

## ¿Puede mentir la inteligencia artificial?

**La inteligencia artificial es, sin duda, una de las revoluciones del siglo XXI**



La inteligencia artificial puede ser engañada con relativa facilidad Shutterstock

Es el actor clave de la llamada **cuarta revolución industrial**, y se aplicó con éxito a campos tan dispares como la medicina, la enseñanza y las ciencias sociales.

### **PENSAMIENTO VISUAL: ESTOS SON LOS BENEFICIOS DE ESTUDIAR HACIENDO DIBUJOS E INFOGRAFÍAS**

Esta disciplina nació a mediados del siglo pasado, aunque no fue hasta hace pocos años cuando logró sus resultados más valiosos e impactantes. **El éxito fue posible gracias a la combinación del fenómeno del *big data* y el aumento en la potencia de cómputo de los ordenadores.**

## ¿Podemos confiar en un sistema inteligente?

El uso de la inteligencia artificial está cada vez más extendido en nuestra vida cotidiana. Debido a **la aparición de casos donde este tipo de sistemas se comportaban de forma racista o sexista**, surge la necesidad de que la inteligencia artificial sea ética y confiable. Así lo demuestran las *Directrices éticas para una inteligencia artificial fiable* de la Unión Europea y la *Estrategia Nacional de Inteligencia Artificial* española.

Los sistemas de inteligencia artificial que más éxitos están cosechando tienen la desventaja de ser, en muchos casos, imposibles de interpretar. En un contexto en el que se nos insta a confiar en ellos, esto hace que nos cuestionemos si este tipo de sistemas pueden ser

fácilmente engañados. **O incluso peor: preguntarnos si tienen el poder de engañarnos a nosotros.**

Es el actor clave de [la llamada cuarta revolución industrial](#), y se aplicó con éxito a campos tan dispares como la medicina, **la enseñanza y las ciencias sociales.**

#### PENSAMIENTO VISUAL: ESTOS SON LOS BENEFICIOS DE ESTUDIAR HACIENDO DIBUJOS E INFOGRAFÍAS

Esta disciplina nació a mediados del siglo pasado, aunque no fue hasta hace pocos años cuando logró sus resultados más valiosos e impactantes. **El éxito fue posible gracias a la combinación del fenómeno del *big data* y el aumento en la potencia de cómputo de los ordenadores.**



Las máquinas, por su parte, aprenden lo que les enseñamos los humano

#### ¿Podemos engañar a un sistema inteligente?

**En los últimos años, cobró importancia una nueva rama dentro de la inteligencia artificial: el aprendizaje adversario.** Los sistemas inteligentes se usan en muchas ocasiones para tomar decisiones importantes que afectan a las personas, por lo que es necesario asegurarse de que no pueden ser engañados con facilidad.

El aprendizaje adversario intenta prevenir posibles ataques, o la introducción de datos falsos que puedan engañar a una máquina, pero no a una persona. Son famosos los casos en los que se añade un mínimo ruido a una imagen, que es imperceptible para el ojo humano. **Este ruido es suficiente para que el algoritmo de inteligencia artificial crea que se trata de una imagen diferente.**

En algunas ocasiones, el engaño es tan difícil de creer como el caso de una impresión 3D de [una tortuga que la mismísima inteligencia artificial de Google clasificó como un rifle](#). **La investigación sobre este tipo de imágenes adversarias, que pueden engañar a un sistema inteligente, es de vital importancia.** Imaginemos las graves consecuencias de que un vehículo autónomo malinterprete señales de tráfico.

#### EL GAMING NO TIENE EDAD Y PUEDE AYUDAR A ENVEJECER CON SALUD

Sabemos, entonces, que la inteligencia artificial puede ser engañada con relativa facilidad. **Pero ¿puede un sistema inteligente engañar a las personas?**

## Aprender a mentir para ganar

*A los humanos nos intentan educar en la honradez, pero el engaño y la mentira son parte de nosotros, así como de algunos animales. Habitualmente mentimos o manipulamos para lograr algún tipo de beneficio. Incluso algunas plantas como la orquídea *Ophrys apifera* tienen flores que simulan ser una abeja reina para atraer a los machos de esta especie y que ayuden a polinizar las flores.*

Las máquinas, por su parte, aprenden lo que les enseñamos los humanos. Un niño de dos años aprende lo que es un perro o un gato porque un adulto se lo enseña. La mayoría de algoritmos de inteligencia artificial aprenden de la misma forma. [El desarrollador le proporciona miles o millones de ejemplos](#) donde se le indica, por ejemplo, qué imágenes corresponden a perros y cuáles a gatos.



Si usamos aprendizaje por refuerzo, y el engaño lleva a un sistema inteligente a conseguir el objetivo perseguido. Getty Images

Recientemente apareció una nueva técnica de inteligencia artificial, el aprendizaje por refuerzo, que construye un sistema de recompensas para el algoritmo. Cuando se consigue un resultado satisfactorio, se refuerza ese comportamiento. Por el contrario, cuando el resultado dista de su objetivo, ese comportamiento se descarta.

Los sistemas basados en aprendizaje por refuerzo son muy exitosos, como el sistema **AlphaZero** de la empresa **DeepMind**.

*Este programa de inteligencia artificial adquirió en solo 24 horas un nivel sobrehumano en varios juegos de mesa, derrotando a otros programas anteriores que habían ganado a campeones humanos.*

*La particularidad de este programa era que no se entrenaba aprendiendo de partidas humanas, sino que lo hacía compitiendo contra sí mismo una y otra vez. Empezó jugando de forma totalmente aleatoria hasta conseguir desarrollar habilidades nunca vistas en estos juegos.*

## ***¿Puede engañarnos un sistema inteligente?***

*Si usamos aprendizaje por refuerzo, y el engaño lleva a un sistema inteligente a conseguir el objetivo perseguido, parece factible que la inteligencia artificial pueda aprender a mentir.*

*Y así ocurrió en 2017, cuando dos investigadores de la Universidad Carnegie Mellon consiguieron que su sistema de inteligencia artificial venciese a los mejores jugadores de no-limit Texas Hold 'Em, una de las variantes más complejas del póker. Para conseguir tal hazaña, [Libratus](#) –así se llamaba– tuvo que aprender a ir de farol.*

### *DISMORFIA DE ZOOM, O CÓMO AFECTA A NUESTRA AUTOESTIMA VERNOS EN LA PANTALLA*

***En el juego del póker, la habilidad de mentir es clave.***

El jugador necesita que sus contrincantes creen que tiene mejores cartas de las que tiene, o todo lo contrario. Libratus fue capaz de perfeccionar el arte del engaño hasta el punto de que uno de los mejores jugadores del mundo, Dong Kim, llegó a pensar que [la máquina podía ver sus cartas](#).

Y consiguió este hito tecnológico sin que nadie le dijera que había que mentir. Sus creadores le proporcionaron la descripción del juego, pero no le dijeron cómo había que jugar. ***Durante meses, Libratus jugó millones de partidas hasta que alcanzó un nivel de juego que desafiaba a los mejores jugadores humanos.***

Ahora que sabemos que la inteligencia artificial es capaz de mentir, y seguramente siga haciéndolo, tenemos que prepararnos para el futuro. ***Cada vez serán más importantes las cuestiones éticas que rodean a este tipo de sistemas inteligentes que, nos guste o no, llegaron para quedarse.***

[Este texto se reproduce de The Conversation bajo licencia Creative Commons.](#)

11 de mayo de 2022

**Verónica Bolón Canedo**

[lanacion.com.ar/tecnologia/puede-mentir-la-inteligencia-artificial-nid11052022/](https://lanacion.com.ar/tecnologia/puede-mentir-la-inteligencia-artificial-nid11052022/)