INTELIGENCIA ARTIFICIAL

**Qué es la misteriosa   
“caja negra” de la IA que desconcierta a los expertos**

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

**Por Julián M. Zappia.  
Diplomatic Week.**

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

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.

Bard ofrece respuestas a preguntas o instrucciones que le dé el usuario, al igual que otros chatbots como ChatGPT.

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.

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.

Esto se lo describe, a veces. como una “caja blanca".

Pero en la IA, los sistemas son realmente muy diferentes. Están más cerca de una caja negra en muchos sentidos, porque no se entiende realmente lo que sucede dentro.

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.

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.

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.

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

**¿Debe preocuparnos?**

Aunque los nuevos chatbots como Bard y ChatGPT parecen ser una tecnología surgida recientemente, en realidad son producto de una década de investigación y desarrollo.

Si nos remontamos a 2012 y comparamos los sistemas que se creaban en ese entonces y los sistemas que se están creando ahora, se ha aumentado de manera muy consistente la cantidad de datos y la potencia de cálculo en el desarrollo de los modelos de IA.

Se ha 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 un programa de televisión de los Estados Unidos, 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 se encuentran funcionando chatbots como Bard.

Ante el cuestionamiento, Sundar Pichai respodió: "Permítanme decirlo de esta manera; tampoco creo que entendamos completamente cómo funciona la mente humana".

Es necesario que en el desarrollo de los sistemas de IA se incluyan no solo a ingenieros, sino también a científicos sociales, especialistas en ética, filósofos, etcétera.

Tenemos que ser muy reflexivos.