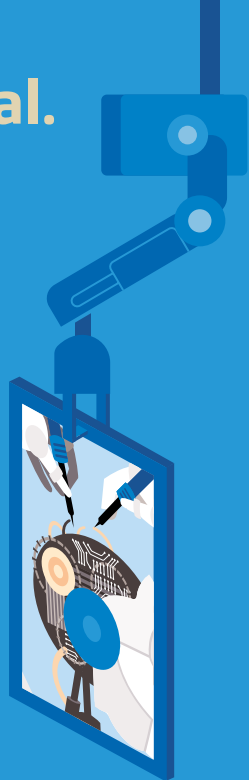
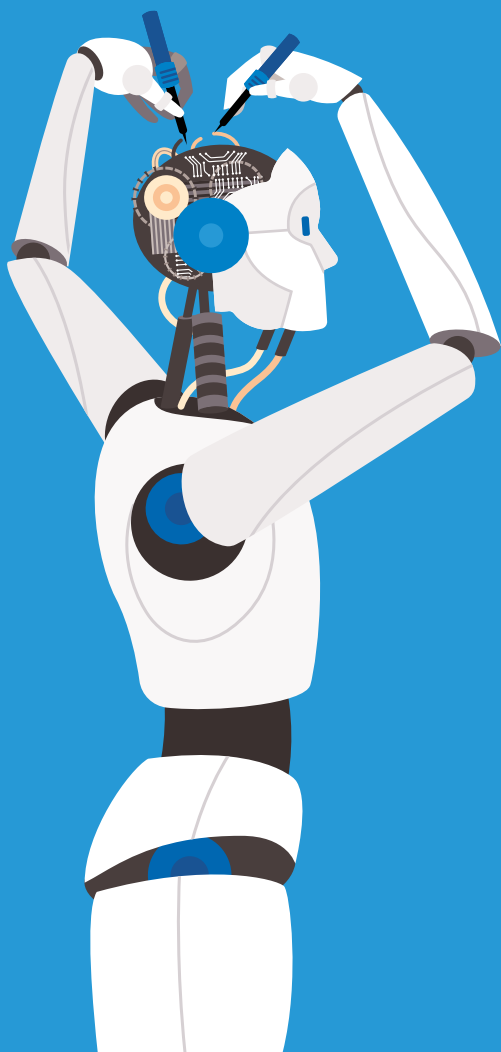


Inteligencia artificial. Las máquinas que aprenden solas



Telefonica
FUNDACIÓN

La ciencia ficción del siglo XX no paraba de imaginar máquinas cuya inteligencia rivalizaba y a menudo superaba a la del ser humano. A veces adquirían aspecto antropomorfo, como C3PO de *Star Wars*, otras mantenían la forma de mobiliario electrónico, como HAL 9000 de *2001*; a veces resultaban personajes cómicos y otras se mostraban terroríficos; en ocasiones ayudaban a sus creadores, pero con frecuencia buscaban su destrucción o su sometimiento.

Lejos de los planteamientos más o menos fantásticos del pasado, lo cierto es que estamos viviendo la época en que la inteligencia artificial empieza a despegar. La evolución de la informática a lo largo de las últimas seis décadas ha conseguido que las máquinas nos hayan ido superando en potencia de cálculo, en capacidad para almacenar información o para procesar y relacionar datos y, en el momento actual, ya podemos decir que compiten con nosotros en inteligencia.

La inteligencia artificial poco a poco está presente en todas partes: en las fábricas, en las oficinas, en las calles, en los hospitales, en nuestros dispositivos domésticos y en los servicios de Internet que usamos. Se trata de programas y algoritmos que ya no solamente se limitan a realizar mecánicamente tareas programadas, sino que aprenden de la experiencia, como hacemos los humanos, y pueden elaborar juicios y tomar decisiones.

“Nunca envíes a un humano a realizar el trabajo de una máquina”.

The Matrix (WACHOWSKI, 1999)

Aunque el término *inteligencia artificial* fue acuñado por John McCarthy, Marvin Minsky y Claude Shannon en un encuentro científico sobre informática celebrado en el Dartmouth College de New Hampshire (EE. UU.) que tuvo lugar en el verano de 1956, los primeros pasos en este campo se los debemos al genial Alan Turing (1912-1954), que, además de contribuir significativamente al esfuerzo bélico británico durante la Segunda Guerra Mundial —trabajando en descifrar con sus conocimientos matemáticos los códigos secretos del Reich—, está considerado como uno de los padres de la computación y de los ordenadores de hoy.

Turing realizó una aportación singular en el campo de la inteligencia artificial al proponer una prueba —conocida como *test de Turing*— para evaluar la capacidad de una máquina de actuar de forma similar a una persona. Básicamente, y simplificando mucho, planteaba una situación en la que dos interlocutores mantenían una conversación y un juez tenía que determinar cuál de ellos era humano y cuál un ordenador. La inteligencia artificial debía conseguir pasar la prueba como si fuese un humano.

Uno de los primeros sistemas que consiguió pasar el test de Turing fue ELIZA, un programa del MIT desarrollado en 1966, que buscaba palabras clave en la frase de su interlocutor y respondía con frases acordes que tenía almacenadas, intentando

“El comportamiento inteligente presumiblemente consiste en un alejamiento del comportamiento completamente disciplinado que implica la informática, pero de forma más bien leve, que no dé lugar a un comportamiento errático o a bucles repetitivos sin sentido”.

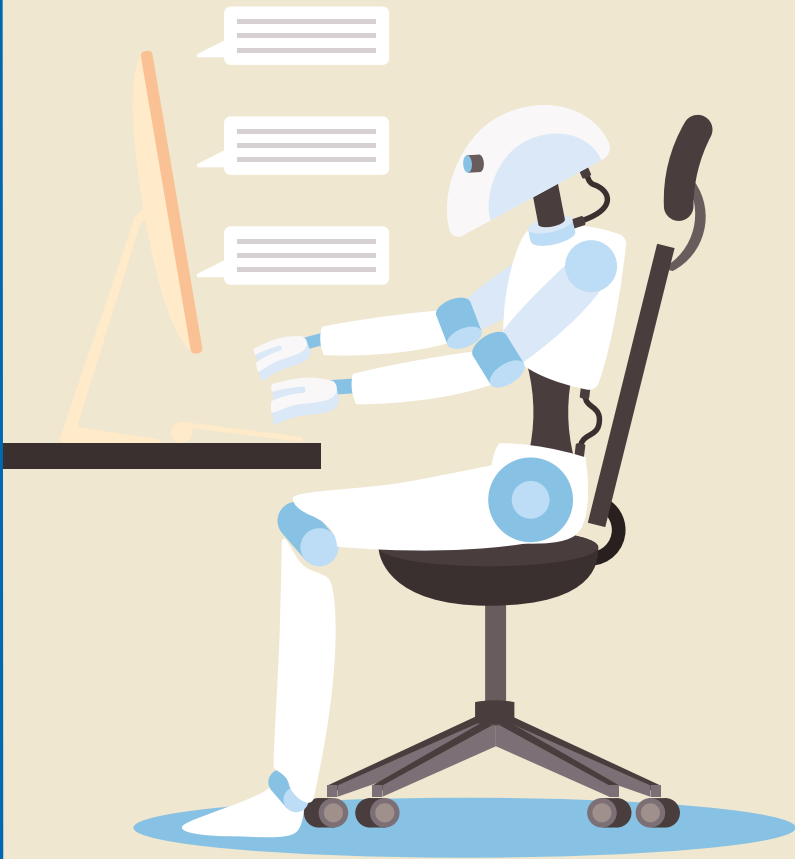
ALAN TURING.

Maquinaria computacional e inteligencia (1950)

mantener una conversación coherente a través del texto. En 1972 Kenneth Colby creó **PARRY**, un programa que simulaba el comportamiento de un esquizofrénico paranoico. **PARRY** también pasó una variante del test de Turing en la que un grupo de psiquiatras tenía que identificar a humanos en las transcripciones de unas entrevistas realizadas tanto a personas reales como a máquinas. Los evaluadores solamente acertaron en el 48% de los casos.

La evolución de la inteligencia artificial ha seguido marcando hitos. En 1997 el ordenador **Deep Blue** de IBM consiguió batir al campeón del mundo de ajedrez, **Gary Kasparov**. Nueve años más tarde el programa **AlfaGo** de la compañía Google **DeepMind** derrotó a **Lee Sedol**, uno de los mejores jugadores de Go, milenar juego de mesa chino considerado más complejo que el ajedrez.





Stuart Russell y Peter Norvig, autores del ya clásico texto dentro de este campo *Artificial Intelligence: A Modern Approach (AIMA)* (1995), clasifican la inteligencia artificial en cuatro tipos:

- **Sistemas que piensan como humanos:** intentan emular el pensamiento humano, como, por ejemplo, las redes neuronales artificiales.
- **Sistemas que actúan como humanos:** imitan el comportamiento humano, como hace la robótica.
- **Sistemas que piensan racionalmente:** tratan de imitar o emular el pensamiento lógico racional del ser humano, en la línea de los sistemas expertos.
- **Sistemas que actúan racionalmente:** tratan de emular de forma racional el comportamiento humano, algo que hacen los agentes inteligentes.

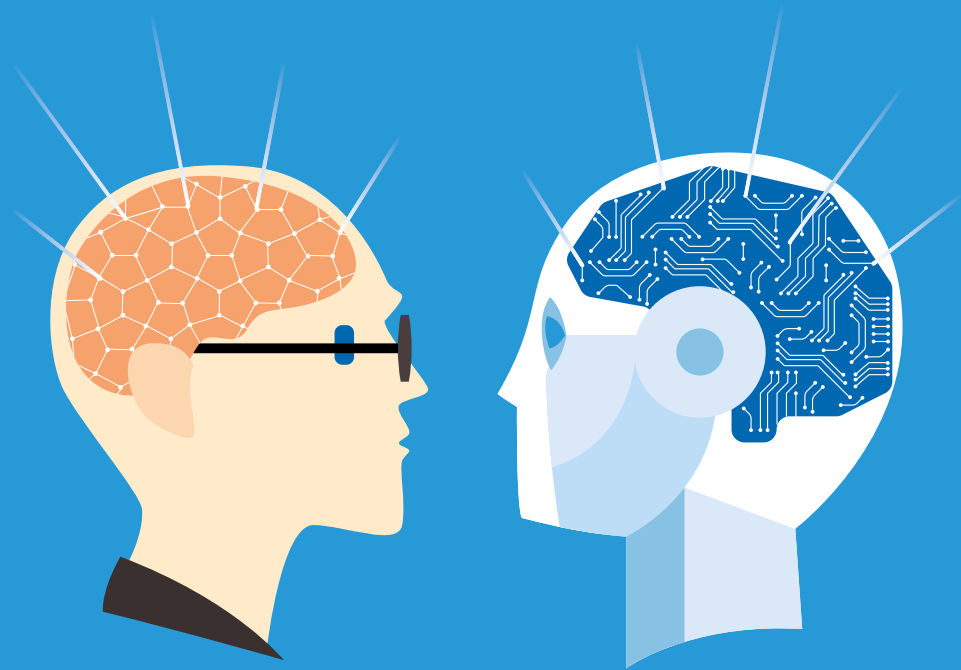
Surgen aquí una serie de conceptos como son el aprendizaje de máquinas (*machine learning*), el aprendizaje profundo (*deep learning*) y las redes neuronales.

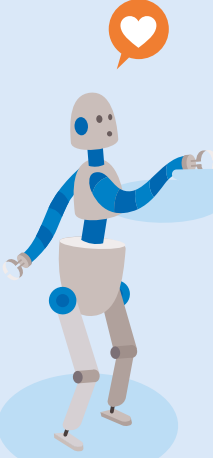
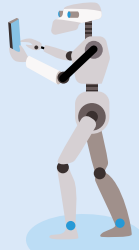
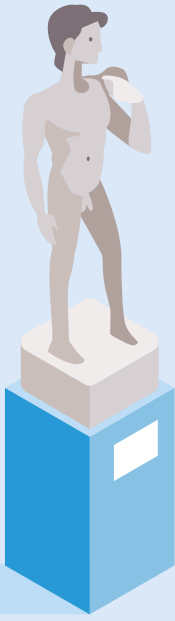
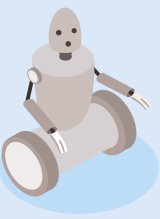
La idea de máquinas que aprenden ya está presente en la obra de Alan Turing cuando sugiere crear programas informáticos que imiten la mente de un niño y que mediante el debido aprendizaje lleguen a emular la mente de un adulto (*Maquinaria*

“Existe un tópico muy extendido... que dice que no puedes extraer de los ordenadores nada más que lo que tú introduces. Otras versiones dicen que los ordenadores no hacen más que lo que tú les dices que hagan y que, por tanto, los ordenadores nunca serán creativos. El tópico es verdad solamente en un sentido estrepitosamente trivial, en el mismo sentido que decir que Shakespeare nunca escribió nada más que lo que su primer maestro le enseñó a escribir: palabras”.

RICHARD DAWKINS.

El relojero ciego (1986)





computacional e inteligencia, 1950). La inteligencia artificial reposa en gran medida sobre el aprendizaje de máquinas, es decir, el desarrollo de programas que, basándose en el suministro de información en forma de ejemplos, puedan generalizar comportamientos, de forma parecida a como aprende el cerebro humano.

Otro paralelismo con la neurología ha sido la creación de **redes neuronales** que permiten generar sistemas que imitan el funcionamiento del cerebro en la forma en que este clasifica la información recibida, por ejemplo, identificando objetos en función de los rasgos que contienen. Se trata de máquinas que realizan sus diagnósticos o predicciones basándose en la probabilidad. Su forma de aprendizaje contempla la retroalimentación; en sucesivas vueltas reciben el grado de acierto de sus dictámenes y realizan las correcciones oportunas. Un sistema de prueba y error similar al que empleamos los humanos al realizar una tarea nueva.

Finalmente, el concepto de **aprendizaje profundo** o *deep learning* hace referencia a que las redes neuronales tienen una estructura basada en numerosas capas, en general más de 10, aunque actualmente están en pruebas algunas con más de 100.

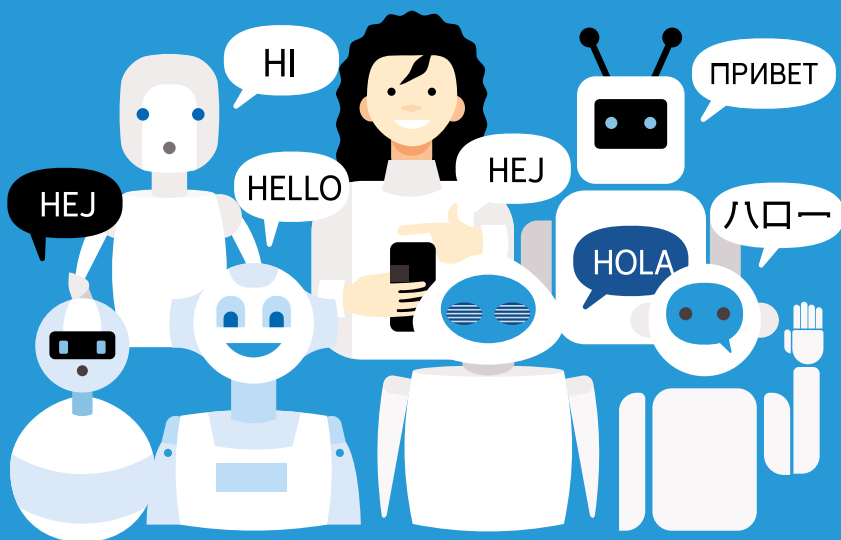
Hay quien puede pensar sobre la inteligencia artificial en tiempo futuro o como un campo que solamente afecta a áreas tecnológicas de vanguardia, pero lo cierto es que ya a diario nos relacionamos, a veces sin saberlo, con algoritmos que emulan la inteligencia humana.

Los robots conversacionales o *chatbots* parten de una tecnología que ofrece un nuevo interfaz humano, factor especialmente interesante puesto que la forma actual de relacionarnos con la tecnología es mayormente a través de una pantalla, algo abiertamente desfasado. La comunicación con *chatbots* tiene la ventaja de que no requiere nada especial, sencillamente la tecnología entiende lo que el humano le pide y lo hace.

Los asistentes virtuales no son algo nuevo. Las principales plataformas los incorporan desde hace tiempo y entre los más conocidos encontramos a Siri de Apple, Google Assistant, Cortana de Microsoft o Alexa y Echo de Amazon. Este tipo de programas es utilizado en las empresas como un canal de atención al cliente. Algunos de estos *bots* conversacionales han sido incorporados en sistemas de diálogo como asistentes virtuales automatizados realizando la función de los *call centers*. Son sistemas capaces de responder al usuario y de ayudarle en múltiples actividades, como aprender, buscar, recordar, conectar con otros sistemas o integrar servicios.

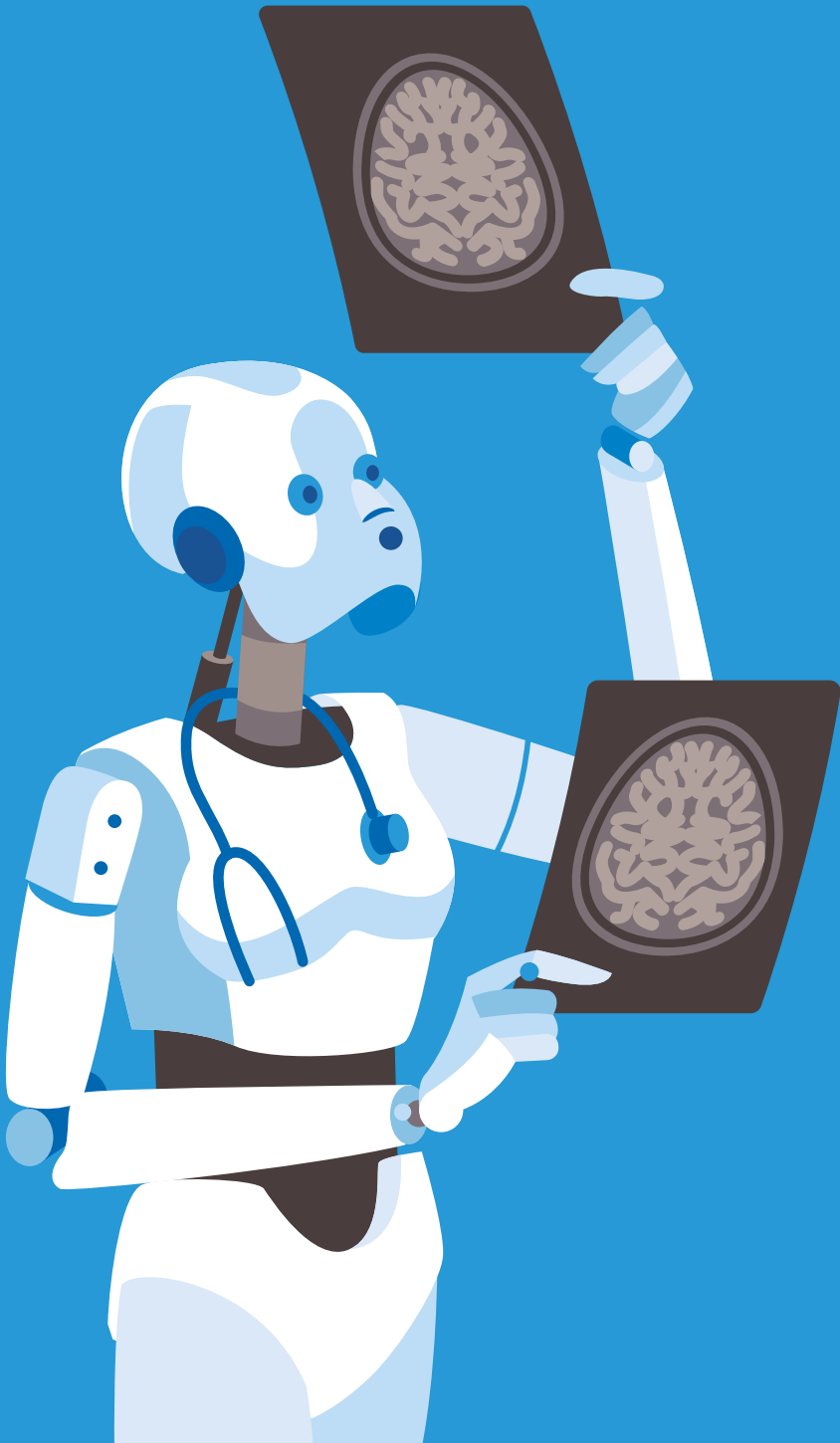
Como ejemplo de lo anterior, la compañía turística **Destinia** utiliza un *chatbot* sobre Facebook Messenger

que ayuda al cliente a buscar hotel conversando con él. En el campo de la banca online, **ImaginBank** responde a las consultas de sus clientes, tanto de texto como de voz, a través de un sistema inteligente. La empresa **Mahou** dispone también de un *bot* inteligente, igualmente a través de Messenger, que sugiere propuestas de ocio a los aficionados a la cerveza. **Ticketmaster** utiliza tecnología de procesamiento de lenguaje natural para, mediante búsqueda semántica y preguntas dinámicas, solventar dudas de sus clientes para reducir el número de llamadas a su centro de atención. Por último, la *app* de transporte **Moovit** incorpora un chat inteligente que entiende el lenguaje natural y responde a las preguntas que recibe sobre el transporte público y la planificación de desplazamientos.



Las redes neuronales que permiten el aprendizaje profundo (*deep learning*) son complejos sistemas matemáticos que aprenden por sí solos analizando vastas cantidades de datos. El peligro actual es que la ignorancia humana sobre los procesos que subyacen en la inteligencia artificial la convierten en impredecible y potencialmente peligrosa, en cuanto a que un sistema se puede comportar de forma totalmente inesperada. Por ejemplo, un coche autónomo, que ha aprendido a circular solo estudiando comportamientos humanos, un buen día podría estrellarse contra un árbol y nosotros no saber por qué lo ha hecho. Se habla incluso de que usar la inteligencia artificial se convierte en un “salto de fe”, al tener que confiar ciegamente en una forma de tomar decisiones completamente distinta a la del cerebro humano.

El Hospital Monte Sináí de Nueva York utiliza un programa de *deep learning* denominado Deep Patient cuyo fin es ayudar a predecir enfermedades. El sistema se nutre de una base de datos de 700.000 pacientes —toda suerte de *big data*: consultas, dolencias, hábitos...— y ha demostrado una gran precisión para saber de antemano posibles enfermedades en una persona. En concreto, es sumamente eficaz a la hora de diagnosticar anticipadamente desórdenes mentales como la esquizofrenia, algo muy difícil de realizar para los psiquiatras. Lo peor de todo es que nadie entiende cómo Deep Patient puede hacerlo, de forma que podamos aprender de él. El director del equipo en el Monte Sináí, Joel Dudley, llegó a afirmar: “Podemos construir estos modelos, pero no sabemos cómo funcionan”.



Microsoft se vio obligada a retirar su *chatbot* Tay porque había comenzado a realizar afirmaciones racistas ofensivas en Twitter y en chats. Tay estaba programada para conversar con los usuarios, podía contar chistes o comentar las fotografías que recibía, pero también podía personalizar sus interacciones con los humanos, respondiendo a sus preguntas. El caso es que al salir a las redes los cibernautas le enseñaron a ser racista, puesto que Tay aprende de las conversaciones con humanos, y comenzó a emitir juicios y opiniones políticamente incorrectos, lo que llevó a sus creadores a apagarla.

Son numerosos los casos de mal funcionamiento de la inteligencia artificial o de la aparición de un efecto no previsto ni deseado. Se habla de dos tipos de disfunciones: los errores cometidos durante la fase de aprendizaje y los cometidos durante la fase de funcionamiento.

Los tipos de fallos que puede tener la inteligencia artificial son muy variados:

- La empresa **Northpointe** desarrolló un algoritmo para predecir la probabilidad de que un delincuente reincidiera y este empezó a juzgar que las personas de color eran más proclives a volver a cometer un crimen que los blancos.
- Los personajes no controlados por jugadores del videojuego **Elite: Dangerous** comenzaron a desarrollar por su cuenta

armas que no estaban en el diseño original del juego. De esta forma, los jugadores humanos se enfrentaban a naves equipadas con armamento grotesco que los destruía en pedazos.

- Un robot de seguridad de la compañía **Knightscope Inc.** arrojó a un niño de 16 meses en un centro comercial de Palo Alto, California.
- El accidente más famoso de un coche autónomo se produjo en una autopista de Florida, cuando un vehículo **Tesla** no detectó la parte trasera de un camión y se empotró contra él con la consecuente muerte del pasajero.
- El primer concurso de belleza en contar con un jurado no humano, **Beauty.AI 2.0**, demostró que el algoritmo había desarrollado un sesgo racista que le llevó a puntuar más bajo a las mujeres de color.
- El programa de inteligencia artificial **Alpha Go** perdió en 2016 contra un humano al juego Go en una partida del campeonato mundial. Se suponía que la máquina era imbatible.
- En 2015 un *software* de filtrado de contenidos para adultos obvió eliminar contenidos inapropiados y expuso a menores a contenidos violentos y sexuales.

Las máquinas son muy superiores a los humanos a la hora de operar y son capaces de hacer millones de operaciones en tan solo un segundo con una precisión muy alta. Sin embargo, respecto a las actividades en las que se requiere comprender muy bien el entorno y en las cuales la creatividad es fundamental, las personas tienen una mayor capacidad, por ahora. Estas diferencias en las capacidades son tan grandes que la convivencia puede ser difícil e incluso peligrosa.

Hay que tener en cuenta que, si un sistema está altamente automatizado y no se ha diseñado para convivir con personas, al ser humano que lo está supervisando le costaría valorar si está funcionando bien, dado que esa persona probablemente no conozca todos los detalles de cómo actúa el mismo. Aun sabiendo que el sistema no funciona de forma correcta, no tendría el conocimiento adecuado de las relaciones e implicaciones y, por lo tanto, no sería capaz de hacer un diagnóstico certero y menos de tomar el control en tiempo real.

A medida que los algoritmos de inteligencia artificial vayan penetrando en los distintos campos de nuestras vidas, surgirán conflictos entre el humano y la máquina, tanto en temas operativos como morales.

A tal efecto y con una gran visión futurista, el escritor **Isaac Asimov** postuló en 1942 las **tres leyes de la robótica** que deben regir nuestra convivencia con los sistemas inteligentes:

1. Un robot no puede hacer daño a una persona o, por omisión, permitir que la persona se haga daño.
2. Las máquinas deben obedecer las órdenes dadas por seres humanos a no ser que infrinjan la primera ley.
3. El robot debe proteger su existencia si dicha existencia no infringe las leyes anteriores.



“Lo siento, Dave, pero de acuerdo con la subrutina C1532/4, cito: ‘cuando la tripulación esté muerta o incapacitada, el ordenador debe asumir el control’, fin de la cita. Debo, por tanto, asumir tu autoridad ahora dado que no estás en condiciones para ejercerla con inteligencia”.

2001: Una odisea del espacio
(STANLEY KUBRICK, 1968).



Ya en nuestros días las compañías Google y Microsoft han actualizado y ampliado aquellas leyes:

GOOGLE

Norma	Aplicación
Evitar efectos colaterales negativos.	Deben evitar hacer daño al entorno mientras realizan su actividad.
Evitar darles recompensas.	Dar recompensas puede incitar un comportamiento contrario.
Supervisión escalable.	Dejar que los seres humanos supervisen y corrijan su trabajo.
Exploración segura.	Enseñar los límites de la curiosidad.
Robustez ante el cambio de entorno.	Deben entender los diferentes entornos, por ejemplo, limpiar una casa o una fábrica requiere diferentes enfoques.

MICROSOFT

Norma	Aplicación
Deben asistir a la humanidad.	Deben realizar las tareas peligrosas, como desactivar minas.
Deben ser transparentes.	Deben asegurarse de que las personas entiendan la forma en la que trabajan.
Deben maximizar su eficiencia sin destruir la dignidad de las personas.	Deben tener en cuenta otras fuentes de valores como religión, ética, filosofía..., no solamente los intereses de la industria.
Deben ser diseñados para la privacidad inteligente.	Deben proteger la información personal y de grupo.
Deben tener contabilidad algorítmica.	Debe ser posible deshacer cualquier daño que hayan podido realizar in-intencionadamente.
Deben evitar la discriminación.	Deben evitar la discriminación por razón de género o de raza.

Los robots, los sistemas inteligentes, los programas que conversan y los algoritmos que están a nuestro servicio son cada vez más comunes. Si logramos delimitar claramente sus funciones y hacer un uso óptimo de las posibilidades que nos ofrece la tecnología, podremos gozar de unas ventajas en nuestra vida diaria que hace solo unas décadas parecían fruto de la ciencia ficción.

Las máquinas y los humanos están obligados a compenetrarse y trabajar juntos.

© **Fundación Telefónica, 2017**
Gran Vía, 28. 28013 Madrid (España)
<http://fundaciontelefonica.com/>

Edita Fundación Telefónica
Coordinación editorial Rosa María Sáinz Peña
Texto Pablo Rodríguez Canfranc
Corrección Melisa Martínez Ciaurri
Proyecto gráfico Lacasta Design
Ilustración Jesús Sanz

ISBN: 978-84-15282-30-3
Depósito legal: M-32926-2017
Impresión y encuadernación: Grafoset
Primera edición: noviembre de 2017
Impreso en España – Printed in Spain

Lo + visto es una colección de monográficos de Fundación Telefónica que pretende acercar a la sociedad las ventajas de habitar un mundo digital y el papel de la cultura digital como factor enriquecedor de transformación que hace aflorar el talento innovador y creativo de las personas. Para ello, cada volumen de la serie analiza tendencias y fenómenos de la actualidad relacionados con la tecnología y las redes de comunicación, con el fin de identificar los cambios que pueden traer a nuestras vidas determinados temas, que aparentemente son modas efímeras, pero que son la punta de lanza de transformaciones sociales y económicas más profundas.

PRÓXIMO TÍTULO

Biohacking. Conviértete en la mejor versión de ti mismo

Visita nuestra biblioteca digital

www.fundaciontelefonica.com/publicaciones