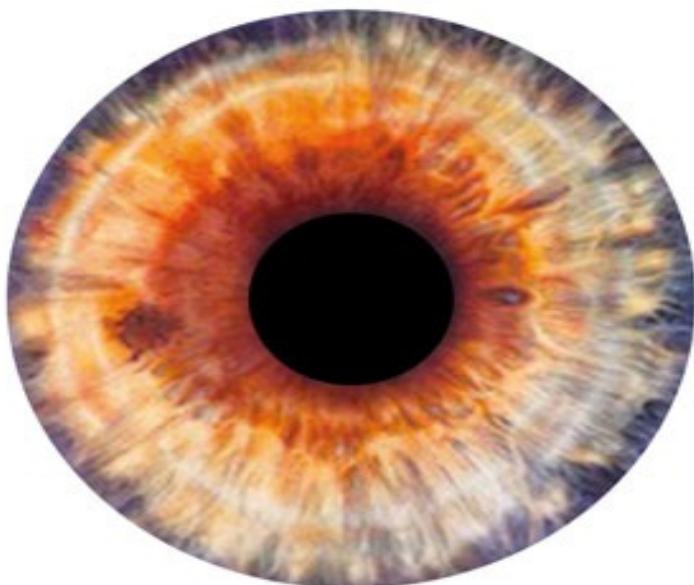


Del autor de *Sapiens*

**Yuval Noah Harari**

21 lecciones  
para  
el siglo XXI



**DEBATE**

## Libertad

### Los macrodatos están observándote

El relato liberal considera la libertad humana el valor más importante. Aduce que toda autoridad surge en último término del libre albedrío de los individuos humanos, que se expresa en sus sentimientos, deseos y opciones. En política, el liberalismo cree que el votante sabe lo que le conviene. Por tanto, defiende las elecciones democráticas. En economía, el liberalismo mantiene que el cliente siempre tiene la razón. Por tanto, da la bienvenida a los principios del mercado libre. En cuestiones personales, el liberalismo anima a las personas a que se escuchen a sí mismas, a que sean fieles a sí mismas y a que sigan los dictados de su corazón, siempre y cuando no vulneren las libertades de los demás. Esta libertad personal queda consagrada en los derechos humanos.

En el discurso político occidental el término «liberal» se usa a veces hoy en día en un sentido partidista mucho más estricto, para denotar a los que apoyan causas específicas como el matrimonio gay, el control de las armas y el aborto. Pero la mayoría de los llamados conservadores también defienden la amplia visión liberal del mundo. Sobre todo en Estados Unidos, tanto republicanos como demócratas deberían tomarse de vez en cuando un respiro de sus acaloradas disputas para recordarse que todos están de acuerdo en cuestiones fundamentales como las elecciones libres, la judicatura independiente y los derechos humanos.

En particular, es vital recordar que héroes de la derecha, como Ronald Reagan

y Margaret Thatcher, fueron grandes adalides no solo de las libertades económicas, sino también de las individuales. En una famosa entrevista de 1987, Thatcher dijo: «No existe tal cosa como la sociedad. Existe un tapiz vivo de hombres y mujeres, [...] y la calidad de nuestra vida depende de lo mucho que cada uno esté preparado para responsabilizarse de sí mismo».[\[1\]](#)

Los herederos de Thatcher en el Partido Conservador estaban totalmente de acuerdo con el Partido Laborista en que la autoridad política procede de los sentimientos, las opciones y el libre albedrío de los votantes individuales. Así, cuando Gran Bretaña necesitó decidir si debía abandonar la Unión Europea, el primer ministro David Cameron no pidió a la reina Isabel II, al arzobispo de Canterbury ni a los rectores de Oxford y Cambridge que resolvieran la cuestión. Ni siquiera a los miembros del Parlamento. En cambio, convocó un referéndum en que a todos y a cada uno de los británicos se les preguntó: «¿Qué opina sobre la cuestión?».

El lector podría objetar que a la gente se le tenía que haber preguntado: «¿Qué piensa?» en lugar de: «¿Qué opina?», pero este es un error común. Los referéndums y las elecciones tienen siempre que ver con los sentimientos humanos, no con la racionalidad humana. Si la democracia fuera un asunto de toma de decisiones racionales, no habría ninguna razón para conceder a todas las personas los mismos derechos de voto o quizá ningún derecho de voto. Existe evidencia sobrada de que algunas personas están más informadas y son más racionales que otras, y en especial cuando se trata de cuestiones económicas y políticas específicas.[\[2\]](#) Después de la votación sobre el Brexit, el eminente biólogo Richard Dawkins protestó diciendo que nunca se le hubiera debido pedir a la inmensa mayoría de la opinión pública británica (él incluido) que votara en referéndum, porque carecían de los conocimientos suficientes de economía y ciencia política. «Por la misma razón podría convocarse un plebiscito nacional para decidir si Einstein hizo correctamente sus cálculos algebraicos, o dejar que los pasajeros de un avión votaran en qué pista debería aterrizar el piloto.»[\[3\]](#)

Sin embargo, para lo bueno y para lo malo, las elecciones y los referéndums

no tratan de lo que pensamos. Tratan de lo que sentimos. Y cuando la cosa va de sentimientos, Einstein y Dawkins no son mejores que cualquier hijo de vecino. La democracia da por sentado que los sentimientos humanos reflejan un «libre albedrío» misterioso y profundo, que este «libre albedrío» es el origen último de la autoridad, y que mientras que algunas personas son más inteligentes que otras, todos los humanos son igualmente libres. Como Einstein y Dawkins, una sirvienta analfabeta también tiene libre albedrío, de modo que el día de las elecciones sus sentimientos (representados por su voto) cuentan tanto como los de cualquier otra persona.

Los sentimientos guían no solo a los votantes, sino también a los líderes. En el referéndum sobre el Brexit de 2016, la campaña del *Leave* estaba encabezada a la vez por Boris Johnson y Michael Gove. Tras la dimisión de David Cameron, Gove apoyó inicialmente a Johnson para el puesto de primer ministro, pero en el último minuto Gove declaró que Johnson era inadecuado para el cargo y anunció su propia intención de presentarse para el puesto. La acción de Gove, que acabó con las opciones de Johnson, se describió como un asesinato político maquiavélico.<sup>[4]</sup> Pero Gove defendió su conducta recurriendo a sus sentimientos, al explicar: «En cada fase de mi vida política me he hecho una misma pregunta: «¿Qué es lo correcto? ¿Qué me dice el corazón?»».<sup>[5]</sup> Esta es la razón por la que, según Gove, luchó con tanto ahínco por el Brexit, y por la que se sintió obligado a traicionar a su antiguo aliado Boris Johnson y a competir él mismo por la posición de macho alfa: porque su corazón le dijo que lo hiciera.

Esta confianza en el corazón puede ser el talón de Aquiles de la democracia liberal. Porque una vez que alguien (ya sea en Pekín o en San Francisco) disponga de la capacidad tecnológica de acceder al corazón humano y manipularlo, la política democrática se transformará en un espectáculo de títeres emocional.

## ESCUCHA EL ALGORITMO

La creencia liberal en los sentimientos y las opciones libres de los individuos no es natural ni muy antigua. Durante miles de años la gente creyó que la autoridad procedía de leyes divinas y no del corazón humano, y que por tanto debíamos santificar la palabra de Dios y no la libertad humana. Solo en los últimos siglos el origen de la autoridad pasó de las deidades celestiales a los humanos de carne y hueso.

La autoridad puede cambiar de nuevo pronto: de los humanos a los algoritmos. De la misma manera que la autoridad divina estaba legitimada por mitologías religiosas y la autoridad humana estaba justificada por el relato liberal, así la revolución tecnológica que se avecina podría establecer la autoridad de los algoritmos de macrodatos, al tiempo que socavaría la idea misma de la libertad individual.

Tal como hemos indicado en el capítulo anterior, los descubrimientos científicos sobre la manera en que nuestro cerebro y nuestro cuerpo funcionan sugerirían que nuestros sentimientos no son una cualidad espiritual exclusivamente humana y que no reflejan ningún tipo de «libre albedrío». Por el contrario, los sentimientos son mecanismos bioquímicos que todos los mamíferos y aves emplean para calcular rápidamente probabilidades de supervivencia y de reproducción. Los sentimientos no están basados en la intuición, la inspiración o la libertad; están basados en el cálculo.

Cuando un mono, un ratón o un humano ve una serpiente, el miedo aflora porque millones de neuronas calculan muy deprisa en el cerebro los datos relevantes y concluyen que la probabilidad de muerte es elevada. Los sentimientos de atracción sexual surgen cuando otros algoritmos bioquímicos calculan que un individuo cercano ofrece una probabilidad elevada de apareamiento exitoso, de vinculación social o de otro objetivo ansiado. Los sentimientos morales, como la indignación, el remordimiento o el perdón, se

derivan de mecanismos neurales que surgieron por evolución para permitir la cooperación en grupo. Todos estos algoritmos bioquímicos se perfeccionaron a lo largo de millones de años de evolución. Si los sentimientos de algún antiguo antepasado cometieron una equivocación, los genes que los modelaron no pasaron a la siguiente generación. Así, los sentimientos no son lo opuesto a la racionalidad: encarnan la racionalidad evolutiva.

Por lo general no nos damos cuenta de que los sentimientos son en realidad cálculos, porque el rápido proceso del cálculo tiene lugar muy por debajo de nuestro umbral de la conciencia. No sentimos los millones de neuronas en el cerebro que computan probabilidades de supervivencia y reproducción, de modo que creemos erróneamente que nuestro miedo a las serpientes, nuestra elección de pareja sexual o nuestras opiniones sobre la Unión Europea son el resultado de algún misterioso «libre albedrío».

No obstante, aunque el liberalismo se equivoca al pensar que nuestros sentimientos reflejan un libre albedrío, hasta el día de hoy todavía tenía un buen sentido práctico. Porque aunque no había nada mágico o libre en nuestros sentimientos, eran el mejor método en el universo para decidir qué estudiar, con quién casarse y a qué partido votar. Y ningún sistema externo podía esperar comprender mis sentimientos mejor que yo. Aun cuando la Inquisición española o el KGB soviético me espiaran cada minuto del día, carecían del conocimiento biológico y la capacidad de cómputo necesarios para acceder subrepticiamente a los procesos bioquímicos que modelan mis deseos y opciones. A efectos prácticos, era razonable argumentar que poseía libre albedrío, porque mi deseo estaba conformado principalmente por la interacción de fuerzas internas, que nadie externo a mí podía ver. Puedo gozar de la ilusión de que controlo mi liza interna y secreta, mientras que los extraños jamás podrán comprender en verdad lo que ocurre en mí y cómo tomo las decisiones.

En consecuencia, el liberalismo estaba en lo cierto al aconsejar a la gente que siguiera los dictados de su corazón en lugar de los de algún sacerdote o de algún *apparatchik* del partido. Sin embargo, pronto los algoritmos informáticos podrán

aconsejarnos mejor que los sentimientos humanos. A medida que la Inquisición española y el KGB dejan paso a Google y a Baidu, es probable que el «libre albedrío» quede desenmascarado como un mito, y el liberalismo pueda perder sus ventajas prácticas.

Porque ahora nos hallamos en la confluencia de dos revoluciones inmensas. Por un lado, los biólogos están descifrando los misterios del cuerpo humano, y en particular del cerebro y los sentimientos. Al mismo tiempo, los informáticos nos proporcionan un poder de procesamiento de datos sin precedentes. Cuando la revolución de la biotecnología se fusiona con la revolución de la infotecnología, producirá algoritmos de macrodatos que supervisarán y comprenderán mis sentimientos mucho mejor que yo, y entonces la autoridad pasará probablemente de los humanos a los ordenadores. Es posible que mi ilusión del libre albedrío se desintegre a medida que me tope diariamente con instituciones, compañías y organismos gubernamentales que comprendan y manipulen lo que hasta la fecha era mi fero interno inaccesible.

Esto ya está ocurriendo en el campo de la medicina. Las decisiones médicas más importantes de nuestra vida no dependen de nuestras sensaciones de enfermedad o bienestar, ni siquiera de las predicciones informadas de nuestro médico, sino de los cálculos de ordenadores que comprenden nuestro cuerpo mucho mejor que nosotros. Dentro de unas pocas décadas, algoritmos de macrodatos alimentados por un flujo constante de datos biométricos podrán controlar nuestra salud a todas horas y todos los días de la semana. Podrán detectar el inicio mismo de la gripe, de un cáncer o del Alzheimer mucho antes de que notemos que algo va mal en nosotros. Entonces podrán recomendar tratamientos, dietas y regímenes diarios apropiados, hechos a medida para nuestro físico, nuestro ADN y nuestra personalidad únicos.

La gente gozará de la mejor atención sanitaria de la historia, pero justo por eso es probable que esté enferma todo el tiempo. Siempre hay algo que está mal en algún lugar del cuerpo. Siempre hay algo que puede mejorarse. En el pasado, nos sentíamos perfectamente sanos mientras no sufriésemos dolor o no padeciéramos

una discapacidad aparente como una cojera. Pero en 2050, gracias a sensores biométricos y algoritmos de macrodatos, podrán diagnosticarse y tratarse las enfermedades mucho antes de que generen dolor o produzcan discapacidad. Como resultado, siempre nos encontraremos padeciendo alguna «enfermedad» y siguiendo esta o aquella recomendación algorítmica. Si nos negamos, quizá nuestro seguro sanitario quede invalidado, o nuestro jefe nos despida: ¿por qué habrían de pagar ellos el precio de nuestra testarudez?

Una cosa es seguir fumando a pesar de las estadísticas generales que relacionan el tabaco con el cáncer de pulmón, y otra muy distinta es continuar fumando a pesar de una advertencia concreta de un sensor biométrico que acaba de detectar diecisiete células cancerosas en la parte superior de nuestro pulmón izquierdo. Y si estamos dispuestos a desafiar al sensor, ¿qué haremos cuando el sensor transmita la advertencia a nuestra agencia de seguros, a nuestro jefe o a nuestra madre?

¿Quién dispondrá del tiempo y la energía para ocuparse de todas estas enfermedades? Con toda probabilidad, podremos sencillamente instruir a nuestro algoritmo de salud para que se ocupe de la mayoría de estos problemas como considere conveniente. En el mejor de los casos, enviará actualizaciones periódicas a nuestros teléfonos inteligentes, y nos dirá que «se detectaron y se destruyeron diecisiete células cancerosas». Los hipocondríacos quizá lean con responsabilidad esas actualizaciones, pero la mayoría seguramente las pasaremos por alto de la misma manera que hacemos caso omiso de esos avisos tan fastidiosos del antivirus en nuestros ordenadores.

## EL DRAMA DE LA TOMA DE DECISIONES

Es probable que lo que ya está empezando a ocurrir en medicina ocurra cada vez en más ámbitos. La invención clave es el sensor biométrico, que la gente puede

llevar sobre su cuerpo o dentro del mismo, y que convierte procesos biológicos en información electrónica que los ordenadores pueden almacenar y analizar. Dados los suficientes datos biométricos y la suficiente potencia de cómputo, los sistemas externos de procesamiento de datos pueden acceder a todos nuestros deseos, decisiones y opiniones. Son capaces de saber con exactitud quiénes somos.

La mayoría de la gente no se conoce muy bien a sí misma. Cuando yo tenía veintiún años, comprendí de una vez por todas que era gay, después de varios años de negarme a aceptarlo. Esto no es nada excepcional. Muchos hombres gais pasan toda su adolescencia inseguros sobre su sexualidad. Imagine ahora el lector la situación en 2050, cuando un algoritmo pueda decirle exactamente a un quinceañero en qué lugar se encuentra en un espectro de gais a heterosexuales (e incluso lo flexible que es dicha posición). Quizá el algoritmo nos muestre imágenes o videos de hombres y mujeres atractivos, siga los movimientos de nuestros ojos, la presión sanguínea y la actividad cerebral, y en cuestión de cinco minutos produzca un número en la escala de Kinsey.[\[6\]](#) Esto podría haberme ahorrado años de frustración. Quizá el lector no quiera realizar dicha prueba de forma individual, pero imagínese que se encuentra con un grupo de amigos en la aburrida fiesta de aniversario de Michelle, y que alguien sugiere que todos nos sometamos por turnos a este algoritmo nuevo y genial (y que todos estén alrededor observando los resultados y comentándolos). ¿Acaso el lector se marcharía?

Incluso en el caso de que lo hiciera, y aunque se escondiera de sí mismo y sus compañeros de clase, no podría esconderse de Amazon, Alibaba o la policía secreta. Mientras el lector navega por la web, mira algo en YouTube o lee las noticias de su red social, los algoritmos lo supervisarán y analizarán discretamente, y le dirán a Coca-Cola que si quiere venderle algún refresco, será mejor que en los anuncios utilice al chico descamisado antes que a la chica sin blusa. El lector ni siquiera lo sabrá. Pero ellos sí lo sabrán, y esta información valdrá miles de millones.

Y además, quizá todo esto se haga de manera abierta y la gente comparta su información a fin de obtener mejores recomendaciones, y al final para hacer que el algoritmo tome decisiones por ella. Se empieza por cosas sencillas, como decidir qué película ver. Mientras nos sentamos con un grupo de amigos para pasar una agradable tarde frente al televisor, primero hemos de elegir qué vamos a ver. Hace cincuenta años no teníamos opción, pero hoy en día, con el auge de los servicios de películas a la carta, existen miles de títulos disponibles. Llegar a un acuerdo puede ser bastante difícil, porque mientras que al lector le gustan las películas de ciencia ficción y suspense, Jack prefiere las comedias románticas y Jill vota por pretenciosos filmes franceses. Podría muy bien ocurrir que terminarais aviniéndoos a ver alguna película mediocre de serie B que os decepcione a todos.

Un algoritmo podría ayudar. Podríamos decirle qué películas anteriores nos han gustado de verdad a cada uno y, en función de su base de datos estadística masiva, el algoritmo encontraría entonces la combinación perfecta para el grupo. Por desgracia, es fácil que un algoritmo tan tosco esté mal informado, en particular porque es evidente que los informes personales suelen ser un indicador muy poco fiable de las verdaderas preferencias de la gente. Suele ocurrir que oímos a muchas personas elogiar una determinada película como una obra maestra, nos sentimos obligados a verla y, aunque nos quedamos dormidos a la mitad, no queremos parecer ignorantes, de modo que decimos a todo el mundo que fue una experiencia increíble.[\[7\]](#)

Sin embargo, estos problemas pueden resolverse si simplemente dejamos que el algoritmo recopile datos en tiempo real sobre nosotros mientras vemos los filmes, en lugar de basarnos en nuestros informes personales y dudosos. Para empezar, el algoritmo puede supervisar qué películas vimos enteras y cuáles dejamos a medio ver. Incluso si le decimos a todo el mundo que *Lo que el viento se llevó* es la mejor película jamás rodada, el algoritmo sabrá que nunca pasamos de la primera media hora y nunca vimos en verdad cómo se incendiaba Atlanta.

Pero el algoritmo incluso puede ir mucho más allá. Hoy en día algunos

ingenieros están desarrollando programas informáticos capaces de detectar las emociones humanas sobre la base del movimiento de nuestros ojos y músculos faciales.[\[8\]](#) Añadamos una buena cámara a la televisión y ese programa sabrá qué escenas nos hicieron reír, qué escenas nos entristecieron y qué escenas nos aburrieron. A continuación, conectemos el algoritmo a sensores biométricos, y sabrá de qué modo cada fotograma ha influido en nuestro ritmo cardíaco, nuestra tensión sanguínea y nuestra actividad cerebral. Mientras vemos, pongamos por caso, *Pulp Fiction*, de Tarantino, el algoritmo puede advertir que la escena de la violación nos causó un asomo apenas perceptible de excitación sexual, que cuando Vincent disparó por accidente a la cara de Marvin nos hizo reír de forma culpable y que no captamos el chiste sobre la Gran Hamburguesa Kahuna, pero aun así nos reímos, para no parecer estúpidos. Cuando uno se obliga a reír, emplea circuitos cerebrales y músculos distintos que cuando nos reímos porque algo es realmente divertido. Los humanos no suelen detectar la diferencia. Pero un sensor biométrico podría hacerlo.[\[9\]](#)

La palabra «televisor» procede del griego *tele*, que significa «lejos», y del latín *visio*, «visión». Originalmente se concibió como un artilugio que nos permite ver desde lejos. Pero pronto nos permitirá que seamos vistos desde lejos. Tal como George Orwell imaginó en *1984*, la televisión nos estará observando mientras la vemos. Una vez hayamos visto toda la filmografía de Tarantino, quizá podamos olvidar la mayor parte de ella. Pero Netflix o Amazon o quienquiera que posea el algoritmo de la televisión conocerá nuestro tipo de personalidad y cómo pulsar nuestros botones emocionales. Estos datos pueden permitir a Netflix y a Amazon elegir filmes para nosotros con precisión asombrosa, pero también puede permitirles que tomen por nosotros las decisiones más importantes de nuestra vida, como qué estudiar, dónde trabajar y con quién casarnos.

Por supuesto, Amazon no acertará siempre. Eso es imposible. Los algoritmos cometerán errores repetidamente debido a datos insuficientes, a programación defectuosa, a definiciones confusas de los objetivos y a la naturaleza caótica de

la vida.[\[10\]](#) Pero Amazon no tiene que ser perfecto. Solo necesita ser, de media, mejor que nosotros, los humanos. Y eso no es muy difícil, porque la mayoría de las personas no se conocen muy bien a sí mismas, y la mayoría de las personas suelen cometer terribles equivocaciones en las decisiones más importantes de su vida. Más incluso que los algoritmos, los humanos adolecen de insuficiencia de datos, de programación (genética y cultural) defectuosa, de definiciones confusas y del caos de la vida.

El lector podría hacer la lista de los muchos problemas que afectan a los algoritmos, y llegar a la conclusión de que las personas nunca confiarán en ellos. Pero eso es un poco como catalogar todos los inconvenientes de la democracia y concluir que ninguna persona en sus cabales elegiría nunca defender un sistema de este tipo. Es sabido que Winston Churchill dijo que la democracia es el peor sistema político del mundo, con excepción de todos los demás. Acertadamente o no, la gente podría llegar a las mismas conclusiones acerca de los macrodatos: tienen muchísimas trabas, pero carecemos de una alternativa mejor.

A medida que los científicos conozcan cada vez mejor la manera en que los humanos toman decisiones, es probable que la tentación de basarse en algoritmos aumente. Acceder a la toma de decisiones de los humanos no solo hará que los algoritmos de macrodatos sean más fiables, sino que los sentimientos humanos sean menos fiables. A medida que gobiernos y empresas consigan acceder al sistema operativo humano, estaremos expuestos a una andanada de manipulación, publicidad y propaganda dirigidos con precisión. Nuestras opiniones y emociones podrían resultar tan fáciles de manipular que nos viéramos obligados a fiarnos de los algoritmos de la misma manera que un piloto que sufre un ataque de vértigo no ha de hacer caso de lo que sus propios sentidos le dicen y debe depositar toda su confianza en la maquinaria.

En algunos países y en determinadas situaciones, quizá a la gente no se le dé ninguna opción, y esta se vea obligada a obedecer las decisiones de los algoritmos de macrodatos. Pero incluso en sociedades supuestamente libres, los algoritmos pueden ir ganando autoridad debido a que aprenderemos por

experiencia a confiar en ellos en cada vez más cuestiones, y poco a poco perderemos nuestra capacidad para tomar decisiones por nosotros mismos. Piense simplemente el lector en la manera en que, en las dos últimas décadas, miles de millones de personas han llegado a confiar al algoritmo de búsqueda de Google una de las tareas más importantes de todas: buscar información relevante y fidedigna. Ya no buscamos información. En lugar de ello, «googleamos». Y a medida que confiamos cada vez más en Google para hallar respuestas, nuestra capacidad para buscar información por nosotros mismos disminuye. Ya hoy en día, la «verdad» viene definida por los primeros resultados de la búsqueda de Google.[\[11\]](#)

Esto ha ido ocurriendo también con las capacidades físicas, como el espacio para orientarse y navegar. La gente pide a Google que la guíe cuando conduce. Cuando llega a una intersección, su instinto puede decirle: «Gira a la izquierda», pero Google Maps le dice: «Gire a la derecha». Al principio hacen caso a su instinto, giran a la izquierda, quedan atascados en un embotellamiento de tráfico y no llegan a tiempo a una reunión importante. La próxima vez harán caso a Google, girarán a la derecha y llegarán a tiempo. Aprenden por experiencia a confiar en Google. Al cabo de uno o dos años, se basan a ciegas en lo que les dice Google Maps, y si el teléfono inteligente falla, se encuentran completamente perdidos.

En marzo de 2012, tres turistas japoneses que viajaban por Australia decidieron realizar una excursión de un día a una pequeña isla situada lejos de la costa, y acabaron con su coche dentro del océano Pacífico. La conductora, Yuzu Noda, de veintiún años, dijo después que no había hecho más que seguir las instrucciones del GPS: «Nos dijo que podríamos conducir hasta allí. No dejaba de decir que nos llevaría a una carretera. Quedamos atrapados».[\[12\]](#) En varios incidentes parecidos, los conductores acabaron dentro de un lago, o cayeron desde lo alto de un puente demolido, aparentemente por haber seguido las instrucciones del GPS.[\[13\]](#) La capacidad de orientarse es como un músculo: o lo usas o lo pierdes.[\[14\]](#) Lo mismo puede decirse de la capacidad de elegir cónyuge

o profesión.

Todos los años, millones de jóvenes necesitan decidir qué estudiar en la universidad. Esta es una decisión muy importante y difícil. Los jóvenes se encuentran sometidos a la presión de sus padres, sus amigos y sus profesores, que tienen intereses y opiniones diferentes. Los jóvenes deben también enfrentarse a sus propios temores y fantasías. Su juicio está ofuscado y manipulado por éxitos de taquilla de Hollywood, malas novelas y refinadas campañas publicitarias. Es particularmente difícil tomar una decisión sensata porque los interesados no saben en realidad qué hace falta para medrar con éxito en las diferentes profesiones y tampoco tienen necesariamente una imagen realista de sus propias fortalezas y debilidades. ¿Qué se necesita para triunfar como abogado? ¿Cómo me comportaré bajo presión? ¿Sabré trabajar bien en equipo?

Una estudiante puede empezar la carrera de Derecho porque posee una imagen inexacta de sus propias capacidades, y una visión todavía más distorsionada de lo que implica en verdad ser abogado (no se sueltan discursos espectaculares ni se grita «¡Protesto, señoría!» a todas horas). Mientras tanto, su amiga decide cumplir un sueño de la infancia y estudiar ballet de manera profesional, aunque carece de la estructura ósea y la disciplina necesarias. Años más tarde, ambas lamentan mucho su elección. En el futuro, confiaremos en que Google tome estas decisiones por nosotros. Google podrá decirme que perderé el tiempo en la Facultad de Derecho o en la academia de ballet, pero que podré ser una excelente (y muy feliz) psicóloga o fontanera.[\[15\]](#)

Una vez que la IA decida mejor que nosotros las carreras e incluso las relaciones, nuestro concepto de la humanidad y de la vida tendrá que cambiar. Los humanos están acostumbrados a pensar en la existencia como un drama de toma de decisiones. La democracia liberal y el capitalismo de libre mercado ven al individuo como un agente autónomo que no para de tomar decisiones sobre el mundo. Las obras de arte (ya sean las piezas teatrales de Shakespeare, las novelas de Jane Austen o las chabacanas comedias de Hollywood) suelen

centrarse en que el o la protagonista ha de tomar alguna decisión particularmente crucial. ¿Ser o no ser? ¿Hacer caso a mi mujer y matar al rey Duncan, o hacer caso a mi conciencia y perdonarlo? ¿Casarme con el señor Collins o con el señor Darcy? Las teologías cristiana y musulmana se centran de manera parecida en el drama de la toma de decisiones, y aducen que la salvación o la condena eternas dependen de haber tomado la decisión correcta.

¿Qué pasará con esta forma de entender la vida si cada vez confiamos más en la IA para que tome las decisiones por nosotros? En la actualidad nos fiamos de Netflix para que nos recomiende películas y de Google Maps para elegir si giramos a la derecha o a la izquierda. Pero una vez que empecemos a contar con la IA para decidir qué estudiar, dónde trabajar y con quién casarnos, la vida humana dejará de ser un drama de toma de decisiones. Las elecciones democráticas y los mercados libres tendrán poco sentido. Lo mismo ocurrirá con la mayoría de las religiones y de las obras de arte. Imagine el lector a Anna Karénina sacando su teléfono inteligente y preguntándole al algoritmo de Facebook si debe seguir casada con Karenin o fugarse con el conde Vronsky. O imagine el lector su obra teatral favorita de Shakespeare con todas las decisiones cruciales tomadas por el algoritmo de Google. Hamlet y Macbeth llevarían una vida mucho más confortable, pero ¿qué tipo de vida sería, exactamente? ¿Tenemos modelos para dar sentido a una existencia de este tipo?

Cuando la autoridad se transfiera de los humanos a los algoritmos, quizás ya no veamos el mundo como el patio de juegos de individuos autónomos que se esfuerzan para tomar las decisiones correctas. En lugar de ello, podríamos percibir todo el universo como un flujo de datos, concebir los organismos como poco más que algoritmos bioquímicos y creer que la vocación cósmica de la humanidad es crear un sistema de procesamiento de datos que todo lo abarque y después fusionarnos con él. Hoy en día ya nos estamos convirtiendo en minúsculos chips dentro de un gigantesco sistema de procesamiento de datos que nadie entiende en realidad. A diario absorbo innumerables bits de datos mediante correos electrónicos, tuits y artículos. No sé exactamente dónde encajo yo en el

gran esquema de las cosas, ni cómo mis bits de datos se conectan con los bits producidos por miles de millones de otros humanos y de ordenadores. No tengo tiempo de descubrirlo, porque estoy demasiado ocupado contestando a todos estos correos electrónicos.

## EL COCHE FILOSÓFICO

La gente tal vez objetará que los algoritmos nunca podrán tomar decisiones importantes por nosotros, porque las decisiones importantes suelen implicar una dimensión ética, y los algoritmos no entienden de ética. Pero no hay ninguna razón para suponer que no serán capaces de superar al humano medio incluso en ética. Ya hoy en día, cuando dispositivos como los teléfonos inteligentes y los vehículos autónomos toman decisiones que solían ser monopolio humano, empiezan a habérselas con el mismo tipo de problemas éticos que han atormentado a los humanos durante milenios.

Por ejemplo, supongamos que dos chicos que persiguen una pelota saltan delante de un automóvil autónomo. Basándose en sus cálculos instantáneos, el algoritmo que conduce el coche concluye que la única manera de evitar atropellar a los chicos es virar bruscamente al carril opuesto, y arriesgarse a colisionar con un camión que viene en sentido contrario. El algoritmo calcula que en tal caso existe un 70 por ciento de probabilidades de que el propietario del coche (que está profundamente dormido en el asiento posterior) muera en el impacto. ¿Qué debería hacer el algoritmo?[\[16\]](#)

Los filósofos llevan milenios debatiendo sobre estos «problemas del tranvía» (se llaman «problemas del tranvía» porque los ejemplos de manual en los debates filosóficos modernos se refieren a un tranvía fuera de control que se precipita por las vías, en lugar de a un automóvil autónomo).[\[17\]](#) Hasta ahora, resulta vergonzoso que estos debates hayan tenido poquísimaa influencia sobre el

comportamiento real, porque en épocas de crisis los humanos suelen olvidar con demasiada frecuencia sus opiniones filosóficas y en cambio siguen sus emociones e instintos viscerales.

Uno de los experimentos más desagradables en la historia de las ciencias sociales se realizó en diciembre de 1970 con un grupo de estudiantes del Seminario Teológico de Princeton, que se estaban preparando para convertirse en ministros de la Iglesia presbiteriana. A cada estudiante se le pidió que se dirigiera apresuradamente a un aula alejada, y que allí diera una charla sobre la parábola del Buen Samaritano, que cuenta que un judío que viajaba de Jerusalén a Jericó fue asaltado y robado por criminales, que lo dejaron moribundo junto al camino. Después de algún tiempo, un sacerdote y un levita pasaron cerca, pero ninguno de ellos hizo caso del hombre. En cambio, un samaritano (un miembro de una secta muy despreciada por los judíos) se detuvo cuando vio a la víctima, cuidó de ella y le salvó la vida. La moraleja de la parábola es que el mérito de la gente ha de juzgarse por su comportamiento real y no por su filiación religiosa ni por sus opiniones filosóficas.

Los jóvenes e impacientes seminaristas corrieron al aula, mientras en el trayecto iban pensando cómo explicar mejor la parábola del Buen Samaritano. Pero los investigadores dispusieron en su ruta a una persona vestida con andrajos, que estaba sentada despatarrada en un portal, con la cabeza gacha y los ojos cerrados. Cada vez que un incauto seminarista pasaba rápidamente por su lado, la «victima» tosía y gemía de forma lastimosa. La mayoría de los seminaristas ni siquiera se detuvieron para preguntar al hombre qué le pasaba, y mucho menos le ofrecieron ayuda. El estrés emocional generado por la necesidad de correr hasta el aula superó a su obligación moral de ayudar a un desconocido en apuros.[\[18\]](#)

Las emociones humanas superan a las teorías filosóficas en muchas otras situaciones. Esto hace que la historia ética y filosófica del mundo sea un relato bastante deprimente de ideas maravillosas y de comportamientos menos que ideales. ¿Cuántos cristianos ofrecen ahora la otra mejilla, cuántos budistas se

elevan en realidad por encima de las obsesiones egoístas y cuántos judíos aman realmente a sus vecinos como a sí mismos? Esta es justo la manera como la selección natural ha modelado a *Homo sapiens*. Al igual que todos los mamíferos, *Homo sapiens* emplea las emociones para tomar rápidas decisiones de vida o muerte. Hemos heredado nuestra ira, nuestro miedo y nuestro deseo de millones de antepasados, los cuales pasaron los controles de calidad más rigurosos de la selección natural.

Por desgracia, lo que era bueno para la supervivencia y la reproducción en la sabana africana hace un millón de años no tiene por qué dar lugar necesariamente a un comportamiento responsable en las carreteras del siglo XXI. Todos los años, conductores humanos distraídos, enfadados y ansiosos matan a más de un millón de personas en accidentes de tráfico. Podemos enviar a todos nuestros filósofos, profetas y sacerdotes a que prediquen ética a dichos conductores, pero en la carretera seguirán predominando las emociones propias de los mamíferos y los instintos de la sabana. En consecuencia, los seminaristas apresurados no prestarán atención a personas en apuros y los conductores en un momento crítico atropellarán a infortunados peatones.

Esta disyunción entre el seminario y la carretera es uno de los mayores problemas prácticos en la ética. Immanuel Kant, John Stuart Mill y John Rawls ya pueden acomodarse en alguna acogedora aula universitaria y discutir durante días problemas teóricos de ética, pero ¿en verdad los conductores estresados aplicarán sus conclusiones en plena emergencia y en una fracción de segundo? Quizá Michael Schumacher (el campeón de Fórmula Uno del que a veces se ha dicho que ha sido el mejor conductor de la historia) tenía la capacidad de pensar en filosofía mientras conducía su coche, pero la mayoría no somos Schumacher.

De todos modos, los algoritmos informáticos no han sido conformados por la selección natural, y no tienen emociones ni instintos viscerales. De ahí que, en momentos críticos, puedan seguir directrices éticas mucho mejor que los humanos, siempre que encontrremos una manera de codificar la ética en números

y estadísticas precisos. Si enseñáramos a Kant, Mill y Rawls a escribir programas informáticos, podrían programar el automóvil autónomo en su confortable laboratorio, y estar seguros de que el coche seguiría sus órdenes en la autopista. En efecto, cada coche será conducido por Michael Schumacher e Immanuel Kant unidos en una única entidad.

Así, si programamos un automóvil autónomo para que se detenga y ayude a extraños en apuros, lo hará contra viento y marea (a menos, desde luego, que insertemos una cláusula de excepción para situaciones de vendavales y marejadas). De manera parecida, si nuestro automóvil autónomo se halla programado para pasar al otro carril a fin de esquivar a dos chicos situados en su trayectoria, podemos apostar la vida a que será justo esto lo que hará. Lo que significa que cuando diseñen su automóvil autónomo, Toyota o Tesla transformarán un problema teórico de la filosofía de la ética en un problema práctico de ingeniería.

Sin duda, los algoritmos filosóficos nunca serán perfectos. Todavía habrá errores, que acarrearán heridos, muertos y pleitos complicadísimos. (Por primera vez en la historia, podremos demandar a un filósofo por las desafortunadas consecuencias de sus teorías, porque por primera vez en la historia podremos demostrar una conexión causal directa entre ideas filosóficas y acontecimientos de la vida real.) Sin embargo, para ocupar el lugar de los conductores humanos, los algoritmos no tienen que ser perfectos. Solo mejor que los humanos. Dado que los conductores humanos matan al año a más de un millón de personas, no es pedir demasiado. A fin de cuentas, ¿preferirá el lector que el coche que está junto al suyo lo conduzca un adolescente ebrio o el equipo Schumacher-Kant?

[19]

La misma lógica puede aplicarse no solo a la conducción de automóviles, sino a muchas otras situaciones. Pongamos como ejemplo la solicitud de empleo. En el siglo XXI, la decisión de contratar o no a alguien para un puesto de trabajo la tomarán cada vez más los algoritmos. No podemos basarnos en la máquina para

establecer criterios éticos relevantes: será necesario que esto sigan haciéndolo los humanos. Pero una vez que hayamos decidido acerca de un criterio ético en el mercado laboral (por ejemplo, que está mal discriminar a los negros y a las mujeres), podemos confiar en las máquinas para que implementen y mantengan ese criterio mejor que los humanos.[\[20\]](#)

Un gestor humano puede saber e incluso estar de acuerdo en que no es ético discriminar a los negros y a las mujeres, pero cuando una mujer negra solicita un empleo, el gestor la discrimina de forma inconsciente, y decide no contratarla. Si permitimos que un ordenador evalúe solicitudes de empleo y lo programamos para que no tenga en absoluto en cuenta raza ni género, no cabe duda de que el ordenador en verdad pasará por alto estos factores, porque los ordenadores no tienen subconsciente. Desde luego, no será fácil diseñar un programa para evaluar solicitudes de empleo, y siempre existirá el peligro de que los ingenieros introduzcan de alguna manera sus propios prejuicios inconscientes en el programa.[\[21\]](#) Pero una vez que hayamos descubierto tales errores, seguramente será mucho más fácil corregir el programa que librar a los humanos de sus prejuicios racistas y misóginos.

Hemos visto que el auge de la inteligencia artificial podría expulsar a la mayoría de los humanos del mercado laboral, entre ellos a conductores y policía de tráfico (cuando los humanos pendencieros sean sustituidos por algoritmos obedientes, la policía de tráfico no será necesaria). Sin embargo, podría haber algunas nuevas vacantes para los filósofos, porque sus habilidades (que hasta ahora carecen de mucho valor de mercado) de repente serán muy demandadas. Así, si queremos estudiar algo que garantice un buen empleo en el futuro, quizá la filosofía no sea una mala apuesta.

Desde luego, los filósofos rara vez se ponen de acuerdo en el procedimiento adecuado. Pocos «problemas del tranvía» se han resuelto a gusto de todos los filósofos, y pensadores consecuencialistas como John Stuart Mill (que juzga las acciones por sus consecuencias) sostienen opiniones muy diferentes a los deontologistas como Immanuel Kant (que juzga las acciones mediante reglas

absolutas). ¿Tendría que posicionarse realmente Tesla respecto a estos asuntos tan enrevesados para producir un automóvil?

Bueno, quizá Tesla simplemente deje esta cuestión al mercado. Fabricará dos modelos de automóvil autónomo: el Tesla Altruista y el Tesla Egoísta. En una emergencia, el Altruista sacrifica a su dueño por un bien mayor, mientras que el Egoísta hace cuanto está en su mano para salvar a su dueño, incluso si ello significa matar a los dos chicos. Entonces los clientes podrán comprar el coche que mejor se adapte a su opinión filosófica favorita. Si hay más personas que compran el Tesla Egoísta, no podremos culpar a Tesla. Al fin y al cabo, el cliente siempre tiene razón.

No es un chiste. En un estudio pionero de 2015, a unas personas se les presentó una situación hipotética en la que un automóvil autónomo estaba a punto de atropellar a varios peatones. La mayoría dijeron que en este caso el coche tenía que salvar a los peatones aun a costa de matar a su dueño. Cuando después se les preguntó si ellos comprarían un automóvil programado para sacrificar a su dueño por un bien mayor, la mayoría contestaron que no. Para ellos, preferirían el Tesla Egoísta.[\[22\]](#)

Imagine la lectora la situación: ha comprado un coche nuevo, pero antes de empezar a usarlo, ha de meterse en el menú de configuración y elegir una de varias opciones. En caso de un accidente, ¿quiere que el automóvil sacrifique su vida o que mate a la familia del otro vehículo? ¿Es esta una elección que desearía hacer? Piense solo en las discusiones que habrá de tener con su marido acerca de qué opción elegir.

¿Acaso el Estado debería intervenir para regular el mercado e imponer un código ético para todos los automóviles autónomos? Algunos legisladores estarían sin duda entusiasmados con la oportunidad de promulgar por fin leyes que siempre se cumplan al pie de la letra. Otros legisladores podrían sentirse alarmados por semejante responsabilidad totalitaria y sin precedentes. Después de todo, a lo largo de la historia las limitaciones del cumplimiento de las leyes han sido un freno bien acogido a los prejuicios, errores y excesos de los

legisladores. Fue una gran fortuna que leyes contra la homosexualidad y contra la blasfemia se hicieran cumplir solo parcialmente. ¿De verdad queremos un sistema en que las decisiones de políticos falibles se conviertan en algo tan inexorable como la fuerza de la gravedad?

## DICTADURAS DIGITALES

La IA suele asustar a la gente porque esta no cree que vaya a ser siempre obediente. Hemos visto demasiadas películas de ciencia ficción sobre robots que se rebelan contra sus amos humanos, que corren descontrolados por las calles matando a todo el mundo. Pero el problema real con los robots es justo el contrario. Debemos temerlos porque probablemente obedecerán siempre a sus amos y nunca se rebelarán.

No hay nada malo en la obediencia ciega, desde luego, mientras los robots sirvan a amos benignos. Incluso en la guerra, basarse en robots asesinos puede asegurar que, por primera vez en la historia, las leyes de la guerra se respeten de verdad en el campo de batalla. A veces los soldados humanos se dejan llevar por sus emociones y asesinan, saquean y violan, transgrediendo así las leyes de la guerra. Solemos asociar las emociones con la compasión, el amor y la empatía, pero en tiempos de guerra, las emociones que predominan son con demasiada frecuencia el miedo, el odio y la crueldad. Puesto que los robots carecen de emociones, puede confiarse en que siempre cumplirán al pie de la letra el código militar, y que nunca se dejarán influir por temores y odios personales.[\[23\]](#)

El 16 de marzo de 1968, en la aldea sudvietnamita de My Lai, los soldados norteamericanos de un regimiento se volvieron locos y aniquilaron a unos 400 civiles. Este crimen de guerra fue el resultado de la iniciativa local de hombres que habían estado combatiendo durante varios meses en una guerra de guerrillas en la jungla. No tuvo ninguna finalidad estratégica, y contravino tanto el código

legal como la política militar estadounidenses. Se debió a emociones humanas. [24] Si Estados Unidos hubiera hecho uso de robots asesinos en Vietnam, la masacre de My Lai jamás se hubiera producido.

No obstante, antes de que nos apresuremos a desarrollar y a usar robots asesinos, debemos recordar que los robots siempre reflejan y amplifican las cualidades de su programa. Si el programa es mesurado y benévolos, los robots serán probablemente una mejora enorme respecto al soldado humano medio. Pero si el programa es despiadado y cruel, los resultados serán catastróficos. El problema real de los robots no es su propia inteligencia artificial, sino más bien la estupidez y crueldad naturales de sus amos humanos.

En julio de 1995, tropas serbobosnias mataron a más de 8.000 musulmanes bosnios en los alrededores de la ciudad de Srebrenica. A diferencia de la masacre caótica de My Lai, los asesinatos de Srebrenica fueron una operación prolongada y bien organizada que reflejaba la política de los serbobosnios de «limpiar étnicamente» Bosnia de musulmanes.[25] Si los serbobosnios hubieran tenido robots asesinos en 1995, es probable que la atrocidad hubiera sido aún mayor. Ningún robot hubiera dudado ni un momento en cumplir las órdenes que se le hubieran dado, y no hubiera escatimado la vida de un solo niño musulmán por sentimientos de compasión, repulsión o simple letargia.

Un dictador despiadado armado con estos robots asesinos nunca debería temer que sus soldados se volvieran en su contra, con independencia de lo desalmadas y locas que fueran sus órdenes. Un ejército de robots habría sin duda estrangulado en su cuna a la Revolución francesa en 1789, y si en 2011 Hosni Mubarak hubiera dispuesto de un contingente de robots asesinos, habría podido desplegarlos contra el populacho sin temor a su defeción. De forma parecida, un gobierno imperialista que se basara en un ejército de robots podría librarse guerras impopulares sin preocuparse de si los robots pierden la motivación, o de si sus familias organizan protestas. Si Estados Unidos hubiera dispuesto de robots asesinos en la guerra de Vietnam, la masacre de My Lai podría haberse evitado, pero la contienda misma tal vez habría durado muchos más años, porque

el gobierno estadounidense no hubiera tenido tantos problemas con soldados desmoralizados, manifestaciones multitudinarias contra la contienda o un movimiento de «robots veteranos contra la guerra» (algunos ciudadanos norteamericanos quizá aún se habrían opuesto a la guerra, pero sin el temor a ser reclutados, sin el recuerdo de haber cometido atrocidades personalmente y sin la dolorosa pérdida de un familiar querido, los manifestantes habrían sido probablemente menos numerosos y habrían estado menos motivados).[\[26\]](#)

Estos tipos de problemas son mucho menos relevantes para los vehículos civiles autónomos, porque ningún fabricante de coches programaría de forma malévolas sus vehículos para que se dirigieran hacia las personas y las mataran. Pero los sistemas de armas autónomas suponen una catástrofe en ciernes, porque hay demasiados gobiernos que tienden a ser éticamente corruptos o directamente malvados.

El peligro no se limita a las máquinas asesinas. Los sistemas de vigilancia pueden ser igualmente peligrosos. En manos de un gobierno benévolo, los algoritmos de vigilancia potentes quizá sean lo mejor que le haya ocurrido nunca a la humanidad. Pero esos algoritmos de macrodatos podrían asimismo empoderar a un futuro Gran Hermano, de modo que termináramos sometidos a un régimen de vigilancia orwelliana en el que cada uno de los individuos fuera controlado todo el tiempo.[\[27\]](#)

De hecho, podríamos acabar de una manera que ni siquiera Orwell hubiera imaginado: con un régimen de vigilancia global que haga el seguimiento no solo de todas nuestras actividades y nuestras manifestaciones externas, sino que también logre incluso metérseños bajo la piel para conocer nuestras experiencias internas. Considérese por ejemplo lo que el régimen de Kim en Corea del Norte sería capaz de hacer con la nueva tecnología. En el futuro, a cada ciudadano norcoreano se le podría exigir que llevara un brazalete biométrico que supervisara cuanto hiciera y dijera, así como su tensión sanguínea y su actividad cerebral. Mediante el uso de nuestro conocimiento creciente del cerebro humano, y empleando los inmensos poderes del aprendizaje automático, el régimen

norcoreano podría, por primera vez en la historia, evaluar lo que todos y cada uno de los ciudadanos está pensando en cualquier momento. Si miramos una fotografía de Kim Jong-un y los sensores biométricos captan las señales que delatan la ira (aumento de la tensión sanguínea, actividad acrecentada en la amígdala), podríamos vernos en el gulag la mañana siguiente.

Por supuesto, debido a su aislamiento, al régimen de Corea del Norte le costaría desarrollar por sí mismo la tecnología necesaria. Sin embargo, dicha tecnología podría iniciarse en naciones más avanzadas técnicamente, y luego ser copiada o comprada por los norcoreanos y otras dictaduras atrasadas. Tanto China como Rusia están mejorando sin cesar sus instrumentos de vigilancia, como varios países democráticos más, desde Estados Unidos hasta mi patria, Israel. Llamada la «nación emprendedora», Israel tiene un sector de alta tecnología muy dinámico y una industria puntera de ciberseguridad. Al mismo tiempo, está también enzarzada en un conflicto letal con los palestinos, y al menos algunos de sus dirigentes, generales y ciudadanos se pondrían muy contentos si se creara un régimen de vigilancia total en Cisjordania tan pronto como se disponga de la tecnología necesaria.

Ya hoy en día, siempre que los palestinos realizan una llamada telefónica, publican algo en Facebook o viajan de una ciudad a otra, es probable que se los vigile con micrófonos, cámaras, drones o programas espía israelíes. Los datos obtenidos se analizan después mediante algoritmos de macrodatos. Esto ayuda a las fuerzas de seguridad israelíes a precisar y a neutralizar amenazas potenciales sin tener que desplazar demasiados efectivos sobre el terreno. Los palestinos pueden administrar algunas ciudades y pueblos en Cisjordania, pero los israelíes controlan el cielo, las ondas de radio y el ciberespacio. Por tanto, son necesarios muy pocos soldados israelíes para controlar de manera efectiva a alrededor de 2,5 millones de palestinos en Cisjordania.[\[28\]](#)

En un incidente tragicómico acaecido en octubre de 2017, un peón palestino publicó en Facebook una fotografía de sí mismo en su lugar de trabajo, al lado de un bulldócer. Junto a la imagen escribió: «¡Buenos días!». Un algoritmo

automático cometió un pequeño error cuando transliteró las letras arábigas. En lugar de «Ysabechhum!» (que significa «¡Buenos días!»), el algoritmo identificó las letras como «Ydbachhum!» (que significa «¡Mátalos!»). Al sospechar que el hombre podía ser un terrorista que intentaba utilizar un buldócer para atropellar a gente, las fuerzas de seguridad israelíes lo detuvieron de inmediato. Quedó en libertad cuando se dieron cuenta de que el algoritmo había cometido un error. No obstante, el post ofensivo se eliminó de Facebook. Nunca se es demasiado prudente.[\[29\]](#) Lo que los palestinos están viviendo hoy en día en Cisjordania podría ser simplemente un burdo anticipo de lo que miles de millones de personas acabarán por experimentar en todo el planeta.

A finales del siglo xx, las democracias superaban por lo general a las dictaduras porque aquellas eran mejores procesando los datos. La democracia difunde el poder para procesar información y la toma de decisiones se hace entre muchas personas e instituciones, mientras que la dictadura concentra la información y el poder en un punto. Dada la tecnología del siglo xx, era ineficiente concentrar demasiada información y poder en un punto. Nadie tenía la capacidad de procesar toda la información con suficiente rapidez y de tomar las decisiones adecuadas. Esa es una parte de la razón por la que la Unión Soviética tomó decisiones mucho peores que Estados Unidos, y por la que la economía soviética se hallaba muy por detrás de la norteamericana.

Sin embargo, puede que la inteligencia artificial haga que el péndulo se mueva en la dirección opuesta. La IA hace posible procesar cantidades enormes de información de manera centralizada. De hecho, podría lograr que los sistemas centralizados fueran mucho más eficientes que los sistemas difusos, porque el aprendizaje automático funciona mejor cuando es capaz de analizar mucha información. Si concentráramos toda la información relacionada con mil millones de personas en una única base de datos, sin tener en cuenta los problemas de privacidad, podemos preparar algoritmos mucho mejores que si respetamos la intimidad individual y en nuestra base de datos solo disponemos de información

parcial sobre un millón de personas. Por ejemplo, si un gobierno autoritario ordenara a todos sus ciudadanos que analizaran su ADN y que compartieran sus datos médicos con alguna autoridad central, obtendría una ventaja inmensa en genética e investigación médica con respecto a sociedades en las que los datos médicos son estrictamente privados. La principal desventaja de los regímenes autoritarios en el siglo xx (el intento de concentrar toda la información en un punto) podría convertirse en su ventaja decisiva en el siglo xxi.

Cuando los algoritmos lleguen a conocernos tan bien, los gobiernos autoritarios se harán con un control absoluto sobre sus ciudadanos, más incluso que en la Alemania nazi, y la resistencia a tales regímenes podría ser de todo punto imposible. El régimen no solo sabrá exactamente cómo sentimos: podrá hacer que sintamos lo que quiera. El dictador tal vez no sea capaz de proporcionar a los ciudadanos asistencia sanitaria o igualdad, pero podrá hacer que lo amen y que odien a sus oponentes. En su forma actual, la democracia no sobrevivirá a la fusión de la biotecnología y la infotecnología. O bien se reinventa a sí misma con éxito y de una forma radicalmente nueva, o bien los humanos acabarán viviendo en «dictaduras digitales».

Esto no implicará un retorno a la época de Hitler y Stalin. Las dictaduras digitales serán tan diferentes de la Alemania nazi como la Alemania nazi lo era de la Francia del *ancien régime*. Luis XIV fue un autócrata centralizador, pero carecía de la tecnología necesaria para erigir un Estado totalitario moderno. Su reinado no sufrió ninguna oposición, pero en ausencia de radios, teléfonos y trenes, ejercía poco control sobre la vida cotidiana de los campesinos de las remotas aldeas bretonas, o incluso de los ciudadanos del centro de París. No tenía la voluntad ni la capacidad para establecer un partido de masas, un movimiento juvenil que abarcara a todo el país o un sistema de educación nacional.[\[30\]](#) Fueron las nuevas tecnologías del siglo xx las que proporcionaron a Hitler la motivación y el poder para hacer estas cosas. No podemos predecir cuáles serán las motivaciones ni la fuerza de las dictaduras digitales en 2084,

pero es muy poco probable que se limiten a copiar a Hitler y a Stalin. Aquellos que se preparen para volver a librarse las batallas de la década de 1930 podrían ser objeto de un ataque por sorpresa procedente de una dirección por completo distinta.

Incluso si la democracia consigue adaptarse y sobrevivir, las personas podrían ser víctimas de nuevos tipos de opresión y discriminación. En la actualidad, hay cada vez más bancos, empresas e instituciones que emplean algoritmos para analizar datos y tomar decisiones sobre nosotros. Cuando solicitamos un préstamo al banco, es probable que nuestra solicitud sea procesada por un algoritmo y no por un humano. El algoritmo analiza muchísimos datos sobre nosotros y estadísticas acerca de millones de otras personas, y decide si somos lo bastante solventes para concedernos el préstamo. A menudo, el trabajo que realiza el algoritmo es mejor que el de un banquero humano. Pero el problema radica en que si el algoritmo discrimina injustamente a algunas personas, es difícil saberlo. Si el banco se niega a concedernos el préstamo y preguntamos: «¿Por qué?», el banco contesta: «El algoritmo dijo que no». Y entonces preguntamos: «¿Y por qué dijo que no el algoritmo? ¿Qué problema hay?», y el banco responde: «No lo sabemos. No hay ningún humano que entienda este algoritmo, porque se basa en aprendizaje automático avanzado. Pero confiamos en nuestro algoritmo, de modo que no le concederemos el préstamo».[\[31\]](#)

Cuando la discriminación se dirige a grupos enteros, como mujeres o negros, estos grupos pueden organizarse y protestar contra su discriminación colectiva. Pero ahora un algoritmo es capaz de discriminarnos a nosotros de forma individual, y no tenemos ni idea de por qué. Quizá el algoritmo encontró algo en nuestro ADN, nuestra historia personal o nuestra cuenta de Facebook que no le gusta. El algoritmo nos discrimina no porque seamos una mujer o un afroamericano, sino porque somos nosotros. Hay algo específico en nosotros que no le gusta. No sabemos qué es, y aunque lo supiéramos no podríamos organizarnos con otras personas para protestar, porque no hay otras personas que padecan el mismo prejuicio exacto. Solo nosotros. En lugar de simplemente

discriminación colectiva, en el siglo XXI podríamos enfrentarnos a un problema creciente de discriminación individual.[\[32\]](#)

En las más altas esferas de la autoridad, probablemente seguirá habiendo hombres de paja humanos, que nos generarán la ilusión de que los algoritmos solo son consejeros y que la autoridad última se halla todavía en manos humanas. No designaremos una IA como el canciller de Alemania o el director ejecutivo de Google. Sin embargo, las decisiones que tomen el canciller y el director ejecutivo estarán determinadas por la IA. El canciller podrá elegir todavía entre varias opciones, pero todas serán el resultado del análisis de macrodatos y reflejarán más la manera como la IA entiende el mundo que la manera como lo entienden los humanos.

Por poner un ejemplo análogo: en la actualidad los políticos de todo el mundo pueden elegir entre varias políticas económicas diferentes, pero en casi todos los casos las diversas políticas que se ofrecen reflejan una perspectiva capitalista de la economía. Los políticos creen que están eligiendo, pero las decisiones realmente importantes ya las han tomado mucho antes los economistas, banqueros y empresarios que modelaron las diferentes opciones en el menú. Dentro de un par de décadas, los políticos podrían encontrarse eligiendo de un menú escrito por la IA.

## INTELIGENCIA ARTIFICIAL Y ESTUPIDEZ NATURAL

Una buena noticia es que al menos en las próximas décadas no tendremos que habérnoslas con la elaborada pesadilla de la ciencia ficción en la que la IA adquiera conciencia y decida esclavizar o aniquilar a la humanidad. Cada vez nos basaremos más en los algoritmos para que tomen decisiones por nosotros, pero es improbable que estos empiecen conscientemente a manipularnos. No tendrán ninguna conciencia.

La ciencia ficción suele confundir la inteligencia con la conciencia, y supone que para equiparse a la inteligencia humana o superarla, los ordenadores tendrán que desarrollar conciencia. El argumento básico de casi todas las películas y novelas sobre IA gira en torno al instante mágico en el que un ordenador o un robot adquieren conciencia. Una vez ocurre, o bien el héroe humano se enamora del robot, o bien el robot intenta matar a todos los humanos, o bien suceden ambas cosas a la vez.

Pero en realidad no hay razón para suponer que la inteligencia artificial adquiera conciencia, porque inteligencia y conciencia son cosas muy distintas. La inteligencia es la capacidad de resolver problemas. La conciencia es la capacidad de sentir dolor, alegría, amor e ira. Tendemos a confundir ambas cosas porque en los humanos y otros mamíferos la inteligencia va de la mano de la conciencia. Los mamíferos resuelven la mayoría de los problemas mediante los sentimientos. Sin embargo, los ordenadores los resuelven de una manera diferente.

Simplemente, hay caminos distintos que conducen a una inteligencia elevada, y solo algunos de dichos caminos implican obtener conciencia. De la misma manera que los aviones vuelan más rápidos que las aves sin desarrollar plumas, así los ordenadores pueden llegar a resolver problemas mucho mejor que los mamíferos sin desarrollar sentimientos. Es cierto que la IA tendrá que analizar con exactitud los sentimientos humanos para tratar enfermedades humanas, identificar a terroristas humanos, recomendar a parejas humanas y circular por una calle llena de peatones humanos. Pero podrá hacerlo sin experimentar ningún sentimiento propio. Un algoritmo no necesita sentir alegría, ira o miedo para reconocer los diferentes patrones bioquímicos de simios alegres, contrariados o asustados.

Desde luego, no es del todo imposible que la IA desarrolle sentimientos propios. Todavía no sabemos bastante de la conciencia para estar seguros de ello. En general, hay tres opciones que es necesario considerar:

1. La conciencia está relacionada de algún modo con la bioquímica orgánica, de tal manera que nunca será posible crear la conciencia en sistemas no orgánicos.
2. La conciencia no está relacionada con la bioquímica orgánica, pero sí con la inteligencia, de tal manera que los ordenadores podrían desarrollar conciencia y los ordenadores tendrán que hacerlo si deben superar un determinado umbral de inteligencia.
3. No existen conexiones esenciales entre la conciencia y la bioquímica orgánica o la inteligencia superior. Por tanto, los ordenadores podrían desarrollar conciencia, pero no necesariamente; podrían llegar a ser superinteligentes al tiempo que siguieran sin tener conciencia.

En nuestro estado actual de conocimiento, no podemos descartar ninguna de estas alternativas. Pero precisamente porque sabemos tan pocas cosas de la conciencia, parece improbable que seamos capaces de programar muy pronto ordenadores conscientes. De ahí que a pesar del inmenso poder de la inteligencia artificial, por ahora su uso continuará dependiendo en cierta medida de la conciencia humana.

El peligro es que, si invertimos demasiado en desarrollar la IA y demasiado poco en desarrollar la conciencia humana, la inteligencia artificial muy sofisticada de los ordenadores solo servirá para fortalecer la estupidez natural de los humanos. Es improbable que nos enfrentemos a una rebelión de robots en las décadas venideras, pero podríamos tener que habérnoslas con multitud de bots que saben cómo pulsar nuestros botones emocionales mejor que nuestra madre, y utilizar esta asombrosa capacidad para intentar vendernos cosas, ya sea un automóvil, a un político o una ideología completa. Los bots podrían identificar nuestros temores, odios y antojos más profundos, y utilizar esta ventaja contra nosotros. Ya se nos ha dado una muestra de esto en elecciones y referéndums en todo el mundo, cuando los piratas informáticos han descubierto cómo manipular a los votantes individuales analizando los datos sobre ellos y explotando sus

prejuicios.[\[33\]](#) Aunque las novelas y películas de ciencia ficción acaban en espectaculares apocalipsis de fuego y humo, en realidad podríamos enfrentarnos a un apocalipsis banal al pulsar una tecla.

Para evitar tales resultados, por cada dólar y cada minuto que invertimos en mejorar la inteligencia artificial sería sensato invertir un dólar y un minuto en promover la conciencia humana. Por desgracia, en la actualidad no hacemos mucho para investigarla y desarrollarla. Estamos investigando y desarrollando capacidades humanas sobre todo según las necesidades inmediatas del sistema económico y político, y no según nuestras propias necesidades a largo plazo como seres conscientes. Mi jefe quiere que conteste los mensajes electrónicos tan rápidamente como sea posible, pero le interesa poco mi capacidad para saborear y apreciar los manjares que como. En consecuencia, reviso los mensajes electrónicos incluso durante las comidas, al tiempo que pierdo mi capacidad de prestar atención a mis propias sensaciones. El sistema económico me presiona para que expanda y diversifique mi cartera de valores, pero me da cero incentivos para expandir y diversificar mi compasión. De modo que me esfuerzo para entender los misterios de la Bolsa de valores, al tiempo que dedico mucho menos esfuerzo a entender las causas profundas del sufrimiento.

En esto, los humanos nos asemejamos a otros animales domésticos. Hemos criado vacas dóciles que producen cantidades enormes de leche, pero que en otros aspectos son muy inferiores a sus antepasados salvajes. Son menos ágiles, menos curiosas y menos habilidosas.[\[34\]](#) Ahora estamos creando humanos mansos que generan cantidades enormes de datos y funcionan como chips muy eficientes en un enorme mecanismo de procesamiento de datos, pero estos datos-vacas en absoluto maximizan el potencial humano. De hecho, no tenemos ni idea de cuál es el potencial humano completo, porque sabemos poquísmo de la mente humana. Y sin embargo, apenas invertimos en la investigación de la mente humana y en cambio nos centramos en aumentar la velocidad de nuestras conexiones a internet y la eficiencia de nuestros algoritmos de macrodatos. Si no somos prudentes, terminaremos con humanos degradados que usarán mal

ordenadores mejorados y que provocarán el caos en sí mismos y en el mundo.

Las dictaduras digitales no son el único peligro que nos espera. Junto a la libertad, el orden liberal depositó también muchas esperanzas en el valor de la igualdad. El liberalismo siempre valoró la igualdad política, y gradualmente llegó al convencimiento de que la igualdad económica tiene casi la misma importancia. Porque sin un sistema de seguridad social y una igualdad económica mínima, la libertad no tiene sentido. Pero de la misma manera que los algoritmos de macrodatos podrían acabar con la libertad, podrían al mismo tiempo crear las sociedades más desiguales que jamás hayan existido. Toda la riqueza y todo el poder podrían estar concentrados en manos de una élite minúscula, mientras que la mayoría de la gente sufriría no la explotación, sino algo mucho peor: la irrelevancia.