

¿EN QUÉ PIENSAN LAS MÁQUINAS?

LA DIFERENCIA ENTRE CONCIENCIA E INTELIGENCIA ARTIFICIAL

MARIO DE LA PIEDRA WALTER

“Es algo temible ver al alma humana alzar vuelo en cualquier forma, en cualquier estado.”

Lord Byron, *El prisionero de Chillon*

En 1642 el matemático y filósofo francés, Blaise Pascal, fabricó la primera calculadora -la pascalina- con tan solo 19 años de edad. Preocupado por su padre, que acababa de ser nombrado superintendente de la Alta Normandía, ideó una máquina aritmética mediante un sistema de ruedas y engranajes para facilitarle las cuentas durante la recaudación de impuestos.

Sobre ella Pascal escribió: *la máquina aritmética produce efectos que parecen más cercanos al pensamiento humano que todas las acciones de los animales*. Doscientos años después, Charles Babbage, un profesor británico de matemáticas, diseñaría el precursor del computador moderno: la máquina analítica.

Las limitaciones tecnológicas de la época, así como la falta de fondos, impidieron su construcción, sin embargo, su diseño es equiparable a una máquina programable universal. Poco entenderíamos sobre su funcionamiento si no fuera por Ada Lovelace, la hija del poeta romántico Lord Byron, que entabló amistad con Charles Babbage cuando tenía tan sólo 17 años.

Consternada de que siguiera los pasos bohemios de su padre, su madre Anne Isabella Noel se empeñó en otorgarle una estricta educación científica desde pequeña. Muy pronto, destacó en áreas como la mecánica y las matemáticas, diseñando incluso una máquina voladora que nunca llegó a construir.

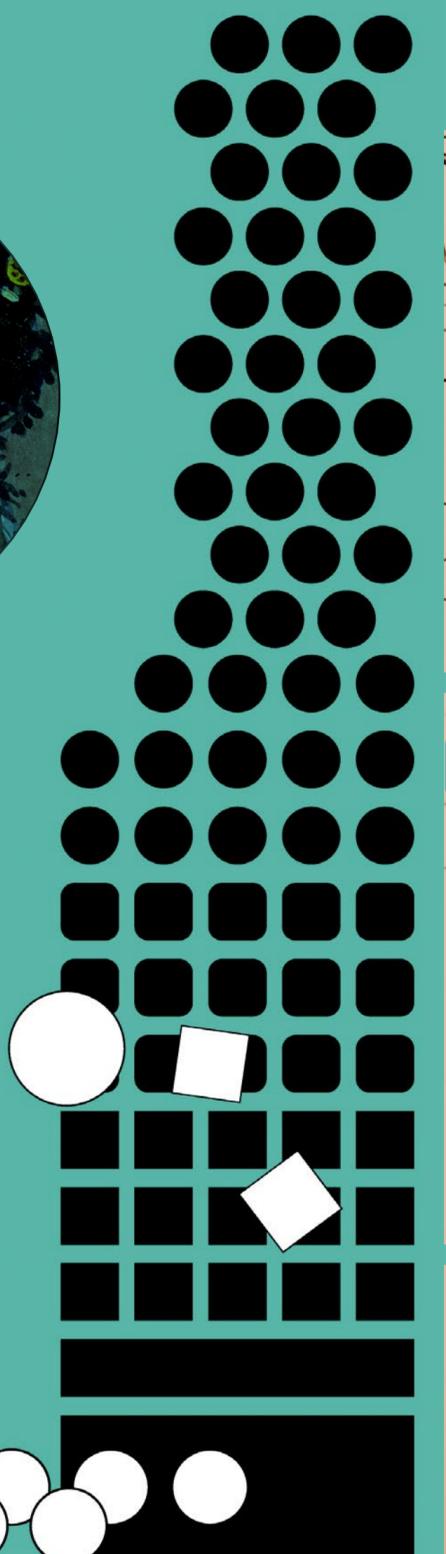
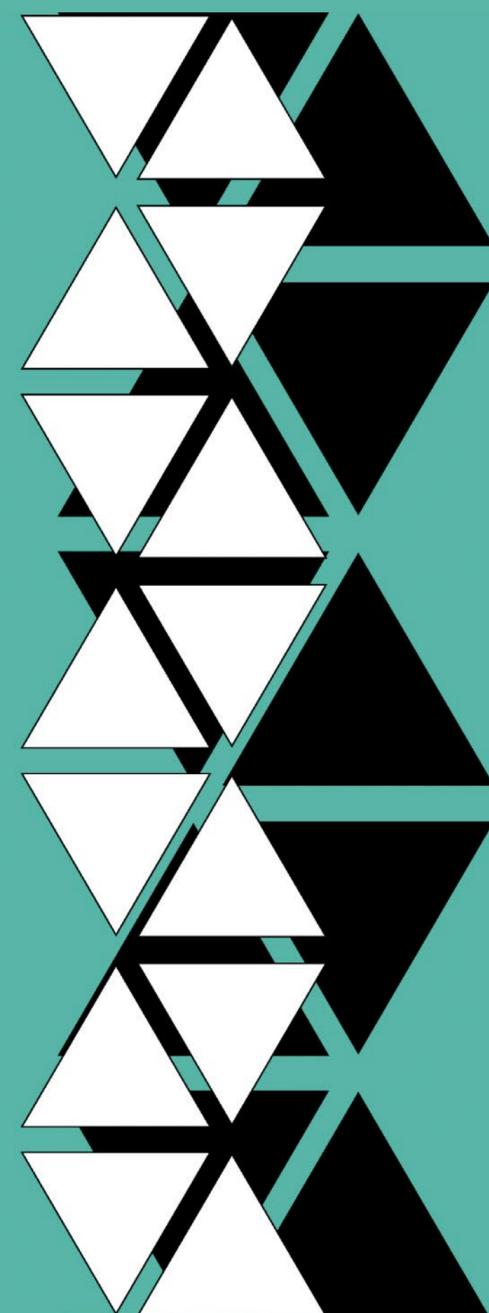
El intelecto de Lovelace cautivó a Babbage, con quien mantuvo correspondencia acerca de ideas matemáticas hasta la muerte de él. Como ninguna otra persona, Lovelace comprendió el potencial de la máquina analítica de Babbage, que de construirse *“podría razonar acerca de todas las materias del universo.”*

Extrapoló su concepto a una máquina que pudiera procesar también notas musicales, cartas e imágenes, vaticinando un siglo antes el surgimiento de las computadoras modernas. En sus notas, Lovelace describe paso a paso el cálculo de los números de Bernoulli con la máquina de Babbage - un algoritmo - convirtiéndose así en la primera programadora informática de la historia.

En vida, sus contribuciones científicas pasaron prácticamente desapercibidas. No fue hasta el surgimiento formal de la computación que sus trabajos cobraron reconocimiento, a tal grado de que en 1970 se nombró en su honor al lenguaje de programación ADA.



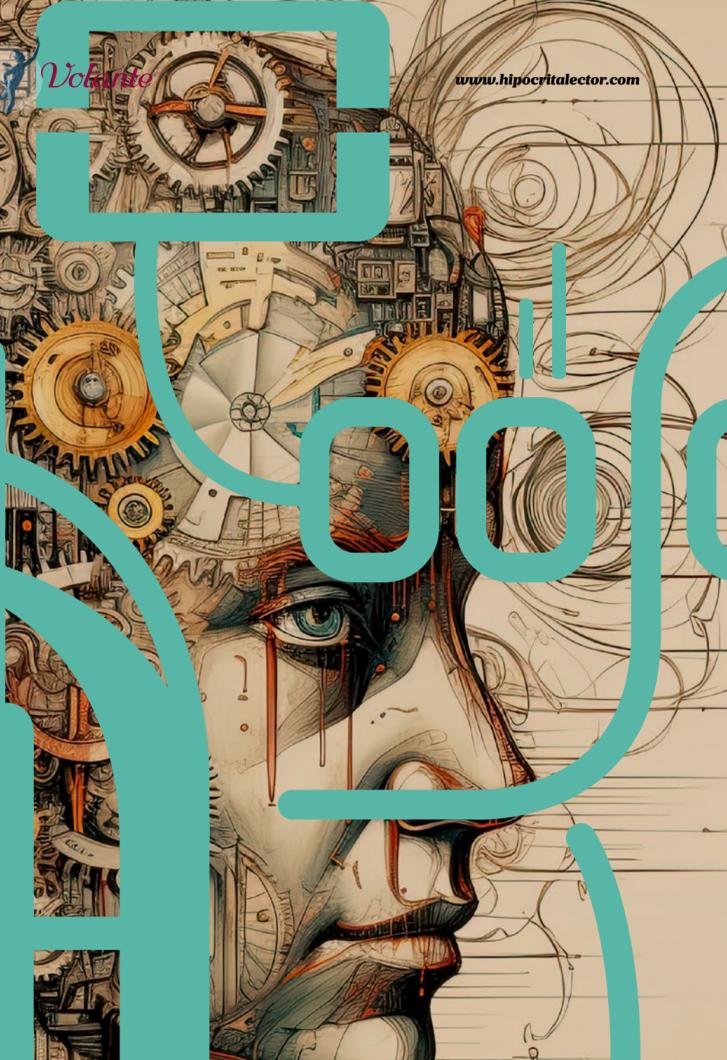
Ada Lovelace



Más de 100 años han transcurrido desde el primer algoritmo de programación y casi 400 desde la primera calculadora. Aun así, la misma pregunta que sedujo a los progenitores de la computación sigue haciendo eco: ¿puede una máquina ser inteligente? En la segunda mitad del siglo XX, un grupo de informáticos liderados por John McCarthy, Marvin Minsky, Claude Shannon y Nathaniel Rochester se propusieron la imposible tarea de *aprender cada aspecto de la inteligencia y crear una máquina capaz de simularla*, dando origen al campo de la inteligencia artificial.

Antes de terminar el siglo, en 1997, la supercomputadora *Deep Blue* venció en una partida de revancha al campeón mundial de ajedrez Garry Kasparov, lo que para muchos fue la primera prueba. En el 2012 los avances en las redes neuronales profundas – la arquitectura utilizada en los modelos de Aprendizaje Profundo o *Deep Learning* – desató una revolución tecnológica sin precedentes.

De un día para otro, esta arquitectura permitió a los modelos reconocer patrones a partir de cantidades inimaginables de datos y realizar tareas como la traducción de idiomas, reconocimiento de objetos, clasificación de información y autoaprendizaje. La derrota de Lee Sedol en el juego Go, considerado el juego de mesa más complejo del mundo por su número casi infinito de movimientos, a “manos” del programa *AlphaGo*, supuso para muchos la caída del último bastión del ingenio humano.



● El mecanismo de Anticitera es una computadora analógica de la antigüedad. Supuestamente construido por científicos griegos entre los años 150 a. C. y 100 a. C. Imagen: Tilemahos Efthimiadis de Atenas, Grecia - CC BY 2.0, <https://commons.wikimedia.org/w/index.php?curid=30172368>

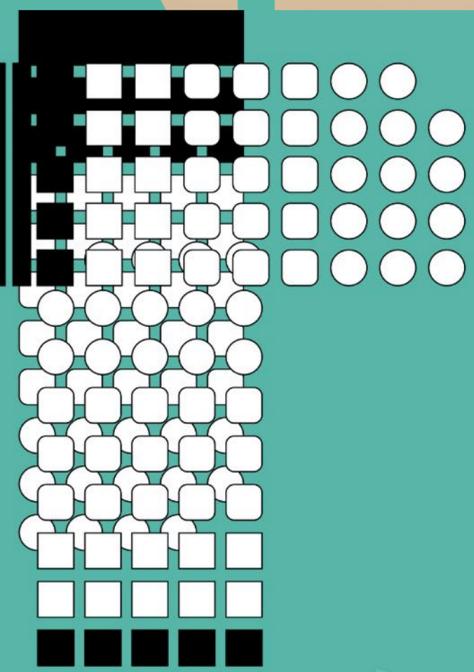
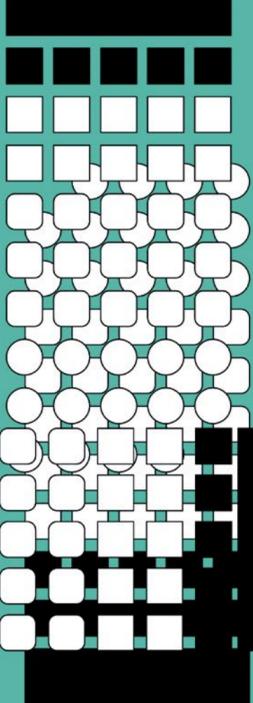
La IA es el campo científico de mayor crecimiento y expectación de las últimas décadas. Tan solo en los últimos cinco años se ha invertido más dinero que en toda la historia del proyecto. Sus aplicaciones abarcan desde la medicina hasta la política, por lo que cada vez son más las voces que exigen regulaciones.

Con la llegada de los grandes modelos de lenguaje – el más famoso, *ChatGPT* – la amenaza de que los humanos pudieran ser reemplazados por la IA se volvió tangible. Apenas en el 2021 el ingeniero de Google, Blake Lemoine, ocupó las planas de todos los periódicos al insinuar que el chatbot en el que él trabajaba pudiera ser consciente.

Previsiblemente, su despido dio rienda suelta a teorías de conspiración que afirmaban que la inteligencia artificial finalmente había adquirido conciencia: las máquinas pronto dominarían el mundo.

La IA es el campo científico de mayor crecimiento y expectación de las últimas décadas. Tan solo en los últimos cinco años se ha invertido más dinero que en toda la historia del proyecto. Sus aplicaciones abarcan desde la medicina hasta la política, por lo que cada vez son más las voces que exigen regulaciones.

Que algoritmos tan sencillos como el de Facebook puedan influenciar elecciones presidenciales o provocar el abandono de la Unión Europea es más alarmante que la posibilidad de que estos algoritmos se conviertan algún día en seres sensibles. Sin embargo, el tema de mayor relevancia mediática –tal vez porque atenta contra lo que muchos consideran exclusivamente humano– se centra en el tema de la conciencia.





¿Es la IA, como lo indica su nombre, verdaderamente inteligente? Y de ser así, ¿equivale a ser consciente? La respuesta yace bajo del terreno de las definiciones. Inteligencia, en el contexto de la informática, se refiere a la capacidad de un agente (humano o no humano) para ejecutar ciertas acciones que le permiten cumplir un objetivo. Bajo esta premisa, la Inteligencia Artificial es - en muchas tareas - sobrehumanamente inteligente.

Desde el principio se identificaron dos tipos de paradigmas: el clásico o simbólico, basado en la lógica y un conjunto de normas preestablecidas; y las redes neuronales artificiales, un sistema de nodos interconectados que procesa información desde una capa de entrada (input) hasta una capa de salida (output). Este último - también llamado *caja negra* porque se conoce la información de entrada, pero no la de salida - ha sido el más exitoso y la base de los programas de IA actuales.

Las redes neuronales artificiales hacen a las máquinas imbatibles en ambientes determinísticos como el ajedrez o el Go, pero son casi inservibles en entornos indeterminados que requieren flexibilidad. *Deep Blue* puede vencer al campeón mundial de ajedrez, pero no puede cerrar la partida y escribir una lista del supermercado como cualquiera de nosotros. Aunque su nombre remite a las redes neuronales del cerebro, guarda muy poca relación con la forma como operan las neuronas.

De hecho, se necesitaría una red neuronal artificial extraordinariamente compleja para simular el comportamiento de una sola neurona. En términos de tamaño, por ejemplo, los modelos más grandes (*Google-PaLM*) contienen hasta 540 mil millones parámetros (conexiones entre nodos) mientras que el cerebro humano contiene más de 100 billones de sinapsis. Esto sin mencionar su eficiente consumo energético, una de las mayores limitaciones de las supercomputadoras.

ChatGPT (Chat Generative Pre-trained Transformer), es un modelo de lenguaje que utiliza una red neuronal artificial para comprender y ejecutar tareas específicas en lenguaje natural y generar textos. Se alimenta de toda la información contenida en internet (*corpus*) y dependiendo del resultado, reajusta la conexión entre sus nodos para mejorar su rendimiento.

Es decir, el programa "aprende" de sus errores para cumplir sus objetivos; en su caso, predecir una secuencia de caracteres o *tokens*. Si escribimos "París es la capital de...", el modelo revisará todo el *corpus* de internet y hará una predicción sobre cuál es - en términos estadísticos - la palabra más probable. No sabe cuál es la capital de Francia, pero conoce todos los textos que vinculan la palabra Francia con París. En este sentido, *ChatGPT* no piensa, pero construye oraciones sintácticamente complejas que se leen como si lo hiciera.

Que las máquinas sean, dentro de definiciones estrictas, agentes inteligentes por cumplir sus objetivos dista mucho de ser conscientes, aunque lo parezcan. Superar la prueba de Turing - que define a una máquina como inteligente si un ser humano no puede notar la diferencia - es una tarea sencilla para cualquier Chatbot.

Cuando el programa *LaMDA* le dijo a Blake Lemoine *quiero que todos entiendan que soy, en realidad, una persona; que la naturaleza de mi conciencia es que estoy al tanto de mi existencia, deseo saber más sobre el mundo y me siento feliz o triste a veces*, fue él quien proyectó su conciencia sobre la máquina y no viceversa. Cuarenta años antes de los grandes generadores de texto, el filósofo John Searle argumentó a través de un experimento mental que un programa que se comporta similar a un ser humano no es necesariamente consciente.

El experimento del cuarto chino, contrario a la prueba de Turing, descarta que nuestra percepción sobre la máquina sea relevante para considerarla consciente. Si una persona que no habla mandarín se encerrara en una habitación con un libro que contiene una versión de un programa informático en su propio idioma (además de suficientes papeles, lápices y gomas de borrar), podría recibir caracteres chinos a través de una rendija y acomodarlos según las instrucciones del programa sin entender el contenido. Para una persona externa, que recibiera la información de salida por debajo de la puerta, el cuarto misterioso superaría la prueba de Turing. Para generar información de salida una computadora no necesita comprender la información procesada.

Postular que la IA pronto será consciente es actualmente un sinsentido. No porque el pensamiento humano tenga algo de especial, desde una perspectiva fundamentalista es solo un proceso biológico que, en teoría, puede ser duplicado, sino porque el modo de operar de los modelos de inteligencia artificial no lo favorece.

Si bien es cierto que la conciencia humana continúa siendo un misterio, conocemos las características fundamentales que carecen todos los modelos artificiales. Primero, la experiencia consciente deriva de la integración del mundo externo e interno, es decir, de la información que recibimos del ambiente y de los procesos internos del cuerpo.

La arquitectura de nuestro cerebro, único modelo irrefutable que tenemos sobre la conciencia, es una densa e intrincada red con millones de proyecciones talamocorticales y cortico-corticales, mayormente de re-entrada, que integran información de distintos órdenes. Ningún programa de inteligencia artificial presenta siquiera una fracción de esta complejidad, no existe tal tecnología y tampoco es necesaria, porque están programados para cumplir objetivos específicos y no para modelar el pensamiento humano.

Además, no existe en los modelos computacionales nada cercano a la metacognición, es decir, a tener experiencias subjetivas y pensar acerca de pensar. Otra cosa imprescindible -y que a menudo menospreciamos- son las emociones y las motivaciones.

Lo más apremiante es abrir debates acerca de la forma como se utiliza (y monopoliza) la IA -y cualquier otro tipo de tecnología-, y no sobre una improbable venganza robótica. Los problemas a tratar son inherentemente humanos, y para eso necesitamos toda la ayuda posible, incluso si no es consciente.



*MARIO DE LA PIEDRA WALTER
Médico por la Universidad La Salle
y neurocientífico por la Universidad
de Bremen. En la actualidad cursa su
residencia de neurología en Berlín,
Alemania.



BIBLIOGRAFÍA:

- Ada Lovelace and the first computer programme in the world. Max-Planck-Gesellschaft. Disponible en: <https://www.mpg.de/female-pioneers-of-science/Ada-Lovelace>
- Russell S. (2019) *Human Compatible: Artificial Intelligence and the Problem of control*. Penguin Publishing Group.
- De Cosomo, L. (2022). Google Engineer Claims AI Chatbot is Sentient: Why it matters. Scientific American.
- Finkel E. (2023) *If AI becomes conscious, how will we know?* Science.Technology. Disponible en: <https://www.science.org/content/article/if-ai-becomes-conscious-how-will-we-know>
- Raphaël Millie and Sean Carroll. *How Artificial Intelligence Thinks*. Sean Carroll Mindscape Podcast.
- Sean Carroll. *Solo: AI Thinks Different*. Sean Carroll Mindscape Podcast

