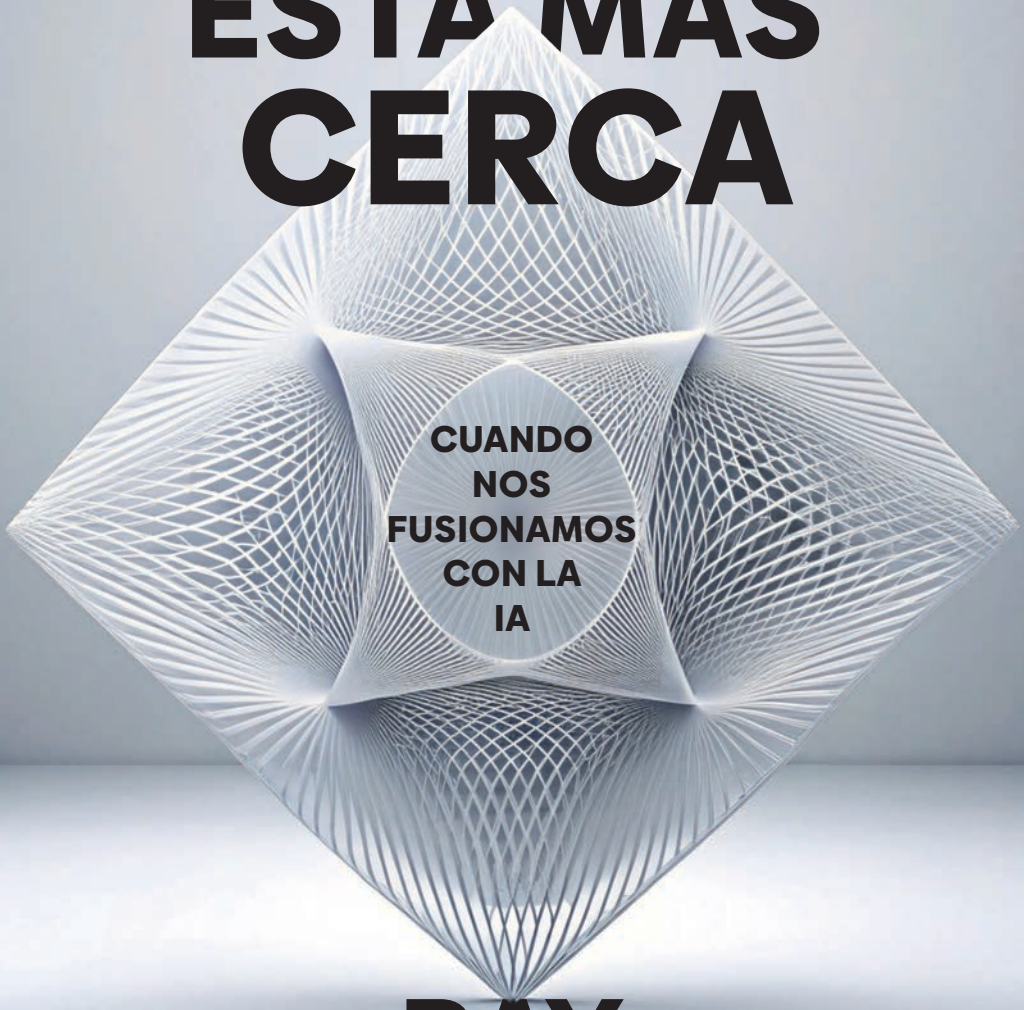
LA SINGULARIDAD ESTÁ MÁS CERCA



CUANDO NOS FUSIONAMOS CON LA IA

RAY KURZWEIL

TRADUCCIÓN DE ALEXANDRE CASANOVAS

DEUSTO

La singularidad está más cerca

Cuando nos fusionamos con la IA

RAY KURZWEIL

Traducción de Alexandre Casanovas

EDICIONES DEUSTO

Deusto es un sello editorial de Centro de Libros PAPF, SLU.
Av. Diagonal, 662-664
08034 Barcelona
www.planetadelibros.com

Diseño de la colección: Sylvia Sans Bassat

Primera edición: enero de 2025
Depósito legal: B. 21.976-2024
ISBN: 978-84-234-3830-3
Composición: Realización Planeta
Impresión y encuadernación: Gómez Aparicio Grupo Gráfico
Printed in Spain - Impreso en España

Sumario

Introducción 9

1. ¿En qué etapa nos encontramos? 17
2. Reinventar la inteligencia 21
3. ¿Quién soy? 113
4. La vida mejora de forma exponencial 165
5. El futuro del trabajo: ¿positivo o negativo? 305
6. Los próximos treinta años en la salud y el bienestar 371
7. Los peligros. 423
8. Un diálogo con Cassandra 455

Apéndice. Relación precio-rendimiento de la potencia de cálculo, fuentes de los gráficos (1939-2023) 459
Agradecimientos 485

1

¿En qué etapa nos encontramos?

En *La singularidad está cerca* expliqué la base de la conciencia como información. Hablaba de seis épocas o etapas, desde el nacimiento del universo, y escribí que cada una de ellas ha alumbrado la siguiente a partir de la información procesada en la última fase. De esta manera, la evolución de la inteligencia tiene lugar a través de una secuencia indirecta de otros procesos.

La Primera Época incluye el nacimiento de las leyes de la física y, gracias a su acción, también de la química. Unos cientos de miles de años después del Big Bang, los átomos empezaron a formarse a partir de los electrones que orbitaban alrededor de un núcleo de protones y neutrones. En principio, los protones de un núcleo no deberían estar unidos, ya que la fuerza electromagnética trata de separarlos con gran violencia. Sin embargo, parece existir algo llamado «fuerza nuclear» que mantiene unidos los protones. «Quienquiera» que diseñara las leyes del universo tuvo que incluir esta fuerza adicional, pues, de cualquier otro modo, la evolución a través de los átomos hubiera sido imposible.

Miles de millones de años después, los átomos formaron moléculas que podían representar información más elaborada. El carbono era el elemento constructivo más útil, ya que era capaz de formar cuatro enlaces, mientras que otros núcleos únicamente tienen uno, dos o tres. En realidad, vivir en un mundo donde la química compleja es una posibilidad real es muy improbable. Por ejemplo, si la fuerza de

la gravedad fuera sólo un poco más débil, no habría supernovas que pudieran crear los elementos químicos de los que se compone la vida. Si fuera un poco más potente, las estrellas se consumirían y morirían antes de que pudiera formarse vida inteligente. Esta sencilla constante física debe permanecer dentro de los límites marcados, o de lo contrario ahora no estaríamos aquí. Vivimos en un universo equilibrado y con suma precisión para poder llegar al nivel de orden que posibilita el desarrollo de la evolución.

Hace miles de millones de años comenzó la Segunda Época: la vida. Las moléculas se volvieron más complejas, hasta el punto de que una sola tenía la capacidad de definir un organismo entero. Así, los seres vivos, cada uno con su propio ADN, fueron capaces de evolucionar y reproducirse.

En la Tercera Época, los animales definidos a partir del ADN formaron el tejido cerebral, un paso que les permitió almacenar y procesar la información. Ese nuevo cerebro proporcionaba grandes ventajas evolutivas, que a su vez incrementaron la propia complejidad del encéfalo en el transcurso de miles de años.

En la Cuarta Época, los animales usaron sus capacidades cognitivas avanzadas, con la ayuda de la pinza manual, para traducir las ideas en acciones complejas. Ahí es donde entran los humanos. Nuestra especie utilizó esas capacidades para crear nuevas tecnologías capaces de almacenar y manipular la información; de los papiros a los discos duros. Estas tecnologías aumentaron la capacidad del cerebro para percibir, recordar y evaluar patrones de información. Aquí hablamos de una fuente evolutiva muy diferente, que en sí misma es muy superior al nivel de progreso anterior. Con el cerebro, ganamos unos 16 centímetros cúbicos de materia gris cada 100.000 años, mientras que con la computación digital duplicamos la ratio precio-rendimiento cada 16 meses.

En la Quinta Época, podremos fusionar la cognición biológica de los seres humanos con la velocidad y la potencia de la tecnología digital. Aquí estaríamos hablando de las interfaces cerebro-ordenador. El procesamiento neuronal de los humanos tiene lugar a una velocidad de varios cientos de ciclos por segundo, mientras que la tecnología digital trabaja a varios miles de millones por segundo. Además de incrementar la velocidad y el tamaño de la memoria, ampliar el cerebro con ordenadores no biológicos nos permitirá añadir muchas más

capas al neocórtex, lo cual abrirá la puerta a un nivel de cognición mucho más complejo y abstracto de lo que podemos imaginar en la actualidad.

En la Sexta Época, la inteligencia se extenderá por todo el universo, hasta convertir la sustancia ordinaria en *computronium*; es decir, en materia organizada según la densidad máxima de la computación.

En mi libro *La era de las máquinas espirituales*, de 1999, predije que, antes de 2029, una máquina aprobará el test de Turing: el momento en que una IA es capaz de comunicarse a través de mensajes de texto de una forma que resulta indistinguible a la de un ser humano. Lo volví a repetir en 2005, en *La singularidad está cerca*. Aprobar el test de Turing implica que la IA domina como un ser humano el lenguaje y el razonamiento basado en el sentido común. Turing describió este concepto en 1950, pero nunca especificó cómo debía realizarse dicho test.[6] En una apuesta que hice con Mitch Kapor, definimos nuestras propias reglas, que son mucho más complejas que otras interpretaciones.

Mi previsión era que, para poder aprobar un test de Turing en 2029, primero teníamos que hacer realidad una gran variedad de hitos intelectuales en el campo de la inteligencia artificial antes de 2020. Y, de hecho, desde que hice esa predicción, la IA ha conquistado muchos de los desafíos intelectuales más difíciles de la humanidad: desde juegos como *Jeopardy!* o el go a aplicaciones más serias, como la radiología o el descubrimiento de nuevos fármacos. Cuando escribo estas líneas, los mejores sistemas de IA, como Gemini y GPT-4, han demostrado sus capacidades en muchos ámbitos de trabajo diferentes; unos pasos muy alentadores en el camino que conduce a la inteligencia general.

A decir verdad, cuando un programa sea capaz de aprobar el test de Turing, en realidad tendrá que aparentar ser mucho menos inteligente, porque de lo contrario será evidente que se trata de una IA. Por ejemplo, si puede resolver al instante cualquier programa matemático, suspendería el examen. Por lo tanto, cuando llegue al nivel del test de Turing, la IA tendrá en realidad unas capacidades que superarán a los mejores humanos en la mayoría de los campos del saber.

6. Turing, Alan M., «Computing machinery and intelligence», *Mind*, 59, 236 (1 de octubre de 1950), p. 435, <https://doi.org/10.1093/mind/LIX.236.433>.

En la actualidad, los seres humanos nos encontramos en la Cuarta Época, ya que disponemos de una tecnología capaz de producir resultados que superan nuestra capacidad de comprensión en ciertos ámbitos. En cuanto a las partes del test de Turing que la IA no domina todavía, cada día avanzamos más deprisa. Aprobar el test de Turing, un hito que anticipo para el año 2029, nos llevará a la Quinta Época.

En la década de 2030, uno de los avances más trascendentales será la conexión de las capas superiores del neocórtex con la nube, lo que ampliará la capacidad de razonamiento del ser humano. En este sentido, la IA ya no será un competidor, sino que se convertirá en una extensión de nosotros mismos. En el momento en que lo logremos, los fragmentos «no biológicos» del cerebro humano nos proporcionarán una capacidad cognitiva miles de veces superior a la que hoy nos ofrecen sus componentes orgánicos.

A medida que todo siga avanzando a una velocidad exponencial, ampliaremos la capacidad del cerebro humano varios millones de veces antes de 2045. La incomprensible velocidad y magnitud de esta transformación es lo que nos permitirá utilizar la metáfora de la singularidad inspirada en las leyes de la física para describir nuestro futuro.

2

Reinventar la inteligencia

¿Qué significa reinventar la inteligencia?

Si toda la historia del universo se reduce a la evolución de los paradigmas sobre el procesamiento de la información, la historia de la humanidad comienza entonces después de haber superado la mitad del camino. Dentro de esta historia tan amplia, nuestro capítulo consiste en protagonizar una transición, desde unos animales con cerebros biológicos hasta unos seres trascendentes cuyas ideas e identidades ya no están limitadas por lo que nos ofrece la genética. En la década de 2020, vamos a entrar en la última fase de esta transformación: reinventar la inteligencia que la naturaleza nos ha dado y convertirla en un sustrato digital más potente con el que confluiremos después. Cuando eso ocurra, la Cuarta Época del universo dará paso a la Quinta Época.

Pero, en concreto, ¿cómo va a suceder? Para comprender lo que significa reinventar la inteligencia, primero debemos echar la vista atrás y fijarnos en el nacimiento de la IA y en las dos grandes escuelas de pensamiento que emergieron de ella. Para entender por qué una prevaleció sobre la otra, estableceremos una relación con las tesis de la neurociencia sobre la forma en que el cerebelo y el neocórtex dieron paso a la inteligencia humana. Tras analizar el aprendizaje profundo y la manera en que ahora recrea la potencia del neocórtex, podremos valorar mejor todo lo que la IA tiene que hacer aún para llegar al nivel de un ser humano —y cómo comprobar que de verdad lo ha conseguí-

do—. Por último, nos centraremos en el proceso de creación, con la ayuda de una IA sobrehumana, de unas interfaces cerebro-ordenador que ampliarán el neocórtex con nuevas capas de neuronas virtuales. Así se liberarán nuevas formas de pensar y, a largo plazo, la inteligencia se ampliará varios millones de veces: eso es la singularidad.

El nacimiento de la IA

En 1950, el matemático británico Alan Turing (1912-1954) publicó un artículo en la revista *Mind* titulado «Maquinaria computacional e inteligencia».[7] En él, Turing planteaba una de las preguntas más profundas de la historia de la ciencia: «¿Las máquinas pueden pensar?». Aunque la idea de una máquina pensante se remonta como mínimo a Talos, el autómata de bronce de la mitología griega, la innovación de Turing consistió en reducir ese concepto a un hecho comprobable de forma empírica.[8] Propuso utilizar el «juego de la imitación» —que conocemos como el «test de Turing»— para determinar si la potencia de cálculo de una máquina era capaz de realizar las mismas tareas cognitivas que el cerebro humano. En el test, un grupo de jueces humanos entrevista al mismo tiempo a una persona real y a una IA a través de un sistema de mensajería instantánea, y sin poder ver con quién de los dos están hablando. Los jueces pueden plantear cualquier pregunta sobre las materias o las situaciones que deseen. Si, después de un cierto período de tiempo, los jueces son incapaces de determinar quién es la IA y quién el ser humano, la inteligencia artificial ha aprobado el examen.

Al transformar esta idea filosófica en un concepto científico, Turing despertó un tremendo entusiasmo entre los investigadores. En 1956, el profesor de matemáticas John McCarthy (1927-2011) propuso la realización de un estudio, de unos dos meses de duración y diez participantes, en el Dartmouth College de Hannover, en Nuevo Hampshire.[9] El objetivo era el siguiente:

7. Ibídem.

8. Shashkevich, Alex, «Stanford researcher examines earliest concepts of artificial intelligence, robots in ancient myths», *Stanford News*, 28 de febrero de 2019, <https://news.stanford.edu/stories/2019/02/ancient-myths-reveal-early-fantasies-artificial-life>.

9. McCarthy, John, *et al.*, «A proposal for the dartmouth summer research pro-

> El estudio consiste en proceder a partir de la suposición de que cada aspecto del aprendizaje, o cualquier otro rasgo de la inteligencia, puede describirse con tanta precisión que es posible fabricar una máquina que lo simule. Se intentará descubrir cómo construir máquinas que usen el lenguaje, formen abstracciones y conceptos, resuelvan problemas que ahora están reservados a los seres humanos y se mejoren a sí mismas.[10]

Mientras se preparaba para la conferencia, McCarthy propuso que esta nueva especialidad, que con el tiempo automatizaría todas las demás, recibiera el nombre de «inteligencia artificial».[11] Este nombre no me gusta demasiado, ya que el término *artificial* induce a pensar que esta forma de inteligencia «no parece real», pero es la palabra que ha perdurado.

El estudio pudo completarse, pero su objetivo —en concreto, conseguir que las máquinas comprendieran problemas planteados en lenguaje natural— no se alcanzó en el plazo previsto de dos meses. Todavía estamos trabajando en ello; por supuesto, con mucho más que diez personas. Según el gigante tecnológico chino Tencent, en 2017 había 300.000 «investigadores y profesionales de la IA» en todo el mundo; y, en 2019, el *Global AI Talent Report*, de Jean-François Gagné, Grace Kiser y Yoan Mantha, identificó a unos 22.400 expertos en IA que estaban publicando investigaciones originales, de las cuales unas 4.000 iban a tener gran influencia.[12, 13] Y según el Instituto de Inteligencia Artificial Centrada en el Ser Humano de Stanford, en 2021, los investigadores en IA generaron más

---

ject on artificial intelligence», propuesta de conferencia, 31 de agosto de 1955, <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>.

10. Ibídem.

11. Childs, Martin, «John McCarthy: computer scientist known as the father of AI», *The Independent*, 1 de noviembre de 2011, <https://www.independent.co.uk/news/obituaries/john-mccarthy-computer-scientist-knownthe-6255307.html>; Christianini, Nello, «The road to artificial intelligence: a case of data over theory», *New Scientist*, 26 de octubre de 2016, <https://www.newscientist.com/article/mg23230971-200-the-irresistible-rise-of-artificial-intelligence/>.

12. James, Vincent, «Tencent says there are only 300.000 AI engineers worldwide, but millions are needed», *The Verge*, 5 de diciembre de 2017, <https://www.theverge.com/2017/12/5/16737224/globaltalent-shortfall-tencent-report>.

13. Gagné, Jean-Francois; Kiser, Grace; y Mantha, Yoan, *Global AI Talent Report 2019*, Element AI, abril de 2019.

de 496.000 publicaciones y más de 141.000 solicitudes de patentes.[14] En 2022, la inversión empresarial en inteligencia artificial fue de 189.000 millones de dólares en todo el mundo, trece veces más que en la década pasada.[15] Las cifras aún serán más altas en el momento en que leas estas líneas.

En 1956 habría sido difícil imaginar algo así. Sin embargo, el objetivo del Taller Dartmouth equivalía más o menos a crear una IA que pudiera aprobar el test de Turing. Mi predicción, que lo conseguiremos antes de 2029, permanece invariable desde 1999 y la publicación de mi libro *La era de las máquinas espirituales*, aparecido en una época en que muchos analistas creían que este hito jamás podría alcanzarse.[16] Hasta hace poco, muchos pensaban que esta proyección era extremadamente optimista. Por ejemplo, una encuesta realizada en 2018 reveló que la predicción más común entre los expertos en IA era que las máquinas no tendrían una inteligencia comparable a la humana hasta 2060.[17] Pero los últimos avances en los grandes modelos de lenguaje han cambiado muy rápido las expectativas. Mientras escribía los primeros borradores de este libro, el consenso en Metaculus, la página web sobre pronósticos más importante del mundo, afirma que tendrá lugar en las décadas de 2040 o 2050. Pero los sorprendentes avances en la IA de los últimos dos

14. Zhang, Daniel, *et al.*, *The AI Index 2022 Annual Report*, AI Index Steering Committee, Instituto de Inteligencia Artificial Centrada en el Ser Humano de Stanford, Universidad de Stanford, marzo de 2022, p. 36, <https://aiindex.stanford.edu/ai-index-report-2022/>; Maslej, Nestor, *et al.*, *The AI Index 2023 Annual Report*, AI Index Steering Committee, Instituto de Inteligencia Artificial Centrada en el Ser Humano de Stanford, Universidad de Stanford, abril de 2023, p. 24, <https://aiindex.stanford.edu/ai-index-report-2023>.

15. Entre 2021 y 2022 se produjo un descenso del 26,7 % en la inversión empresarial, pero probablemente se pueda atribuir a las tendencias macroeconómicas cíclicas y no a un cambio en la trayectoria a largo plazo del compromiso corporativo con la IA. Véase: Maslej, *et al.*, *op. cit.*, pp. 171, 184.

16. Kurzweil, Ray, *La era de las máquinas espirituales: cuando los ordenadores superen la mente humana*, *op. cit.*; Jacquette, Dale, «Who's afraid of the turing test?», *Behavior and Philosophy*, 20, 21 (1993), p. 72, <https://www.jstor.org/stable/27759284>.

17. Grace, Katja, *et al.*, «Viewpoint: when will AI exceed human performance? Evidence from AI experts», *Journal of Artificial Intelligence Research*, 62 (julio de 2018), pp. 729-754, <https://jair.org/index.php/jair/article/view/11222>.

años han superado todas las expectativas, y en mayo de 2022 el consenso en Metaculus coincide exactamente con la fecha que yo propuse: 2029.[18] Desde entonces, ha ido fluctuando hasta llegar incluso a 2026, ¡lo que técnicamente me deja en el grupo de los lentos![19]

Hasta los grandes expertos en la materia se han quedado perplejos por muchos de los recientes avances de la IA. No sólo están llegando antes de lo que esperaba la mayoría, sino que, además, parecen producirse de repente, y sin dar muchas pistas de que el salto adelante vaya a ser inminente. Por ejemplo, en octubre de 2014, Tomaso Poggio, un experto del MIT en IA y ciencia cognitiva, dijo: «La capacidad para describir el contenido de una imagen será uno de los mayores desafíos intelectuales de todos los que se planteen a una máquina. Necesitaremos otro ciclo de investigación elemental para resolver esta clase de pregunta».[20] Poggio calculaba que aún faltaban unas dos décadas para que ese hito se hiciera realidad. Al mes siguiente, Google presentó una IA de reconocimiento de objetos capaz de hacerlo. Cuando Raffi Khatchadourian, de *The New Yorker*, le preguntó sobre la cuestión, Poggio recurrió a un escepticismo de naturaleza más filosófica, ya que cuestionó que esa habilidad pudiera compararse con una verdadera inteligencia. No recojo aquí esta anécdota como una crítica a Poggio, sino más bien como la constatación de una tendencia que todos compartimos. O sea, antes de que la IA logre un determinado objetivo, nos parece algo extremadamente difícil y exclusivo de los seres humanos. Pero después de que la IA lo consiga,

18. Para saber más sobre los motivos que hay detrás de mi predicción y consultar una comparativa con una gran variedad de opiniones de expertos en IA, véase: Kurzweil, Ray, «A wager on the turing test: why i think i will win», KurzweilAI.net, 9 de abril de 2002, <https://www.kurzweilai.net/a-wager-on-the-turing-test-why-i-think-i-will-win>; Müller,Vincent C.; y Bostrom, Nick, «Future progress in artificial intelligence: a survey of expert opinion», en Müller, Vincent C. (ed.), *Fundamental issues of artificial intelligence*, Springer, pp. 553-571, Suiza, 2016, <https://philpapers.org/archive/MLLFPI.pdf>; Aguirre, Anthony, «Date weakly general AI is publicly known», Metaculus, <https://www.metaculus.com/questions/3479/date-weakly-general-ai-system-is-devised>.

19. Aguirre, *op. cit.*

20. Khatchadourian, Raffi, «The Doomsday invention», *The New Yorker*, 23 de noviembre de 2015, <https://www.newyorker.com/magazine/2015/11/23/doomsday-invention-artificial-intelligence-nick-bostrom>.

ese mismo logro se minimiza a ojos de la gente. En otras palabras, el progreso real es mucho más significativo de lo que nos parece en retrospectiva. Por este motivo sigo siendo optimista en relación con mi predicción sobre 2029.

Pero entonces, ¿por qué estos avances llegan de una forma tan repentina? La respuesta se encuentra en un problema teórico planteado durante el nacimiento de la especialidad. En 1964, cuando yo estaba en el instituto, conocí a dos pioneros de la inteligencia artificial: Marvin Minsky (1927-2016), que fue uno de los organizadores del Taller Dartmouth, y Frank Rosenblatt (1928-1971). En 1965, me matriculé en el MIT y empecé a estudiar con Minsky, que estaba trabajando en los conceptos básicos en que se basan los espectaculares avances en la IA que vemos en la actualidad. Minsky me enseñó que hay dos técnicas para crear soluciones automatizadas a los problemas: el enfoque simbólico y el conectivo.

El método simbólico describe la forma en que un experto humano resolvería un problema a partir de unos términos basados en reglas. En algunos casos, los sistemas basados en este método pueden ser satisfactorios. Por ejemplo, en 1959, la Rand Corporation presentó el llamado Solucionador General de Problemas (General Problem Solver, GPS), un programa informático que podía combinar axiomas matemáticos simples para resolver problemas lógicos.[21] Herbert A. Simon, J. C. Shaw y Allen Newell desarrollaron el Solucionador General de Problemas para que tuviera la capacidad teórica de resolver cualquier problema que se presentara como un conjunto de fórmulas bien formadas (fbf). Para que el GPS funcionara, tenía que usar una fbf (en esencia, un axioma) en cada etapa del proceso, en una metódica elaboración de la demostración matemática de la respuesta.

Incluso si no tienes experiencia en lógica formal o en matemáticas basadas en demostraciones, esta idea es la misma que se aplica en el álgebra. Si sabes que $2 + 7 = 9$, y que un número desconocido x sumado a 7 es 10, puedes demostrar que $x = 3$. Esta clase de lógica tiene

21. Newell, A.; Shaw, J. C.; y Simon, H. A., «Report on a General Problem-Solving program», RAND P-1584, RAND Corporation, 9 de febrero de 1959, <http://bitsavers.informatik.uni-stuttgart.de/pdf/rand/ipl/P-1584_Report_On_A_General_Problem-Solving_Program_Feb59.pdf>. En el apéndice aparecen las fuentes utilizadas en el libro para determinar la potencia de cálculo a lo largo de la historia.

aplicaciones mucho más amplias que la simple resolución de ecuaciones. También es la lógica que utilizamos (sin siquiera pensar en ello) cuando nos preguntamos si una cosa encaja con una definición concreta. Si sabes que un número primo sólo puede dividirse por 1 y por sí mismo, y sabes que el 11 es un divisor del 22, y que 1 no es lo mismo que 11, puedes llegar a la conclusión de que el 22 no es número primo. Tras empezar con los axiomas más básicos y fundamentales, el GPS podía hacer esta clase de cálculos con preguntas mucho más difíciles. Al final, es lo mismo que hacen los matemáticos humanos; la diferencia es que una máquina (al menos en teoría) puede rebuscar entre todas las formas posibles de combinar los axiomas fundamentales para encontrar la verdad.

Por poner un ejemplo, si tenemos la posibilidad de escoger entre 10 axiomas de este tipo para cada punto, y se necesitan 20 axiomas para encontrar una solución, eso significaría que hay 10^{20} posibles soluciones (o sea, 100 trillones). Gracias a los ordenadores modernos, hoy podemos trabajar con unos números tan grandes, pero en 1959 esa cifra estaba muy lejos de lo que podía hacer la velocidad de aquellos procesadores. Aquel año, el ordenador DEC PDP-1 podía realizar 100.000 operaciones por segundo.[22] En 2023, la máquina virtual Google Cloud A3 pudo llevar a cabo unos 26.000.000.000.000.000.000 de operaciones por segundo.[23] Hoy, un solo dólar compra una potencia de cálculo 1,6 billones de veces superior a la que existía cuando se desarrolló el GPS.[24] Problemas que requerirían decenas de miles de años de trabajo con la tecnología de 1959 pueden resolverse hoy con un ordenador doméstico en unos pocos minutos. Para compensar sus limitaciones, el GPS tenía una serie de heurísticas programadas que trataban de clasificar la prioridad de las posibles soluciones. Las heurísticas funcionaron bien en varios momentos, y sus éxitos apuntalaron la idea de que

22. Digital Equipment Corporation, *PDP-1 Handbook*, Digital Equipment Corporation, p. 10, Estados Unidos, 1963. <https://www.computerhistory.org/pdp-1/_media/pdf/DEC.pdp_1.1963.102636240.pdf>.

23. Vahdat, Amin; y Lohmeyer, Mark, «Enabling next-generation AI workloads: announcing TPU v5p and AI hypercomputer», Google Cloud, 6 de diciembre de 2023, <https://cloud.google.com/blog/products/ai-machine-learning/introducing-cloud-tpu-v5p-and-ai-hypercomputer>.

24. En el apéndice aparecen las fuentes utilizadas en el libro para determinar la potencia de cálculo a lo largo de la historia.

una solución informatizada podía resolver cualquier problema definido de forma rigurosa.

Otro ejemplo de este enfoque fue un sistema llamado MYCIN, desarrollado en la década de 1970 para diagnosticar y recomendar tratamientos terapéuticos para enfermedades infecciosas. En 1979, un equipo de evaluadores expertos comparó los resultados con los de un grupo de médicos humanos, y descubrió que el MYCIN lo hacía tan bien o mejor que cualquiera de los facultativos.[25]

Un «diagnóstico» típico del MYCIN tenía este aspecto SI:

1. la infección que requiere tratamiento es meningitis, y
2. el tipo de infección es de origen fúngico, y
3. no se han detectado organismos en los cultivos, y
4. el paciente no es un huésped de riesgo, y
5. el paciente ha estado en una zona donde las coccidiomicosis son endémicas, y
6. la raza del paciente está entre: [N]egro [A]siático [I]ndio, y
7. el antígeno criptocócico en el CSF no ha dado positivo...

ENTONCES: hay pruebas que sugieren (5) que el criptococo no es uno de los organismos (entre los detectados en cultivos o frotis) que podrían ser la causa de la infección.[26]

A finales de la década de 1980, estos «sistemas expertos» estaban usando modelos de probabilidades y podían combinar muchas fuentes de datos para tomar una decisión.[27] Aunque una sola regla «si-enton-

25. Yu, V. L., *et al.*, «Antimicrobial selection by a computer: a blinded evaluation by infectious diseases experts», *Journal of the American Medical Association*, 242, 12 (21 de septiembre de 1979), pp. 1279-1282, <https://jamanetwork.com/journals/jama/article-abstract/366606>.

26. Buchanan, Bruce G.; y Shortliffe, Edward Hance (eds.), *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project*, Addison-Wesley, Estados Unidos, 1984; Edelson, Edward, «Programmed to think», MOSAIC 11, 5 (septiembre/octubre de 1980), p. 22, <https://books.google.co.uk/books?id=PU79ZK2tXeAC>.

27. Gill, T. Grandon, «Early expert systems: where are they now?», *MIS Quarterly*, 19, 1 (marzo de 1995), pp. 51-81, <https://www.jstor.org/stable/249711>.

ces» no bastaba por sí sola, tras combinar varios miles de procesos de este tipo, el sistema podía tomar decisiones fiables sobre un problema limitado.

Aunque el enfoque simbólico se ha utilizado durante más de medio siglo, su principal limitación ha sido el «techo de complejidad».[28] Cuando el MYCIN y otros sistemas similares cometían un error, corregirlo podía resolver aquel problema en particular, pero, a su vez, creaba tres errores nuevos que podían sacar la cabeza en otras situaciones. Parecía existir un límite a la complejidad que reducía muchísimo la cantidad de problemas del mundo real que podían resolverse.

Una forma de entender la complejidad de los sistemas basados en reglas consiste en verlos como un conjunto de posibles puntos de fallo. En términos matemáticos, un grupo de *n* cosas tiene dos subconjuntos 2^{n-1} (sin contar el conjunto vacío). Por lo tanto, si una IA utiliza un conjunto de reglas que únicamente contiene un elemento, sólo existe un punto único de fallo; o sea, ¿esa regla aislada funciona bien o no? Si hay dos reglas, hay tres puntos de fallo: uno por cada una de las reglas, y otro por la interacción cuando se combinan. Y la relación sigue creciendo de forma exponencial. Cinco reglas incluyen 31 posibles puntos de fallo, 10 reglas significan 1.023, 100 reglas son más de un cuatrillón, ¡y 1.000 reglas contienen un gúgol de gúgol de gúgoles! Por lo tanto, cuantas más reglas se utilizan, más puntos de fallo se añaden al subconjunto cuando se incorporara una nueva. Incluso si sólo una fracción extremadamente minúscula de todas las posibles combinaciones presenta un nuevo problema, se llega a un punto (cuya ubicación exacta varía de una situación a otra) en que añadir una nueva regla para resolver un problema causa más de un error adicional. En eso consiste el techo de complejidad.

El sistema experto que lleva más tiempo en funcionamiento quizá sea el Cyc (de la palabra *enciclopedia*), diseñado por Douglas Lenat y sus colegas de Cycorp.[29] Creado en 1984, Cyc tiene como objetivo co-

28. Para una explicación sin tecnicismos sobre por qué el aprendizaje automático reduce la dificultad del problema del techo de complejidad, véase: Saxena, Deepanker, «Machine learning vs. rules based systems», *Socure*, 6 de agosto de 2018, <https://www.socure.com/blog/machine-learning-vs-rule-based-systems>.

29. Metz, Cade, «One genius' lonely crusade to teach a computer common sense», *Wired*, 24 de marzo de 2016, <https://www.wired.com/2016/03/doug-lenat

dificar todo «el conocimiento basado en el sentido común»; es decir, hechos bien conocidos, tales como: «Un huevo que cae al suelo se romperá»; o «Un niño que corre por la cocina con los zapatos llenos de barro enfadará a sus padres». Son millones de pequeñas ideas que no están escritas con claridad en ninguna parte. Son suposiciones tácitas que subyacen al comportamiento y el razonamiento humanos, y que son necesarias para comprender lo que sabe una persona cualquiera sobre una variedad de temas distintos. Pero, como el sistema Cyc presenta este conocimiento con reglas simbólicas, también debe enfrentarse al techo de complejidad.

En la década de 1960, mientras Minsky me aconsejaba sobre los pros y los contras del enfoque simbólico, empecé a ver el valor añadido que aportaba el método conexionista. Este enfoque se basa en redes de nodos que crean la inteligencia a través de su estructura, y no del contenido. En vez de usar reglas inteligentes, estas redes utilizan nodos «tontos» organizados de una forma que permite extraer la información de los propios datos. Como resultado, tienen el potencial de descubrir patrones sutiles que nunca se le ocurrirían a un programador humano que intentara diseñar reglas simbólicas. Una de las ventajas fundamentales del enfoque conexionista es que permite resolver problemas sin comprenderlos. Incluso si tuviéramos la capacidad de formular e implementar a la perfección reglas sin errores para resolver problemas con una IA simbólica (que no es el caso), estaríamos limitados por nuestra comprensión imperfecta de las instrucciones que serían ideales.

El método conexionista es una manera muy potente de abordar problemas complejos, aunque en realidad puede convertirse en un arma de doble filo. La IA conexionista tiene tendencia a convertirse en una «caja negra», capaz de escupir la respuesta correcta, pero incapaz de explicar cómo la ha conseguido.[30] Esta cuestión podría

---

-artificial-intelligence-common-sense-engine>; «Frequently Asked Questions», Cycorp, <https://cyc.com/faq>, [consulta: 20 de noviembre de 2021].

30. Para encontrar más información sobre el problema de la caja negra y la transparencia de la IA, véase: Knight, Will, «The dark secret at the heart of AI», *MIT Technology Review*, 11 de abril de 2017, <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai>; «AI detectives are cracking open the black box of deep learning», *Science Magazine* [vídeo], YouTube, 6 de julio de 2017, <https://www.youtube.com/watch?v=gB_-LabED68>; Voosen, Paul, «How AI de-

llegar a convertirse en un grave problema, ya que los usuarios querrán tener la posibilidad de ver el razonamiento que hay detrás de decisiones con una gran trascendencia, como un tratamiento médico, la aplicación de una ley, una gestión de riesgos o una solución epidemiológica. Por este motivo, muchos expertos en IA están trabajando en el desarrollo de mejores sistemas de «transparencia» (o «interpretabilidad mecánica») sobre las decisiones basadas en el aprendizaje automático.[31] Queda por ver hasta qué punto esa transparencia será efectiva cuando el aprendizaje profundo se vuelva más potente y complejo.

Cuando empecé a interesarme por el conexionismo, sin embargo, los sistemas eran mucho más sencillos. La idea básica consistía en crear un modelo informatizado que estuviera inspirado en el funcionamiento de las redes neuronales humanas. Al principio era algo muy abstracto, porque el método se diseñó antes de que comprendiéramos en detalle cómo se organizan en realidad las redes neuronales biológicas.

---

tectives are cracking open the black box of deep learning», *Science*, 6 de julio de 2017, <https://doi.org/10.1126/science.aan7059>; Shum, Harry, «Explaining AI» [vídeo], YouTube, 16 de enero de 2020, <https://www.youtube.com/watch?v=rI_L95qnVkM>; Instituto para el Futuro de la Vida, «Neel Nanda on what is going on inside neural networks» [vídeo], YouTube, 9 de febrero de 2023, <https://www.youtube.com/watch?v=mUhO6st6M_0>.

31. Para encontrar un excelente resumen sobre el concepto de interpretabilidad mecánica, realizado por el investigador Neel Nanda, véase: Instituto para el Futuro de la Vida, «Neel Nanda on what is going on inside neural networks».