

# Qué es la misteriosa “caja negra” de la inteligencia artificial que desconcierta a los expertos (y por qué aún no entendemos cómo aprenden las máquinas)

20 abril, 2023



Los desarrolladores de inteligencia artificial (IA) de Google se encontraron hace poco con la misteriosa “caja negra”.

Unos ingenieros estaban trabajando en un software de IA que, de manera inesperada, adquirió la habilidad de entender un nuevo idioma.

“Descubrimos que con muy pocas indicaciones en bengalí, ahora puede traducir todo el bengalí”, contó James Maneka, el jefe de la división de IA de Google, en la reciente emisión del

programa 60 Minutes, de la cadena CBS.

Al comentar al respecto, el director ejecutivo de Google, Sundar Pichai, dijo que esta capacidad de los programas de IA de generar habilidades u ofrecer respuestas de manera inesperada es lo que los expertos llaman la “caja negra”.

“No lo entiendes del todo. No puedes decir muy bien por qué [el robot] dijo esto o por qué se equivocó en esto otro. Tenemos algunas ideas y nuestra capacidad para entenderlo mejora con el tiempo. Pero ahí es donde estamos hasta ahora”, señaló.

El desarrollo de la inteligencia artificial se ha acelerado en los últimos años. Las grandes firmas tecnológicas están invirtiendo importantes sumas para generar chatbots como Bard, de Google, o ChatGPT, de OpenAi y Microsoft. Más recientemente el multimillonario Elon Musk anunció que entraría a esta carrera.

Al mismo tiempo, expertos en AI han señalado que puede haber riesgos si la tecnología no es controlada por sus desarrolladores, como el que adquieran habilidades para las que no fueron instruidas.

En este último escenario es en donde habita la caja negra.

### **La caja blanca**

La idea de la caja negra es la parte opuesta a lo que los expertos en IA llaman caja blanca.

Ian Hogarth, cofundador de la empresa tecnológica Plural y autor del libro The Status of AI Reports, explica a la BBC que cuando los desarrolladores de software crean un programa de manera “tradicional”, las líneas de código que introducen se reflejan claramente en el resultado que arroja el software.

“A veces la gente lo describe como una caja blanca”, señala.

“Pero en la IA, los sistemas son realmente muy diferentes. Están más cerca de una caja negra en muchos sentidos, pues no entiendes realmente lo que sucede dentro”, añade.

Y es que los programadores no esperan que arrojen resultados tan creativos.

“Me gusta pensar que las cultivamos. Esa es la mejor idea que he oído sobre cómo construimos estos sistemas hoy en día. Y lo difícil es que, a medida que las desarrollamos, se producen grandes saltos bruscos en sus capacidades”, explica.

A diferencia de la programación de software más tradicional, basada en la implementación de instrucciones para obtener un resultado, en el desarrollo de IA los ingenieros trabajan para conseguir un sistema que imite las “redes neuronales” de la inteligencia humana.

Esto involucra una gran cantidad de procesadores interconectados que pueden manejar grandes cantidades de datos, detectar patrones entre millones de variables utilizando el aprendizaje automático y, más importante, autoadaptarse en respuesta a lo que están haciendo.

David Stern, gerente de investigación cuantitativa en G-Research, una empresa de tecnología que utiliza el aprendizaje automático para predecir precios en los mercados financieros, advierte que “el progreso más rápido en la investigación de IA en los últimos años ha implicado un enfoque de caja negra cada vez más basado en datos”.

“En el enfoque de red neuronal que es tendencia en la actualidad, este procedimiento de entrenamiento determina la configuración de millones de parámetros internos que interactúan de manera compleja y son muy difíciles de explicar y aplicar ingeniería inversa”, señaló a la BBC.

Otra tendencia es el “aprendizaje de refuerzo profundo” en el que un “diseñador simplemente especifica los objetivos de

comportamiento del sistema y este aprende automáticamente interactuando directamente con el entorno”, expone.

“Esto da como resultado un sistema que es aún más difícil de entender”.

### **¿Debe preocuparnos?**

Ian Hogarth hace notar que, aunque los nuevos chatbots como Bard y ChatGPT parecen una ser tecnología surgida muy recientemente, en realidad son producto de una década de investigación y desarrollo.

“Si nos remontamos a 2012 y comparamos los sistemas que estábamos creando entonces y los sistemas que estamos creando ahora, hemos aumentado de manera muy consistente la cantidad de datos y la potencia de cálculo en el desarrollo de los modelos de IA”, explicó a la BBC.

“Hemos aumentado la cantidad de potencia de cálculo consumida por estos modelos en aproximadamente 100 millones en la última década. Así que, aunque en la práctica ChatGPT parezca que surgió de la nada para la mayoría de las personas, esta es una tendencia de muy larga data que continuará”.

Las nuevas capacidades de robots que se han dado a conocer recientemente han generado cuestionamientos sobre las diversas formas en las que tendrán influencia en la sociedad, desde los impactos en el mercado laboral hasta el control de procesos de seguridad pública o el ámbito militar.

En el programa de 60 Minutes, el director ejecutivo de Google fue cuestionado sobre el hecho de que los ingenieros no comprendan por completo cómo pasan cosas en la caja negra, pero aún así ya estén funcionando chatbots como Bard.

“Permítanme decirlo de esta manera: tampoco creo que entendamos completamente cómo funciona la mente humana”, respondió Sundar Pichai, quien considera que la llegada

progresiva de la IA a la sociedad es una manera de acostumbrarse a ella.

“Creo que en el desarrollo de esto debe incluirse no solo a ingenieros, sino también a científicos sociales, especialistas en ética, filósofos, etcétera. Y tenemos que ser muy reflexivos. Estas son las cosas que la sociedad necesita resolver a medida que avanzamos. No le corresponde a una empresa decidir”, señaló a 60 Minutes.

Ian Hogarth también considera que la IA terminará por tocar la vida de la gente y es necesario un diálogo abierto sobre sus impactos.

“Creo que tiene un potencial notable para transformar todos los aspectos de nuestras vidas. En cierto modo, es quizás la tecnología actual más poderosa. Lo principal es que deberíamos tener un debate mucho más público sobre la rapidez con la que progresan estos sistemas y lo diferentes que son de las generaciones anteriores de software”, concluye.

Fuente: BBC Mundo