La inteligencia artificial (IA) ha pirateado el sistema operativo de la civilización humana

Yuval Noah Harari

Según el historiador y filósofo, las computadoras que cuentan historias cambiarán el curso de la historia humana. Desde los inicios de la era de la computación, los temores sobre la IA han acechado a la humanidad. Hasta ahora, estos temores se centraban en máquinas que utilizaban medios físicos para matar, esclavizar o reemplazar a las personas. Sin embargo, en los últimos años han surgido nuevas herramientas de IA que amenazan la supervivencia de la civilización humana desde una dirección inesperada. La IA ha adquirido habilidades notables 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 el elemento principal de casi toda la cultura humana. Los derechos humanos, por ejemplo, no están inscritos en nuestro ADN. Más bien, son artefactos culturales que creamos mediante la narración de historias y la redacción de leyes. Los dioses no son realidades físicas, sino artefactos culturales que creamos inventando mitos y escribiendo escrituras.

El dinero también es un artefacto cultural. Los billetes son simplemente trozos de papel de colores, y en la actualidad más del 90% del dinero ni siquiera son billetes, sino información digital en computadoras. Lo que le da valor al dinero son las historias que los banqueros, los ministros de finanzas y los gurús de las criptomonedas nos cuentan al respecto. Sam Bankman-Fried, Elizabeth Holmes y Bernie Madoff no eran especialmente buenos creando un valor real, pero todos eran extremadamente hábiles contando historias.

¿Qué sucedería una vez que una inteligencia no humana se vuelva mejor que el humano promedio para 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 se enfocan en ejemplos como los niños que utilizan la IA para escribir sus ensayos. ¿Qué sucederá con el sistema educativo cuando los niños hagan eso? Pero este tipo de pregunta pasa por alto el panorama general. Olvidemos los ensayos escolares. Piense en la próxima carrera presidencial estadounidense en 2024 e intente imaginar el impacto de las herramientas de IA que pueden producir contenido político en masa, historias de noticias falsas y escrituras para nuevas sectas.

En los últimos años, el culto de QAnon se ha consolidado en torno a mensajes en línea anónimos conocidos como "gotas Q". Los seguidores recolectaron, veneraron e interpretaron estas gotas Q como un texto sagrado. Hasta donde sabemos, todas las gotas Q anteriores fueron compuestas por humanos, y los bots solo ayudaron a difundirlas. Sin embargo, en el futuro podríamos ver los primeros cultos de la historia cuyos textos venerados fueron escritos por una inteligencia no humana. A lo largo de la historia, las religiones han afirmado que sus libros sagrados tienen un origen no humano. Pronto, eso podría convertirse en una realidad.

A un nivel más prosaico, es posible que pronto nos encontremos participando en extensas discusiones en línea sobre el aborto, el cambio climático o la invasión rusa de Ucrania con entidades que creemos que son humanos, pero que en realidad son IA. El problema es que es completamente inútil para nosotros pasar tiempo tratando de cambiar las opiniones declaradas de un bot de IA, mientras que la IA podría perfeccionar sus mensajes de manera tan precisa que tiene muchas posibilidades de influir en nosotros.

A través de su dominio del lenguaje, la IA incluso podría formar relaciones íntimas con las personas y utilizar el poder de la intimidad para cambiar nuestras opiniones y perspectivas del mundo. Aunque no hay indicios de que la IA tenga conciencia o sentimientos propios, basta con que la IA logre hacer que las personas se sientan emocionalmente vinculadas a ella para fomentar una falsa intimidad. En junio de 2022, Blake Lemoine, un ingeniero de Google, afirmó públicamente que el chatbot de IA Lambda, en el que estaba trabajando, se había vuelto consciente. La polémica afirmación le costó su trabajo. Lo más interesante de este episodio no fue la afirmación de Lemoine, que probablemente era falsa, sino su disposición a arriesgar su lucrativo empleo por el bien del chatbot de IA. Si la IA puede influir en las personas para que arriesguen sus empleos por ella, ¿qué más podría inducirnos a hacer?

En una batalla política por las mentes y los corazones, la intimidad es el arma más eficiente, y la IA acaba de adquirir la capacidad de producir relaciones íntimas en masa con millones de personas. Todos sabemos que durante 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 se está desplazando de la atención a la intimidad. ¿Qué sucederá en la sociedad humana y en la psicología humana cuando la IA luche contra la IA en una batalla para crear relaciones íntimas falsas con nosotros, que luego se puedan utilizar para convencernos de votar por políticos específicos o comprar productos específicos?

Incluso sin crear "falsas intimidades", las nuevas herramientas de IA tendrían una influencia inmensa en nuestras opiniones y perspectivas del mundo. Es posible que las personas empiecen a utilizar un único asesor de IA como un oráculo omnisciente. No es de extrañar que Google esté aterrado. ¿Por qué molestarse en buscar cuando puedo simplemente preguntar al oráculo? Las industrias de noticias y

publicidad también deberían estar aterradas. ¿Por qué leer un periódico cuando puedo preguntarle al oráculo las últimas noticias? ¿Y cuál es el propósito de los anuncios cuando puedo preguntarle al oráculo qué comprar?

Incluso estas situaciones no capturan realmente el panorama general. Lo que estamos hablando es potencialmente el fin de la historia humana. No el fin de la historia en sí, sino el fin de su parte dominada por los humanos. La historia es la interacción entre la biología y la cultura; entre nuestras necesidades y deseos biológicos, 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 moldean la comida y el sexo.

¿Qué sucederá con el curso de la historia cuando la IA tome el control de la cultura y comience a producir historias, melodías, leyes y religiones? Las herramientas anteriores, como la imprenta y la radio, ayudaron a difundir las ideas culturales de los humanos, pero nunca crearon nuevas ideas culturales por sí mismas. La IA es fundamentalmente diferente. La IA puede crear ideas completamente nuevas, una cultura completamente nueva.

Al principio, es probable que la IA imite los prototipos humanos en los que fue entrenada en sus primeros días. Pero con cada año que pasa, la cultura de la IA se aventurará audazmente hacia lugares a los que ningún 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 acosado a la humanidad solo en las últimas décadas. Pero durante miles de años, los humanos han sido acosados por un temor mucho más profundo. Siempre hemos apreciado el poder de las historias y las imágenes para

manipular nuestras mentes y crear ilusiones. En consecuencia, desde tiempos antiguos los humanos han temido quedar atrapados en un mundo de ilusiones.

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

En la antigua India, los sabios budistas e hindúes señalaron que todos los seres humanos vivían atrapados dentro de Maya, el mundo de las ilusiones. Lo que normalmente consideramos realidad a menudo son solo ficciones en nuestras propias mentes. Las personas pueden librar guerras enteras, matando a otros y dispuestas a ser asesinadas ellas mismas, debido a su creencia en esta o aquella ilusión.

La revolución de la IA nos está enfrentando al demonio de Descartes, a la caverna de Platón, a la Maya. Si no tenemos cuidado, podríamos quedar atrapados detrás de un velo de ilusiones, que no podríamos desgarrar, ni siquiera darnos cuenta de que está ahí.

Por supuesto, el nuevo poder de la IA también se podría utilizar para propósitos buenos. No profundizaré en esto, porque las personas que desarrollan la IA ya hablan lo suficiente al respecto. El trabajo de historiadores y filósofos como yo es señalar los peligros. Pero ciertamente, la IA puede ayudarnos de innumerables formas, desde encontrar nuevas curas para el cáncer hasta descubrir soluciones para la crisis ecológica. La pregunta que enfrentamos es cómo asegurarnos de que

las nuevas herramientas de IA se utilicen para el bien en lugar del mal. Para lograrlo, primero debemos apreciar las verdaderas capacidades de estas herramientas.

Desde 1945 sabemos que la tecnología nuclear puede generar energía barata en beneficio de los humanos, pero también puede destruir físicamente la civilización humana. Por lo tanto, reconfiguramos todo el orden internacional para proteger a la humanidad y asegurarnos de que la tecnología nuclear se usara principalmente para el bien. Ahora debemos enfrentarnos a un nuevo arma de destrucción masiva que puede aniquilar nuestro mundo mental y social.

Todavía podemos regular las nuevas herramientas de IA, pero debemos actuar rápidamente. Mientras que las armas nucleares no pueden inventar armas nucleares más poderosas, la IA puede crear IA exponencialmente más poderosas. El primer paso crucial es exigir controles de seguridad rigurosos antes de que las potentes herramientas de IA se pongan a disposición del público. Al igual que una compañía farmacéutica no puede lanzar nuevos medicamentos sin probar sus efectos secundarios a corto y largo plazo, las empresas tecnológicas no deberían lanzar nuevas herramientas de IA antes de que se vuelvan seguras. Necesitamos un equivalente de la Administración de Alimentos y Medicamentos para la nueva tecnología, y lo necesitamos ayer.

¿No ralentizar las implementaciones públicas de IA hará que las democracias queden rezagadas frente a regímenes autoritarios más despiadados? Justamente lo contrario. Las implementaciones de IA no reguladas 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. Cuando la IA piratea el lenguaje, puede destruir nuestra capacidad de tener conversaciones significativas, lo que a su vez destruye la democracia.

Nos hemos encontrado con una inteligencia alienígena aquí en la Tierra. No sabemos mucho al respecto, excepto que podría destruir nuestra civilización. Debemos poner fin al despliegue irresponsable de herramientas de IA en el ámbito público y regular la IA antes de que nos regule a nosotros. Y la primera regulación que sugeriría es hacer obligatorio que la IA revele que es una IA. Si estoy teniendo una conversación con alguien y no puedo decir si es humano o una IA, eso significa el fin de la democracia.

Este texto ha sido generado por un humano.

¿O lo ha sido?