

Daniel Innerarity

Una teoría crítica de la inteligencia artificial



Galaxia Gutenberg

DANIEL INNERARITY

Una teoría crítica
de la inteligencia artificial

III Premio de Ensayo Eugenio Trías

Galaxia Gutenberg



**Universitat
Pompeu Fabra**
Barcelona

CEFET
Centro de Estudios Filosóficos
Eugenio Triás

Con la colaboración de la Fundación "la Caixa".

Un jurado presidido por Victoria Camps e integrado por Marina Garcés, Antonio Monegal, Miguel Triás, Joan Tarrida y David Triás concedió a esta obra el 26 de noviembre de 2024 el III Premio de Ensayo Eugenio Triás, que convoca Galaxia Gutenberg junto con el Centro de Estudios Filosóficos Eugenio Triás (CEFET) de la Universidad Pompeu Fabra

Publicado por
Galaxia Gutenberg, S.L.
Av. Diagonal, 361, 2.º I.ª
08037-Barcelona
info@galaxiagutenberg.com
www.galaxiagutenberg.com

Primera edición: marzo de 2025
Segunda edición: abril de 2025
Tercera edición (primera en este formato): febrero de 2026

© Daniel Innerarity, 2025
© Galaxia Gutenberg, S.L., 2025

Preimpresión: Fotocomposición gama, sl
Impresión y encuadernación: Sagrafic
Depósito legal: B 3350-2026
ISBN: 979-13-88019-68-5

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra sólo puede realizarse con la autorización de sus titulares, aparte de las excepciones previstas por la ley. Dirijase a CEDRO (Centro Español de Derechos Reprográficos) si necesita fotocopiar o escanear fragmentos de esta obra (www.conlicencia.com; 91 702 19 70 / 93 272 04 45)

*A Jürgen Habermas,
por su magisterio y amistad*

He escrito la mayor parte de este libro en estos últimos cuatro años, como titular de la Cátedra de Inteligencia Artificial y Democracia del Instituto Universitario Europeo de Florencia. Quisiera agradecer aquí a los miembros de su Consejo asesor internacional (Amparo Alonso Betanzos, Piergiorgio Donatelli, Emilia Gómez, Stephan Lessenich, Sofia Näsström y Helen Margetts) así como a los investigadores de la Cátedra (Lucía Bosoer, Marta Cantero, Ioannis Galariotis, Stefania Milan, Helga Nowotny y Júlia Pareto). Lo he terminado en el Institut für Sozialforschung, sede de la que fue la célebre Escuela de Fráncfort, de cuyo Consejo formo parte. En el que fue el despacho de Adorno he pensado muchas veces qué teoría crítica habría escrito él de haber conocido el actual despliegue de la inteligencia artificial.

Fráncfort, noviembre de 2024

Índice

Introducción. Crítica de la razón algorítmica . . .	17
1. Una moratoria artificial.	24
2. El complemento ético	30
3. La perspectiva de la teoría crítica.	33

PRIMERA PARTE

Teoría de la razón algorítmica

I. La inteligencia de la inteligencia artificial. . .	41
1. Historia de dos inteligencias	42
2. La actual encrucijada de la inteligencia artificial	48
3. La especificidad del conocimiento humano.	55
a. Sentido común	56
b. Reflexividad	62
c. Conocimiento implícito.	66
d. Inexactitud	71
e. Aprendizaje.	81
f. Economía	85
g. Inteligencia corporal	87
4. <i>Techies & fuzzies</i>	92

2. Arte. El sueño de la máquina creativa	95
1. Naturaleza y límites de la creatividad humana.	97
2. Naturaleza y límites de la creatividad artificial	100
3. La obra de arte en la época de su generatividad artificial.	105
3. Datos. La sociedad de los <i>big data</i>	111
1. Datos para la política	113
2. Datos que no nos representan	118
a. Naturaleza general de los datos.	119
b. El mito de la cantidad	122
c. El mito de la neutralidad.	125
d. La necesidad de interpretación	128
3. El poder de los datos	131
Excurso 1: Nada personal. La privacidad como bien público.	138
Excurso 2: La pandemia de los datos	151
4. Predicción. Crítica de la analítica predictiva	161
1. Los pronósticos aciertan demasiado	163
2. Los pronósticos se equivocan demasiado.	166
3. Los grandes ausentes de la predicción algorítmica	169
a. El individuo inclasificable	170
b. La discontinuidad impredecible.	172
c. Propensión no es causalidad	175
d. El futuro abierto de las sociedades democráticas	178
4. El humanismo de la incertidumbre	182

SEGUNDA PARTE
Pragmática de la razón algorítmica

5.	Tecnología. La infraestructura tecnológica de la sociedad digital	187
1.	El tecnosolucionismo.	189
2.	El tecnoneutralismo.	193
3.	El tecnodeterminismo	198
4.	La condición digital.	202
5.	La política de los artefactos	206
6.	Teoría crítica de la transformación digital	211
6.	Automatización. El sentido de lo que funciona sin nosotros	221
1.	La lógica del automatismo.	223
2.	¿Quién decide cuando decide un algoritmo?.	227
3.	El factor humano, el control y sus límites.	234
a.	Elogio de la inconsciencia	236
b.	El dilema de la automatización	239
c.	Demasiados humanos en el <i>loop</i>	242
4.	Máquinas sin humanidad	246
7.	Máquinas. El nuevo contrato social tecnológico	251
1.	La rebelión de las máquinas.	252
2.	La recuperación humana del control	257
3.	Más allá del control y de la sumisión.	262
4.	Identidad, diferencia, hibridación	266
a.	La indistinción entre los humanos y las máquinas.	267
b.	La diferencia entre los humanos y las máquinas: el error de Turing.	269

c. La hibridación entre humanos y máquinas	274
5. El ecosistema humanos-máquinas	278
6. La nueva delimitación de la humanidad.	281
8. Transparencia. ¿Cuánta opacidad requiere y soporta la inteligencia artificial?	285
1. El usuario sumiso	288
2. Tipos de opacidad	292
a. La opacidad intencional	293
b. La opacidad objetiva.	296
c. La opacidad emergente	300
3. Lo que muestra y esconde un algoritmo.	302
a. El mito de la exclusividad	303
b. El mito de la visibilidad.	304
c. La naturaleza social de los algoritmos	307
d. La naturaleza dinámica de los algoritmos	308
e. La autoría múltiple de los algoritmos	310
4. Una inteligencia artificial explicable	312
5. La comprensión como asunto intersubjetivo	318
a. El individuo sobrecargado.	319
b. La resignación digital	321
c. La confianza digital.	323

TERCERA PARTE

Filosofía política de la razón algorítmica

9. Control. Las máquinas, las instituciones y la democracia	329
1. La delegación del control.	332

2. El control de la delegación.	337
3. La delegación como control.	342
10. Gobernanza. Las expectativas políticas de la inteligencia artificial	351
1. De la burocracia a la gobernanza algorítmica	351
2. Las promesas de la gobernanza algorítmica	355
a. La promesa de objetividad	355
b. La promesa de subjetividad.	357
3. Las limitaciones democráticas de la gobernanza algorítmica.	359
4. La inevitabilidad de decidir	366
Excursio: Paradigmas del espacio digital.	369
1. Ágora	373
2. Mercado	376
3. Burocracia	378
11. Preferencias e intereses. La democracia de las recomendaciones.	381
1. Remedios digitales contra el malestar de la democracia	381
2. La lógica individualista de los sistemas de recomendación	386
3. La construcción algorítmica de las preferencias.	389
4. El poder de las preferencias del pasado	396
5. La protección de las preferencias futuras	401
6. La libertad de las preferencias y la libertad frente a las preferencias	404

12. Justicia. Igualdad algorítmica y democracia deliberativa	409
1. Unos algoritmos frente a otros.	409
2. Errores algorítmicos	412
3. Justicia controvertida	414
4. La agregación imposible	418
5. La autodeterminación deliberativa	422
13. Parlamento. ¿Cómo se representan políticamente los algoritmos?	429
1. Democracia como politización	431
2. La despolitización algorítmica.	433
3. La politización como garantía del pluralismo.	437
4. Parlamentarizar la digitalización.	440
14. Democracia. Razones epistémicas de la resistencia democrática	443
1. Dos modos de pensar y decidir	445
2. La ambigüedad política.	448
3. La contingencia política	452
4. La incertidumbre democrática.	455
Conclusión. El futuro de la democracia en la era digital	457
1. La histeria digital	458
2. El condicionamiento digital de la democracia	463
3. Una idea de control compatible con la complejidad	467
Bibliografía	473

Introducción

Crítica de la razón algorítmica

La tecnología es, actualmente, filosofía encubierta; la cuestión es hacerla abiertamente filosófica.

PHILIP AGRE 1997, 240

La organización política de las sociedades ha tenido siempre una pretensión de automaticidad. En cuanto se supera la simpleza de la familia o la tribu, las organizaciones humanas necesitan datos y procedimientos que permitan gestionar la incipiente complejidad. Hay política propiamente hablando desde el momento en que las sociedades no se pueden divisar con un solo golpe de vista, en cuanto se quiebra la homogeneidad y aparecen intereses contrapuestos, cuando falla la evidencia y debe tenerse en cuenta una dimensión que supera lo inmediato, cuando hay que calcular y organizar, allí donde no basta con la simple espontaneidad adaptativa. Desde este punto de vista, la política no sólo no desaparece cuando se complican los procedimientos para la toma de decisiones colectivas, sino que es allí donde propiamente comienza. Nos preguntamos hoy si la política sobrevivirá a la informática, si es posible la política en un entorno de creciente complejidad, cuando lo cierto es,

más bien, el punto de vista contrario: fue la administración política de las sociedades la que originó la disciplina del cálculo y la protocolización.

El historiador Jon Agar (2003) argumenta que las raíces históricas del ordenador están en la administración pública, frente a la suposición inversa de que la administración hace suyo, ahora, un procedimiento que le sería completamente ajeno. Las prácticas políticas son operaciones que miden, planifican y establecen procesos para la toma de decisiones conforme a cierto orden. Por eso se ha podido afirmar que, en el fondo, las operaciones algorítmicas son «prácticas arcanas» (Mau 2017, 206). El Estado fue definido por Thomas Hobbes como un «*automaton*», como un «hombre artificial» (1969, 9). Hobbes es conocido como «el abuelo de la inteligencia artificial» (Haugeland 1985, 23) por dos razones: porque inventó la idea de razonamiento como computación y porque elaboró la idea de una persona artificial para la política. Su filosofía refleja, como pocas, el universo conceptual de la modernidad: el creciente deseo de calcular y la conciencia de la artificialidad de sus construcciones, como la idea de la representación, la del pueblo o la del soberano, que no se corresponden con ningún personaje real y concreto, sino que constituyen un ideal regulativo.

Desde esta perspectiva, la racionalidad algorítmica, más que representar una ruptura absoluta con el pasado, puede ser analizada de acuerdo con continuidades históricas o posibilidades comparativas. Hay quien ha trazado precedentes interesantes con el cálculo en el imperio de Babilonia (Innis 1986), es decir, siempre que ha habido que establecer un orden en un entorno de complejidad y heterogeneidad. Muchas de las prácticas de control algorítmico por parte de los estados o los acto-

res económicos estaban ya planteadas en el imperio babilónico, en los comienzos del Estado moderno y el primer capitalismo. Encontramos precedentes interesantes en las reglamentaciones de la manufactura francesa o en la Inglaterra victoriana. Así pues, la actual digitalización podría entenderse como una continuación intensiva de prácticas de burocratización y racionalización de siglos anteriores, tal como fueron estudiadas por Sombart y Weber.

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. Parecen dar la razón a Tocqueville cuando, en sus *Consideraciones sobre la Revolución*, afirmaba que «toda la política se reduce a una cuestión aritmética» (2004, 492). Gracias a excelentes estudios, conocemos muy bien la relación entre la formación del Estado moderno, la estadística, la probabilidad y los datos (Porter 1986; Hacking 1990; Desrosières 1998). Destaca, especialmente, ese periodo entre 1820 y 1840 calificado como una «avalancha de números impresos» (Hacking 2015), cuando los Estados comenzaron a contar y clasificar intensivamente (Porter 1986, 11). Los *big data* (macrodatos) pueden situarse en la larga historia de la estadística social.

Diversos sociólogos han subrayado, desde los tiempos de Max Weber, que la organización burocrática del Estado está impulsada por la misma tendencia modernizadora que las empresas industriales. La actual algoritmización de la sociedad podría entenderse como continuidad con el cálculo moderno, con sus estadísticas y sistemas de lógica formal. Las organizaciones de la moderna administración se enfrentaron a la contingencia del mundo numerizando y formalizando el caos de la

realidad. A quien se encuentra frente a un mundo lleno de contingencias, el enfoque probabilístico le ofrece la posibilidad de transformar la contingencia en calculabilidad formalizada. Entonces y ahora, los procedimientos de cálculo y algoritmización prometen neutralizar los prejuicios subjetivos mediante procedimientos exactos de decisión. La mecanización del gobierno comenzó a finales del siglo XVIII, cuando la administración pública del Reino Unido, con el objetivo de dirigir un imperio global, invirtió en recoger y procesar información de todo el mundo. Hay que retrotraerse, no obstante, hasta los años cuarenta del siglo pasado para encontrar, en la joven disciplina de la cibernética, los primeros intentos de pensar un gobierno y una administración automatizados. En cualquier caso, desde las primeras formas elementales de gobierno, organizar políticamente la sociedad equivale a poner en marcha un conjunto de procesos, dispositivos y procedimientos que constituyen la tecnología administrativa de la burocracia.

Como la burocracia para el Estado moderno, la inteligencia artificial parece llamada a ser la lógica de legitimación de las organizaciones y los gobiernos en las sociedades digitales. Los tres elementos que modificarán la política de este siglo son los sistemas cada vez más inteligentes, una tecnología más integrada y una sociedad más cuantificada. Si la política a lo largo del siglo XX giró en torno al debate acerca de cómo equilibrar Estado y mercado (cuánto poder debía conferírsele al Estado y cuánta libertad debería dejarse en manos del mercado), la gran cuestión hoy es decidir si 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 nú-

cleo normativo de la organización democrática de las sociedades. El modo en que configuremos la gobernanza de estas tecnologías va a ser decisivo para el futuro de la democracia; puede implicar su destrucción o su fortalecimiento.

El hecho de que pueda identificarse una continuidad entre las primeras formas de organización política de la complejidad y la actual razón algorítmica no significa que la era digital no represente una dimensión cualitativamente diferente e incluso cierta ruptura respecto de la clásica racionalidad burocrática. El objetivo de generar datos estadísticamente analizables es central en la formación de los Estados (Spittler 1980; Vormbusch 2012), pero las sociedades digitalizadas se diferencian cualitativamente de las anteriores por el hecho de que esas autoobservaciones utilizan tecnologías digitales que las registran de forma binaria, archivable, enlazable y combinable (Baecker 2018). La datificación no es sólo un continuo crecimiento del volumen de los datos sociales sino un cambio cualitativo del proceso de constitución de la realidad social.

Si se examinan desde una perspectiva histórica, lo específico de los *big data* no es la gran acumulación de datos, aunque esta cantidad sea un elemento considerable de ese hilo genealógico. Lo decisivo es el modo en el que el concepto de *big data* ha acumulado poder comercial, organizativo y económico. Deberíamos centrarnos, más bien, en el análisis de los métodos y los modos de pensar que acompañan a esta manera de concebir la realidad, en la idea que tenemos de los datos, en cómo los concebimos y presentamos, y no tanto en la infraestructura tecnológica. Con el incremento de la complejidad social aumenta el número de las tareas que es necesario confiar a procedimientos de cálculo, pero la actual automatización no es un aumento cuantitativo sino que

implica –en una medida que deberemos analizar– un salto cualitativo; la algoritmización de tantas decisiones políticas no obedece a una lógica instrumental (humanos que emplean instrumentos) sino, hasta cierto punto, a una lógica de remplazo (humanos que renuncian a decidir) y autonomización (máquinas capaces de decidir por sí mismas). Este salto implica un rediseño institucional de la sociedad, una enorme promesa y una no menos inquietante amenaza. Ciertas decisiones –quién sabe si demasiadas o tal vez todas, en el futuro– no son adoptadas sólo por los seres humanos sino confiadas en todo o en parte a sistemas que procesan datos y dan lugar a un resultado que no es plenamente pronosticable. La paradoja es que los artefactos que inventamos para hacer calculable el mundo nos introducen, al mismo tiempo, en otro tipo de imprevisibilidad.

Así pues, la sociedad digital no se caracteriza tanto porque haya simplemente más datos que en las tradicionales; la explosión de los datos exige que dirijamos nuestra atención a las herramientas de elaboración de esos datos, que no podemos considerar como meros artefactos técnicos, sino también como métodos para adquirir conocimiento y orientar nuestras decisiones y que, por tanto, debemos examinar de forma crítica. El modelo utópico de las élites digitales se traduce en una sociedad en la que cualquier desafío puede resolverse en el marco del modelo de negocio digital, como si todas las crisis fueran traducibles en problemas técnicos cuya solución simplemente requiere procedimientos matemáticos. El supuesto implícito es que la realidad social es plenamente calculable: «regular lo que es regulable y hacer regulable lo que no se puede regular», según la célebre formulación cibernética (Schmitt 1941, 41). Con las palabras que abren la *Dialéctica de la Ilustración*: «hacer el mun-

do calculable» (Horkheimer / Adorno 2002, 4). Debido a su eficacia en el manejo de la complejidad y a su incompatibilidad con nuestra intervención reflexiva, la tecnología algorítmica parece haber abandonado completamente una época en la que era posible la crítica, es decir, la capacidad de cuestionar la construcción social de las categorías mediante las cuales percibimos y evaluamos el mundo (Rouvroy 2020).

Los humanos siempre hemos aspirado a que algún procedimiento mecánico nos haga menos dependientes de la voluntad de los otros. La racionalidad algorítmica parece prometerlo, pero ¿es realmente así? ¿Cómo interactúan y convergen la digitalidad y los modos de gobierno? Al confiar en los procedimientos algorítmicos combatimos la arbitrariedad y el subjetivismo, pero ¿cómo hacerlo sin renunciar a ese «derecho a una decisión humana» (Huq 2020) que corresponde a nuestro deseo de libre autodeterminación y que, al mismo tiempo, es la causa de tanta dominación? ¿Es posible promover la intervención de procedimientos de decisión algorítmicos sin sacrificar nuestro poder a una nueva forma de dominación?

El problema fundamental de la inteligencia artificial es la creciente externalización de decisiones humanas en ella. La automatización generalizada plantea el problema de qué lugar corresponde a la decisión humana, si se trata simplemente de un suplemento, de una modificación o un remplazo. Decía el historiador Agar que los ordenadores vendrían a ser un «sistema nervioso periférico» de las organizaciones del gobierno (2003); la pregunta que deberíamos hacernos es si pueden y deben llegar a ser su sistema central, hasta qué punto las decisiones humanas pueden ser sustituidas por procesos informáticos controlados por algoritmos, de manera que

government machine sea algo más que una metáfora. El hecho de que automaticemos ciertas decisiones individuales o colectivas implica grandes beneficios en términos de efectividad. Sin embargo, este potencial puede constituir una amenaza si implica una rendición absoluta de nuestra soberanía. Determinar el tipo de poder que tenemos los humanos cuando automatizamos y después de automatizar procesos es una cuestión que requiere que contemplemos hasta qué punto es deseable tener el poder de decidir o si puede ser ventajoso no necesitar hacerlo, qué tipo de toma de decisiones es un algoritmo y si el trabajo crítico de los algoritmos puede entenderse como una recuperación de nuestra capacidad de decidir. La respuesta a todas estas cuestiones permitiría convertir la informática en una disciplina política (Reichl / Welzer 2020, 48). En definitiva, ¿quién decide cuando, aparentemente, nadie decide?

Hay tres respuestas posibles a este conjunto de problemas planteados por el creciente protagonismo de la razón algorítmica debido a la delegación de decisiones en la inteligencia artificial: 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 o, si se prefiere, de «humanización»; debemos examinar ahora sus aportaciones y sus limitaciones.

I. UNA MORATORIA ARTIFICIAL

En el año 2023, al estupor, el entusiasmo o el pánico provocados por el ChatGPT y sus fabulosas prestacio-

nes los siguió una *Carta abierta* en la que científicos y empresarios de la industria tecnológica pedían una moratoria digital, con tan buenas razones como confusos procedimientos. Quienes tienen una gran responsabilidad en la creación y desarrollo de la inteligencia artificial dibujaban un escenario apocalíptico con una autoridad en materia de prospectiva que no tiene por qué ir asociada a sus éxitos tecnológicos. Continuó con la prohibición del Chat en Italia, por posibles vulneraciones de la ley de protección de datos. Las peticiones de control y regulación también han aumentado en Estados Unidos ante la Comisión Federal de Comercio, que apelan a la legislación comercial, la seguridad pública y la privacidad. Es evidente que cuanto más sofisticada es una tecnología, mayores son sus prestaciones pero también sus riesgos. Los seres humanos exploramos ese territorio en parte desconocido mediante la reflexión, que es una forma de pausar los procesos y adelantarse a los posibles problemas antes de que se produzcan. En el contexto de los actuales progresos de la inteligencia artificial se están haciendo presentes ciertos peligros como la discriminación, la pérdida de control, la precariedad laboral o la desinformación, todos ellos de tal envergadura que parecen hacer aconsejable pausar el desarrollo tecnológico todo lo que se pueda, con el fin de disponer de un enfoque regulador, ponernos de acuerdo sobre los criterios éticos y políticos, y establecer autoridades de supervisión y certificación. Suponiendo que la tecnología no iba a esperar, los autores de aquella *Carta abierta* exigían una moratoria de seis meses, lo que, de entrada, suscita la sospecha de que pueda ser, a la vez, demasiado tiempo y demasiado poco, que haya otros intereses en la petición, que no sea factible o que implique otro tipo de riesgos.

Semejante parón tecnológico se reclama en favor de la humanidad en su conjunto, pero a nadie se le oculta que implicaría ganancias y perjuicios desigualmente repartidos. Podemos sospechar que se trata de una alianza de los perdedores, que sacarían alguna ventaja de frenar la carrera tecnológica, pero también es cierto que tal moratoria concedería ventajas a quienes ya disponen de los grandes modelos de lenguaje (*Large Language Models*, LLM). Algo tan drástico como detener sectores tecnológicos dinámicos y competitivos plantea muchas dudas en cuanto a su viabilidad, tanto en lo referido a los Estados como en el sector privado. En la actual configuración geoestratégica del mundo, tan fragmentada, y donde la carrera tecnológica se ha convertido en uno de los principales escenarios de competencia, es inimaginable una regulación vinculante y de obligado cumplimiento. Tampoco hay ningún motivo para que las empresas dominantes asuman voluntariamente un freno que podría poner en peligro su posición. Revela mucha ingenuidad creer que todos los programadores van a cerrar sus computadoras y que gobiernos del mundo entero se sentarán durante seis meses con el objetivo de aprobar unas normas vinculantes para todos.

Pero el principal argumento en contra de una moratoria artificial es que con la pretensión de evitar ciertos riesgos acentúe otros. ¿Estamos tan seguros de que no mejorar los modelos de procesamiento durante un tiempo es menos arriesgado que seguir mejorándolos? Es cierto que los actuales sistemas plantean muchos riesgos, pero también es peligroso retrasar la aparición de sistemas más inteligentes. Uno de esos posibles efectos indeseados sería la pérdida de transparencia. Si se decidiera esa moratoria, nadie podría asegurar que el trabajo de formación de tales modelos no continuara de for-

ma encubierta. Esto plantearía el peligro de que su desarrollo –que con anterioridad había sido, en gran medida, abierto y transparente– se volviera más inaccesible y opaco. Además, ¿qué es exactamente lo que habría que parar: la investigación o su aplicación? y ¿qué campos, si es que no son todos? La inteligencia artificial en medicina, por ejemplo, es una gran oportunidad de salvar más vidas o de reducir el sufrimiento, como también de promover el ahorro energético y luchar contra un cambio climático que no se detiene, por lo que sería un grave error interrumpir la investigación en estas y otras áreas. ¿Cómo establecer una distinción entre lo que habría o no que parar, teniendo en cuenta, además, que hay muchas investigaciones básicas que son útiles en distintas áreas?

La idea de la moratoria evidencia una falta de comprensión acerca de la naturaleza de la tecnología, de su articulación con los humanos y, concretamente, de las potencialidades de la inteligencia artificial en relación con la inteligencia humana, a mi juicio, menos amenazada de lo que suponen quienes temen al supremacismo digital. Por supuesto que nos encontramos con un desfase cada vez más inquietante entre la rapidez de la tecnología y la lentitud de su regulación. Los debates políticos o la legislación son, sobre todo, reactivos. Una moratoria tendría la ventaja de que el marco regulatorio podría adoptarse de forma proactiva antes de que la investigación siga avanzando. Pero las cosas no funcionan así, menos aún con este tipo de tecnologías tan sofisticadas. La petición de moratoria describe un mundo ficticio porque, por un lado, considera posible la victoria de la inteligencia artificial sobre la humana, y por otro, sugiere que la inteligencia artificial sólo necesitaría algunas actualizaciones técnicas durante seis meses

de congelación de su desarrollo. ¿En qué quedamos? ¿Cómo es que la amenaza es tan grave y al mismo tiempo bastan seis meses de moratoria para neutralizarla?

Si pasamos de la política ficción a la política real nos encontramos un escenario bien distinto. La Unión Europea es el ámbito político en el que todo esto se está regulando con mayor eficacia y rapidez. Pues bien, la propuesta *Artificial Intelligence Act* de la Comisión Europea tardó cuatro años en aprobarse y estableció un periodo de otros dos para su implementación en los Estados de la Unión Europea. Más que una prueba de irresponsabilidad o lentitud injustificada es una confirmación de la complejidad del asunto, de que no es posible acelerar los procesos de regulación y detener el desarrollo tecnológico cuando hay que poner de acuerdo a muchos actores, incluidos los propios sectores tecnológicos que se pretende regular. En el mundo real, las moratorias son difíciles, parciales y de dudosa eficacia; las cosas suelen arreglarse de otra manera. El filósofo austriaco Otto Neurath sugería la metáfora de arreglar el barco en alta mar, en medio de la travesía y sin poder llevarlo a puerto para las reparaciones. Era una crítica a la analogía fundacionalista de Descartes, para quien el modelo del conocimiento era, más bien, la demolición de un edificio y su completa reconstrucción. Dada la actual sofisticación de las tecnologías y el hecho de que su desarrollo se lleva a cabo en un entorno con muy diversos actores, es poco realista pensar que la evolución de la tecnología puede obedecer a un plan, lo cual no nos debería eximir del esfuerzo por hacer compatible este desarrollo con cierta anticipación y la mejor regulación posible. Tenemos que aprender a utilizar los sistemas de inteligencia artificial con más cuidado en vez de frenar la investigación. La teoría crítica de la razón algorítmica

ca hay que hacerla al mismo tiempo que el desarrollo tecnológico y en diálogo con sus protagonistas.

La idea de una moratoria alimenta malentendidos y falsas percepciones sobre la inteligencia artificial. Sugiere capacidades completamente exageradas y la presenta como una herramienta más poderosa de lo que en realidad es. De este modo, contribuye a distraer la atención de los problemas realmente existentes, sobre los que tenemos que reflexionar ahora y no en un hipotético futuro. Los escenarios apocalípticos nos están distrayendo de los problemas que nos plantea una inteligencia artificial todavía no general ahora mismo y no en un futuro que podría irrumpir. El énfasis constante en los grandes riesgos sirve para generar atención, ya que los mensajes extremos son siempre más interesantes que las propuestas parciales y provisionales. Anticipar como inexorable un determinado futuro no es científico e impide que se tomen medidas concretas que son importantes en el presente para adaptar y regular los sistemas de inteligencia artificial. No hace falta que la inteligencia artificial nos supere para que nos plantee problemas y desafíos que es necesario abordar.

La principal aportación de las peticiones de una moratoria es concienciar a segmentos amplios de la población de que, en efecto, hay cuestiones relevantes en juego. Lo más valioso de estos llamamientos es su mensaje performativo, a saber, que subrayan la gravedad de lo que tienen entre manos la ciencia, la tecnología, la economía, la política, las instituciones educativas y el público en general, y la petición de que se forjen las alianzas necesarias. Son actos retórico-performativos para llamar la atención sobre un problema extremadamente urgente e importante. El problema no es que la inteligencia artificial sea –ahora o en el futuro– demasiado

inteligente, sino que lo será demasiado poco mientras no hayamos resuelto su integración equilibrada y justa en el mundo humano y en el entorno natural. Y eso no se conseguirá parando nada sino con más reflexión, investigación, inteligencia colectiva, debate democrático, supervisión ética y regulación.