

RICHARD SUSSKIND

DESCIFRANDO

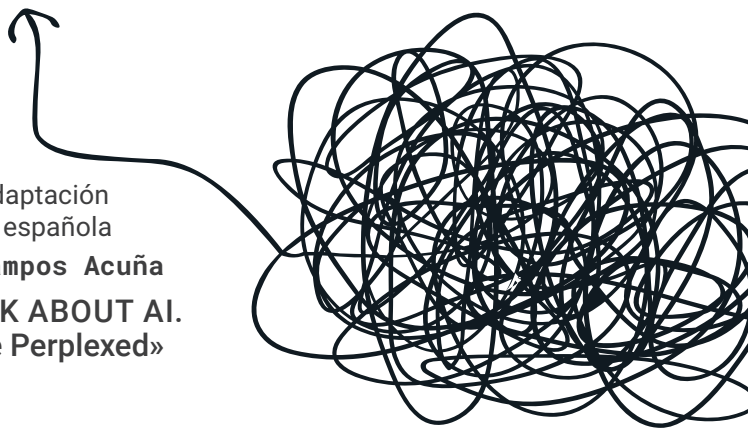
LA IA

CÓMO USARLA
SIN PERDER EL JUICIO

Revisión y adaptación
de la edición española

Concepción Campos Acuña

«HOW TO THINK ABOUT AI.
A Guide for the Perplexed»



III LA LEY

© Richard Susskind, 2026
© ARANZADI LA LEY, S.A.U.

Primera edición en lengua española: Abril 2026 | Traducción de la obra original «**How To Think About AI: A Guide For The Perplexed**», 1ª edición | Oxford University Press, 2025.

© Concepción Campos Acuña. Revisión y adaptación de la traducción a la lengua española.
© Oxford University Press, 1.ª edición en inglés, 2025.

ARANZADI LA LEY, S.A.U.

C/ Collado Mediano, 9

28231 Las Rozas (Madrid)

www.aranzadilaley.es

Atención al cliente: <https://areacliente.aranzadilaley.es/publicaciones>

Primera edición: Abril 2026

Depósito Legal: M-6573-2026

ISBN versión impresa: 979-13-88078-13-2

ISBN versión electrónica: 979-13-88078-14-9

Diseño, Preimpresión e Impresión: ARANZADI LA LEY, S.A.U.

Printed in Spain

© **ARANZADI LA LEY, S.A.U.** Todos los derechos reservados. A los efectos del art. 32 del Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba la Ley de Propiedad Intelectual, ARANZADI LA LEY, S.A.U., se opone expresamente a cualquier utilización del contenido de esta publicación sin su expresa autorización, lo cual incluye especialmente cualquier reproducción, modificación, registro, copia, explotación, distribución, comunicación, transmisión, envío, reutilización, publicación, tratamiento o cualquier otra utilización total o parcial en cualquier modo, medio o formato de esta publicación.

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la Ley. Dirijase a **Cedro** (Centro Español de Derechos Reprográficos, www.cedro.org) si necesita fotocopiar o escanear algún fragmento de esta obra.

El editor y los autores no asumirán ningún tipo de responsabilidad que pueda derivarse frente a terceros como consecuencia de la utilización total o parcial de cualquier modo y en cualquier medio o formato de esta publicación (reproducción, modificación, registro, copia, explotación, distribución, comunicación pública, transformación, publicación, reutilización, etc.) que no haya sido expresa y previamente autorizada.

El editor y los autores no aceptarán responsabilidades por las posibles consecuencias ocasionadas a las personas naturales o jurídicas que actúen o dejen de actuar como resultado de alguna información contenida en esta publicación.

ARANZADI LA LEY no será responsable de las opiniones vertidas por los autores de los contenidos, así como en foros, chats, o cualesquiera otras herramientas de participación. Igualmente, ARANZADI LA LEY se exime de las posibles vulneraciones de derechos de propiedad intelectual y que sean imputables a dichos autores.

ARANZADI LA LEY queda eximida de cualquier responsabilidad por los daños y perjuicios de toda naturaleza que puedan deberse a la falta de veracidad, exactitud, exhaustividad y/o actualidad de los contenidos transmitidos, difundidos, almacenados, puestos a disposición o recibidos, obtenidos o a los que se haya accedido a través de sus PRODUCTOS. Ni tampoco por los contenidos prestados u ofertados por terceras personas o entidades.

ARANZADI LA LEY se reserva el derecho de eliminación de aquellos contenidos que resulten inveraces, inexactos y contrarios a la ley, la moral, el orden público y las buenas costumbres.

ÍNDICE GENERAL

ELOGIOS A <i>DESCIFRANDO A LA IA</i>	9
INTRODUCCIÓN	15
NOTA.....	17
NOTA PERSONAL	23
INTRODUCCIÓN	25

PARTE UNO DESCIFRANDO LA IA

CAPÍTULO 1	
EL VERANO DE LA IA	33
CAPÍTULO 2	
SOBRE LA TECNOLOGÍA	53

PARTE DOS PENSAR DE MANERA DIFERENTE

CAPÍTULO 3	
PENSAMIENTO ORIENTADO AL PROCESO Y PENSAMIENTO ORIENTADO AL RESULTADO	75

CAPÍTULO 4	
CONFUSIONES	83

CAPÍTULO 5	
NO TENEMOS LAS PALABRAS	97

PARTE TRES HACER QUE LA IA FUNCIONE

CAPÍTULO 6	
AUTOMATIZACIÓN, INNOVACIÓN, ELIMINACIÓN	113

CAPÍTULO 7	
CAMBIO ESTRUCTURAL RADICAL	125

PARTE CUATRO AFRONTAR LOS RIESGOS

CAPÍTULO 8	
CATEGORÍAS DE RIESGO	135

CAPÍTULO 9	
APROVECHAR LA IA	157

PARTE CINCO CONTEMPLAR EL FUTURO

CAPÍTULO 10	
¿MÁQUINAS CONSCIENTES?	187

CAPÍTULO 11	
PRÓXIMAMENTE	203

CAPÍTULO 12	
LA GRAN ESCISIÓN	219
CONCLUSIÓN	237
NOTAS	239
LOS TREINTA MEJORES LIBROS DE IA	269

SOBRE LA TECNOLOGÍA

Para comprender la importancia a largo plazo de la IA, sea generativa o de otro tipo, resulta útil tomar distancia y obtener una visión general de lo que está ocurriendo de manera más amplia en el mundo de la tecnología, entendida aquí como tecnología digital. También considero útil situar los avances tecnológicos y de la IA en el contexto más amplio de las distintas etapas del desarrollo humano y social. Asimismo, ayuda sacar a la luz un par de sesgos que pueden distorsionar nuestra comprensión del avance tecnológico. Con estas herramientas, podemos empezar a responder a la pregunta: ¿hasta qué punto podrían llegar a ser avanzados los sistemas de IA con el tiempo?

Tecnología digital

Los tecnólogos disponen de muchas formas de explicar la extraordinaria era en la que vivimos y trabajamos. Yo lo hago a través de cinco ideas principales.

La primera y más importante es que nuestras máquinas son cada vez más capaces¹. Dejemos de lado por un momento el aprendizaje automático, la computación cuántica y todas las demás tecnologías y conceptos habilitadores de los que tanto se habla. Centrémonos, en cambio, en el hecho de que nuestros sistemas y máquinas, sencilla-

mente, hacen cada vez más cosas. Esta observación es esencial. Casi a diario parece que escuchamos noticias sobre avances notables, nuevos sistemas, herramientas, tecnologías, aplicaciones y empresas emergentes. Gran parte de estas noticias se refieren a sistemas que realizan tareas que, hasta hace poco, considerábamos exclusivas de los seres humanos: desde herramientas de diagnóstico médico hasta sistemas capaces de producir pinturas, desde sistemas que detectan si una sonrisa humana es falsa hasta aquéllos que pueden predecir cuántos años de vida nos quedan.

Desde hace mucho tiempo aceptamos que excavadoras, motosierras y grúas pueden superarnos en tareas manuales realizadas en jardines o en obras. Hemos asumido que los coches viajan más rápido de lo que podemos correr y, desde hace generaciones, sabemos que las cosechadoras pueden recolectar cultivos con mayor rapidez que los trabajadores humanos. En las fábricas, desde hace décadas, sabemos que las máquinas son mucho más productivas y seguras que los seres humanos en muchas actividades.

Estamos mucho menos dispuestos a conceder que las máquinas puedan rivalizar o superar a los humanos en actividades que requieren intelecto, incluidas nuestras capacidades para pensar, razonar, resolver problemas, formular hipótesis y planificar. Sin embargo, en la era digital, la trayectoria general es clara: es probable que muchas o la mayoría de las actividades y tareas que creíamos que solo podían realizar los humanos sean asumidas, con el tiempo, de forma autónoma por sistemas capaces, o por humanos menos expertos apoyados por dichos sistemas. Algunos ejemplos se exponen en el Capítulo 1. Los logros de los sistemas actuales y emergentes sugieren no solo que nuestros sistemas son cada vez más capaces, sino que avanzan a un ritmo que aparentemente se acelera. En muchos ámbitos, ya superan a los seres humanos.

Mi segunda observación fundamental sobre la tecnología es que la notable variedad de técnicas y tecnologías habilitadoras que impulsan nuestras máquinas avanza de forma exponencial. Ya hablemos de potencia de procesamiento, memoria, ancho de banda, capacidad de almacenamiento de datos u otros muchos habilitadores, el ritmo de cambio es explosivo y, a menudo, difícil de comprender. Pensemos en las redes neuronales, la tecnología que sustenta la mayoría de los sistemas de IA actuales. La potencia computacional disponible para «entrenar» estos sistemas se duplicó aproximadamente cada seis meses entre 2010 y 2022². El avance en la tecnología de chips ha sido espectacular³. Esto significa que, a este ritmo, cada diez años aproximadamente, la velocidad de computación mejorará en torno a un millón de veces. Aunque podemos y debemos asombrarnos ante desarrollos recientes como ChatGPT y otros sistemas de IA generativa, quizá deberíamos quedar aún más impresionados por el hecho de que la tecnología que los sustenta pueda avanzar, por sí misma, un millón de veces en la próxima década.

Vivimos en una era de mayor cambio tecnológico que cualquier otra conocida por la humanidad. Más aún, como ya se ha señalado, el propio ritmo del cambio se está acelerando. Ray Kurzweil, destacado inventor y futurista, formaliza esta aceleración en su «ley de los rendimientos acelerados». Según esta idea, «las tecnologías de la información, como la computación, se abaratan exponencialmente porque cada avance facilita el diseño de la siguiente etapa de su propia evolución⁴». Los propios sistemas emergentes desempeñan un papel en el diseño y desarrollo de los siguientes avances, y así sucesivamente. El ejemplo reciente más claro se encuentra en los sistemas de IA generativa capaces de escribir código: sistemas que desarrollan sistemas. Esto tiene una importancia profunda para el futuro de la tecnología y de la humanidad: cuando los sistemas pueden autocorregirse, auto-mejorarse y autoextenderse, debemos esperar avances aún más rápidos

de los que hemos visto hasta ahora. El impacto afectará tanto a los sistemas actuales como a la propagación de sistemas completamente nuevos, por ejemplo, en la creación de enormes mundos virtuales en línea en constante evolución.

En conjunto, en el avance de las tecnologías habilitadoras, hemos superado el «codo» de la curva exponencial y deberíamos esperar avances explosivos adicionales en todo el ámbito de la tecnología digital.

En tercer lugar, no parece existir una línea de meta en la competición global por desarrollar tecnologías digitales. Nadie en las principales empresas tecnológicas del mundo ni en los laboratorios de investigación de Estados Unidos, China o Corea del Sur espera que el trabajo tecnológico termine pronto. No hay un destino final a la vista. Se trata, más bien, de una carrera implacable y sin fin.

La IA ha capturado la imaginación del mercado y ha cautivado a la comunidad inversora. Se están inyectando niveles extraordinarios de capital. Las grandes empresas tecnológicas y de IA están destinando sumas desorbitadas a investigación, los gobiernos están desembolsando cantidades masivas, la mayoría de las grandes empresas asignan fondos significativos a proyectos piloto y desarrollo, y la vibrante y talentosa comunidad de start-ups de IA se beneficia de grandes inyecciones de capital riesgo. Las firmas de capital privado también se han sumado a la contienda, en busca de las mayores oportunidades. Un estudio de la firma de auditoría PwC situó en 2017 la contribución potencial de la IA a la economía global en 2030 en 15,7 billones de dólares, antes incluso de la llegada de la IA generativa⁵. En 2023, la consultora McKinsey concluyó que solo la IA generativa «podría añadir hasta 4,4 billones de libras a la economía global, anualmente»⁶. Podemos discutir sobre el número exacto de billones, pero el sentimiento del mercado es claro: la IA es un fenómeno de enorme valor

económico potencial. A su vez, muchos de los mejores graduados y científicos más brillantes no han podido resistirse a la atracción de la financiación para investigación, los salarios elevados y el capital de inversión que acompañan a los proyectos y empresas de IA.

La rápida aparición de nuevos sistemas es prueba de la magnitud de la actividad. Pensemos en ChatGPT, buque insignia de la última oleada de sistemas de IA. En los cuatro meses siguientes a su lanzamiento apareció una actualización importante (v.4), junto con toda una serie de competidores creíbles. Para cuando el lector tenga este libro en sus manos, espero que GPT-5 ya esté disponible.

También están floreciendo nuevas técnicas. Un ejemplo es el trabajo sobre «datos sintéticos». Suena algo circular, pero la idea es que las máquinas pueden generar grandes volúmenes de datos que luego pueden alimentar futuros sistemas de IA. Esto reduciría la dependencia del proceso jurídicamente cuestionable de extraer datos de la web y de bases de datos existentes⁷. Otro ejemplo es el desarrollo de sistemas que ya no requieren que sus usuarios sean «ingenieros de prompts». En su lugar, los propios sistemas son capaces de formular preguntas a los usuarios de manera sistemática, para ayudar a concretar, categorizar y clasificar los problemas o cuestiones sobre los que se solicita orientación. O existe la posibilidad de sistemas de IA autónomos capaces de comunicarse entre sí, tomar decisiones o acceder y desplegar recursos informáticos o físicos sin intervención humana directa (a veces denominados «IA agéntica»).

Todo esto me lleva, en cuarto lugar, a una noción que encuentro a la vez estimulante y profundamente inquietante. Suponiendo que la humanidad no sea aniquilada en alguna catástrofe planetaria (autoinfligida o por factores externos), el lanzamiento de las tecnologías digitales más poderosas e influyentes probablemente se sitúe en los años venideros. A éstas las denomino tecnologías aún no inventadas. La

tecnología digital solo nos acompaña desde hace unos ochenta y cinco años. Todavía estamos calentando motores. Dentro de no muchos años, nuestras tecnologías actuales parecerán primitivas, del mismo modo que hoy nos resulta anticuado el equipamiento de los años ochenta.

En el futuro, las vidas humanas se verán transformadas por tecnologías que aún no han sido inventadas ni siquiera imaginadas. Los sistemas de mayor impacto esperan todavía ser descubiertos y desarrollados. No puedo probarlo, por supuesto, pero nuestra experiencia de las últimas décadas (desde la aparición de la web hasta las redes sociales y la IA generativa) respalda firmemente esta afirmación.

Bajo este epígrafe, deberíamos admitir, al menos conceptualmente, tanto sistemas de IA concebibles como inconcebibles. En una ocasión escuché al escritor y pensador Edward de Bono explicar informalmente una distinción similar. Dijo —y parafraseo de memoria— que, desde que los seres humanos podían alzar la vista al cielo, al menos podían imaginar caminar sobre la luna. Eso era concebible. Lo que resultaba inconcebible hasta el segundo cuarto del siglo XX era que, el día en que los humanos caminaran realmente sobre la luna, esto sería visto por más de 650 millones de personas en pequeñas cajas situadas en la esquina de sus salones. Con la IA ocurre algo similar: hay mucho que hoy podemos imaginar y concebir, pero habrá mucho más que está simplemente más allá de nuestra contemplación actual.

Muchos recién llegados al campo de la IA parecen creer que los inventos recientes —ChatGPT y otros— representan el final del camino, el punto máximo del avance tecnológico. En realidad, son nuestros primeros y titubeantes pasos infantiles. La mayoría de las predicciones sobre el futuro son, a mi juicio, irremediabilmente defectuosas porque ignoran las tecnologías aún no inventadas.

En quinto lugar, y lo más desconcertante de todo en este resumen de la era digital, ya no comprendemos plenamente cómo funcionan nuestros sistemas de IA más avanzados. Y es probable que este estado de incomprensión aumente. Esto se hizo evidente para mí al leer *What Is ChatGPT Doing... and Why Does It Work?*⁸, de Stephen Wolfram, uno de los científicos de la computación más destacados de nuestro tiempo. Me llamó la atención cuántas veces admite que no puede explicar cómo funciona ChatGPT (y eso era GPT-3.5; desde entonces se han lanzado versiones más nuevas). Una y otra vez, Wolfram describe el rendimiento de ChatGPT como «notable» y sus resultados como «inesperados»⁹. Va más allá. Sostiene que nuestra comprensión de cómo funcionan estos sistemas sigue siendo preteórica: no disponemos de formas generales, sistemáticas y científicas para explicar cómo ChatGPT y sistemas similares hacen lo que hacen. Wolfram incluso escribe sobre circunstancias en las que «empieza a colarse un poco de vudú»¹⁰. Habla del «arte» de entrenar redes neuronales¹¹ y de una técnica clave sobre la que afirma: «No creo que haya ninguna ciencia particular detrás de esto. Simplemente se han probado distintas cosas y esta es una que parece funcionar», y ello «sin necesidad de "entender a nivel de ingeniería" cómo ha acabado configurándose la red neuronal»¹². En resumen, Wolfram afirma que estamos «descubriendo que, sí, ahí dentro es complicado y no lo entendemos, aunque al final esté produciendo un lenguaje humano reconocible»¹³.

Esto debería hacer humildes a todos los que creen saber qué está ocurriendo cuando utilizan ChatGPT y sistemas similares. Cuando los sistemas alcanzan cierto nivel de complejidad, sus operaciones cotidianas individuales (a diferencia de su diseño general) a menudo desafían la explicación y la comprensión humanas. Vimos esto a finales del milenio con el problema del año 2000. Los observadores no téc-

nicos preguntaban razonablemente por qué los especialistas informáticos no podían decirnos qué ocurriría cuando el reloj marcara la medianoche. Nuestros sistemas, incluso entonces, parecían demasiado complejos para ser comprendidos. Los sistemas de IA llevan esta incompreensión a un nuevo nivel.

En resumen de la era digital: nuestros sistemas y máquinas son cada vez más capaces; las tecnologías subyacentes avanzan a un ritmo exponencial y explosivo; no existe una línea de meta aparente; aún no hemos inventado la mayoría de las tecnologías que transformarán nuestras vidas futuras; y, ya hoy, no comprendemos plenamente el funcionamiento de algunos de nuestros sistemas de IA.

Cuatro, cinco o incluso seis etapas del progreso humano

Otra forma de analizar el impacto de la tecnología digital es en referencia al modo principal en que la información se captura, comparte y comunica en la sociedad¹⁴. Sigo aquí a antropólogos como Walter Ong, que observan que los seres humanos progresan a través de distintas etapas: la era de la oralidad, en la que la comunicación estaba dominada por la palabra hablada; la era de la escritura; luego la de la imprenta; y hoy un mundo en el que la comunicación está en gran medida posibilitada por diversas tecnologías digitales¹⁵. Dependiendo de cómo se desarrollen los acontecimientos, es probable que exista una quinta etapa —quizá la época del transhumanismo (véanse los Capítulos 11 y 12)— en la que converjan humanidad, tecnología, robótica, nanotecnología molecular, realidad virtual, biología sintética e ingeniería genética. Y podría existir, al menos de forma concebible, una sexta etapa, en la que los sistemas de IA podrían reemplazarnos por completo (véase el Capítulo 12).

A mi juicio, actualmente nos estamos acercando al final de una fase de transición entre la tercera y la cuarta etapas de desarrollo,

entre una sociedad industrial basada en la imprenta y una sociedad digital habilitada por la IA. No se trata de un cambio menor. Algunos comparan su impacto con la magnitud del cambio provocado por la invención de la imprenta, la tecnología que permitió el paso de la segunda a la tercera etapa del progreso humano. Veo la analogía en que las máquinas de tipos móviles dieron lugar a una revolución en la educación, en la difusión de ideas y en el mercado global de la lectura. A medida que avanzamos hacia una sociedad digital plenamente basada en la IA, sustentada por internet y la web, y difundida a través de dispositivos móviles ubicuos (y pronto mediante interfaces cerebro-ordenador; véase el Capítulo 11), se producirá, sin duda, un nuevo cambio colosal en la forma en que creamos, almacenamos, difundimos y consumimos información. Pero la materia prima de la IA es mucho más que datos e información. Los sistemas de IA no se limitan a almacenar y distribuir conocimiento: aplican ese conocimiento.

Se ha producido otro cambio en la transición de la sociedad basada en la imprenta a la digital. Durante los últimos ochenta y cinco años aproximadamente, los científicos de la computación y los diseñadores de sistemas han ido interponiendo capa tras capa entre los humanos y los unos y ceros de las máquinas digitales. Al principio, los ordenadores solo podían ser operados directamente por ingenieros que escribían en lenguajes de programación de bajo nivel. Con el tiempo, se desarrollaron lenguajes de programación de más alto nivel, más fáciles de usar y ejecutar. Para los usuarios, avances como las interfaces gráficas y el reconocimiento de voz han permitido que personas sin conocimientos tecnológicos utilicen ordenadores, incluidas tabletas y dispositivos portátiles. Gran parte de la industria informática se ha dedicado a hacer la tecnología accesible y utilizable tanto para los mortales comunes como para el sacerdocio de la informática. Con los avances en IA y la posibilidad de comunicarnos con nuestras máquinas en



Richard Susskind es una de las voces más influyentes dentro del panorama jurídico internacional actual. Se dedica a explicar de una manera clara y rigurosa cómo y por qué está cambiando la profesión jurídica. No escribe teoría abstracta, sino que traduce la tecnología, la innovación y la evolución social en impactos concretos sobre abogados, jueces, notarios y despachos profesionales.

A lo largo de este libro se analiza el profundo impacto que ya está teniendo la inteligencia artificial en la sociedad, en los negocios y en el mundo jurídico. «Descifrando la IA» apunta directamente a la necesidad de reflexión y discernimiento ante una tecnología claramente disruptiva. Susskind examina dilemas morales, como la toma de decisiones automatizada, la privacidad y el sesgo algorítmico, y propone formas de gestionar estos riesgos.

Leer a Susskind es adelantarse al futuro inmediato del Derecho, ya que te ayudará a entender qué prácticas desaparecerán, cuáles se transformarán y dónde está el verdadero valor del jurista en un entorno digital cada vez más dominado por la IA y cada vez más exigente. Sus libros no alarman ni idealizan, proporcionan herramientas para adaptarse y seguir siendo relevante.

El autor invita a no reaccionar con miedo o entusiasmo irracional, sino a desarrollar un pensamiento crítico y estructurado sobre la IA. Aquel profesional del Derecho que se adentre en Descifrando la IA descubrirá una lectura que incomoda lo justo y aporta mucho.

ISBN: 979-13-88078-13-2



9 791388 078132



ER-02802005



GA-20200102

LA LEY