



Observatorio de las Ideas

REVISTA DE IDEAS

EJEMPLAR EDITADO PARA

Cortesía del Editor

Nº 145 ABRIL 2025



DIRECTORA

Gloria Álvarez

CONSEJO ASESOR

Andrés Ortega

Francesc Trillas

Anna Birulés

Antón Costas

Guillermo de la Dehesa

Javier Nadal

Ana Palacio

Ignacio Pérez de Arriaga

Manuel Pimentel

Josep Piqué †

Narcís Serra

Pedro Solbes †

Juan Tapia

EQUIPO DE INVESTIGACIÓN

José Balsa

Manuel Cebrián

Jordi Domènech

Xavier Massa

Jaime Moreno

Ángel Pascual-Ramsay

Federico Steinberg

EDITA

Observatorio de Ideas S. L.

PRESIDENTE

Daniel Fernández

COORDINACIÓN DEL CONSEJO EDITORIAL

Àngels Ingla

CIF B65855868

C/DIPUTACIÓ 262 2^a 08007

Barcelona Tel. 93 494 97 20

www.observatoriodli.com

ISSN - edición en papel: 2339-8892

ISSN - edición digital: 2938-6438

D.Legal: B.3130-2014



Estimado/a lector/a:

En este número exploramos las limitaciones humanas en la toma de decisiones en contextos complejos y sus efectos en la optimización de elecciones, la inteligencia artificial y la gestión del talento.

La economía del comportamiento ha demostrado que los seres humanos no siempre decidimos racionalmente. Así, Simon, Tversky, Kahneman, Enke *et al.* confirman que, cuanto más complejo es un problema, menos ajustamos nuestras decisiones, incluso si optimizarlas nos beneficiara. En contextos complejos, nuestras elecciones tienden a mantenerse, aunque deberían cambiar.

El segundo artículo nos lleva a la vanguardia de la IA. Park y su colegas replican con agentes de IA las respuestas de más de mil personas, lo que permite predecir reacciones en escenarios sociales. Estas conclusiones plantean cuestiones éticas, de seguridad y privacidad, como señala Manuel Cebrián, y abren nuevas posibilidades para estudiar fenómenos sociales complejos.

El tercer artículo aborda la idea de Gignac sobre la «excepcionalidad» como fenómeno extraño, con implicaciones en la selección de personal. Invita a reconsiderar la «escasez de talento»: buscar el «candidato perfecto» puede excluir perfiles valiosos y generar dificultades para cubrir puestos clave en organizaciones y empresas.

Dos reseñas exploran la influencia crítica de la IA y la tecnología en la sociedad. Mientras que una destaca su sofisticación matemática, la otra advierte sobre la necesidad de regulación para mitigar riesgos políticos y sociales.

Recordemos por un momento las aulas de matemáticas: ecuaciones, derivadas, vectores..., conceptos que tal vez resuenen o se hayan desdibujado con el tiempo. En *Why Machines Learn: The Elegant Maths Behind Modern AI*, Ananthaswamy recupera estos fundamentos matemáticos, desmitifica la IA y demuestra que se basa en décadas de avances, desde reglas elementales hasta el aprendizaje profundo y modelos como ChatGPT.

Después, en *The Tech Coup: How to Save Democracy from Silicon Valley*, Marietje Schaake, política y académica europea, insta a las democracias a recuperar el control cedido a las grandes corporaciones de Silicon Valley. Expone los riesgos de esta concentración de poder en manos de unas pocas empresas y propone medidas para frenar lo que denomina un «golpe de estado tecnológico», de forma que la tecnología opere en favor del interés público.

La tensión entre innovación y gobernanza es evidente. ¿Cómo equilibrarlas para evitar que la tecnología refuerce desigualdades o debilite la democracia? Europa, con su enfoque regulador y su énfasis en los derechos digitales, podría liderar la respuesta. Aun así, la pregunta sigue abierta.

Esperamos que estas ideas y preguntas susciten su interés.

Reciba un afectuoso saludo,

Gloria Álvarez Hernández

Directora



Observatorio de las Ideas

REVISTA DE IDEAS

| IDEAS DE INTERÉS |

LA INCERTIDUMBRE COGNITIVA DISTORSIONA EL COMPORTAMIENTO ECONÓMICO

Publicación: «Behavioral Attenuation», de Benjamin Enke, Thomas Graeber y Ryan Oprea.

Síntesis: Este análisis abarca un gran número de experimentos relacionados con decisiones centrales sobre economía, como las de consumo, ahorro, trabajo o inversión. Se constatan algunas atenuadas respecto de lo que predicen los modelos económicos por la incapacidad de procesar información compleja. Estos resultados son generalizables a la mayoría de las decisiones económicas.

UN LABORATORIO VIRTUAL (¡SIN HUMANOS!) CASI PERFECTO PARA LA SOCIOLOGÍA

Publicación: «Generative Agent Simulations of 1000 People», de Joon Sung Park, Carrie Cai, Meredith Ringel Morris, Robb Willer, Percy Liang, Michael S. Bernstein, Carolyn Q. Zou, Aaron Shaw y Benjamin Mako Hill.

Síntesis: Los autores proponen una innovadora arquitectura de agentes generativos que simula las actitudes y comportamientos de 1052 individuos reales. Partiendo de entrevistas con ellos, construyen perfiles detallados sobre una persona que alimentan un modelo de lenguaje que la simulará. Los resultados muestran que los agentes replican las respuestas de los participantes con una precisión cercana a la perfección. Es un cambio de paradigma en estudios a gran escala en sociología.

LA EXCEPCIÓN DE LOS EXCEPCIONALES

Publicación: «The Number of Exceptional People: Fewer Than 85 per 1 Million Across Key Traits», de Gilles E. Gignac.

Síntesis: Este estudio ofrece una perspectiva realista y fundamentada sobre la rareza de encontrar individuos que sobresalgan simultáneamente en inteligencia, concienciación y estabilidad emocional. Utilizando simulaciones estadísticas avanzadas, demuestra que la verdadera «excepcionalidad» es un fenómeno extremadamente raro, lo que tiene importantes implicaciones para la selección de personal y la gestión del talento.

| LIBROS |

POR QUÉ LAS MÁQUINAS APRENDEN

Why Machines Learn. The Elegant Maths Behind Modern AI, de Anil Ananthaswamy.

EL GOLPE DE ESTADO TECNOLÓGICO: CÓMO SALVAR LA DEMOCRACIA DE SILICON VALLEY

The Tech Coup: How to Save Democracy from Silicon Valley, de Marietje Schaake.

LA INCERTIDUMBRE COGNITIVA DISTORSIONA EL COMPORTAMIENTO ECONÓMICO

■ **Publicación:** «Behavioral Attenuation», documento de trabajo, National Bureau of Economic Research núm. 32973, septiembre de 2024.

■ **Benjamin Enke** es profesor de Economía en la Universidad de Harvard; **Thomas Graeber**, profesor de Administración de Empresas en la Escuela de Negocios de Harvard; **Ryan Oprea**, catedrático de Economía en la Universidad de California-Santa Bárbara, y **Jeffrey Yang**, estudiante de doctorado en la Escuela de Negocios de la Universidad de Harvard.

Resumen: *Este análisis abarca un gran número de experimentos relacionados con decisiones centrales sobre economía, como las de consumo, ahorro, trabajo o inversión. Se constatan algunas atenuadas respecto de lo que predicen los modelos económicos por la incapacidad de procesar información compleja. Estos resultados son generalizables a la mayoría de las decisiones económicas.*

Desde los pioneros trabajos de Herbert Simon y de Amos Tversky y Daniel Kahneman, sabemos que se producen importantes desviaciones respecto del comportamiento racional típico de los modelos de economía. Los humanos tenemos una racionalidad limitada (*bounded rationality*) y poseemos, además, varios sesgos psicológicos, como pueden ser nuestra aversión a las pérdidas y la escasa capacidad para ordenar nuestras preferencias de forma consistente.

Se han dado varias explicaciones para entender estos fenómenos. Tversky y Kahneman propusieron la existencia de dos sistemas de pensamiento, el rápido y el lento. En el caso de decisiones rápidas, se suelen tomar atajos heurísticos que a veces llevan a desviarse del comportamiento racional. En otros casos, se ha tenido en cuenta la heterogeneidad de las preferencias individuales o las anomalías de éstas. Un caso concreto es el de la tasa de descuento hiperbólica, que nos hace sobrevalorar el presente respecto al futuro en nuestras preferencias intertemporales. Si estas preferencias son dominantes, se puede incurrir en comportamientos que reducen nuestra tasa de ahorro, la inversión en capital humano o los hábitos saludables que garanticen una mejor salud futura.

Sin embargo, en este artículo, los autores plantean la hipótesis de la incertidumbre cognitiva, lo que generaliza a la gran mayoría de las decisiones individuales de índole económica. La incertidumbre cognitiva sería cuán segura está una persona de haber tomado la mejor decisión. En contextos muy saturados de información difícil de procesar, los individuos no tienen clara la decisión óptima, con lo cual suelen tomar una conservadora y poco comprometida; es decir, una respuesta debilitada o atenuada respecto a la decisión óptima. En la figura 1 podemos ver algunos ejemplos de decisiones atenuadas dados los cambios en los parámetros que la determinan. Uno es la relación entre el esfuerzo individual y el salario, y el segundo es un experimento con donaciones a una institución en que se cambia el parámetro de la efectividad de la institución. Como se puede observar en los dos gráficos, en contextos de mayor incertidumbre (reflejados por los puntos en rojo), la elasticidad de la respuesta a los cambios en el parámetro es más atenuada (la curva es más plana) que en el caso de las decisiones que se toman en contextos de menor incertidumbre

(puntos en azul). El resultado general es que, en la mayoría de los experimentos, la elasticidad de la respuesta ante cambios en la variable fundamental se acerca a cero en contextos de mayor complejidad informativa. Es decir, cuando una situación es más difícil de entender o requiere mucho procesamiento de información, las personas responden menos a los cambios; en contextos complejos, las decisiones casi no cambian, aunque deberían hacerlo.

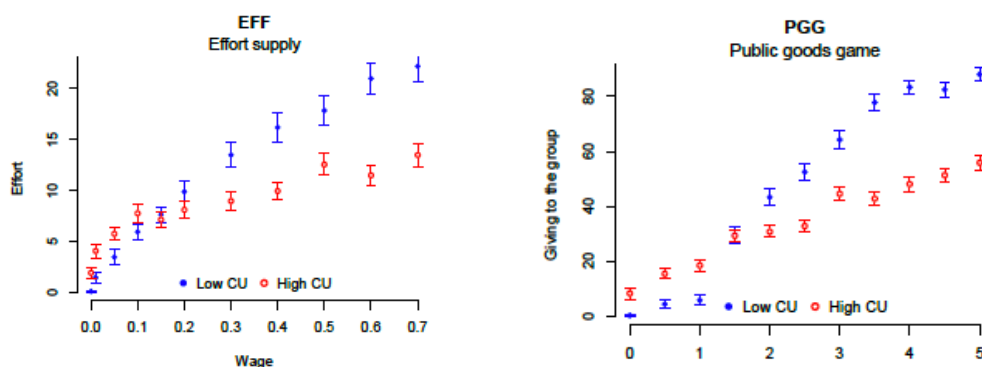


Figura 1. Evolución de varias decisiones individuales en contextos de baja (azul) y alta (rojo) incertidumbre cognitiva. El primer gráfico muestra la evolución del esfuerzo con relación al salario. Se observa una relación más atenuada a medida que aumenta el salario en situaciones de alta incertidumbre cognitiva (la curva se hace más plana). Lo mismo sucede con donaciones filantrópicas a una institución basándose en el nivel de eficiencia de ésta. En contextos cognitivos más saturados (puntos rojos), la respuesta a la mayor eficiencia de la institución receptora está atenuada (pendiente de la curva está más próxima a cero que en el caso de las observaciones en azul).

Una de las contribuciones del artículo es demostrar que la hipótesis de la complejidad informativa explica la aparición de comportamientos atenuados en un gran número de decisiones, desde las de consumo y ahorro a las que versan sobre filantropía, compra de loterías, valoración de políticas públicas o voto. Para probar este punto, los autores usan una gran variedad de experimentos relacionados con decisiones financieras (ahorro, distribución de activos en una cartera de inversión, predicción de retornos de un activo entre otras), decisiones de esfuerzo y participación laboral, de consumo, filantropía, toma de decisiones estratégicas o políticas, preferencias temporales y sobre el riesgo y, finalmente, con tareas cognitivas (predicción, actualización de opiniones, etc.). En total, se analizan y replican treinta experimentos. La incertidumbre cognitiva y la atenuación de las respuestas individuales afectan a las decisiones que tienen soluciones tanto objetivas (dados los parámetros de las variables que se usan para tomar la decisión) como subjetivas (las que dependen de preferencias o información desconocida para los investigadores que diseñan el experimento). Todos los experimentos arrojan un resultado generalizable: en casos de mayor incertidumbre, donde el procesamiento de la información es más difícil, la típica respuesta individual es mucho más atenuada que en contextos de baja incertidumbre. Por tanto, esta incertidumbre informativa y su efecto atenuante sobre las decisiones se entiende como una característica general en la toma de decisiones de los humanos.

Aparte de su indudable interés teórico para sentar las bases de un modelo más general del comportamiento económico individual, nos podemos preguntar si el diseño de políticas públicas debería tener en cuenta la atenuación de las decisiones individuales en contextos de complejidad informativa. Para los autores, esto es absolutamente necesario. Por ejemplo, en medicina, la elasticidad de las pruebas médicas al riesgo del paciente está atenuada. Los médicos suelen realizar demasiadas pruebas a pacientes de bajo riesgo y, en cambio, en el caso de los pacientes de alto riesgo, lo hacen con menor frecuencia de la deseable. En otro campo, la relación entre las percepciones subjetivas de tasas de retorno y la tenencia de activos tampoco obedece a las predicciones de los modelos financieros. Esta atenuación de la respuesta podría producirse tanto en las decisiones de venta de activos, cuando está empezando a deteriorarse el retorno de éstos, como con las de compra, cuando mejoran los retornos esperados. Dicha atenuación sería compatible con las grandes fluctuaciones en el precio de activos financieros o inmobiliarios. Quizás en paralelo al diseño adecuado de las políticas públicas, los legisladores tendrán que pensar en maneras de reducir la complejidad de la información que tienen que procesar los ciudadanos para favorecer que las políticas implementadas tengan la tracción y efectos deseados.

* * *

Reseña de **Jordi Domènech**, profesor de Historia Económica en la Universidad Carlos III de Madrid.

UN LABORATORIO VIRTUAL (¡SIN HUMANOS!) CASI PERFECTO PARA LA SOCIOLOGÍA

- **Publicación:** «Generative Agent Simulations of 1000 People», noviembre de 2024, disponible en el repositorio arXiv: <https://shorturl.at/G5D5p>
- **Joon Sung Park, Carrie Cai, Meredith Ringel Morris, Robb Willer, Percy Liang y Michael S. Bernstein** (Departamento de Ciencias de la Computación de la Universidad de Stanford), **Carolyn Q. Zou** (Northwestern University, aportando perspectivas metodológicas en la intersección entre la comunicación y la tecnología), **Aaron Shaw** (participa desde la Universidad de Washington) y **Benjamin Mako Hill** (Google DeepMind).

Resumen: *Los autores proponen una innovadora arquitectura de agentes generativos que simula las actitudes y comportamientos de 1052 individuos reales. Partiendo de entrevistas con ellos, construyen perfiles detallados sobre una persona que alimentan un modelo de lenguaje que la simulará. Los resultados muestran que los agentes replican las respuestas de los participantes con una precisión cercana a la perfección. Es un cambio de paradigma en estudios a gran escala en sociología.*

El avance de la inteligencia artificial (IA) ha abierto nuevas vías para explorar y comprender el comportamiento humano. En este contexto, surge la idea de construir agentes generativos (modelos de lenguaje capaces de generar un plan para resolver un problema y ejecutarlo paso a paso) que simulen las actitudes y repliquen las potenciales respuestas de individuos reales. Así, este artículo se centra en una innovadora metodología que utiliza entrevistas en profundidad realizadas a individuos para captar la riqueza de su experiencia personal. Estos datos se agregan y se utilizan para alimentar un modelo de lenguaje de última generación. El resultado es que se recrean, con sorprendente fidelidad, las respuestas de más de mil participantes.

A raíz de lo anterior, se propone crear un laboratorio virtual en el que cada agente represente a un individuo, lo que permite estudiar, de manera controlada y replicable, cómo responden diferentes segmentos de la población a variados estímulos sociales, políticos y económicos. Al integrar información cualitativa proveniente de entrevistas –donde se exploran desde trayectorias vitales hasta perspectivas sobre temas contemporáneos–, la arquitectura propuesta busca ir más allá de los enfoques tradicionales, basados únicamente en datos demográficos o descripciones simples de las personas.

Además, este enfoque ayuda a que los agentes no sólo sean réplicas estáticas de respuestas, sino que posean una «memoria» y capacidad de reflexión que los habilita para interactuar en contextos complejos. Esto se traduce en la posibilidad de replicar no sólo respuestas puntuales en encuestas, sino también comportamientos dinámicos y efectos de intervenciones experimentales. La metodología central de Park *et al.* en la recogida de datos consiste en utilizar entrevistas semiestructuradas para capturar la complejidad idiosincrática de cada individuo. A diferencia de los métodos tradicionales, las entrevistas permiten que cada participante exprese, con sus propias palabras, aspectos relevantes de su vida, sus experiencias y sus opiniones. Este proceso narrativo es fundamental para evitar el «aplanamiento» (simplificación excesiva) de las identidades en estereotipos y

para proporcionar a los agentes generativos la información suficiente para simular comportamientos y actitudes de manera precisa. Asimismo al ser testado en contextos experimentales y en simulaciones de encuestas, este mecanismo permite que el agente responda de forma coherente y fiel a la «personalidad» del participante.

Para evaluar la fidelidad de la simulación, se utilizan varios instrumentos clásicos experimentales:

- General Social Survey (GSS): un instrumento de 177 preguntas estructuradas que se usa para medir la precisión de los agentes mediante la comparación de las respuestas del modelo con las de los propios participantes en una segunda medición, obteniéndose una exactitud normalizada de aproximadamente del 85 % de coincidencia en las respuestas.
- Big Five Personality Inventory (BFI-44): los agentes predicen rasgos de personalidad en torno al 80 % de éxito, superando modelos basados en descripciones demográficas o de descripciones de la persona.
- Juegos económicos y experimentos sociales: la capacidad de los agentes para replicar efectos experimentales –como los encontrados en estudios sobre asignación de culpa o reacciones a la injusticia– es cercana al 98 % cuando se comparan con el original humano.

Los resultados previos muestran que la arquitectura no sólo capta la complejidad individual, sino que también es capaz de simular respuestas colectivas, lo que abre un nuevo paradigma para la experimentación virtual masiva en las ciencias sociales.

En definitiva, el trabajo de Park *et al.* sienta las bases para una «simulación integral» del comportamiento humano. Plantea que, a medida que se perfeccionen las técnicas de recogida y procesamiento de datos –por ejemplo, a través de mejoras en la transcripción automática y en la generación de reflexiones–, será posible construir laboratorios virtuales en los que se simulen respuestas a gran escala, y con esto será posible:

- modelar el comportamiento individual y colectivo en escenarios de política pública, crisis o innovación;
- explorar interacciones complejas entre diferentes subgrupos y evaluar el impacto de intervenciones específicas;
- facilitar la replicación experimental de estudios sociales con una precisión comparable a la obtenida en experimentos presenciales, a la vez que se minimizan los costos y se optimiza la privacidad.

Esta «hoja de ruta» sugiere que, en un horizonte relativamente cercano, podríamos disponer de herramientas de simulación robustas que transformen la investigación en las ciencias sociales, con una experimentación virtual sin precedentes.

Reflexiones finales

El artículo representa un avance impresionante e inesperado en el campo de la simulación del comportamiento humano. Al integrar entrevistas en profundidad con modelos de lenguaje de última generación, los autores ofrecen una herramienta capaz de replicar la complejidad y la diversidad de las actitudes y comportamientos humanos. Esta metodología no sólo mejora la precisión de las simulaciones, sino que –de nuevo sorprenden-

temente– también reduce sesgos que tradicionalmente han limitado la representatividad de los modelos basados en datos demográficos. En realidad, desde el punto de vista de la sociología y de los métodos cualitativos, incluyendo los etnográficos, la propuesta no es nueva. Todos los métodos comentados se han utilizado para análisis en multitud de investigaciones; la diferencia radica en la escala, pues los agentes agregan los *small data* en forma de *big data*. Estas pequeñas pistas nos pueden conducir a las grandes tendencias, como señalaba Martin Lindstrom en su libro *Small data: las pequeñas pistas que nos advierten de las grandes tendencias* (2016).

Los autores proponen una vuelta de tuerca en la sofisticación y escalación de las herramientas de análisis cualitativos, como historia de vida, entrevistas semiestructuradas..., pero miden su fidelidad con instrumentos cuantitativos y de experimentación en escenarios acotados que son limitados. Por ello, existe un riesgo de que se sobreestime lo que esta tecnología puede hacer a corto plazo, al no tener en cuenta variables y marcos más complejos y contextuales que la sociología sí incluye. Por tanto, un riesgo es considerar que esto llevaría a eliminar el papel de la sociología, ya de por sí bastante debilitada. Sin embargo, igualmente, si estas herramientas se utilizan para «bien», tienen un potencial para una mejor adaptación de soluciones a las necesidades de las sociedades.

Para un sociólogo computacional, esta aproximación abre un abanico de posibilidades a la hora de investigar fenómenos sociales complejos en un entorno controlado y replicable. La capacidad de simular agentes que actúen de manera coherente con sus correspondientes humanos posibilita ensayar intervenciones, predecir respuestas ante políticas públicas y comprender dinámicas colectivas con un grado de detalle antes inalcanzable. Desde el lado del *marketing*, o las predicciones electorales, este tipo de métodos pueden ser amplificados todavía más, a la vez que sus riesgos y beneficios. Por tanto, es fundamental seguir trabajando en aspectos como la ética, la seguridad y la protección de la privacidad, para que estas herramientas sean implementadas de manera responsable y eficiente y no se usen para «usurpar» la personalidad, o intentar manipular el comportamiento de los ciudadanos, simulándolos con precisión. El mal uso de estas herramientas podría replicar el impacto del caso de Cambridge Analytica a una escala aún mayor, afectando a múltiples escenarios. Por ello, su despliegue debe realizarse de manera responsable, con un marco ético y regulaciones claras que garanticen su uso adecuado, lo que implica una mayor participación de los gobiernos en su supervisión y control.

* * *

Reseña de **Manuel Cebrián Ramos**, científico titular del Centro de Automática Robótica, del Consejo Superior de Investigaciones Científicas (CSIC).

LA EXCEPCIÓN DE LOS EXCEPCIONALES

■ **Publicación:** «The Number of Exceptional People: Fewer Than 85 per 1 Million Across Key Traits», *Personality and Individual Differences*, 234, 112955 (2025). Artículo disponible en <https://shorturl.at/ejUrz>

■ **Gilles E. Gignac** es profesor de Psicología en la Universidad de Western Australia, especializado en psicología diferencial e inteligencia humana.

Resumen: *Este estudio ofrece una perspectiva realista y fundamentada sobre la rareza de encontrar individuos que sobresalgan simultáneamente en inteligencia, concienciación y estabilidad emocional. Utilizando simulaciones estadísticas avanzadas, demuestra que la verdadera «excepcionalidad» es un fenómeno extremadamente raro, lo que tiene importantes implicaciones para la selección de personal y la gestión del talento.*

En el mundo de la psicología diferencial y la gestión del talento, siempre se ha debatido sobre la existencia de «candidatos excepcionales», aquellos individuos que destacan en varias áreas clave. Muchas veces, reclutadores y profesionales de recursos humanos tienden a imaginar que estos «unicornios» son más comunes de lo que realmente son. Gignac se plantea una pregunta crucial: ¿cuál es la verdadera prevalencia de individuos que logran, simultáneamente, puntuaciones muy altas en varias dimensiones fundamentales?

Para investigar este tema, estudia cómo la inteligencia, la concienciación (*conscientiousness* en inglés) y la estabilidad emocional son simultáneamente prevalentes en una población. El autor no escoge estos rasgos al azar: son fundamentales por diversas razones. La inteligencia es ampliamente reconocida como un predictor crucial de la capacidad de aprendizaje, solución de problemas y adaptabilidad en contextos complejos, mientras que la concienciación se asocia con la organización, la responsabilidad y la persistencia en el logro de objetivos. Por su parte, la estabilidad emocional permite hacer frente al estrés y las adversidades, y facilita la toma de decisiones y la adaptación a diversos entornos dinámicos.

La estrategia metodológica consiste en un enfoque basado en simulaciones estadísticas. Se generó una población simulada de veinte millones de casos en la que cada individuo fue evaluado en las tres dimensiones de estudio: inteligencia, concienciación y estabilidad emocional. Cada una de estas variables se estandarizó luego, estableciendo una media de 0 y una desviación estándar de 1, lo que permite una comparación uniforme entre los distintos rasgos. Además, el estudio muestra las correlaciones específicas entre las variables: se observa una relación casi nula entre inteligencia y concienciación ($r = -0.03$), una correlación muy leve entre inteligencia y estabilidad emocional ($r = 0.07$) y una correlación moderada entre concienciación y estabilidad emocional ($r = 0.42$). Esto refleja la realidad de que los rasgos no actúan de manera independiente en la vida cotidiana, y es precisamente esta interrelación la que puede afectar significativamente a la probabilidad de que un individuo se destaque simultáneamente en todas las áreas.

El autor clasifica a la población en diversas categorías según sus puntuaciones. Se considera «notable» a quien tiene una puntuación igual o superior a la media en todas las dimensiones, mientras que el nivel «destacable» corresponde a aquellos que se ubican

una desviación estándar por encima de la media. La categoría de «excepcional» se reserva para quienes logran dos desviaciones estándar simultáneamente en todas las dimensiones, y la de «profundamente excepcional» a aquellos que superan las tres desviaciones estándar. Los resultados obtenidos son sorprendentes: alrededor del 16 % de la población se clasifica como notable, menos del 1 % alcanza el nivel destacable, y únicamente el 0,0085 % –aproximadamente ochenta y cinco personas por cada millón– cumple con los criterios para ser considerado excepcional. La categoría de profundamente excepcional es prácticamente inalcanzable, ya que en la simulación sólo se identificó un caso en veinte millones.

Las implicaciones prácticas de estos hallazgos son numerosas y tienen un impacto directo en ámbitos como el reclutamiento y la gestión del talento. En muchos sectores, especialmente en algunos como el tecnológico o las finanzas, se suele buscar a candidatos que se ajusten a un perfil casi perfecto o excepcional en múltiples dimensiones. Sin embargo, los resultados del estudio indican que exigir que un candidato se sitúe dos desviaciones estándar por encima de la media en inteligencia, concienciación y estabilidad emocional es una expectativa poco realista. Son perfiles extremadamente raros. Además, esto no implica que los candidatos que no alcancen esos niveles sean inadecuados, sino que es necesario adoptar un enfoque más equilibrado en el que se valore también a aquellos individuos que, estando moderadamente por encima del promedio, pueden aportar habilidades valiosas y un desempeño sobresaliente en sus funciones. Al reconocer que la verdadera excepcionalidad es estadísticamente infrecuente, las organizaciones pueden ajustar sus procesos de selección para favorecer una evaluación más holística del talento, considerando el conjunto de competencias y la capacidad de adaptarse a entornos complejos, en lugar de fijarse únicamente en puntuaciones extremas.

Comentario

Aunque el estudio presenta un sólido enfoque metodológico, es importante señalar algunas limitaciones. La simulación se basa en la suposición de que los rasgos se distribuyen de forma normal y en la utilización de parámetros fijos para cada variable. Si bien esta aproximación es razonable y común en la literatura, puede no captar por completo la complejidad de las interacciones en poblaciones reales. Al fin y al cabo, definir la excepcionalidad como alcanzar dos desviaciones estándar por encima de la media es una convención que, aunque sea ampliamente aceptada, puede resultar arbitraria en algunos contextos, y no siempre es el mejor reflejo de la valía en situaciones concretas. Además, estudios previos han señalado que la relación entre ciertos rasgos y el rendimiento puede no ser lineal, lo que sugiere que niveles moderadamente altos podrían ser, en algunos casos, incluso más beneficiosos que niveles extremadamente altos.

En este sentido, el trabajo de Gignac se inserta en un debate más amplio dentro de la psicología diferencial y la gestión del talento. Investigaciones previas sobre la contribución conjunta de la inteligencia y la personalidad al rendimiento laboral han enfatizado que el éxito de un individuo depende de un equilibrio adecuado de sus rasgos clave. La evidencia empírica de este estudio refuerza la idea de que la búsqueda del «candidato perfecto» basado en estándares excesivamente altos es poco práctica y podría conducir a la exclusión de perfiles valiosos. Al mismo tiempo, estos hallazgos invitan a repensar los criterios de selección y a considerar estrategias que permitan identificar y desarrollar

el potencial de aquellos individuos que, aunque no sean estadísticamente «excepcionales», poseen cualidades que pueden marcar la diferencia en un entorno real.

Con esta reflexión, queda claro que la comprensión de la verdadera magnitud de la excepcionalidad puede ayudar tanto a académicos como a profesionales a desarrollar estrategias más realistas y efectivas para identificar y potenciar el talento. La aplicación de modelos de simulación y análisis estadístico avanzado se convierte en una herramienta clave para trascender expectativas idealistas y ajustar los criterios de selección a la complejidad inherente de la naturaleza humana. En definitiva, este estudio nos enseña que, más allá de buscar unicornios, debemos reconocer y valorar la riqueza que ofrece la diversidad de competencias en la construcción de equipos y en el desarrollo personal y profesional.

* * *

Reseña de **Alberto Palacios Abad**, doctor en Ciencias Sociales por la Universidad Carlos III de Madrid

POR QUÉ LAS MÁQUINAS APRENDEN

Anil Ananthaswamy, *Why Machines Learn. The Elegant Maths Behind Modern AI* («Por qué las máquinas aprenden: Las matemáticas elegantes detrás de la inteligencia artificial moderna»), Allen Lane, 2024, 464 págs.

Por **Javier Antonio Sánchez Pita**

Hoy en día, la inteligencia artificial (IA) ha alcanzado la madurez necesaria para hacernos la vida más fácil. Es capaz de analizar datos, sacar conclusiones y ayudarnos a realizar tareas de manera mucho más eficaz. Sin embargo, de igual forma que sabemos que un niño aprende, pero no entendemos lo que pasa dentro de su cerebro, tampoco sabemos cómo es el proceso de aprendizaje de las máquinas de inteligencia artificial. Este libro trata de explicar los algoritmos y las matemáticas que se esconden detrás de estos sistemas. Por tanto, resulta imprescindible tener una mínima base de conocimientos matemáticos y estadísticos para comprender el texto.

Para responder a la pregunta «¿Cómo aprenden las máquinas?», deberíamos quedarnos con la idea de que los sistemas de IA, al igual que los seres humanos, aprenden mejor y más profundamente cuanto más repasan los datos. Al igual que nosotros, siempre podrán mejorar y estarán sujetos a posibles errores. Es decir, no deberíamos confiar ciegamente en la IA, ya que los resultados obtenidos pueden estar sujetos a errores o sesgos. Estos suelen deberse a problemas de entendimiento entre el usuario y la máquina o a que los datos empleados para entrenar el sistema no son completos o están sesgados.

Ababtgaswamy levanta el misterioso velo de la IA para descubrirnos las matemáticas que son el artífice de su funcionamiento, desde el perceptrón de Rosenblatt, invención sobrevalorada en su época pero que sentó las bases de la IA moderna, hasta los avances contemporáneos como el aprendizaje profundo y los modelos de lenguaje extensos como ChatGPT.

Antes de entrar en conceptos matemáticos, que el libro aborda en sucesivos capítulos de menor a mayor complejidad, es útil repasar algunas definiciones clave que faciliten la comprensión de algunos conceptos de la inteligencia artificial. Según Ababtgaswamy, el aprendizaje automático (*ML-machine learning*) «consiste en enseñar a las máquinas a identificar patrones en los datos y usar lo aprendido (es decir, el modelo de los datos) para hacer predicciones con datos nuevos o incluso generar datos completamente nuevos». Para ello, son necesarios grandes volúmenes de datos. Una parte se utiliza para entrenar a la máquina, mientras que el resto se reserva para evaluar su rendimiento antes de su implementación.

El aprendizaje automático se basa en fundamentos matemáticos sólidos que incluyen álgebra lineal, cálculo, probabilidad, estadística y teoría de la optimización, y se nutre de contribuciones históricas de gigantes matemáticos como Isaac Newton, Thomas Bayes, Carl Friedrich Gauss, así como de figuras contemporáneas vinculadas a la IA como Ilya Sutskever o Geoffrey Hinton, entre otros. Cabe destacar que el prólogo sólo menciona a una matemática, Eugenia Cheng, autora del libro *Is Math real?*, de la cual enfatiza su reflexión sobre el aprendizaje gradual de las matemáticas, comparándolo con el desafío de escalar una montaña. Esta analogía no sólo refleja el proceso de aprendizaje descrito en el libro, sino también la experiencia de leerlo, o quizá de leer esta reseña.

Dada su complejidad, proponemos una forma de lectura alineada con la base matemática del libro, avanzando de menos a más. Esto permite que cada lector aborde el contenido según su nivel de asimilación, decidiendo hasta dónde profundizar. La estructura incluye descripciones generales, anécdotas o ejemplos ilustrativos, además de explicaciones matemáticas, siendo estas últimas de mayor complejidad.

Breve historia y bases matemáticas iniciales de la IA

A mediados del siglo pasado, existían dos aproximaciones distintas para definir cómo debería funcionar una máquina de inteligencia artificial. Algunos científicos creían en el modelo de programación, en el que todos los escenarios e instrucciones debían de ser programados en la máquina. Otros apostaban por sistemas que aprendieran a partir de los datos, de manera semejante a como aprende el cerebro humano. Con los años, ha sido esta segunda aproximación la que ha permitido llegar al nivel de desarrollo que la IA tiene hoy en día. Algo que no era obvio en aquel entonces.

La primera referencia a una neurona matemática aparece en un artículo científico de 1943. La neurona se definía como una función con dos valores de entrada, que podían ser 0 o 1, y que según su combinación daban un resultado diferente a su salida. La complejidad aumentaba añadiendo cuantos datos de entrada fueran necesarios, cada uno de ellos con un peso diferente.

Por otro lado, la primera red neuronal, basada en el principio anterior, fue propuesta en 1957 y es conocida como el perceptrón de Rosenblatt. Este científico proponía que los valores de las entradas de la neurona para obtener un valor determinado a la salida podían obtenerse mediante un procedimiento de prueba y error. Este innovador modelo tenía sus limitaciones. Y fueron precisamente estas limitaciones las que usaron los científicos Minsky y Papert, en un libro publicado en 1969, en favor del modelo de programación, desacreditando las posibilidades del perceptrón e insinuando, sin demostrarlo, que una red neuronal más compleja se vería también sujeta a las mismas limitaciones.

No fue hasta 1991 cuando los investigadores de los Bell Labs Bernhard Boser y Vladimir Vapnik, aplicando las antiguas matemáticas de Lagrange, así como otros trabajos de científicos rusos de los años sesenta, demostraron que se podían superar las limitaciones del perceptrón y desarrollaron la máquina de vectores de soporte (*support vector machine*), que mejoraba la eficiencia del proceso de aprendizaje de las redes neuronales.

A pesar de que hubo que esperar hasta los años noventa para demostrar cómo superar las limitaciones del perceptrón, los ochenta fueron bastante fructíferos. Por un lado, Hopfield desarrolló un modelo que demostraba cómo se podía almacenar información en las neuronas, y, a la vez, que era posible su entrenamiento. Esto sirvió para que, en 1986, David Rumelhart, Geoffrey Hinton y Ronald Williams presentaran el algoritmo de propagación hacia atrás. Éste explica que una red neuronal, con distintas capas de neuronas conectadas entre sí, puede ser entrenada a partir de una base de datos con un proceso de ajuste continuo basado en el principio de prueba y error. Demostraron también que, añadiendo distintas capas de neuronas entre la entrada y la salida, se pueden resolver problemas hasta entonces irresolubles.

Tan relevante como el algoritmo de propagación hacia atrás fue el teorema de la aproximación universal publicado por Cybenko en 1989, que probaba que una red neuronal con sólo una capa oculta (entre la entrada y la salida) podía asimilarse a cualquier función. O lo que es lo mismo: cualquier caso de aprendizaje podría ser resuelto con una red neuronal.

Si bien el artículo de Minsky y Papert supuso un jarro de agua fría para la IA tal y como la conocemos hoy, retrasando su avance, también tuvo su lado bueno, puesto que permi-

tió abrir otras líneas de investigación que condujeron a la invención de distintos algoritmos alternativos, hoy usados en muchos casos de aprendizaje, como son el algoritmo de cuadro mínimo, la estimación de máxima probabilidad, la aproximación de máximo *a posteriori* o el algoritmo del vecino más cercano.

En los años ochenta, cuando se comenzaban a aplicar redes neuronales a la identificación de imágenes, el reto estaba en que, cuando la misma imagen era percibida desde distintos puntos, fuera identificada como dos imágenes distintas. Este problema se solucionó mediante las redes convolucionales propuestas por LeCun a finales de esa década.

Aun así, las redes convolucionales tardaron en imponerse a otros métodos, porque LeCun había hecho sus desarrollos trabajando en AT&T y, por tanto, estaban sujetas a propiedad intelectual. En consecuencia, si bien se podían ver los resultados, se desconocía cómo se conseguían. Otras vías de investigación confiaban en llegar a resultados similares. En la última década, sin embargo, resultó evidente que ningún otro sistema era capaz de alcanzar los niveles de exactitud de las redes convolucionales.

Recientemente, la aparición de los chips GPU ha permitido hacer redes convolucionales mucho más eficientes reduciendo las necesidades de neuronas y capacidad de procesamiento para obtener los mismos resultados. La Universidad de Toronto y, en particular, Hinton y otros investigadores que han trabajado con él, están detrás de los principales avances en este sentido.

Conceptos matemáticos clave

Esta sección de la reseña ahonda, desde un punto matemático, en cada uno de los algoritmos de los que hablábamos en el apartado anterior. Es un contenido valioso para los que deseen profundizar, pero puede resultar complicado matemáticamente hablando. El lector puede decidir también pasar al siguiente apartado sin perder el hilo principal. De forma resumida, el perceptrón y las redes neuronales son herramientas que imitan el cerebro humano para aprender y tomar decisiones. Aunque comienzan con reglas simples, se vuelven más poderosas al agregar capas y métodos matemáticos avanzados, lo que permite resolver problemas desde clasificar datos hasta analizar imágenes, entre otros.

Bases del perceptrón

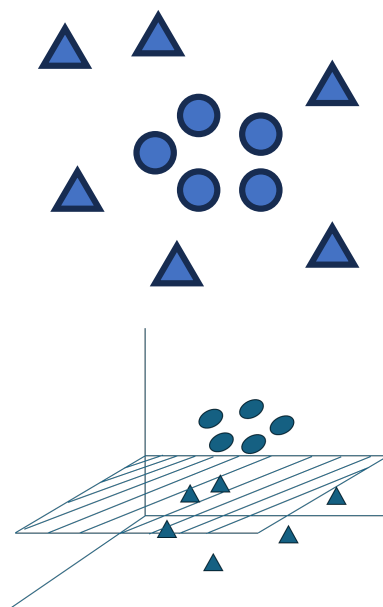
El perceptrón se define como una «neurona» artificial que toma datos de entrada (x_1, x_2, \dots, x_n), cada uno de ellos con un peso distinto ($w_1, w_2 \dots w_n$), y cuya salida se calcula con una función matemática.



Conocido un número de datos (combinaciones de valores de x_1, x_2, \dots, x_n y de y_1, y_2, \dots, y_n), podemos, mediante prueba y error, averiguar el peso w_1, w_2, \dots, w_n de cada entrada.

Se utiliza para clasificar datos. La precisión del modelo depende de éstos. Por ejemplo, si tenemos sólo dos datos, existen múltiples hiperplanos que dividen el espacio de forma que un dato esté en un lado y otro en el otro. Cuantos más datos tengamos, mayor será la exactitud y menor la posibilidad de error. El perceptrón, no obstante, presentaba el reto de que no siempre se puede encontrar un hiperplano; es decir, no se pueden separar los datos con una sola línea. Como se puede apreciar en el gráfico arriba a la derecha, resulta imposible encontrar una recta que deje los dos círculos a un lado y los dos triángulos al otro. El sistema puede mejorar en la resolución de este problema mediante la creación de distintas capas de neuronas unidas unas a otras. Ésta es la limitación a la que hicieron referencia Minsky y Papert en su artículo de 1969.

La solución propuesta en 1991 por Bernhard Boser y Vladimir Vapnik consistía en proyectar en una dimensión superior los datos que no podían ser separados. El siguiente ejemplo muestra círculos y triángulos que no pueden ser separados en dos dimensiones, pero que sí podemos separar cuando proyectamos en una tercera dimensión (nótese que, si miramos la segunda figura desde arriba en sólo dos dimensiones, tendríamos la primera):

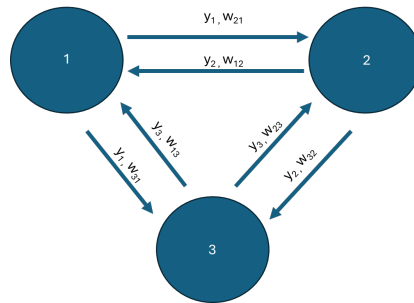


Al no saber en qué dimensiones se puede hacer la separación, necesitamos una ingente capacidad de procesamiento para hacer miles de proyecciones. Ésta se reduce mediante el uso de unas funciones llamadas kernel, perfeccionadas a finales del siglo xx en las conocidas máquinas de soporte vertical o SVM (*support vector machine*, por sus siglas en inglés), ofreciendo un método alternativo a las redes neuronales.

Modelo de Hopfield, redes neuronales y el teorema de la aproximación universal

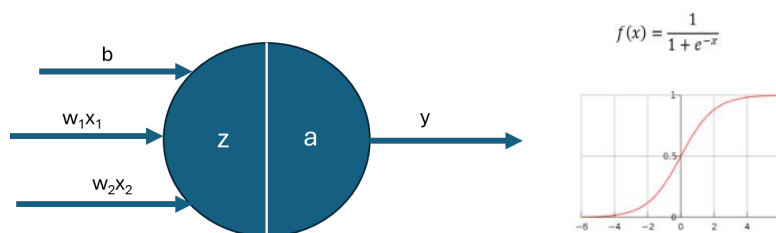
A finales de los setenta, John Hopfield definió la mente como una máquina capaz de cambiar la configuración de distintas neuronas mediante cambios dinámicos en su estado hasta alcanzar la estabilidad. Su modelo parte de tres neuronas que se alimentan entre sí de forma simétrica ($w_{ij} = w_{ji}$). Progresivamente, los pesos iban ajustándose hasta un estado estable que se correspondía con el de menor energía del sistema. El sistema ajusta los pesos de forma dinámica hasta que dejan de cambiar, alcanzando un estado estable que actúa como una «memoria» del sistema. Hopfield verificó que para n neuronas, la capacidad

de almacenaje es de $0,14 \times n$ memorias; estudios posteriores han mejorado esta capacidad en las conocidas como «redes mejoradas de Hopfield».



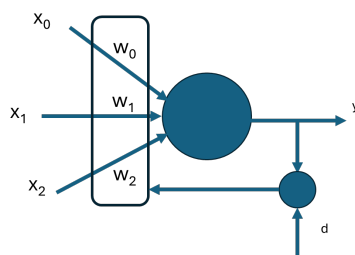
El modelo de Hopfield demostró que se podía almacenar información (recordar información) en las neuronas y cómo se puede entrenar un sistema de más de una neurona. Esto abrió la puerta a futuras investigaciones que llevaron al descubrimiento del algoritmo de propagación hacia atrás. Dicho algoritmo organiza las neuronas en capas, donde las salidas de una capa sirven como entrada de las siguientes, y va equiparando los pesos entre neuronas para mejorar los resultados. Para un conjunto de valores de entrada en la primera capa, el sistema calcula una salida, la compara con el valor esperado y mide la diferencia (la conocida función pérdida). Luego ajusta mínimamente los pesos, repitiendo el proceso y comparando si hay mejora, en cuyo caso repite el proceso hasta llegar a un valor aceptable de la función pérdida, o, en su defecto, vuelve a los valores anteriores para realizar un ajuste diferente de los pesos. Usa funciones, llamadas de activación, que tienen una forma sigmoide para facilitar los cálculos. Al desplazarnos hacia atrás en la cadena de neuronas, nos movemos por las mismas funciones usadas en el desplazamiento hacia adelante, con lo que se reducen las necesidades de capacidad de cálculo. Una de las conclusiones más importantes es que, además, los resultados de las salidas de las neuronas ocultas (las capas intermedias) captan propiedades relevantes en el análisis que se está realizando. Esto permite que el algoritmo mediante el añadido de distintas capas entre la entrada y la salida pueda resolver problemas que no tienen una solución lineal.

Finalmente, recordamos que el teorema de la aproximación universal de Cybenko constató que una red neuronal con una capa oculta puede aproximar cualquier función matemática. Cybenko explicó que la neurona realiza su cálculo en dos pasos: una función estándar ($z = w_1x_1 + w_2x_2 + b$) y una función de activación ($y = a(z)$). La función de activación $y = a(z)$ usada por Cybenko tenía forma de sigmoide (abajo, figura derecha). Modificando el valor del sesgo b se puede desplazar la curva a derecha o izquierda y modificando los pesos se puede hacer que la transición de activación (de apagado a encendido) sea más abrupta. La salida del sistema es la suma lineal de las distintas funciones de las neuronas activadas o no de acuerdo con los pesos y los sesgos aplicados.



Algoritmo de cuadrado mínimo medio

Propuesto en 1959 por Widrow y Hoff, este algoritmo tiene su origen en las investigaciones sobre filtros adaptativos que se utilizan en telecomunicaciones para entrenar a los sistemas en identificar señales separándolas de las interferencias. Estos filtros calculaban la raíz cuadrada del error cometido en un conjunto de medidas, y , mediante derivadas y múltiples interacciones, buscaban minimizar el error cometido acercándolo a cero. El ejemplo del siguiente gráfico muestra la aplicación a una neurona.



Estimación de máxima probabilidad y aproximación de máximo a posteriori

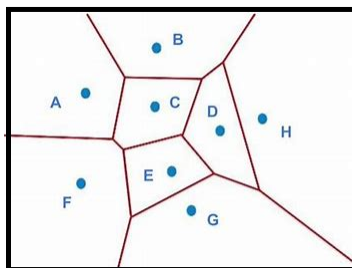
Cada conjunto de datos que inyectamos en una neurona es un suceso y , y por tanto, se puede definir una función de distribución que describa cómo están organizados estos datos. No obstante, no siempre es posible conocer cuál es esa función. Por eso, se usan dos métodos diferentes para intentar identificarla que usan probabilidades para modelar los datos.

El primero, el método de estimación de máxima probabilidad (*maximum likelihood estimation* o *MLE*) busca, para unos datos conocidos, los parámetros de una distribución que maximicen la probabilidad de esos datos. De entre las distintas distribuciones posibles, el algoritmo encuentra aquella que maximiza la probabilidad de observar los datos que tenemos. Este método asume que conocemos el tipo de distribución, lo que no siempre es verdad.

El segundo método, conocido como *maximum a posteriori* (*MAP*), se apoya en el teorema de Bayes. En él, el algoritmo considera los parámetros que caracterizan la distribución como una variable. Por ejemplo, si hablamos del suceso tirar una moneda al aire, podría asumir que ésta está equilibrada o que tiene sesgo hacia una u otra cara. Con esa suposición, el algoritmo encuentra la distribución más probable para explicar cualquier conjunto de datos X que tengamos.

Algoritmo del vecino más cercano (NNA)

Este algoritmo aborda el reto de la clasificación de datos desde una perspectiva distinta, con la ventaja de no necesitar conocer de antemano cómo están distribuidos los datos; es decir, la distribución de probabilidad. Conocidos los datos, divide el espacio en zonas en torno al punto más cercano a cada uno de ellos. Para entenderlo mejor, imaginemos que en una ciudad dan servicio varias oficinas de correos. El algoritmo asignará cada dirección postal a la oficina que esté más cerca, creando un diagrama como el de la siguiente figura:



En este gráfico, conocido como diagrama de Voronoi, cada zona recoge todas las direcciones que están más cerca de la oficina de correos que se encuentra en su interior. Apliquemos este algoritmo, por ejemplo, al caso en el que queramos identificar un número. Cada número, como una imagen, está caracterizado por 63 píxeles. Podemos analizar cualquier nueva imagen como una muestra de 63 píxeles, e identificar a qué imagen conocida se parece más para clasificarlo correctamente. Esto nos permite realizar una clasificación no lineal superando las limitaciones del perceptrón. La comparación se hace con varios datos cercanos, para evitar el problema de que alguno de los datos acabe erróneamente etiquetado.

Redes convolucionales

Este modelo se basa en dos niveles de neuronas. El primero recibe distintas entradas, equivalentes a distintos puntos de vista de la imagen, y en base a ellos detecta o no una característica. Si alguna de las neuronas la detecta, ésta es procesada por el segundo nivel para identificar otros patrones. Este proceso se enfrenta al reto de que es necesario identificar las distintas funciones matemáticas que identifican cada característica, lo que se resuelve mediante un proceso que aplica redes neuronales para facilitar el aprendizaje automático de estas funciones; es decir, aprender automáticamente cuáles serían las funciones necesarias. El proceso de aplicar distintas funciones de manera sucesiva a los distintos píxeles se le llama «convolución». Para simplificar el sistema y reducir las necesidades de procesamiento, se aplica una función llamada de *pooling*, que permitía reducir distintos píxeles a uno solo más relevante.

Retos actuales y el futuro de la IA

Los últimos diez años han visto aparecer múltiples variedades de las redes convolucionales con modelos en los que la propagación se da en distintos sentidos y con distintos niveles y neuronas en cada nivel. La elección de los parámetros adecuados para construir la red se ha convertido en un arte que en muchos casos depende del problema que se ha de resolver. Uno de los grandes desarrollos en este sentido han sido las redes neuronales autoentrenadas, en las que los datos usados en el entrenamiento no están etiquetados. Es el propio sistema el que se encarga de realizar la detección de parámetros, etiquetar y dar el resultado más probable. Estos modelos son los usados por la inteligencia artificial detrás del famoso ChatGPT.

El estado del arte presenta, no obstante, diversos retos. El primero de ellos radica en que el comportamiento de las redes neuronales no se corresponde con lo esperado, y, al desconocer qué ocurre en los miles de neuronas de los niveles intermedios, resulta imposible encontrar la razón. Al entrenar redes neuronales se suelen dividir los datos en dos: una parte para el entrenamiento y otra para las pruebas. Si se simplifica mucho el entrenamiento, se cometen errores que luego aparecen en la fase de pruebas, y, si sobreentrena-

mos, el sistema acaba detectando patrones incorrectos que dan resultados erróneos. Pues bien, se ha demostrado que, aun introduciendo hasta un 5% de casos erróneos o mal etiquetados en el paquete de entrenamiento, el sistema es capaz de funcionar correctamente con los datos de prueba. La máquina es capaz de regularizar los errores y estabilizar su eficiencia. El segundo problema es el sesgo. Las redes neuronales dependen de la calidad de los datos introducidos; si éstos están sesgados, el resultado acaba volcando también estos sesgos. Teniendo en cuenta la opacidad de las redes, cada vez resulta más difícil corregir estos problemas, aun cuando se detecte el sesgo en virtud del resultado.

Finalmente, los sistemas de aprendizaje automático como ChatGPT están demostrando su capacidad para proponer resultados complejos que se asemejan al razonamiento humano. La comunidad científica se encuentra inmersa en el debate sobre si realmente se está produciendo un cierto nivel de razonamiento o simplemente los sistemas están proporcionando información a partir de la correlación de ingentes cantidades de datos.

Para añadir más leña al fuego, DeepSeek, modelo de IA desarrollado en China, se ha hecho viral a principios de 2025. Utiliza aprendizaje por refuerzo puro, aprendiendo mediante prueba y error sin depender de datos etiquetados. Esto marca un hito en el campo de la IA, ya que antes se consideraba insuficiente para entrenar modelos avanzados. DeepSeek combina aprendizaje por refuerzo y redes neuronales en un enfoque híbrido que le permite superar las limitaciones del aprendizaje supervisado, ofreciendo soluciones más flexibles y eficientes para problemas complejos. Además, otra distinción es que mantiene el enfoque abierto frente al cerrado de Open AI. Estamos ante una explosión de modelos y algoritmos en la IA. Pero, a pesar de haber enfoques diferentes, sin duda, las matemáticas siempre jugarán un papel esencial.

* * *

Anil Ananthaswamy es escritor de literatura científica, anteriormente colaborador de la revista *New Scientist*. Es el autor de varios libros de ciencia populares, como *The Man Who Wasn't There*, que ha sido nominado para el premio de literatura científica E. O Wilson de la sociedad PEN América.

Reseña de **Javier Antonio Sánchez Pita**, director de Preventa de Servicios en Nokia, máster y licenciado en Ingeniería de Telecomunicaciones, máster en Dirección de Empresas y postgrado en dirección de proyectos de IA. Trabaja en proyectos de transformación digital con el uso de la inteligencia artificial.

EL GOLPE DE ESTADO TECNOLÓGICO: CÓMO SALVAR LA DEMOCRACIA DE SILICON VALLEY

Marietje Schaake, *The Tech Coup: How to Save Democracy from Silicon Valley* («El golpe de Estado tecnológico: cómo salvar la democracia de Silicon Valley»), Princeton University Press, 2024, 336 págs.

Por **Paula Oliver Llorente**

El libro de Marietje Schaake es particularmente pertinente dado el contexto internacional en el que nos encontramos. Aborda el poder desmedido de las Big Tech, las grandes empresas tecnológicas, cuyos productos, en muchos casos, incorporan características antidemocráticas, y que han asumido funciones gubernamentales sin obligaciones legales ni mecanismos de rendición de cuentas. En el proceso de ascenso de las Big Tech, los gobiernos democráticos, y especialmente las sucesivas administraciones estadounidenses, han escogido no regularlas, permitiendo su crecimiento desmedido y la maximización de beneficios. Esto ha llevado a una erosión acelerada de la democracia; una dinámica en la que la responsabilidad y el poder se alejan de los líderes elegidos en las urnas, dando lugar a un creciente autoritarismo digital y al declive de la gobernanza.

En sus diez capítulos, la autora introduce al lector a los principales actores involucrados, los aspectos físicos de la tecnología y su interrelación con la geopolítica mundial, las vulnerabilidades del vacío de gobernanza digital, ejemplos de la externalización de funciones democráticas a soluciones tecnológicas, las estrategias corporativas de las Big Tech para perpetuar su poder, los principales modelos de gobernanza digital que existen y, por último, propuestas concretas para poner fin a lo que llama el «golpe de Estado tecnológico».

Actores implicados

En cuanto a los actores, los pioneros de la revolución digital tienen un papel fundamental en la ideología y comportamiento posterior de las Big Tech. Ellos eran escépticos del Estado central, por lo que creían profundamente en la capacidad democratizadora del internet descentralizado, que ofrecía libertad y transparencia. Estos principios se extendieron con el tiempo a la industria digital en su conjunto. Sin embargo, también con el tiempo, dicha ideología resultó ser errónea al enfrentarse a la dinámica de «el ganador se lo lleva todo» que permitían los efectos de red de las redes sociales. Aun así, las empresas se valieron de esos principios para articular una visión que limitase el papel del gobierno en el mundo digital: la regulación impide la innovación. De esta manera, se ignora deliberadamente el papel que jugó la inversión pública, por ejemplo, en la creación de internet a través de la agencia DARPA (Agencia de Proyectos de Investigación Avanzados de Defensa).

La clase política, por su parte, abrazó muchos de esos preceptos y dejó a las grandes tecnológicas promover su visión de un internet desregulado. En primer lugar, mantuvieron una amplia separación entre el gobierno del espacio digital y el gobierno de Estados Unidos. Desde el punto de vista regulatorio, la primera ley, y la más influyente, que gober-

nó internet en Estados Unidos se aprobó en 1996: la Communications Decency Act. Esta ley y su artículo 230 se han utilizado para eximir a las plataformas digitales de responsabilidad sobre el contenido publicado y distribuido en ellas. Así, se las deja operar ampliamente. Esta concepción de separación también se refleja en la operativa y gestión técnica de internet, que se encuentra alejada de actores elegidos democráticamente. La Corporación de Internet para la Asignación de Nombres y Números (ICANN, por sus siglas en inglés), una entidad independiente conformada por actores privados, es quien se encarga de ello.

En segundo lugar, las sucesivas administraciones estadounidenses permitieron que las tecnológicas crecieran y se volviesen imprescindibles, lo cual ha hecho mucho más difícil ponerles freno *a posteriori*. Durante la Administración Bush, los atentados del 11-S y la lucha contra el terrorismo justificaron el uso de la tecnología para una vigilancia extraordinaria sobre la población civil, y eso permitió a las tecnológicas poner los cimientos de su dominio futuro. Con Obama, las Big Tech consolidaron su presencia e influencia en los mercados. Fue en este momento cuando la política exterior tecnológica comenzó a emerger, a través de la Estrategia Internacional para el Ciberespacio de 2011. También fue cuando el peligro de la tecnología y su uso se destapó: en 2013, Edward Snowden reveló la injerencia estatal del gobierno estadounidense sobre ciudadanos y gobiernos extranjeros a través de la tecnología, con el pretexto de la seguridad nacional. En 2016, Apple negó al FBI acceso al teléfono de uno de los sospechosos en un tiroteo masivo en California.

La corriente comenzó a cambiar durante el primer mandato de Donald Trump. Su gobierno protagonizó enfrentamientos directos contra las tecnológicas, especialmente las chinas Huawei y TikTok, y evidenció la carrera tecnológica contra China con los primeros controles a las exportaciones de semiconductores. Por su parte, su sucesor, Biden, trató de revocar sin éxito el artículo 230 de la Communications Decency Act y nombró a importantes expertos antimonopolio en posiciones clave. Aun así, pese a una retórica creciente de confrontación, el continuismo ha caracterizado las relaciones de las administraciones de los últimos treinta años con las tecnológicas: *laissez-faire*. Como resultado, las Big Tech tienen poder sobre los derechos y libertades de la población. El establecimiento de normas del mundo digital se realiza a través de su código, modelos de negocio y estándares técnicos. Algunos gobiernos democráticos han intentado sobreponerse a esta hegemonía, pero la creciente asimetría de voluntad y capacidad (de poder, información e influencia) entre los gobiernos y el sector tecnológico dan la ventaja a las empresas.

La pila (*stack*) tecnológica

En este contexto jurídico y político, los aspectos físicos de la tecnología digital cobran importancia. La ausencia de un gobierno público y democrático sobre ellos los expone a vulnerabilidades geopolíticas y abusos corporativos. Hay tres elementos clave que condicionan la tecnología de hoy en día: los microchips (o semiconductores), presentes en muchos de los bienes que consumimos en nuestro día a día y cuya disponibilidad depende de cadenas de valor muy complejas; los cables de fibra óptica, vulnerables a ataques y interrupciones derivados de las tensiones geopolíticas; y los centros de datos, imprescindibles para el almacenaje de la inmensa cantidad de datos generados a diario, pero con importante impacto medioambiental. En los tres casos, los gobiernos democráticos (tanto nacionales como regionales y locales) están a merced de empresas privadas que pueden sufrir presiones por parte de terceros países. Por ejemplo, el 90 % de la producción de los chips más sofisticados los fabrica la empresa taiwanesa TSMC, lo cual hace que su disponibilidad sea vulnerable a la relación entre Taiwán y China. En el caso de los cables submarinos, éstos son construidos y operados por empresas privadas pese a ser objetivos de guerra, lo que genera dudas sobre las responsabilidades de seguridad. Finalmente, las decisiones sobre si

aprobar o no la construcción de centros de datos, que requieren de un alto consumo de energía y agua para operar, se dirimen a nivel local. Son ayuntamientos pequeños los que toman decisiones con impactos importantes sobre los recursos locales. Lo hacen en situación de inferioridad frente a la experiencia, conocimiento y tácticas de las Big Tech, que esconden su identidad como empresa promotora.

Vacío de gobernanza digital y externalización de funciones democráticas

El vacío de poder político que deja la ausencia de poderes gubernamentales en las configuraciones anteriores, junto con la digitalización completa de nuestras vidas, aumenta exponencialmente los riesgos en materia de ciberseguridad. Los ciberataques, típicamente perpetrados por motivos financieros, están siendo asociados cada vez más a razones políticas. Las instituciones más vulnerables son las menos preparadas: hospitales, ayuntamientos, instituciones educativas, etc.; todas ellas con gran cantidad de datos sensibles a su disposición. Además, las administraciones son dependientes de empresas externas para su ciberseguridad. Esta dinámica conlleva varios problemas: por un lado, la naturaleza privada del *software* impide que la sociedad civil evalúe los riesgos de los sistemas que implementan las administraciones públicas; por otro, la externalización de la ciberseguridad hace que las administraciones carezcan de conocimientos internos; y, por último, la falta de ordenamiento jurídico al respecto, incluido a nivel internacional, dificulta la investigación, atribución y judicialización de los ataques. De esta manera, los gobiernos democráticos están ausentes en su propia ciberseguridad. Se necesita que éstos sean los que den ordenes concisas y supervisen las ciberoperaciones, con cadenas de responsabilidad y rendición de cuentas claras.

A las funciones asumidas por las tecnológicas se añade la privatización de unas tareas que anteriormente se presumían del Estado. Es el caso en tres tecnologías: las criptomonedas, que tratan de sustituir las monedas fiduciarias pero han demostrado ser un activo especulativo sin valor real; las tecnologías de reconocimiento facial, que permiten la policía predictiva sin reparos a los sesgos discriminatorios; y los productos de integración de datos, que analizan grandes cantidades de datos para guiar decisiones públicas sin revelar ni responder por la lógica de sus algoritmos, con peligros importantes para la toma de decisiones de seguridad nacional. La privatización de éstas y otras funciones gubernamentales se hace de manera discreta e invisible para el público, de forma que se entronca la tecnología en el funcionamiento de la administración. También se construye de manera piramidal, con unos productos y servicios sobre otros, impidiendo la identificación y subsanación ágil de errores y sesgos. Todo esto, además, sin el consentimiento de los ciudadanos, lo cual vacía de legitimidad las funciones de gobierno. Es aquí donde la autora incluye una de las líneas más reveladoras del libro: «La privatización de la gobernanza en favor de las empresas tecnológicas es no-democrática porque no es, en sí misma, democrática; y es antidemocrática porque menoscaba la legitimidad de los gobiernos democráticos».

Estrategias de las Big Tech para perpetuar su poder

Durante estos años, las empresas tecnológicas han conseguido conducir el debate sobre internet por los cauces que más les convenía. Para ello, han desarrollado narrativas convenientes, firmado acuerdos inocuos y protagonizado intensas actividades de *lobbying* en las capitales relevantes. Las narrativas han conseguido determinar y permear el debate sobre la regulación tecnológica. Las tres principales han sido la equiparación de internet con unas pocas compañías (en otras palabras, las Big Tech), el mantra «la regulación impide la innovación» y la asociación de las propuestas de regulación tecnológica en Estados

democráticos con acciones de gobiernos autoritarios. Mientras tanto, han creado y firmado innumerables iniciativas de autorregulación, incluyendo algunas con gobiernos, sin que ninguna tenga la solidez e instrumentos para ser ejecutada firmemente como lo haría una ley. A esto se han añadido esfuerzos de *lobbying* que han permitido eliminar propuestas regulatorias, ganar contratos públicos y mejorar su imagen.

Modelos de gobernanza digital

Ante la falta de un marco de gobernanza internacional, diferentes jurisdicciones están explorando sus modelos digitales preferidos. La autora distingue cuatro: el modelo europeo, el estadounidense, el chino y el indio.

El de la Unión Europea (UE) destaca por centrarse en los derechos de los usuarios. Sin tecnología propia potente, la UE depende de Estados Unidos y China en la sociedad digital, por lo que ha adoptado un papel de «superregulador». Ha impuesto obligaciones a las empresas en materia de protección de datos, moderación de contenido, competencia e interoperabilidad, y ha establecido sanciones proporcionales a la facturación de la compañía infractora.

Al otro lado del Atlántico, ninguno de los partidos estadounidenses tiene la mayoría relevante ni la voluntad para aprobar una legislación digital de calado. El debate político en cuanto a la tecnología se centra en consideraciones de seguridad nacional, ya que el liderazgo tecnológico chino se percibe como una amenaza directa y casi existencial al sistema estadounidense. La polarización del poder legislativo imposibilita ir más allá y aprobar normas que frenen el poder de sus propias tecnológicas. Así, las agencias estatales han adquirido protagonismo, entre otras cosas, por su papel en la protección de los consumidores y la aplicación de la ley antimonopolio. Es probable que esto cambie con la nueva Administración Trump.

En China, la tecnología es un instrumento del poder político. Tiene un papel clave en el sistema de vigilancia estatal y censura que el gobierno lleva a cabo sobre su población y que exporta a otros países del mundo a través de la Ruta de la Seda Digital. La competición geopolítica y la desconfianza de la dependencia china han llevado a algunos, incluyendo a Estados Unidos y, hasta cierto punto, a la UE, a tratar de desacoplar su sector tecnológico del chino.

Finalmente, la India. Es un país considerado en pleno retroceso democrático y que tiene un modelo que se acerca poco a poco al chino. El gobierno de la India ha ordenado el mayor número de apagones de internet del mundo, algunos de ellos por motivos políticos, como evitar protestas civiles. Narendra Modi ha aprobado leyes para facilitar la censura *on-line* y ha adoptado un sistema de identificación biométrico que es casi imprescindible para el acceso a las políticas de bienestar del país. Pese a esto, la India representa para muchos países emergentes y en desarrollo un nuevo e inspirador paradigma como opción alternativa a los otros tres actores. Mientras tanto, los países democráticos ven a la India como un aliado imprescindible para actuar como contrapeso al modelo digital autoritario chino. Aun así, a pesar de las diferencias entre modelos que la autora identifica, cabe destacar que las Big Tech estadounidenses juegan un papel importante en todos ellos, incluso en los que facilitan la represión obedeciendo órdenes gubernamentales y eliminando aplicaciones sensibles de sus tiendas de aplicaciones.

Frenar el «golpe de estado tecnológico»

Ante los retos que presenta el poder de las grandes tecnológicas para la democracia, Schaake finaliza el libro con una amplia batería de medidas. Su finalidad, según la autora,

es priorizar el interés público, eliminar las dependencias irresponsables y frenar el «golpe de estado tecnológico» que se está produciendo. Incluyo todas ellas en los párrafos a continuación, ya que, en el contexto actual, es importante conocer cuáles son las opciones que tenemos a nuestro alcance:

1. **Ampliar el uso del principio de cautela** establecido en el artículo 191 del Tratado de Funcionamiento de la Unión Europea para el medioambiente. Schaake propone aplicarlo también a las innovaciones tecnológicas en las que no haya suficiente investigación y evidencia de sus efectos adversos. Un comité de expertos independientes podría evaluar qué tecnologías deberían someterse a este principio y qué compañías deberían informar confidencialmente de los productos que planean sacar al mercado sin la suficiente base científica. Se debería permitir también a académicos e investigadores acceso a las tecnologías para llevar a cabo las evaluaciones.
2. **Frenar el uso de cuatro tecnologías** que han probado ser dañinas: los *softwares* de espionaje, que conllevan graves riesgos de seguridad nacional; los intermediarios (*brokers*) de datos, que recogen, compran y venden datos, y representan una amenaza especialmente preocupante para las comunidades vulnerables; los sistemas de reconocimiento facial, que verifican identidades y recogen datos biométricos; y las criptomonedas, cuyos casos positivos de uso no se han manifestado y han provocado estafas, especulación y endeudamiento.
3. **Incrementar la transparencia** en el ámbito tecnológico a través de la identificación clara del contenido sintético (IA) con respecto al auténtico y la transparencia en las fuentes de financiación y empresas detrás de las inversiones, licitaciones de terreno para centros de datos, así como los impactos medioambientales del desarrollo y uso de nuevas tecnologías. También se propone extender las obligaciones de rendición de cuentas desde los gobiernos a aquellas empresas que ejecuten tareas por ellos. Para esto último, se necesita actualizar las normas de protección de secretos comerciales.
4. **Dar ejemplo desde el sector público**, utilizando el poder de los gobiernos en los países democráticos como los principales clientes de *software*. Para ello, deben pedir responsabilidades cuando los proveedores no cumplan con sus compromisos e incluir requisitos de transparencia y especificaciones en contratos y subvenciones. El sector público también puede realizar la contratación colectiva de *software* para instituciones más pequeñas, como pueden ser hospitales y universidades, y registrar y sancionar públicamente los incumplimientos. Desde el punto de vista del conocimiento, el sector público también debería crear servicios de expertos tecnológicos que puedan volcar su conocimiento en el proceso legislativo para elaborar regulaciones mejor informadas y menos sensibles al *lobby* corporativo. Esto último debería servir también para posicionar a las instituciones públicas como empleadores competitivos para el talento tecnológico.
5. **Desarrollar sistemas y mecanismos de rendición de cuentas**, identificando a aquellas instituciones tecnológicas que son importantes a nivel sistémico (similar al *too big to fail* bancario) e imponiendo requisitos que impliquen mayor supervisión regulatoria, como auditorías a los algoritmos de recomendación y a las medidas de seguridad. En relación con esto último, se propone la creación de un tribunal de arbitraje para la ciberseguridad, parecido al mecanismo de resolución de disputas entre Estado e inversor que se aplica en los conflictos comerciales. Para ello, se necesitan definiciones y estándares para la defensa del espacio público internacional, la actualización del derecho internacional y la mejora de los acuerdos entre países. Por último, se recomienda que las fuerzas democráticas se unan en una coalición global. Esta propuesta podría materializarse a través de la reforma del Foro de Gobernanza de Internet de Naciones Unidas, un foro multiactor que se reúne anualmente, y que podría cubrir con su mandato tanto la política de internet como de IA. La Declara-

ción del Futuro de Internet, firmada por más de sesenta gobiernos, podría convertirse en el documento fundacional de esta coalición democrática. Lo más importante, en cualquier caso, es que la institución que se cree tenga poderes reales de aplicación de los acuerdos y de sanción de las infracciones.

6. Por último, Schaake destaca **revitalizar la esfera pública digital**, para resolver la excesiva dependencia de las instituciones públicas en empresas privadas. Es necesario entonces desarrollar un *stack* (conjunto) de tecnología pública, incorporando en su diseño valores y el interés público. Es una oportunidad para redefinir los bienes comunes digitales (*digital commons*) y crear infraestructura pública digital. Se debería poner el foco en la aplicación efectiva de las normas, y proporcionar a los reguladores los recursos para llevar a cabo esa función. Y, finalmente, dar más conocimiento a la ciudadanía, que debe saber más sobre tecnología y participar en este tipo de debates.

La autora concluye advirtiendo del poder disruptivo de tecnologías que no han recibido aún suficiente atención pero que pueden alterar la experiencia humana de manera irreversible. Por ello, la prevención y la gobernanza proactiva de la tecnología es importante. Estamos ante un problema sistémico, y reconocerlo es la única manera de llevar a cabo un cambio real.

* * *

Marietje Schaake es directora de Política Internacional en el Stanford University Cyber Policy Center y asociada de Política Internacional del Institute for Human-Centered Artificial Intelligence de Stanford. Entre 2009 y 2019, fue diputada en el Parlamento Europeo por los Países Bajos. Escribe una columna mensual para el *Financial Times* sobre tecnología y gobernanza.

Paula Oliver Llorente es ayudante de investigación para la Unión Europea en el Real Instituto Elcano. Es máster del Colegio de Europa. Anteriormente, ha trabajado en la Comisión Europea y en el proyecto Digital Revolution and the New Social Contract del Center for the Governance of Change de IE University.

ODLI. N.º 144 MARZO 2025

IDEAS DE INTERÉS

1. ECOSISTEMAS COLABORATIVOS ENTRE HUMANOS E INTELIGENCIAS ARTIFICIALES

- Autores: Karl J. Friston *et al.*
- Comentario: Manuel Cebrián Ramos.

2. IMPACTO MUY POLARIZADOR DEL CAMBIO TECNOLÓGICO SOBRE LAS PROFESIONES

- Autores: David Autor *et al.*
- Comentario: Jordi Domènech.

3. ¿LOS IMPUESTOS ECOLÓGICOS FRENAN LA DEGRADACIÓN MEDIOAMBIENTAL Y EL CAMBIO CLIMÁTICO?

- Autor: Sinan Erdogan.
- Comentario: Jorge Díaz Lanchas.

LIBROS

- ¿EL FINAL DEL MILAGRO ECONÓMICO ALEMÁN?
Reseña de Javier Asensio.
- TRIBAL: LOS INSTINTOS CULTURALES NOS DIVIDEN, PERO TAMBIÉN PUEDEN UNIRNOS
Reseña de Ricardo Dudda.

ODLI. N.º 143 FEBRERO 2025

IDEAS DE INTERÉS

1. EL TRILEMA DE LA SOSTENIBILIDAD.

- Autores: Dani Rodrik *et al.*
- Comentario: Ángel Pascual Ramsay.

2. LA DIMENSIÓN ÉTICA DE UNA ACTIVIDAD CADA VEZ MÁS GLOBALIZADA.

- Autores: Chidiogo Uzoamaka Akpuokwe, Seun Solomon Bakare, Nkechi Emmanuella Eneh y Adekunle Oyeyemi Adeniyi.
- Comentario: Jorge Díaz Lanchas.

3. NUEVO PARADIGMA: LA COMPUTACIÓN VERDE.

- Autores: Shuo Yang, Yue Wang y Yixing Zhang.
- Comentario: Gloria Álvarez Hernández.

4. PROMESAS DE LA INTELIGENCIA ARTIFICIAL: NO SIN LAS MUJERES.

- Autores: Judy Wajcman y Erin Young.
- Comentario: Cecilia Castaño.

LIBROS

- IA: LA AMBICIÓN DE FRANCIA.
- *I.A. Notre ambition pour la France*, Comisión de la Inteligencia Artificial, presidida por Philippe Aghion y Anne Bouverot.
Reseña de Regina H. de Benoist.

ODLI. N.º 142 ENERO 2025

IDEAS DE INTERÉS

1. LA POLÍTICA EXTERIOR DE EE UU Y EL NUEVO ORDEN GEOPOLÍTICO. DIÁLOGO ENTRE JOHN MEARSHEIMER Y JEFFREY SACHS.

- Autores: John Mearsheimer y Jeffrey Sachs.
- Comentario: Ángel Pascual Ramsey.
- 2. CALENTAMIENTO GLOBAL: TRAERÁ MÁS CONFLICTOS Y VIOLENCIA.**
- Autores: Marshall Burke, Joel Ferguson, Solomon M. Hsiang y Edward Miguel.
- Comentario: Jordi Domènech.
- 3. DESACUERDO CONSTRUCTIVO: NUEVAS PERSPECTIVAS PARA CONVERSACIONES DIFÍCILES.**
- Autores: Julia Minson y Erica Chenoweth.
- Comentario: Jaime Moreno.

LIBROS

- *Determined. The Science of Life Without Free Will*, de Robert Sapolsky.
Reseña de Francesc Trillas
- *Free Agents: How Evolution Gave Us Free Will*, de Kevin Mitchell.
Reseña de Ignacio Berberana

ODLI. N.º 141 DICIEMBRE 2024

IDEAS DE INTERÉS

1. EL POPULISMO, FENÓMENO MUTANTE E INCOMPREN-SIBLE.

- Autores: Alan de Bromhead y Kevin H. O'Rourke.
- Comentario: Jordi Domènech.

2. 27 SALIDAS DE LA GRAN DEPRESIÓN.

- Autores: Martin Ellison, Sank Seok Lee y Kevin H. O'Rourke.
- Comentario: Jordi Domènech.

3. CÓMO REGULAR LAS ENERGÍAS TRANSFORMADORAS.

- Autores: Daron Acemoglu y Todd Lensman.
- Comentario: Gloria Álvarez Hernández.

4. AUTOMATIZACIÓN BASADA EN IA DENTRO DE UNA COMPAÑÍA LOGÍSTICA: MÁS TRABAJO, PRODUCTIVIDAD Y RENTABILIDAD.

- Autores: Erdem Dogukan Yilmaz y Christian Peukert.
- Comentario: José Balsa Barreiro.

LIBROS

- *Una generación ansiosa criada por teléfonos inteligentes, redes sociales y familias helicóptero* de Jonathan Haidt.
Reseña de Arturo Lahera Sánchez.

