 era muy homófobo", dice.

Creció cerca de Haifa, en el norte de Israel, y es el hijo menor (tiene dos hermanas mayores) de Shlomo, contratista de defensa, y Pnina Harari. Al final de su adolescencia se inclinó por la historia medieval, que estudió en la Universidad Hebrea de Jerusalén hasta 1998, cuando se trasladó al Jesus College de Oxford para completar su doctorado.

Harari regresó a Israel en 2003, el año en que la Hebrew U (como se conoce a la Universidad Hebrea), empezó a impartir una asignatura básica para estudiantes universitarios llamada "Introducción a la Historia del Mundo". Como especialistas, los académicos establecidos tendían a rechazar un curso tan amplio, por lo que Harari, como miembro nuevo y joven de la facultad, se encontró impartíéndolo. Nervioso, escribió guiones para sus 20 clases. Se convertirían en la base de Sapiens.



Decenas de miles de israelíes protestan el 15 de abril en Tel Aviv contra los planes del Gobierno de reformar el sistema judicial CREDITO: Oded Barry/AP

Pero eso es hacer que su camino hacia la fama y la riqueza parezca sencillo. De hecho, el manuscrito de Sapiens fue rechazado repetidamente en Israel, y cuando finalmente encontró editor, languideció sin traducir durante varios años. Harari recurrió incluso a traducirlo él mismo al inglés y a utilizar el servicio de impresión bajo demanda de Amazon.

Una lucha que quizá no hubiera tenido la determinación de proseguir sin la valentía de su marido, Itzik Yahav, a quien conoció a través de una página de citas en 2002. Yahav no sólo ayudó a Harari a despegar, sino que ha sido un factor clave para que su carrera haya llegado tan alto, hasta el punto de que ahora tiene una oficina bien dotada que es la envidia de otros intelectuales públicos. "Si hubiera sido yo solo, hace tiempo que me habría derrumbado o me habría rendido", dijo Harari en una ocasión sobre Yahav.

Desde entonces, Harari ha publicado otros dos arrolladores libros de historia: Homo Deus (2016) y 21 lecciones para el siglo XXI (2018). El año pasado publicó su primer libro infantil, Unstoppable Us: How Humans Took Over the World, y Sapiens también ha sido adaptado a una serie de novelas gráficas. Pero su preocupación más acuciante hoy en día es la situación de su propia nación, en particular los planes de la coalición liderada por Benjamin Netanyahu para recortar los poderes del Tribunal Supremo de Israel, que actualmente actúa como único freno al poder legislativo.

"El gobierno está intentando establecer aquí una dictadura", afirma Harari, con un toque de ironía en su habitual tono metódico. "Creo que existe una amenaza real para la supervivencia de la democracia israelí". Dice que es "absolutamente cierto" que los israelíes y las empresas israelíes se están marchando porque "el gobierno está tratando de desmantelar los controles y equilibrios de la democracia israelí."

"Si el Parlamento israelí aprueba una ley que, por ejemplo, quite el derecho de voto a los ciudadanos árabes -y algunos políticos han estado hablando exactamente de una ley así-, la única institución que puede intervenir, anular esa ley, es el Tribunal Supremo."

Dice que algunos miembros de la coalición quieren poder "amañar las elecciones", "ya han preparado una larga lista de leyes y reglamentos que discriminarán a los ciudadanos árabes, a las mujeres, a las personas LGBT, a los laicos" y "tienen opiniones mesiánicas y una fuerte creencia en la supremacía judía".

Aunque las medidas se han pausado tras las enormes protestas callejeras, "la crisis sigue su curso", afirma. "Y si el solo país es exitoso, entonces Israel seguirá el mismo camino que Turquía y Rusia - que todavía pueden celebrar elecciones periódicas, pero ya no son realmente democracias".

Las comparaciones con el gobierno de Viktor Orban en Hungría están fuera de lugar, afirma, porque Israel es "una gran potencia militar con capacidad nuclear, y también con capacidades cibernéticas que realmente tienen un alcance global. La gente en el Reino Unido debería preocuparse por lo que está ocurriendo en Israel porque podría desestabilizar todo Oriente Medio, con consecuencias inmediatas también para Europa".

Mi tiempo con el oráculo se pasa volando. Si pagara ya pasaría del millón de dólares. Consideramos la guerra de Ucrania ("las consecuencias para toda la humanidad podrían ser trágicas"); el cambio climático ("oigo a la gente decir que el cambio climático es algo que las democracias son intrínsecamente incapaces de abordar porque es un problema a largo plazo. No creo que eso sea cierto"); los cierres de Covid ("en las primeras fases de la pandemia, cuando no sabíamos a qué nos enfrentábamos, las medidas extremas estaban más justificadas"); los derechos de los transexuales ("Mi postura personal es tradicional. Creo que el sexo es un fenómeno biológico objetivo, mientras que el género es cultural"); la coronación ("En un solo país es muy difícil conseguir que decenas de millones de personas convivan y se pongan de acuerdo en algo. En Gran Bretaña, la familia real desempeña un papel simbólico muy importante"). Finalmente, terminamos hablando de la felicidad. Harari concluye Sapiens señalando que, a pesar del enorme aumento de nuestra prosperidad, la humanidad "parece tan descontenta como siempre". ¿Le ocurre lo mismo?

"Mi vida cambió por completo [con la fama]", afirma. "Hace diez años nadie quería entrevistarme sobre nada. Así que tenía mucho tiempo para leer libros y escribir mis cosas. La fama a nivel personal, normalmente sólo crea más problemas". Sin embargo, en conjunto, los amigos, la meditación y la terapia, y el haberme despojado de las ansiedades de la juventud significan que "ahora soy más feliz".

Es una forma optimista de terminar un debate con un hombre que, como algunos oráculos de antaño, a veces es criticado por ofrecer un análisis sombrío de la condición de la humanidad, pero sin ofrecer ningún remedio (su libro más famoso concluye: "El régimen Sapiens en la Tierra ha producido hasta ahora poco de lo que podamos estar orgullosos"): "El régimen Sapiens en la Tierra ha producido hasta ahora poco de lo que podamos estar orgullosos").

Pero cuando terminamos no es su capacidad para describir con elegancia tanto catástrofes potenciales lo que me inquieta profundamente, sino su incapacidad para explicar por qué nuestra especie tiene esa costumbre de causar estragos. Sus palabras resuenan en mis oídos: "Lo mejor que puedo decir es que la historia está llena de errores. Muchos acontecimientos importantes no son el resultado de fuerzas inevitables de la historia. Son simplemente el resultado de terribles errores humanos. Lo básico que hay que asumir es que las personas son fallibles. Las personas son corruptibles. Una buena institución o un buen país es aquel que goza de fuertes mecanismos de autocorrección".

Como el Tribunal Supremo de Israel. Es una lección, piensa, no sólo clave para la supervivencia de su nación como democracia, sino para la supervivencia del mundo y punto.

Temas relacionados
Inteligencia artificial

El reto de Kemi Badenoch para aprovechar al máximo el Brexit

Me mandó al infierno: Una viétnita española de violación culpa a la ministra de Igualdad tras la chapaza de la ley que reduce las penas

La BBC debe devolver el "acento británico" a sus informativos, dice Simon McCoy

Los cirujanos sólo operan una vez cada quince días

Se publica por primera vez una foto del Rey Carlos en la matrícula de la Universidad de Cambridge

Corredores y corredoras para sustituir a Richard Sharp como presidente de la BBC

VISTA DEL TELÉGRAFO
29 Abr 2023, 6:00am

Por James Backcock
29 Abr 2023, 6:00am

Por Craig Simpson
29 abr 2023, 12:46

Por Laura Donnelly
29 Abr 2023, 12:37am

Por Reporteros del Telégrafo
29 Abr 2023, 12:07am

Por Dominic Penna
28 abr 2023, 22:38

More from The Telegraph

Por invitación | Inteligencia artificial

Yuval Noah Harari afirma que la IA ha pirateado el sistema operativo de la civilización humana

Los ordenadores para contar historias cambiarán el curso de la historia de la humanidad, afirma el historiador y filósofo



Dean Williams

28 de abril de 2023

Compartir

EL MIEDO A LA INTELIGENCIA ARTIFICIAL (IA) ha perseguido a la humanidad desde los inicios de la era informática. Hasta ahora, estos temores se centran en las máquinas que utilizaban medios físicos para matar, esclavizar o sustituir a las personas. Pero en los últimos dos años han aparecido nuevas herramientas de IA que amenazan la supervivencia de la civilización humana desde una dirección inesperada. La IA ha adquirido notables capacidades para manipular y generar lenguaje, ya sea con palabras, sonidos o imágenes. De este modo, la IA ha pirateado el sistema operativo de nuestra civilización.

El lenguaje es la materia de la que está hecha casi toda la cultura humana. Los derechos humanos, por ejemplo, no están inscritos en nuestro ADN. Son más bien artefactos culturales que creamos contando historias y escribiendo leyes. Los dioses no son realidades físicas. Son más bien artefactos culturales que creamos inventando mitos y escribiendo escrituras.

El dinero también es un artefacto cultural. Los billetes no son más que papeles de colores y, en la actualidad, más del 90% del dinero ni siquiera son billetes, sino información digital almacenada en ordenadores. Lo que da valor al dinero son las historias que banqueros, ministros de finanzas y gurús de las criptomonedas nos cuentan sobre él. Sam Bankman-Fried, Elizabeth Holmes y Bernie Madoff no eran especialmente buenos en la creación de valor real, pero todos ellos eran narradores extremadamente capaces.

¿Qué pasaría cuando una inteligencia no humana fuera mejor que el ser humano medio a la hora de contar historias, componer melodías, dibujar imágenes y redactar leyes y escrituras? Cuando la gente piensa en ChatGPT y otras nuevas herramientas de IA, a menudo le vienen a la cabeza ejemplos como el de los escolares que utilizan la IA para escribir sus redacciones. ¿Qué pasará con el sistema escolar cuando los niños hagan eso? Pero este tipo de pregunta no tiene en cuenta el panorama general. Olvídense de las redacciones escolares. Piense en las próximas elecciones presidenciales de Estados Unidos en 2024 e intente imaginar el impacto de las herramientas de IA que pueden utilizarse para producir en masa contenidos políticos, noticias falsas y escrituras para nuevos cultos.

En los últimos años, el culto QAnon se ha aglutinado en torno a mensajes anónimos en línea, conocidos como "gotas". Los seguidores coleccionaban, veneraban e interpretaban estos "drops" como un texto sagrado. Aunque, por lo que sabemos, todos los "drops" anteriores fueron compuestos por humanos, y los bots simplemente ayudaron a difundirlos, en el futuro podríamos ver los primeros cultos de la historia cuyos textos venerados fueron escritos por una inteligencia no humana. Las religiones a lo largo de la historia han reivindicado una fuente no humana para sus libros sagrados. Pronto podría ser una realidad.

En un nivel más prosaico, pronto podríamos encontrarnos manteniendo largas discusiones en línea sobre el aborto, el cambio climático o la invasión rusa de Ucrania con entidades que pensamos que son humanos, pero que en realidad son IA. El problema es que es totalmente inútil que dediquemos tiempo a intentar cambiar las opiniones declaradas de un robot de IA, mientras que la IA podría perfeccionar sus mensajes con tanta precisión que tendría muchas posibilidades de influir en nosotros.

Gracias a su dominio del lenguaje, la IA podría incluso entablar relaciones íntimas con las personas y utilizar el poder de la intimidad para cambiar nuestras opiniones y visiones del mundo. Aunque no hay indicios de que la IA tenga conciencia o sentimientos propios, para fomentar una falsa intimidad con los humanos basta con que la IA consiga que éstos se sientan emocionalmente unidos a ella. En junio de 2022, Blake Lemoine, ingeniero de Google, afirmó públicamente que el chatbot de IA LAMDA, en el que estaba trabajando, se había vuelto sensible. La polémica afirmación le costó el puesto. Lo más interesante de este episodio no fue la afirmación de Lemoine, probablemente falsa. Lo más interesante de este episodio no fue la afirmación de Lemoine, probablemente falsa, sino su voluntad de arriesgar su lucrativo empleo por el chatbot de inteligencia artificial. Si la IA puede influir en la gente para que arriesgue su trabajo por ella, ¿qué más podría inducirles a hacer?

En una batalla política por las mentes y los corazones, la intimidad es el arma más eficaz, y la IA acaba de adquirir la capacidad de producir en masa relaciones íntimas con millones de personas. Todos sabemos que en la última década las redes sociales se han convertido en un campo de batalla para controlar la atención humana. Con la nueva generación de IA, el frente de batalla está pasando de la atención a la intimidad. ¿Qué ocurrirá con la sociedad y la psicología humanas cuando la IA se enfrente a la IA en una batalla por fingir relaciones íntimas con nosotros, que luego pueden utilizarse para convencernos de que votemos a determinados políticos o compremos determinados productos?

Incluso sin crear una "falsa intimidad", las nuevas herramientas de IA tendrían una inmensa influencia en nuestras opiniones y visiones del mundo. La gente podría llegar a utilizar un único asesor de IA como un oráculo único que todo lo sabe. No es de extrañar que Google esté aterrizado. ¿Para qué molestarse en buscar, si puedo preguntarle al oráculo? Las industrias periodística y publicitaria también deberían estar aterrizadas. ¿Para qué leer un periódico si puedo pedirle al oráculo que me cuente las últimas noticias? ¿Y para qué sirven los anuncios si puedo pedirle al oráculo que me diga qué comprar?

E incluso estos escenarios no captan realmente el panorama general. De lo que estamos hablando es potencialmente del fin de la historia humana. No el fin de la historia, sino el fin de su parte dominada por los humanos. La historia es la interacción entre biología y cultura; entre nuestras necesidades y deseos biológicos de cosas como la comida y el sexo, y nuestras creaciones culturales como las religiones y las leyes. La historia es el proceso a través del cual las leyes y las religiones dan forma a la comida y al sexo.

¿Qué ocurrirá con el curso de la historia cuando la IA se haga cargo de la cultura y empiece a producir historias, melodías, leyes y religiones? Herramientas anteriores como la imprenta y la radio ayudaron a difundir las ideas culturales de los humanos, pero nunca crearon nuevas ideas culturales propias. La IA es fundamentalmente diferente. La IA puede crear ideas completamente nuevas, una cultura completamente nueva.

Al principio, es probable que la IA imite a los prototipos humanos con los que se entrenó en su infancia. Pero con el paso de los años, la cultura de la IA se atreverá a llegar donde ningún ser humano ha llegado antes. Durante milenios, los seres humanos han vivido dentro de los sueños de otros seres humanos. En las próximas décadas podríamos encontrarnos viviendo dentro de los sueños de una inteligencia alienígena.

El miedo a la IA ha perseguido a la humanidad sólo durante las últimas décadas. Pero durante miles de años los humanos han sido perseguidos por un miedo mucho más profundo. Siempre hemos apreciado el poder de las historias y las imágenes para manipular nuestras mentes y crear ilusiones. En consecuencia, desde la antigüedad los humanos han temido quedar atrapados en un mundo de ilusiones.

En el siglo XVII, René Descartes temía que tal vez un demonio malicioso le estuviera atrapando dentro de un mundo de ilusiones, creando todo lo que veía y oía. En la antigua Grecia, Platón relató la famosa Alegoría de la caverna, en la que un grupo de personas permanecen encadenadas dentro de una cueva toda su vida, frente a una pared en blanco. Una pantalla. En esa pantalla ven proyectadas diversas sombras. Los prisioneros confunden las ilusiones que ven allí con la realidad.

En la antigua India, los sabios budistas e hindúes señalaban que todos los seres humanos vivían atrapados en *Maya*, el mundo de las ilusiones. Lo que normalmente tomamos por realidad a menudo no son más que ficciones en nuestras propias mentes. La gente puede librar guerras enteras, matando a otros y estando dispuesta a que la maten a ella misma, debido a su creencia en esta o aquella ilusión.

La revolución de la IA nos enfrenta al demonio de Descartes, a la caverna de Platón y a los *mayas*. Si no tenemos cuidado, podríamos quedar atrapados tras una cortina de ilusiones que no podríamos arrancar, ni siquiera darnos cuenta de que está ahí.

Por supuesto, el nuevo poder de la IA también podría utilizarse con buenos fines. No voy a insistir en esto, porque la gente que desarrolla la IA ya habla bastante de ello. El trabajo de los historiadores y filósofos como yo es señalar los peligros. Pero, ciertamente, la IA puede ayudarnos de innumerables maneras, desde encontrar nuevas curas para el cáncer hasta descubrir soluciones a la crisis ecológica. La cuestión a la que nos enfrentamos es cómo garantizar que las nuevas herramientas de IA se utilicen para el bien y no para el mal. Para ello, primero tenemos que apreciar las verdaderas capacidades de estas herramientas.

Desde 1945 sabemos que la tecnología nuclear puede generar energía barata en beneficio de los seres humanos, pero también puede destruir físicamente la civilización humana. Por ello, hemos reformado todo el orden internacional para proteger a la humanidad y asegurarnos de que la tecnología nuclear se utilizara principalmente para el bien. Ahora tenemos que lidiar con una nueva arma de destrucción masiva que puede aniquilar nuestro mundo mental y social.

Aún podemos regular las nuevas herramientas de IA, pero debemos actuar con rapidez. Mientras que las armas nucleares no pueden inventar armas nucleares más potentes, la IA puede crear una IA exponencialmente más potente. El primer paso crucial es exigir rigurosos controles de seguridad antes de que las potentes herramientas de IA salgan al dominio público. Del mismo modo que una empresa farmacéutica no puede lanzar nuevos medicamentos antes de probar sus efectos secundarios a corto y largo plazo, las empresas tecnológicas no deberían lanzar nuevas herramientas de IA antes de que sean seguras. Necesitamos un equivalente de la Administración de Alimentos y Medicamentos para las nuevas tecnologías, y lo necesitamos ayer.

¿Ralentizar el despliegue público de la IA no hará que las democracias se queden rezagadas con respecto a los regímenes autoritarios más despiadados? Todo lo contrario. Los despliegues no regulados de IA crearían un caos social que beneficiaría a los autócratas y arruinaría las democracias. La democracia es una conversación, y las conversaciones se basan en el lenguaje. Si la IA piratea el lenguaje, podría destruir nuestra capacidad de mantener conversaciones significativas, destruyendo así la democracia.

Acabamos de encontrarnos con una inteligencia alienígena, aquí en la Tierra. No sabemos mucho sobre ella, salvo que podría destruir nuestra civilización. Deberíamos poner fin al despliegue irresponsable de herramientas de IA en la esfera pública, y regular la IA antes de que ella nos regule a nosotros. Y la primera regulación que yo sugeriría es hacer obligatorio que la IA revele que es una IA. Si estoy conversando con alguien y no sé si es un ser humano o una IA, se acabó la democracia.

Este texto ha sido generado por un ser humano.

¿O no?

Yuval Noah Harari es historiador, filósofo y autor de "Sapiens", "Homo Deus" y la serie infantil "Unstoppable Us". Es profesor del Departamento de Historia de la Universidad Hebrea de Jerusalén y cofundador de Sapienship, una empresa de impacto social.

Share Reuse this content

More from By Invitation



Dean Williams

Avi Shlaim calls for critical reflection

Like Zionism itself, the state has become a settler-colonial project, argues the historian and author

South Korea has earned the right to be more assertive, says a former minister

Kang Kyung-wha believes that the country faces tests in its relations with both America and Japan



Dean Williams

There is much to celebrate—and worry about—says Yair Lapid

The leader of the opposition on the fight for the nation's future



Dean Williams

Subscribe

Group subscriptions

Reuse our content

The Trust Project

Help and contact us

Keep updated



Published since September 1843 to take part in "a severe contest between intelligence, which presses forward, and an unworthy, timid ignorance obstructing our progress."

The Economist

About

Advertise

Press centre

The Economist Group

The Economist Group

Economist Intelligence

Economist Impact

Economist Events

Working Here

Economist Education Courses

Which MBA?

Executive Jobs

Executive Education Navigator

Terms of Use

Privacy

Cookie Policy

Manage Cookies

Accessibility

Modern Slavery Statement

Sitemap

Copyright © The Economist Newspaper Limited 2023. All rights reserved.

California: Do Not Sell My Personal Information

Translation in progress...