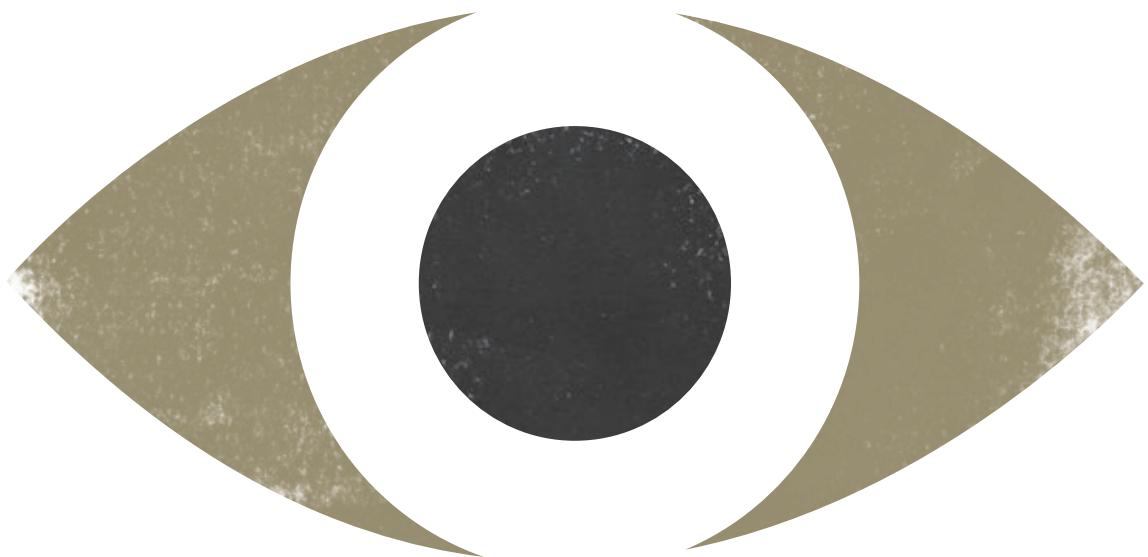


**Paloma Llana**

# **DATA NOMICS**



**TODOS LOS DATOS PERSONALES  
QUE DAS SIN DARTE CUENTA  
Y TODO LO QUE LAS EMPRESAS  
HACEN CON ELLOS**

Claves, consejos y herramientas para proteger tu privacidad

**DEUSTO**

# **Datanomics**

Todos los datos personales que das  
sin darte cuenta y todo lo que las  
empresas hacen con ellos

**PALOMA LLANEZA**



EDICIONES DEUSTO

© Paloma Llana González, 2019

© Editorial Planeta, S.A., 2019

© de esta edición: Centro de Libros PAPP, SLU.

Deusto es un sello editorial de Centro de Libros PAPP, SLU.

Av. Diagonal, 662-664

08034 Barcelona

[www.planetadelibros.com](http://www.planetadelibros.com)

ISBN: 978-84-234-3020-8

Depósito legal: B. 1.373-2019

Primera edición: febrero de 2019

Preimpresión: gama sl

Impreso por Romanyà Valls, S.A.

Impreso en España - *Printed in Spain*

No se permite la reproducción total o parcial de este libro, ni su incorporación a un sistema informático, ni su transmisión en cualquier forma o por cualquier medio, sea éste electrónico, mecánico, por fotocopia, por grabación u otros métodos, sin el permiso previo y por escrito del editor. La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual (Art. 270 y siguientes del Código Penal).

Diríjase a CEDRO (Centro Español de Derechos Reprográficos) si necesita fotocopiar o escanear algún fragmento de esta obra. Puede contactar con CEDRO a través de la web [www.conlicencia.com](http://www.conlicencia.com) o por teléfono en el 91 702 19 70 / 93 272 04 47.

# Sumario

---

Introducción .....	13
1. Dataísmo.....	19
2. Doppelgänger.....	37
3. Todo es percepción .....	53
4. Google, el memorioso .....	69
5. Casandra .....	113
6. <i>Adiction by design</i> .....	141
7. Manipulation by default.....	163
8. GAFA.....	175
9. Familia .....	195
10. Sensorium.....	203
11. Nuestros cuerpos, nosotros .....	217
12. Tracking .....	237
12. Dataveillance .....	247
14. Data breach .....	259
15. La gran mentira.....	273
16. Nada que ocultar .....	281
Bibliografía .....	291

## Dataísmo

### La nueva religión

En el futuro cercano de 2020 y dos décadas después de la primera explosión de la burbuja puntocom, el modelo de negocio impulsado por la publicidad de las principales compañías de internet se desmorona. A medida que las empresas web sobrevaloradas se hunden, todos —grandes y pequeños, delincuentes y compañías decentes— compiten por quedarse con la propiedad de activos de datos subvalorados —y quién sabe si potencialmente valiosos— tras el hundimiento. Estamos ante un futuro distópico, una guerra por los datos con un trasfondo de estrés financiero, pánico, derechos de propiedad ambiguos, mercados opacos y trolls de datos en todas partes. En este mundo, la ciberseguridad y la seguridad de los datos se entrelazan de forma inextricable. Hay dos activos clave que los delincuentes explotan: los propios conjuntos de datos, que se convierten en los principales objetivos de ataque; y los humanos que trabajan en ellos, ya que el colapso de la industria deja a los científicos de datos desempleados en busca de empleo.

Los modelos de negocio basados en datos de la primera mitad de la década de 2010 desaparecerán. El ambiente, sombrío, anuncia el fin de la tercera era de las compañías de internet que tendrán que buscar nuevas fuentes de ingresos. Algunas compañías de hardware comenzarán a cobrar el precio completo por sus dispositivos (por ejemplo, Amazon podría revocar todas las ofertas de precios especiales en su Kindle). Para reducir su dependencia de los datos, las empresas de servicios empezarán a

cobrar. El «Freemium» será una palabra del pasado, y la mayoría de las aplicaciones «gratuitas» que han sido símbolos icónicos de la web 2.0, dejarán de serlo.

Este escenario es uno de los planteados por el Centro de Ciberseguridad a Largo Plazo de la Universidad de Berkeley<sup>1</sup> en 2016. Mirando a un futuro que casi se nos viene encima, parece poco probable que la economía basada en datos o con alta dependencia de éstos vaya a desaparecer de un plumazo.

Siempre se han usado datos para dibujar la realidad o tomar decisiones para mantenerla o cambiarla. La información siempre ha sido poder y no debemos caer en el adanismo de considerar que hemos inventado el análisis de datos y los algoritmos. Lo que es cierto es que nunca antes habíamos tenido de manera conjunta la capacidad de recabar datos con enorme granularidad (hasta el punto de identificar a un individuo por cómo interactúa con un teléfono), almacenarlos durante largo tiempo y sin límite de capacidad, y de analizarlos con rapidez y precisión. Por tanto, no es de extrañar que algunos piensen que los datos han venido a redimirnos.

## **Harari y los humanismos liberales**

Nuestras religiones sostienen que existe un orden sobrehumano que no es producto de caprichos o convenios humanos, sino la base y los principios sobre los que establecemos normas y valores a los que dotamos de valor obligatorio y que, por su origen divino, no admiten crítica. Esta estructura de pensamiento se extiende a ideologías que participan de los mismos elementos constituyentes que le hemos otorgado a la religión, como el humanismo liberal o el dataísmo, otro fenómeno que, junto con el entusiasmo tecnológico, son percibidos en las últimas décadas como un acontecimiento religioso que no admite crítica u oposición.

1. «Cybersecurity Futures 2020.» Center for Long-Term Cybersecurity University of California, Berkeley. 2016. <<https://cltc.berkeley.edu/scenarios/>>.

Los dataístas han encontrado sentido a la sinrazón evolutiva a través de la atractiva idea de que los datos lo explican todo, le dan significado a nuestra existencia. Como resalta Harari,<sup>2</sup> «el dataísmo sostiene que el universo consiste en flujos de datos, y que el valor de cualquier fenómeno o entidad está determinado por su contribución al procesamiento de los mismos [...]. El dataísmo une ambos<sup>3</sup> y señala que las mismas leyes matemáticas se aplican tanto a los algoritmos bioquímicos como a los electrónicos. De esta manera, el dataísmo hace que la barrera entre animales y máquinas se desplome, y espera que los algoritmos electrónicos acaben por descifrar los algoritmos bioquímicos y los superen». Tan potente es la propaganda a favor de los datos y en defensa de la desaparición de la privacidad (como si su reivindicación fuera pecaminosa, propia de personas con hábitos dudosos que prefirieran ocultar) que Harari coloca el dataísmo dentro de los movimientos humanistas de corte religioso. Y es verdad que si uno se atreve a oponerse al tratamiento de los grandes datos es señalado como alguien contrario, en su totalidad, al avance y a la innovación. Continuando con Harari, «para los políticos, los empresarios y los consumidores corrientes, el dataísmo ofrece tecnologías innovadoras y poderes inmensos y nuevos. Para los estudiosos e intelectuales promete asimismo el santo grial científico que ha estado eludiéndonos durante siglos: una única teoría global que unifique todas las disciplinas científicas».

Esta visión elevada de los datos como el GATACA que todo lo explica, aterriza en el frío suelo de los números cuando nos enfrentamos al uso que las empresas dan a los datos que les facilitamos de un modo u otro. Las dotamos del poder de los augures, y les dejamos leer nuestro futuro en los posos de nuestros hábitos. Gracias a nuestros datos, siguen comportamientos y sacan conclusiones que les permiten predecir lo que va a suceder, no el porqué, que hoy ha pasado a ser irrelevante.

2. Harari, Yuval Noah. *Homo Deus. Breve historia del mañana*. Debate, 2016, p. 400.

3. La evolución darwiniana y el procesamiento complejo de Turing.

Esta lectura sólo se puede hacer usando grandes cantidades de datos. Los usuarios suelen proporcionar información de contacto fidedigna para poder disfrutar del servicio o producto que se oferta en la plataforma, y añaden luego más datos con el uso recurrente del servicio.

Porque, como veremos a lo largo de esta obra, somos datos. Datos que nos definen con una precisión que supera la maternal y que, llevado al extremo y sin control, nos puede colocar, como individuos y sociedad, en una precaria situación. El propio Harari lo reconocía en una entrevista al diario *El País*:<sup>4</sup> «El mayor problema político, legal y filosófico de nuestra época es cómo regular la propiedad de los datos. En el pasado, delimitar la propiedad de la tierra fue fácil: se ponía una valla y se escribía en un papel el nombre del dueño. Cuando surgió la industria moderna, hubo que regular la propiedad de las máquinas. Y se consiguió. Pero ¿los datos? Están en todas partes y en ninguna. Puedo tener una copia de mi historial médico, pero eso no significa que yo sea el propietario de esos datos, porque puede haber millones de copias. Necesitamos un sistema diferente. ¿Cuál? No lo sé. Otra pregunta clave es cómo conseguir una mayor cooperación internacional».

Harari coloca la propiedad de los datos entre los tres principales problemas humanos de carácter global, lo que compromete y dificulta su resolución al no depender de las soberanías nacionales sino de un consenso internacional imposible en un mundo de intereses locales egoístas: «Nuestros tres principales problemas son globales. Un sólo país no puede arreglarlos. Hablo de la amenaza de una guerra nuclear, del cambio climático y de la disrupción tecnológica, en especial del auge de la inteligencia artificial y de la bioingeniería. Por ejemplo, ¿qué podría hacer el Gobierno español contra el cambio climático? Aunque España se convirtiera en el país más sostenible y redujera sus emisiones a cero, sin la cooperación de China o Estados Unidos

4. Harari, Yuval Noah. «La tecnología permitirá “hackear” a seres humanos.» *El País Semanal*, 20 de agosto de 2018. <[https://elpais.com/elpais/2018/08/20/eps/1534781175\\_639404.html](https://elpais.com/elpais/2018/08/20/eps/1534781175_639404.html)>.

no serviría de mucho. En cuanto a la tecnología, aunque la UE prohíba experimentar con los genes de una persona para diseñar superhumanos, si Corea o China lo realizan, ¿qué haces? Es probable que Europa acabará creando seres superinteligentes para no quedarse atrás. Es difícil ir en la dirección contraria».

## Los datos como subproducto

La RAE define los datos como «información sobre algo concreto que permite su conocimiento exacto o sirve para deducir las consecuencias derivadas de un hecho». Ésta y otras definiciones conexas inciden en la idea de que la transmisión de la información reflejada en los datos causará efectos, inmediatos o diferidos, en el comportamiento del receptor de la transmisión.

Como indica Pérez-Chirinos,<sup>5</sup> hasta la llegada de los artificios técnicos, los seres vivos hemos transmitido la información por medios bioquímicos: muchos animales son capaces de comunicarse a grandes distancias acústicamente y algunas especies cuentan con lo que pueden considerarse propiamente lenguajes con propósitos específicos, capaces de causar efectos predecibles en sus congéneres receptores.

Además de la capacidad de transmitir información verbalmente de efecto inmediato, los humanos disponemos desde hace milenios de instrumentos que nos permiten almacenar los datos que queremos comunicar diferidamente. Desde la representación primitiva de nuestras ideas en obras de arte hasta las últimas técnicas de almacenamiento digital, tenemos la posibilidad de comunicarnos a través del tiempo, incluso con personas que nacerán cuando ya no estemos vivos.

Para ello recurrimos a codificar el mensaje que queremos transmitir siguiendo las reglas apropiadas al soporte material en el que queremos conservarlo. Este proceso de codificación es conceptualmente idéntico tanto si se trata de escribir un documento, transcribir una canción en su partitura o generar una

5. Pérez-Chirinos, César. *Algoritmos y programación*.

factura electrónica: tan sólo cambia el algoritmo de codificación utilizado para obtener el resultado deseado, que es registrar el mensaje de forma duradera para que pueda ser decodificado posteriormente.

Gracias a esta posibilidad de registrar mensajes de forma persistente, los humanos podemos construir sistemas de información capaces de dirigir la conducta de otros humanos en un futuro distante del momento en que fueron construidos. Esta capacidad de dirigir conductas futuras opera sobre los receptores de estos mensajes del pasado cuando decodifican los mensajes persistentes.

Los sistemas informáticos producen datos de manera constante. Está en su naturaleza que sea así. Los alimentamos con datos para que nos den respuestas y, a cambio, nos vomitan datos estructurados, nos contestan a las preguntas que les hacemos o actúan como les hemos pedido que hagan. En este proceso, además, generan más datos, como un subproducto necesario para que las máquinas, sin que las entendamos, se hablen entre sí. En algunos casos, y por una mala programación, los sistemas, aplicaciones y redes generan más datos, más ruido de fondo del que sería deseable. El número y variedad de datos accesorios producidos por un sistema tenía, hasta hace unos años, una limitación basada en la eficiencia y la eficacia: se generaban los datos que la capa de transporte, la de aplicaciones o los sistemas de almacenamiento pudieran soportar. Con sistemas limitados en alguno de estos aspectos se tenía especial cuidado en no generar datos que no se podían procesar, transportar o almacenar. Eran los tiempos felices de los disquetes y los Amstrad en los que tener 100 megas de memoria era todo un lujo. Sobre esos ordenadores corrían procesadores de texto que eran capaces de guardar versiones de los documentos y los metadatos necesarios para que no se desconfiguraran. Ahora, cualquier procesador guarda una cantidad ingente de información y todas las versiones anteriores de los documentos.

Todo esto ocurría en un entorno cerrado y controlado: el de los grandes centros de datos, conectados en redes propietarias o en nuestro PC de sobremesa. Los datos se gestionaban de mane-

ra interna por principios de necesidad y eficiencia. Y llegó internet y conectamos nuestras máquinas entre sí. Transmitiendo, al principio, la información estrictamente necesaria —porque el ancho de banda no daba para más— y tras una década de movilidad a conexiones 5G que permiten que todas las cosas se conecten en una red infinita en la que se habla de nosotros a nuestras espaldas.

Una vez que conectas un sistema a una red abierta, los datos que produce y los que producimos nosotros al llevarlo en el bolsillo se multiplican y van más allá de las páginas webs visitadas, los anuncios clicados —aunque sea por error— o las palabras tecladas. Todas las máquinas que intervienen en el proceso (nuestro PC, los routers, los servidores de contenidos y comunicaciones, las antenas de comunicaciones...), todos ellos, sin excepción, generan datos pasivos: los navegadores facilitan a las webs qué modelo son, desde qué ordenador se conectan, cuál es su sistema operativo, qué permisos se han dado y cuáles no, la IP y, por tanto, la compañía de telecomunicaciones con la que tenemos contratado el servicio. En muchos casos, sólo con estos datos es posible identificar de manera única el ordenador y, con un poco más, a los usuarios de éste.

Con la movilidad inteligente hemos dado un paso más allá: ha traído la individualidad, la identificación unívoca de todos y cada uno de nosotros. Es un sistema, pues un ordenador puede ser usado por cualquiera o tener usuarios intercambiables, igual que el teléfono fijo familiar. Antes, para hablar con el novio había que pasar por el filtro del padre malhumorado, la hermana chinchona o la madre indiferente. Ahora las relaciones se ahogan en la hiperconectividad y el hipercontrol del doble check de WhatsApp. Al ser el móvil un identificador personal más potente que el DNI, todo lo que se instale, todo lo que se almacene, todo lo que se diga o haga nos identifica de manera unívoca. Controlar los móviles es controlar la identidad. Por eso, no es casual que Google regale el sistema operativo a los fabricantes de móviles que compiten con Apple.

Pero la movilidad no sería nada sin la hiperconectividad. Los teléfonos son ordenadores que, además, hacen llamadas. Los co-

ches ya son gigantescos ordenadores que, además, tienen ruedas y motor para llevarte de un punto a otro. Los inofensivos electrodomésticos son ordenadores que le dicen a la compañía de seguros si bebemos demasiada cerveza o abusamos de los embutidos, o les cuentan a los ladrones cómo son nuestras viviendas y a qué hora no estamos en casa. De hecho, con los vestibles o los ingeribles, nuestros cuerpos pasan a ser un repositorio de datos que se transmiten y analizan para tomar decisiones sobre nosotros. Está cercano el día en que se penalice a los enfermos coronarios con dejarles sin tratamiento por su indolente alimentación y falta de ejercicio. Un adecuado análisis de varias fuentes de datos, por ejemplo, Instagram y una pulsera de ejercicio, junto con la edad y el peso que le habremos facilitado a la aplicación de la pulsera al darnos de alta en el servicio, permitirá saber si hemos comido de manera inadecuada y no hemos hecho el ejercicio necesario. Con estos datos se nos podrá penalizar por nuestro pasado y predecir nuestro futuro estado, invirtiendo en nuestra recuperación en una ponderación coste-beneficio-castigo.

Y todo eso sin haber introducido un único dato, sin haber tecleado más que unos pocos datos de registro y haber sido algo indolentes con los permisos otorgados a las aplicaciones o servicios. En muchos casos, esas aplicaciones o servicios están diseñados para que la búsqueda de los parámetros de configuración de privacidad sean más sencillos que la octava de las doce pruebas de Astérix (sí, la del formulario A-38).

## **Datos por servicios. Pago por datos**

La publicidad lo cambió todo. En 1998, Sergey Brin y Larry Page presentaron un artículo en una conferencia en Australia en el que mantenían que la publicidad corrompía la tecnología del buscador. En un tiempo en que el contenido se indexaba mal y que incluso se vendían agendas para anotar las URL más usadas como si de listines telefónicos se tratasen, los buscadores como Yahoo! funcionaban bajo el concepto de portal de entrada a internet: no tenías que recordar o anotar todas las webs sino sólo

recordar terra.com o yahoo.com para acceder a la red comercial. Todos luchaban por ser la entrada a la ciudad medieval de la web y, como tal, presentaban una home recargada y llena de banners publicitarios. Google.com fue una completa revolución: se presentó como una página en blanco, sin publicidad ni distracciones, tan sólo había una casilla central de búsqueda, y dos botones: «enter» y «voy a tener suerte». Todo ello bajo el eslogan *Don't be evil*. Tras esa puerta inmaculada, que ya no preside ninguna declaración de intenciones, se alza toda una maquinaria orientada a concentrar los servicios más útiles, mejor diseñados y sin intercambio monetario que se hayan visto en los últimos veinte años.

Google, y después Facebook, gracias a una política de competencia pensada en la era Reagan, han crecido en su posición de mercado hasta constituirse en un duopolio en varias áreas, una de ellas indiscutible: la publicidad online. En 2017, ambas ingresaron por este concepto 135 millones de dólares controlando, según las cifras, entre el 60 y el 75 por ciento del mercado global publicitario.

Veremos cómo muchos de estos servicios, los no necesariamente adictivos pero sí prácticos y útiles, representan una concentración del mercado indeseada e indeseable que hace que el idílico internet descentralizado sea un recuerdo romántico, casi tanto como los pitidos agónicos de los primeros módems. Un ejemplo: Facebook estuvo caído durante 45 minutos el 3 de agosto de 2018. Durante ese tiempo, el tráfico web global subió un 2,3 por ciento, lo que supuso un aumento el 11 por ciento del tráfico directo a las páginas web y un 22 por ciento en otras aplicaciones, ampliándose en un 8 por ciento el tráfico de buscadores.

En este contexto, son varios los debates que se abren: el problema de la seguridad de los datos concentrados en los repositorios de unas pocas empresas; la obtención de los datos sin un consentimiento real para un tratamiento poco transparente que ponga en riesgo el derecho fundamental a la intimidad, para empezar, pero también el de la libre expresión, el derecho a la participación política y la libertad de movimientos; y el carácter

ético —o no— de que los titulares de los datos comercien con los mismos, de que obtengan un pago por ellos más allá de recibir servicios útiles pero poco transparentes, en muchos casos diseñados para ser adictivos, y con consecuencias impensadas en la propia democracia representativa.

## **Monetización, dividendo y sindicatos de los trabajadores del dato**

La monetización de los datos es un argumento casi tan recurrente y falaz como poner la responsabilidad del mal uso de los datos en el lado del usuario. Los datos son imprescindibles para establecer tendencias de consumo o estados de ánimo, para detectar tempranamente una enfermedad, o para individualizar y personalizar lo que se ve y lo que se ofrece a un usuario. Sin embargo, el paquete de datos individualizado de un ser humano no tiene gran valor monetario —o no al menos como para vivir únicamente de él—. Tomemos como ejemplo a Jennifer Lyn Morone. La artista estadounidense se registró como una empresa en Delaware para explotar sus datos personales con la finalidad de obtener un pago por los mismos y, de paso, hacer una acción artística. Hizo varios paquetes de datos, subconjuntos que exhibió en una galería de Londres en 2016 ofreciéndolos para su compra. Los paquetes tenían distintos precios, desde 100 libras por unos datos básicos, hasta 700 libras por todos ellos, incluidos los de salud y su número de seguridad social.

Tal vez sus datos eran valiosos, o lo eran en una primera venta (los datos reutilizados se deprecian en la Deep web), pero la media está muy por debajo de las 100 libras de Morone. Haciendo un sencillo cálculo, si Facebook compartiera sus beneficios entre todos sus usuarios, cada uno obtendría sólo nueve dólares al año, cuantía que se acerca bastante a los precios que se manejan en otros entornos. Tal vez los precios de los datos frescos de tarjetas de crédito o de cámaras hackeadas de menores sean superiores en el internet profundo, pero sigue estando muy lejos de lo que un ser humano necesita anualmente para sobrevivir.

Los motivos son numerosos, pero el primero que se nos viene a la cabeza, por evidente, es que los datos valen si se analizan de manera combinada con otros datos. Un voto suelto no hace ganar unas elecciones (por mucho que Delibes y *El disputado voto del señor Cayo* se empeñen), pero lo cierto es que sin cada uno de ellos no se llega al Congreso de los Diputados.

De adoptarse esta aproximación —que los ciudadanos controlasen sus datos y comerciasen con ellos—, ¿cómo sería una economía de datos en la que los gigantes tecnológicos tuvieran que pagar por el acceso a éstos?

Como resalta *The Economist*,<sup>6</sup> no sería la primera vez que un importante recurso económico haya pasado de ser usado a ser poseído y negociado, como ocurrió con la tierra, el agua o, añado, la propia imagen. Sin embargo, este mismo medio considera que la información digital es un candidato improbable para vivir esta transición. A diferencia de los recursos físicos, los datos personales son un ejemplo de los bienes que pueden usarse más de una vez. De hecho, cuanto más se utilizan correctamente y con todas las garantías, mejor es para la sociedad, que dirían los dataístas.

Precisamente porque se pueden reutilizar, nada impide que una empresa los compre una vez y los use cientos de veces. Un sistema de monetización justa requeriría una transparencia total de las compañías que, desde el más profundo conocimiento de cómo funcionan, es improbable. No mencionaremos los problemas que causan las frecuentes filtraciones —que hacen que los datos fluyan como por una cañería averiada— o el hecho de que los sistemas de información estén diseñados desde la vagancia, generando un número desproporcionado de datos que podrían evitarse o minimizarse, y que, en la mayor parte de los casos, son proporcionados por el usuario, que no tiene la más remota idea de que están ahí y que están siendo recogidos, agregados, tratados y analizados.

6. «Data workers of the world, unite. What if people were paid for their data?» *The Economist*. <<https://www.economist.com/the-world-if/2018/07/07/what-if-people-were-paid-for-their-data>>.

Pero no dejemos que una evidencia desanime a una teoría florida. E. Glen Weyl, autor de *Radical Markets*, escrito en colaboración con Eric Posner, de la Universidad de Chicago,<sup>7</sup> considera que en el futuro, como ocurrió con cualquier avance de la Revolución industrial, el uso será tan intensivo que el precio subirá. Tan convencido está que milita por el estudio de un mecanismo para distribuir la riqueza creada por la IA en el futuro, a la vista de la alta concentración actual de los tratamientos de grandes datos. Weyl argumenta que las habilidades necesarias para generar datos valiosos pueden estar más ampliamente diseminadas de lo que se podría pensar, por lo que el trabajo de datos podría afectar la jerarquía estándar del capital humano. En una toma de los cuarteles de invierno digitales, Weyl avisa de una unión de los trabajadores de datos del mundo en una internacional dataísta que acabará con la desigualdad y traerá un mundo soleado, sin lluvia ni replicantes.

Hasta que el sindicato universal de los datos llegue, la BBC, en un reportaje de finales de 2018,<sup>8</sup> evidenció la realidad de las granjas de humanos que entrenan a las IA por cantidades de miseria. Toda la magia que hay tras estas maravillas pensantes se llama Brenda, una madre soltera de veintiséis años que vive en Kibera (Nairobi), el barrio pobre más grande de África, y quizás el vecindario más duro del mundo, donde cientos de miles de personas viven en un espacio no mucho mayor que el Hyde Park de Londres.

Brenda trabaja para Samasource, una compañía con sede en San Francisco que cuenta entre sus clientes a Google, Microsoft, Salesforce y Yahoo. A la mayoría de estas empresas no les gusta hablar sobre la naturaleza exacta de su trabajo con Samasource, pero cabe afirmar que la información en la que trabajan Brenda

7. Posner, Eric A.; y Weyl, E. Glen. «Radical Markets: Uprooting Capitalism and Democracy for a Just Society Hardcover.» Princeton University Press, 15 de mayo de 2018.

8. Lee, Dave. «Why Big Tech pays poor Kenyans to teach self-driving cars.» BBC News, 3 de noviembre de 2018. <<https://www.bbc.com/news/technology-46055595>>.

y sus compañeros en jornadas de ocho horas es una parte crucial de una parte de sus procesos con la IA. La joven carga una imagen en la que busca todo lo que se encuentre en ella y lo marca para que los coches sin conductor sepan lo que están viendo. Etiqueta a las personas, los automóviles, las señales de tráfico, las líneas en el pavimento, hasta el cielo, especificando si está nublado o soleado. Se asegura de que no haya un sólo píxel etiquetado incorrectamente. Su trabajo será revisado por un superior, quien lo devolverá si no está a la altura. Los entrenadores más rápidos y precisos son premiados con el honor de tener su nombre en una de las muchas pantallas de televisión de la oficina y con, lo que es más importante, cupones para la compra. Samasource paga a Brenda y a sus compañeros un salario de unos nueve dólares diarios, digno para Nairobi, miserable para Silicon Valley.

Llegados a este punto, incluso Angela Merkel, la canciller de Alemania —uno de los países más estrictos con la protección de datos personales, pues llega a exigir que se almacenen dentro del territorio nacional—, está considerando que se ponga o se pague un precio por el uso de los datos personales.

Frente a esta opción se alza la cuestión ética y legal de si se puede comerciar con derechos tales como la dignidad o la intimidad, y si el derecho de cancelación de los datos, que no es más que una revocación de uso de los mismos, iría en contra del concepto de venta, que se basa en el carácter irrevocable cuando es legal. Es decir, si vendo una casa no puedo exigir que me la devuelvan por igual precio. En cuanto a los datos, sí puedo venderlos y pedir que se borren al día siguiente. Además, está la discusión de si es ética su venta y, si se me permite, si es práctica. La UE está trabajando en un nuevo concepto, el del «dividendo» de los datos: un reparto de los beneficios del tratamiento entre la empresa que los usa y quien los provee.

Otra opción que está encima de la mesa es la de los trabajadores de los datos, la del pago por entrenar a las Inteligencias Artificiales (IA).

En 1770, el inventor húngaro Wolfgang von Kempelen construyó una máquina de ajedrez conocida como Mechanical Turk. Su objetivo, en parte, era impresionar a la emperatriz María Te-

resa de Austria. Este dispositivo era capaz de jugar ajedrez contra un oponente humano, y tuvo un éxito espectacular al ganar la mayoría de los juegos en sus demostraciones en Europa y América durante casi nueve décadas. Pero Mechanical Turk era sólo una ilusión que permitía a un maestro de ajedrez humano esconderse dentro de la máquina y operarla.

Unos 160 años después, Amazon.com llamó a su plataforma de crowdsourcing basada en micropago con igual nombre. Según Ayhan Aytes, la motivación inicial de Amazon para construir Mechanical Turk surgió después del fracaso de sus programas de inteligencia artificial en la tarea de encontrar páginas duplicadas de productos en su web.<sup>9</sup> Después de una serie de intentos inútiles y caros, los ingenieros del proyecto recurrieron a los humanos para trabajar detrás de los sistemas en red.<sup>10</sup> Amazon Mechanical Turk simulaba los sistemas de inteligencia artificial mediante el control, la evaluación y la corrección de los procesos de aprendizaje automático usando la capacidad intelectual humana: los usuarios tenían la impresión de que era la inteligencia artificial avanzada la que realizaba tareas cuando, en realidad, estaba más cerca de una forma de «inteligencia artificial-artificial», impulsada por una fuerza de trabajo remota, dispersa y mal remunerada que ayudaba al cliente a lograr sus objetivos comerciales. Como observó Aytes, «en ambos casos —tanto el Mechanical Turk de 1770 como la versión contemporánea del servicio de Amazon— la actuación de los trabajadores que animan el artificio queda oscurecida por el espectáculo de la máquina».<sup>11</sup>

Este tipo de trabajo invisible y oculto, externalizado o subcontratado detrás de interfaces, y camuflado dentro de procesos

9. Aytes, Ayhan. «Return of the Crowds: Mechanical Turk and Neoliberal States of Exception.» *Digital Labor: The Internet as Playground and Factory*. Trebor Scholz (ed.). Londres: Routledge, 2012.

10. Pontin, Jason. «Artificial Intelligence, With Help From the Humans.» *The New York Times*, 25 de marzo de 2007. <<https://www.nytimes.com/2007/03/25/business/yourmoney/25Stream.html>>.

11. Aytes, *op. cit.*

algorítmicos es común, sobre todo en los procesos de marcar y etiquetar miles de horas de archivos digitales con el fin de alimentar las redes neuronales. A veces, este trabajo no se paga en absoluto, como en el caso del reCAPTCHA de Google. En una paradoja que muchos de nosotros hemos experimentado: para demostrar que no somos un agente artificial, nos vemos forzados a entrenar el sistema de IA de reconocimiento de imagen de Google de forma gratuita, al seleccionar múltiples cajas que contienen números de calles, automóviles o casas. En otras ocasiones, nos hacemos selfies para ver si nos parecemos más a la Mona Lisa o a la Chica de la Perla, como la aplicación de Google «Arts & Culture», que es un ejemplo perfecto de cómo regalamos un dato biométrico y, al tiempo, entrenamos una herramienta de reconocimiento visual.

Frente a esta situación tan habitual, los defensores de «datos como mano de obra» creen que a los usuarios se les debería pagar por usar servicios en línea.

El trabajo, como los datos, es un recurso difícil de precisar que no ha sido compensado adecuadamente durante la mayor parte de la historia humana. Incluso cuando los humanos pasamos a ser empleados aparentemente libres para decidir, tuvimos que sobrevivir a unas cuantas revoluciones antes de que el pago por nuestra labor alcanzara una cifra digna.

Weyl considera que los datos son como era el carbón, y que las materias primas o los recursos humanos son un elemento productivo más con el que alimentar a lo que ellos prefieren llamar «inteligencia colectiva». La mayoría de los algoritmos de IA necesitan ser entrenados con montones de ejemplos generados por humanos en un proceso llamado «aprendizaje automático». A menos que sepan cuáles son las respuestas correctas (proporcionadas por humanos), los algoritmos no pueden traducir idiomas, entender el habla ni reconocer objetos o imágenes. Desde esa perspectiva economicista y nada filosófica, los datos serían una forma de trabajo con la que alimentar la IA. Sería un trabajo inconsciente o pasivo en el que se generan los datos —aun de manera voluntaria y activa por actividades como publicaciones en medios sociales, escuchar música o recomendar restaurantes,

que generan los datos necesarios para impulsar nuevos servicios—; o un trabajo consciente y activo —en el que se toman decisiones sobre los datos, como etiquetar imágenes o conducir un automóvil a través de una ciudad, que pueden usarse como base para el entrenamiento de sistemas de inteligencia artificial.

Sin embargo, e independientemente de que estos datos se generen de manera activa o pasiva, pocas personas tendrán el tiempo necesario para realizar un seguimiento de toda la información que generan o para estimar su valor. Incluso aquellos que lo hagan carecerán del poder de negociación para obtener un buen trato de las empresas de IA.

Weyl entiende que la historia del trabajo ofrece una pista sobre cómo podrían evolucionar las cosas: gracias a la negociación colectiva a través de la fuerza sindical. De manera similar, espera ver el surgimiento de lo que llama «sindicatos de datos y trabajo», organizaciones que servirían como guardianes de los datos. Al igual que sus predecesores en el mundo laboral, Weyl imagina una Arcadia en la que estos «sindicatos del dato» negocian las tarifas, supervisan el trabajo de datos de los miembros y garantizan la calidad de su producción digital. Los sindicatos podrían, en su visión, canalizar el trabajo de datos especializados a sus miembros e incluso organizar huelgas, por ejemplo, bloqueando el acceso a los datos de sus miembros para ejercer presión sobre una empresa que los emplea sin su permiso o sin respetar sus derechos. Del mismo modo, estos sindicatos podrían canalizar las contribuciones de los datos de sus miembros —al más puro estilo de una cooperativa— al tiempo que hacen labores de seguimiento, identificación, persecución y facturación a las empresas de IA que se beneficien de ellos.

La cuestión es cómo vamos a conseguir que Google o Facebook renuncien a su actual modelo de negocio de datos gratis para vender publicidad (y alguna cosa más). Como veremos cuando hablemos de las GAFAs, ambas se encuentran en una situación de dominio en el mercado porque ninguna autoridad de competencia ha movido un dedo para evitar que se posicionaran ni para que abusaran de esa situación de monopolística. La interferencia en las elecciones presidenciales de 2016 en Estados

Unidos ha motivado que se hayan comenzado a tomar tímidas medidas para paliar ese poder. Medidas desde el completo desconocimiento del funcionamiento interno de estas compañías y de cómo usan los datos que recogen a paladas.

Un par de datos. Como hemos dicho, en 2017 ambas compañías ganaron 135 mil millones en dólares en publicidad. Al respecto, la Comisión Europea sancionó a Google con dos multas que suman algo más de 7.500 millones de euros (por abuso de posición de dominio en el mercado de los buscadores y por instalar su navegador por defecto en su sistema operativo móvil Android); y la Agencia de Protección de Datos española hizo lo mismo con Facebook, a la que multó con 1,2 millones de euros por la filtración de los datos en el escándalo de Cambridge Analytica (septiembre de 2017).

Las multas en el caso de Google son recurribles, molestas pero no irremediables. Sobre todo, no animan a cambiar un modelo de negocio altamente rentable. En el caso de Facebook, la sanción es tan ridícula comparada con su volumen de negocio que seguro que gasta más en vasos de papel en una semana que en lo que ha de pagar por saltarse la normativa europea. Aunque el nuevo reglamento europeo de protección de datos aumenta significativamente la cuantía de las sanciones, acercándolas a las que se imponen en materia de competencia, no parece que éste sea el camino para incentivar a estas compañías en el camino del uso ético y del pago de los datos.

Es más, si tuvieran que compensar a las personas por sus datos, éstos serían mucho menos rentables y, además, seguro que las empresas encontrarían maneras de inducir a la entrega gratuita de los mismos ya que, debido a la concentración de los servicios online, los usuarios (que no clientes) están cautivos.

Hay ya ejemplos de trabajos intensivos en la categorización de datos y plataformas de crowdworking, como la mencionada Mechanical Turk de Amazon (etiquetado de imágenes) o Mighty AI, una startup con sede en Seattle que contrata a miles de trabajadores en remoto para que etiqueten imágenes de escenas callejeras usadas para entrenar los algoritmos que se utilizan en los coches sin conductor.

A medida que los servicios de IA se vuelven más sofisticados, los algoritmos necesitarán alimentarse con una «dieta de información digital» de mayor calidad. *The Economist* mantiene que esto sólo sería posible si los datos son proporcionados por los usuarios previo pago. Me temo que con un internet de cosas conectadas que facilita toda una panoplia de datos en contexto, esta afirmación se tambalea.

En este mundo distópico en el que los humanos —como si de un episodio de *Black Mirror* se tratase—, los trabajadores, comenzarían su jornada laboral comprobando sus tareas para el día, facilitadas por su sindicato de datos, en la que podrían elegir desde mirar publicidad —mientras la cámara del ordenador recogería las reacciones faciales— hasta traducir un texto a un idioma raro, pasando por comprobar lo fácil o difícil que es explorar un edificio virtual, el trabajador tendría un panel de control en el que vería sus ganancias, un histórico de su comportamiento, su ranking e, incluso, sugerencias para adquirir nuevas habilidades.

Para poder comerciar con los datos, los titulares de los mismos habrían de ser conscientes de su valor y de su importancia, algo que, de ocurrir, será más tarde que pronto. Incluso los que dicen ser conscientes de su importancia, los regalan con total inconsciencia, convirtiéndose así en socios fundadores de la «paradoja de la privacidad».

Uno se pregunta si con este panorama distópico quedará alguien a quien venderle los servicios y los productos obtenidos con estos avanzados análisis conductuales.