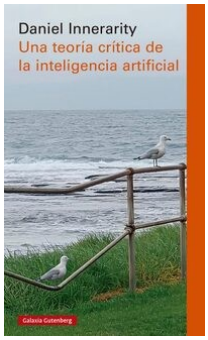

Innerarity, Daniel (2025).
Una teoría crítica de la inteligencia artificial. Galaxia Gutenberg.
ISBN: 978–84–10317–18–5

Eguzki Urteaga

Universidad del Pais Vasco; eguzki.urteaga@ehu.eus; 0000–0002–8789–7580

Daniel Innerarity acaba de publicar su último libro, titulado *Una teoría crítica de la inteligencia artificial,* en la editorial Galaxia Gutenberg. Conviene recordar que el autor es catedrático de filosofía política y social, investigador senior en Ikerbasque y director del Instituto de Gobernanza Democrática. Asimismo, es titular de la Cátedra Inteligencia Artificial y Democracia del Instituto Universitario Europeo de Florencia. Ha sido profesor invitado en varias universidades, tales como la Universidad de la Sorbona, la London School of Economics, la Universidad de Georgetown o el Max Planck Institute de Heidelberg. Ha recibido varios galardones, como pueden ser el Premio de Humanidades, Artes, Cultura y Ciencias Sociales de Eusko Ikaskuntza-Caja Laboral en 2008, el Premio Príncipe de Viana en 2013 y el Premio Nacional de Investigación en Humanidades en 2022. Entre sus obras más recientes figuran *La sociedad del desconocimiento* (2022), *La libertad democrática* (2023) y *Una teoría crítica de la inteligencia artificial* (2025).

En la presente obra, consagrada a la inteligencia artificial, el autor recuerda que «el actual fenómeno de la gobernanza algorítmica forma parte de una tendencia más amplia hacia la matematización y la mecanización de la gobernanza que viene de antiguo» (p. 19). De hecho, la formación del Estado moderno está directamente relacionada con la estadística, la probabilidad y los datos (Porter, 1986; Hacking, 1990; Derosières, 1998). Asimismo, el fenómeno de los *big data*, que alude a macrodatos, se sitúa en la larga historia de la estadística social. Por lo cual,

> la actual algoritmización de la sociedad podría entenderse como continuidad con el cálculo moderno, con sus estadísticas y sistemas de lógica formal. La organización de la moderna administración se enfrentó a la contingencia del mundo numerizando y formalizando el caos de la realidad. (pp. 19–20)

Urteaga, Eguzki (2026). Reseña de Innerarity (2025) Una teoría crítica de la inteligencia artificial. *Athenea Digital*, *26*(1), e3948. https: //doi.org/10.5565/rev/athenea.3948

Fecha de publicación: 08-01-2026

Ante un mundo cada vez más incierto, «el enfoque probabilístico le ofrece la posibilidad de transformar la contingencia en calculabilidad formalizada» (p. 20).

En su búsqueda de eficacia, de cuantificación y de neutralidad, alejada de los prejuicios y errores humanos, «la inteligencia artificial parece llamada a ser la lógica de legitimización de las organizaciones y de los gobiernos en las sociedades digitales» (p. 20), sabiendo que las sociedades contemporáneas están marcadas por «sistemas cada vez más inteligentes, una tecnología más integrada y una sociedad más cuantificada» (p. 20). En semejante contexto, las cuestiones que se plantean son las siguientes: «¿Nuestras vidas deben estar regidas por procedimientos algorítmicos y en qué medida? ¿Cómo articular los beneficios de la robotización, automatización y digitalización con aquellos principios de autogobierno que constituyen el núcleo normativo de la organización burocrática de las sociedades?» (pp. 20–21). En ese sentido, la configuración de la gobernanza de estas tecnologías resultará crucial para el futuro de la democracia.

El hecho de que exista cierta continuidad entre las formas preliminares de gestión política de la complejidad y de la contingencia, no significa que «la era digital no represente una dimensión cualitativamente diferente e incluso cierta ruptura respecto de la clásica racionalidad burocrática» (p. 21), nos dice Innerarity. Más allá de su capacidad a tratar una cantidad ingente de datos, la principal diferencia estriba en el hecho de que las tecnologías digitales registran los datos «de forma binaria, archivable, enlazable y combinable» (Baecker, 2018). Esto confiere a los *big data* un gran poder organizativo, económico y comercial. Con la complejidad creciente de las sociedades y el incremento del número de tareas que exigen una actividad de cálculo, se produce una algoritmización de las decisiones políticos que va en aumento, lo que lleva «a una lógica de remplazo (humanos que renuncian a decidir) y automatización (máquinas capaces de decidir por sí mismas)» (p. 22).

En otras palabras, «el problema fundamental de la inteligencia artificial es la creciente externalización de decisiones humanas en ella» (p. 23). Ante este problema, según Innerarity, existen tres respuestas posibles: «la moratoria, la ética y la crítica política, es decir, la propuesta de que la tecnología sea detenida, al menos por un tiempo, sometida a códigos éticos o examinada de acuerdo con una perspectiva de crítica política. Cada una de ellas presupone un tipo diferente de relación entre los humanos y la tecnología» (p. 24), con sus ventajas y limitaciones.

El autor opta por la tercera vía, a pesar de sus insuficiencias, en la medida en que, «en esencia, una interrogación casi nunca es plenamente satisfecha sobre los supuestos que tendemos a dar por acreditados» (p. 33). Fundamentalmente, la visión crítica consiste en «facilitar una revelación, que muestre una dimensión de la realidad que no es manifiesta» (p. 34). A su vez, aspira a dar cuenta de la complejidad de la racionalidad algorítmica. De hecho, nos dice el autor, «la algoritmización requiere pensar muchas categorías socioculturales, como sujeto, acción, responsabilidad, conocimiento o trabajo» (p. 35). En ese sentido, su objetivo es «desarrollar una teoría de la decisión democrática en un entorno mediado por la inteligencia artificial, elaborar una teoría crítica de la razón automática y algorítmica» (pp. 35–36). En otros términos, propone una filosofía política de la inteligencia artificial.

El libro se divide en 14 capítulos repartidos en 3 partes.

La primera parte, titulada *Teoría de la razón algorítmica*, empieza por un primer capítulo dedicado a la inteligencia de la inteligencia artificial (pp. 41–94). El autor subraya que cualquier artefacto que realiza ciertas actividades humanas suscita cierto temor, a pesar de que la inteligencia humana y la inteligencia artificial sean más diferentes y complementarias que competitivas. No en vano, la inteligencia artificial generativa agudiza los temores de sustitución. En realidad,

> la actual encrucijada de la inteligencia artificial pasa por examinar hasta qué punto es inteligente, qué tipo de inteligencia tiene, cómo se relaciona con la inteligencia humana y, en consecuencia, a qué tipo de reconceptualización de nuestra inteligencia nos están obligando sus espectaculares desarrollos. (p. 42)

El segundo capítulo, que se interesa por el arte, haciendo referencia al sueño de la máquina creativa (pp. 95–109), subraya el hecho de que la inteligencia artificial es capaz de proceder a «la composición musical, la modelación creativa de procesos visuales, las series televisivas, el diseño arquitectónico o la escritura de historias. Son, propiamente, obras producidas por la inteligencia artificial» (p. 95), aunque los programas hayan sido configurados por humanos. Esto refuerza el temor de ser sustituidos por la IA, incluso para realizar tareas creativas. Ello conduce el autor a preguntarse sobre la naturaleza y los límites de la inteligencia humana y de la inteligencia artificial, así como sobre la obra de arte en la época de la inteligencia artificial generativa.

El tercer capítulo, que se titula *La sociedad de los big data* (pp. 111–160), pone de manifiesto que cualquier organización

> necesita datos y su abundancia en la era digital va a tener enormes repercusiones en su manera de gobernarse. Se trata de una tecnología que no solo va a modificar la eficiencia en la provisión de servicios públicos o la precisión de la planificación estratégica, sino también las relaciones entre la ciudadanía y el poder público, así como entre los políticos y el sistema administrativo. (p. 111)

Si el análisis de datos posibilita atender necesidades de los sectores públicos y privados, entraña el riesgo de que se refuercen los sesgos ideológicos, los prejuicios y los estereotipos. En ese sentido, «la actual euforia entre el potencial de análisis de datos debería madurar hacia una comprensión equilibrada de sus fortalezas y limitaciones. Unas y otras requieren una reflexión sobre la naturaleza de los datos y su estatuto epistemológico» (p. 112).

El cuarto capítulo, que concluye esta primera parte, procede a una crítica de la analítica predictiva (pp. 161–183). El autor recuerda que «una de las promesas más importantes del análisis de datos es la capacidad de adelantarse al futuro: dispara la expectativa de que todo puede ser calculable, incluido cualquier incierto futuro» (p. 161). De esa forma, se pretende tener una actitud más proactiva frente a los riesgos y a las amenazas a los que nos enfrentamos en un mundo cada vez más incierto (Urteaga, 2023). La cuestión que se plantea es si las

actuales tecnologías son capaces de integrar en sus modelos las discontinuidades y los cambios incesantes. Desde un punto de vista conceptual y crítico, Innerarity desea responder a tres preguntas:

> ¿Por qué los pronósticos aciertan demasiado? ¿Por qué, simultáneamente, fallan tanto?¿Qué consecuencias tienen el hecho de que nuestros instrumentos de predicción ignoren, al menos, [ciertas] realidades que son necesarias para acertar en los pronósticos, o, como mínimo, para ser conscientes de sus límites? (p. 163)

La segunda parte de la obra, titulada *Pragmáticas de la razón algorítmica*, debuta con el quinto capítulo dedicado a la infraestructura tecnológica de la sociedad digital (pp. 187–219), sabiendo que «hay una estrecha relación entre los modos de comunicación y los tipos de democracia» (p. 187). En la actualidad, la democracia está

> estrechamente vinculada al crecimiento de las sociedades saturadas de multimedia, cuyas estructuras de poder son continuamente cuestionadas por una multitud de mecanismos de control o vigilancia que operan dentro de una nueva galaxia mediática definida por el *ethos* de la abundancia comunicativa» (Keane, 2013, p. 79)

En ese sentido, indica el autor, es preciso «repensar la democracia en la era digital desde el cruce de la teoría democrática, la filosofía de la técnica y las ciencias de la comunicación» (p. 187). Una vez analizada la relación entre tecnología y actividad humana, propone adentrarse en dos cuestiones clave para entender la dimensión política de la digitalización: «el condicionamiento algorítmico de la vida política y la política que hacen los artefactos de este nuevo ecosistema» (p. 188).

El sexto capítulo, dedicado a la automatización (pp. 221–249), permite poner de manifiesto el hecho de que las sociedades contemporáneas «están delegando cada vez más decisiones en los sistemas inteligentes. El desarrollo tecnológico se basa en la creciente autonomía de los sistemas de decisión autonomizada» (p. 221), a los que los humanos confían decisiones que tomaban ellos mismos hasta la fecha. Esto se produce en la producción de conocimiento, la creación de confianza en la activación de los intercambios en los mercados automatizados, en la ejecución de la ley con contratos inteligentes, en la negociación automática, en la consultoría patrimonial o en la cirugía robótica. «La tecnología es tan omnipresente que apenas nos damos cuenta de ello» (p. 222). Esto conduce a una deshumanización paulatina de la decisión.

El séptimo capítulo, que se centra en las máquinas y el nuevo contrato social tecnológico (pp. 251–283), Innerarity incide en la necesidad de establecer un nuevo contrato social sobre la relación que mantienen los humanos con las máquinas en la era de la inteligencia artificial poblada de robots, algoritmos y sistemas automáticos de decisión. Esto implica determinar en qué los humanos se parecen y se distinguen de las máquinas. Solo así es posible «pensar y configurar un ecosistema humano-máquinas en el que los límites de la humanidad quedarán inevitablemente redefinidos» (p. 252).

El octavo capítulo se centra en la transparencia y en el grado de opacidad que soporta la inteligencia artificial (pp. 285–326). El autor constata que «vivimos en una sociedad llena de cajas negras para nosotros; mecanismos, sistemas, algoritmos, robots, códigos, automatismos y dispositivos que usamos y nos afectan, pero cuyo funcionamiento nos es desconocido, parcial o totalmente» (p. 285). Por lo tanto, la prevalencia creciente de los algoritmos hace que aumente «la necesidad de equilibrar las asimetrías cognitivas que de ello resultan», lo que implica una mayor transparencia (Balkin, 2016; Benjamin, 2013; Cohen, 2016; Mehra, 2015), aunque no sea la panacea para solucionar todos los problemas éticos planteados por las nuevas tecnologías (Mittelstadt et al. 2016; Neyland, 2016; Crawford, 2016). En cualquier caso, se antoja ineludible «construir toda una nueva arquitectura de justificación y control, donde las decisiones automáticas puedan ser examinadas y sometidas a revisión crítica» (p. 286).

La tercera y última parte, que se titula *Filosofía política de la razón algorítmica*, se inicia con el noveno capítulo dedicado a la cuestión del control en referencia a las máquinas, las instituciones y la democracia (pp. 329–349). De hecho, en un mundo contemporáneo cada vez más complejo, nos vemos obligados, de manera creciente, a confiar en las máquinas y los algoritmos, y, como consecuencia, «esa confianza se ve decepcionada y se rompe, hasta el punto de desatar el movimiento contrario» (p. 329). En semejante contexto, el autor propone «explorar nuestra voluntad de control para explicar algunos rasgos centrales del paisaje ideológico en el que nos encontramos y hasta qué punto puede esto puede afectar al futuro de la democracia» (p. 329). Más precisamente, explora tres aspectos: 1) la propensión creciente a delegar buena parte de las decisiones a artefactos; 2) la aspiración a controlar esa delegación, así como sus límites y consecuencias indeseadas; y, 3) se pregunta «¿hasta qué punto y bajo qué condiciones podemos entender la delegación como control?» (p. 331).

El décimo capítulo, consagrado a la gobernanza y a la gestión de las expectativas políticas generadas por la inteligencia artificial (pp. 351–379), pone de manifiesto que «el problema de la relación entre gobernanza algorítmica y democracia representa una continuidad y, al mismo tiempo, una ruptura con las clásicas formas de administración burocrática» (p. 351), lo que implica detenerse en las expectativas que genera la IA, «básicamente, la de proporcionar más objetividad a las decisiones políticas y la de adoptarlas con una mayor consideración de nuestra subjetividad como ciudadanos destinatarios de estas decisiones» (p. 351). Los límites de estas promesas conducen el autor a concluir en «la inevitabilidad de la decisión humana, de la política en cualquier entorno tecnológico, también el configurado por las nuevas formas de gobernanza algorítmica» (p. 351).

El decimoprimer capítulo, centrado en la democracia de las recomendaciones (pp. 381–407), recuerda que la tecnología digital ha sido presentada simultáneamente como una amenaza y una oportunidad para la democracia. Así, los sistemas de recomendación «parecen responder al formato democrático, en la medida en que sugieren en vez de imponer y prometen satisfacer nuestras preferencias en lugar de prescribirlas» (p. 381). No en vano,

> ¿hasta qué punto se trata de nuestras preferencias y si la democracia puede considerarse como un sistema que satisface preferencias o, más bien, ofrece un marco intersubjetivo en el que es posible la ponderación reflexiva e in-

> cluso el descubrimiento de nuestras verdaderas preferencias, más allá de lo que se nos recomienda atendiendo, únicamente, a nuestro comportamiento pasado? (p. 381)

El decimosegundo capítulo, que aborda la cuestión de la justicia en alusión a la igualdad algorítmica y la democracia deliberativa (pp. 409–428), se plantea la siguiente pregunta: ¿si la democracia consiste en posibilitar que todas las personas tengan similares oportunidades de influir en las decisiones que les afectan, las sociedades digitales tienen que interrogarse por el modo de conseguir que los nuevos entornos hagan factible esa igualdad? Lo cierto es que la respuesta a esta pregunta no consiste en recurrir a una técnica agregativa, sino que requiere compromisos políticos, de modo que una concepción deliberativa de la democracia parezca «la más apta para conseguir esa igualdad a la que aspiran las sociedades democráticas» (p. 409).

El decimotercer capítulo se pregunta ¿cómo se representan políticamente los algoritmos? (p. 429–442). La tecnología sugiere sin imponer y hace un llamamiento a nuestro *habitus* (Bourdieu, 1990). «Es ese carácter silente el que permite a las tecnologías escapar al cuestionamiento crítico» (p. 429). La razón principal es que «las categorías algorítmicas señalan certidumbre, desalientan exploraciones, alternativas y crean coherencia entre objetos dispares» (Anamy, 2016: 103). A esa escasa reflexividad, «se añade su imagen de neutralidad, debida a que se trata de un sistema vacío que procesa símbolos sin emitir juicios» (p. 430). A su vez, «los artefactos tecnológicos tienen, a menudo, un aura de sofisticación que los hace demasiado complejos para regularlos y demasiado poderosos para rechazarlos» (p. 430).

El último capítulo se interesa por las razones epistémicas de la resistencia democrática (pp. 443–456). Innerarity incide sobre el hecho de que, más allá de la calidad de los datos y de la potencia de los instrumentos de computación, la capacidad de la gobernanza algorítmica de tomar el control de la democracia se enfrenta a un límite epistemológico. De hecho, «hay cosas que la inteligencia artificial no puede hacer porque no es capaz» (p. 413), lo que queda de manifiesto en el ámbito de la decisión política, porque los humanos y las máquinas decidimos de manera muy diferente. «En lo propiamente político de la política es donde este contraste y nuestra mayor idoneidad son más manifiestos» (p. 443).

En el apartado de conclusiones, el autor recuerda que el rápido auge de la IA genera expectativas e inquietudes.

> Dado su actual desarrollo, no estamos en un momento de balance *ex post* sino de expectativas y temores *ex ante*. El desconocimiento de los efectos de una tecnología nueva desata los calificativos más variados: puede ser celebrado como el triunfo de la comodidad y la exactitud, la victoria definitiva sobre los prejuicios o el final de la arbitrariedad, pero también hay quien lamenta la periferización de los humanos en un mundo en el que parece que hubiéramos dejado ya de decidir.» (p. 457)

En realidad, la cuestión primordial consiste en determinar «cómo hemos de interpretar la automatización general y hasta qué punto esta impide decidir el destino personal y colectivo o lo realiza de otro modo» (p. 458).

Al término de la lectura de *Una teoría critica de la inteligencia artificial*, es preciso subrayar la gran actualidad del tema abordado, dado que la IA ha irrumpido rápidamente y con fuerza en las sociedades contemporáneas, y la sutileza y pertinencia con la que analiza esta realidad emergente, asumiendo la complejidad de la misma y el hecho de que sea difícil determinar con precisión cuáles son y serán sus efectos. El autor no se deja llevar por el pesimismo o el optimismo característicos de los partidarios de la tecnofobia y de la tecnofilia, sino que trata de ofrecer un panorama pormenorizado de la inteligencia artificial, haciendo gala de rigor analítico y de lucidez intelectual. En ese sentido, esta obra, que plantea más preguntas que proporciona respuestas categóricas, aspira a ofrecer una teoría crítica de la IA. En definitiva, la lectura de este libro se antoja ineludible para mejorar nuestra comprensión de la IA en las sociedades actuales.

Referencias

Anamy, M. (2016). Toward an Ethics of Algorithms. *Science, Technology & Human Values, 41(*1), 93–117.

Balkin, J. (2016). Information fiduciaries and the first amendment. *UC Davis Law Review, 49*(4), 1183–1234.

Benjamin, S.M. (2013). Algorithms and Speech, *University of Pennsylvania Law Review*, (161), 1445–1493.

Bourdieu, P. (1990). *The logic of practice.* Stanford University Press.

Cohen, J. (2016). The regulatory state in the information age. *Theoretical Inquiries in Law, 17*(2), 369–414.

Crawford, K. (2016). Can an algorithm be agonistic? *Science, Technology & Human Values, 41*(1), 77–92.

Derosières, A. (1998). *The Politics of Large Numbers: A History of Statistical Reasoning.* Harvard University Press.

Hacking, O. (1990). *The Taming of Change.* Cambridge University Press.

Innerarity, D. (2022). *La sociedad del desconocimiento.* Galaxia Gutenberg.

Innerarity, D. (2023). *La libertad democrática.* Galaxia Gutenberg.

Innerarity, D. (2025). *Una teoría critica de la inteligencia artificial.* Galaxia Gutenberg.

Keane, J. (2013). *Democracy and media decadence.* Cambridge University Press.

Mehra, S. (2015). Antitrust and the robot-seller: Competition in the time of algorithms. *Minessota Law Review*, (100), 1323–1375.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Bid Data & Society*, (julio-diciembre), 1–21.

Neyland, D. (2016). Beating Accountable Witness to the Ethical Algorithms System. *Science, Technology & Human Values, 4*(1), 50–76.

Porter, T.M. (1986). *The Rise of Statistical Thinking, 1820–1900.* Princeton University Press.

Urteaga, E. (2023). *La société de l'incertitude.* L'Harmattan.