

Reinventar la empresa en la era digital

Los *big data* y el futuro de los negocios

Kenneth Cukier

Kenneth Cukier defiende que ningún ámbito de la actividad humana ni sector de la industria será inmune a la total reorganización que están a punto de traer los *big data* a medida que transforman la sociedad, la política y los negocios. Tal y como afirma con claridad: «Más no es solo más. Más es nuevo. Más es mejor. Más es distinto».

Aunque sigue habiendo limitaciones (por lo general originadas en problemas de privacidad) sobre lo que uno puede obtener y hacer con los datos, necesitamos revisar nuestras principales ideas sobre los costes de recopilar información y la dificultad de procesarla.

En su opinión, este nuevo mundo de datos afectará a dos áreas de las políticas y regulaciones públicas:

- Empleo: las tecnologías generarán a medio plazo una oleada de desempleo estructural.
- Privacidad: porque los datos se recogen de manera invisible y pasiva, como subproducto de otro servicio.

Según Cukier, los *big data* cambiarán los negocios y los negocios cambiarán la sociedad. Confía en que los beneficios pesen más que las desventajas, pero se muestra cauto, puesto que la sociedad no parece todavía competente en la gestión de los datos que ya tiene capacidad para recoger.

Kenneth Cukier

The Economist



Kenneth Cukier es editor de datos de la revista *The Economist* en Londres y coautor con Viktor Mayer-Schönberger del varias veces premiado libro *Big Data. La revolución de los datos masivos* (2013), que ha estado en la lista de más vendidos del *New York Times* y ha sido traducido a 20 idiomas. Es colaborador habitual de la BBC, la CNN y NPR y miembro del consejo del Foro Económico Internacional sobre desarrollo basado en datos. De 2002 a 2003 fue investigador asociado en el Harvard's Kennedy School of Government. Es miembro del consejo de administración de International Bridges to Justice y del Consejo de Relaciones Exteriores (CFR).

Claves para la empresa del futuro:

Aprovechar los *big data* para emprender innovaciones ambiciosas y revolucionarias

Las compañías suelen fijarse como objetivo el 10% de mejoras. Pero esto supone únicamente una «innovación sostenida»: aprender a partir de datos a hacer lo mismo que se está realizando, solo que ligeramente mejor. Con el tiempo, este enfoque impide aprovechar las oportunidades de lograr verdaderos avances. En el futuro, las compañías inteligentes estarán siempre atentas a cómo usar los *big data* para lograr lo ambicioso y lo revolucionario: innovaciones disruptivas que creen productos y mercados radicalmente nuevos. Estos no se limitan a mejorar lo que ya existe, sino que abren el camino para hacer las cosas de maneras por completo distintas.

Usar los *big data* para conseguir «pequeños logros» que vayan sumando

Lo ambicioso coexiste con su contrario, es decir, usar los datos para identificar mejoras minúsculas que por sí solas son insignificantes, pero que combinadas constituyen un enorme aumento del rendimiento. En el pasado, a una empresa no le compensaba dedicar esfuerzo a logros tan pequeños. Pero a medida que se reducen los costes de gestionar los *big data* y resulta más fácil hacer operativo lo aprendido, la capacidad de encontrar y poner en práctica estos «pequeños logros» se vuelve viable. Esta capacidad será lo que defina la manera en que competirán entre sí las empresas del futuro.

Combinar los *big data* con valores e intuición

Las compañías que triunfarán en el futuro aprenderán de los *big data* e interpretarán sus enseñanzas a modo de recordatorio de las limitaciones cognitivas, las parcialidades, etcétera, de los gestores. Pero al mismo tiempo las compañías no pueden aceptar a ciegas lo que dicen los *big data*, sino que deben integrar estos con los valores humanos de justicia, honradez y sentido común. Los individuos deben seguir siendo los amos de la tecnología, no sus siervos, ya que los datos son siempre un mero simulacro de la realidad, no la realidad misma. Las empresas del futuro han de saber gestionar los datos con un respeto saludable por sus limitaciones.

Los *big data* y el futuro de los negocios

Parte I. Más

La base del emprendimiento empresarial es la información. De hecho, algunas de las primeras manifestaciones de escritura y contabilidad proceden de mercaderes sumerios de alrededor del año 8.000 a.C., que utilizaban pequeñas cuentas de barro para representar artículos para el comercio y después dejaban constancia escrita de sus transacciones. Así que, cuando examinamos el papel de los datos en la actualidad, podemos afirmar que no ha cambiado gran cosa. Ahora tenemos capacidad para recopilar, almacenar y usar más información, pero la naturaleza de los datos y su importancia no han variado esencialmente. De acuerdo con esta interpretación, la expresión *big data* es solo una manera grandilocuente de describir la capacidad de la sociedad de controlar más datos que nunca, pero no altera la naturaleza fundamental del comercio desde la Antigüedad hasta nuestros días.

Es posible, sin embargo, que dicha interpretación sea completamente errónea. Y es que, en muchos ámbitos de la vida, cuando uno cambia la cantidad, está modificando también la forma. Por ejemplo, puesto que ya se habían inscrito símbolos en tablillas de barro y después formado y escrito palabras con tinta en papiros, nadie sugeriría hoy que la imprenta no supuso una revolución sin precedentes cuando se introdujo

alrededor de 1450. Sí, antes ya existían palabras y libros, y sí, había más palabras y más libros. Pero no era lo mismo. Más no era simplemente más; más era distinto.

Las consecuencias de la imprenta fueron un aumento drástico de los materiales escritos y un descenso de los costes de producirlos. La nueva era de «más palabras» trajo consigo enormes transformaciones añadidas. Diluyó la autoridad de la Iglesia y el poder de las monarquías; hizo posible la alfabetización masiva, la democracia, el capitalismo y una sociedad basada en el conocimiento como requisito para el trabajo en lugar de solo la fuerza física.

Hoy en día, el concepto de material escrito —de «libro»— está cambiando de nuevo. Hay tabletas informáticas digitales como el iPad con capacidad para almacenar todos los libros de la biblioteca de una universidad importante en un único dispositivo. Y que además buscan, indexan y permiten que se copien y compartan con facilidad y de forma instantánea pequeñas porciones de contenido. Más es nuevo. Más es mejor. Más es distinto.

Hasta aquí las palabras. Ahora pensemos en las comunicaciones. En el pasado, las sociedades tenían capacidad de enviar mensajes a lugares distantes. En la Antigua Roma ya se empleaban palomas mensajeras. Para comunicarse con sus oficiales, Gengis Khan creó casas de postas para palomas por toda Asia y partes de Europa oriental. En los negocios, en la década de 1880, la familia de banqueros Rothschild mandaba sus mensajes usando palomas, igual que hacía el servicio de noticias sobre mercados financieros Reuters.

Pero cuando llegó el telégrafo nadie pudo negar que los cables y las pulsaciones eléctricas eran una versión mejorada de las palomas mensajeras. Más era distinto. Luego llegó el teléfono y, con él, las grandes comunicaciones. La reducción de costes y la creciente facilidad de uso no eran más de lo mismo. Igual sucedió con la radio. Hoy internet es tan fundamentalmente distinto de las palomas mensajeras que comparar ambos medios de comunicación parece ridículo. Pero eso no hace más que subrayar el grado hasta el cual más no es solo más; más es nuevo, mejor y distinto.

Con los datos ocurre como con las palabras y las comunicaciones. Tenemos más información que nunca. Pero su importancia no reside en el hecho de que podamos hacer más de lo que ya hacemos, o saber más de lo que ya estamos estudiando. Sino que el cambio de escala conduce a un cambio de estado. Y el cambio cuantitativo a un cambio cualitativo. Fundamentalmente, al tener más datos podemos realizar muchas cosas que no podíamos lograr cuando teníamos cantidades menores de información.

De hecho, estamos empezando a aprender lo que son esas cosas, ya que siempre hemos autocensurado nuestra imaginación sobre lo que es posible conseguir con datos. Esto lo hacíamos de manera inconsciente, porque nos resultaba inconcebible disponer de tantos datos, porque ignorábamos que sería tan fácil recopilarlos,

Tenemos más datos que nunca. Pero no significa que podamos hacer o saber más: el cambio de escala conduce a un cambio de estado, y el cambio cuantitativo a uno cualitativo

almacenarlos, procesarlos y difundirlos. ¿Sobre qué base habríamos podido extrapolar una información así?

Ni el hombre más sabio del mundo habría podido concebir, con un ábaco en la mano, la calculadora mecánica con discos para introducir los miles de millones. El experto que manejaba estos discos quizá no concebía la existencia de la computadora electrónica. E incluso cuando se inventó el transistor, años después de las primeras computadoras, habría resultado difícil, salvo para el ingeniero más visionario, imaginar el ritmo estipulado por la ley de Moore. Esta ley establece, como principio de la era digital, que el número de transistores de un chip se duplica cada dos años, lo que, con el tiempo, se traduce en reducciones exponenciales en costes y aumento de potencia.

Los cambios en el grado hasta el cual una sociedad puede recopilar e interactuar con información tienen profundas consecuencias en nuestra comprensión de la economía. La idea misma de economía es un concepto reciente. Cuando el economista clásico surgió en Gran Bretaña a mediados de la década de 1700, su disciplina se llamaba filosofía política; el término «economía» apareció más tarde. Su verdadero padre fundador, Adam Smith, era un filósofo moral cuya obra más importante anterior a *La riqueza de las naciones* fue un libro titulado *La teoría de los sentimientos morales*.

Resulta tentador leer pasajes de los economistas clásicos y suponer que vivían en un mundo basado en la observación y de prosa recargada en el que los asuntos comerciales eran descritos mediante palabras majestuosas antes que con números desnudos, un mundo de ideas donde los datos apenas tenían cabida. Pero esto sería incorrecto. De hecho, *La riqueza de las naciones* de Adam Smith está repleta, página tras página, de datos sobre producción de maíz. Los primeros pensadores económicos del siglo XVIII se apoyaban de forma significativa en datos para elaborar sus ideas.

Sin embargo, cuando llegó el momento de definir los factores que intervienen en la producción, identificaron tres: tierra, mano de obra y capital. No incluyeron

«información» como componente individual, aunque Smith y otros escribieron de forma elocuente sobre cómo los mercados dependen de la información. Resulta sencillo comprender por qué la excluyeron. En aquel tiempo era tan sumamente difícil recopilar, almacenar y usar la información que la idea de que pudiera ser por sí misma una materia

Los cambios en la recopilación e interacción con información tienen profundas consecuencias en nuestra comprensión de la economía

prima en los negocios habría sonado ridícula. Después de todo, los datos tenían que ser recopilados por una persona equipada con una pluma de ganso y un pergamino. Manejar y utilizar la información era caro y laborioso. Recordemos que por entonces no existían ni siquiera los rudimentos de la estadística. Así que, incluso si uno disponía de los datos, no podía hacer gran cosa con ellos.

Obviamente, hoy la situación ha cambiado por completo. Por supuesto que siguen existiendo limitaciones sobre lo que uno puede realizar con datos. Pero

necesitamos cuestionar la mayor parte de nuestras ideas preconcebidas sobre los costes de recopilarlos y la dificultad de procesarlos. Seguimos anclados en una mentalidad de «escasez», como esas personas mayores que corren a ponerse al teléfono cuando les llama un familiar que vive lejos y enseguida cuelgan porque «es conferencia», un comportamiento heredado de los días en que el teléfono era caro, antes de que la liberalización del mercado y las nuevas tecnologías cambiaran para siempre las tarifas de las telecomunicaciones.

Nuestras instituciones siguen asentadas sobre principios de escasez e información de coste elevado. Las cajas negras de los aviones conservan solo una cantidad minúscula de datos, unas cuantas horas de parca información relativa a la mecánica y a la cabina de mandos, un residuo de la época en que fueron diseñadas. La señal que permite recuperarlas es débil y la batería tiene una duración corta, de solo treinta días. El mundo se ha puesto a subsanar estas deficiencias después de la tragedia del vuelo de Malaysia Airlines MH370 que desapareció en marzo de 2014.

Lo cierto es que el planteamiento «caja negra» permitiría ayudar a la sociedad de varias maneras. Podrían instalarse estos dispositivos en vehículos policiales e incorporarse al uniforme de los agentes para ayudar a los tribunales a dictaminar sobre acusaciones de agresión policial frente al uso legítimo de la fuerza. Sin embargo, solo se emplean en pocos lugares. De igual modo, las cajas negras podrían usarse en los quirófanos para que los cirujanos aprendan de sus errores, para que pacientes que han sido víctimas de negligencia consigan obtener su indemnización o para demostrar que una actuación médica ha sido impecable.

Pero los médicos temen que algo así abra la puerta a una avalancha de demandas por negligencia y se resisten a su introducción. Lo cierto es que ni a la policía ni a los médicos les falta razón en sus posturas antidatos: la sociedad necesita tiempo para asimilar cómo aceptar e integrar una tecnología nueva y para desarrollar la nueva cultura que esta requiere. Al fin y al cabo, no hemos hecho más que empezar a sentirnos cómodos con los ordenadores, medio siglo después de que su uso se haya generalizado.

En este sentido, la experiencia de los medios de comunicación sociales resulta instructiva. En su elogiado libro *Delete: the Virtues of Forgetting in the Digital Age* [Borrar. Las virtudes de olvidar en la era digital], Viktor Mayer-Schönberger¹ (coautor de mis dos libros sobre *big data*) recoge anécdotas espeluznantes de personas a las que se les denegó un empleo por cuestiones como una fotografía suya tomada durante una juerga y publicada años atrás en su página de Facebook. Sucesos como este venían a evidenciar el extremo hasta el cual los encargados de recursos humanos no se habían replanteado sus métodos de selección para un mundo en el que nuestro pasado siempre está *online* y en el que resulta más necesario que nunca «pasar por alto» las locuras juveniles de cada uno.

De manera similar, en el mundo de los *big data* se grabarán muchas cosas por defecto simplemente porque existen o porque suceden. Transcurrirá un tiempo antes de que la sociedad encuentre la manera de manejar esto y cambien las prácticas y las actitudes para que sea posible incorporar de forma razonable la tecnología a nuestras vidas, a nuestras instituciones y a nuestros valores.

Es interesante que esta tensión —entre lo que la tecnología es capaz de conseguir y las actitudes y reglas con las que convive— ponga de relieve uno de los principales puntos de fricción al que ha tenido que hacer frente el *establishment* político estadounidense después de que Snowden filtrara información sobre vigilancia masiva por parte de la Agencia Nacional de Seguridad (NSA, por sus siglas en inglés). Dicha tensión inherente se resume así: la ley está diseñada para una época en la que recopilar y analizar datos era difícil y costoso, y por tanto refleja esos supuestos. Con el tiempo, estas acciones —como, por ejemplo, revisar metadatos telefónicos u otras actividades que habrían sido consideradas imposibles o al menos excepcionales en la década de 1970, cuando se aprobaron las leyes— se volvieron fáciles y baratas y, para junio de 2013, cuando se hicieron públicas, eran ya un lugar común.

Transcurrirá un tiempo antes de que cambien las prácticas y actitudes para incorporar de forma razonable la tecnología a nuestras vidas, a nuestras instituciones y a nuestros valores

Desde el punto de vista de los defensores de la privacidad, las actividades de vigilancia a gran escala de la NSA no estaban autorizadas por la ley. Según la NSA, sus programas no eran más que versiones ampliadas de lo que permite la ley. ¿No debería una agencia de seguridad proveerse de las mismas herramientas modernas que emplean sus adversarios para actuar en su contra?, argumenta. A lo que los críticos responden: si quieres tener ese poder, consigue primero autorización legal y luego convéncete de que el público aceptará un Estado policial digital.

Desafortunadamente, el sistema político estadounidense aún no ha sido capaz de mantener un debate maduro y responsable sobre estas cuestiones que le permita alcanzar algún tipo de consenso. Aunque este análisis no pretende en modo alguno exonerar estas actividades, sí supone, quizá, un paso adelante a la hora de explicarlas. Aquí, una vez más, volvemos al tema central de los *big data*. Más no es simplemente más. Más es mejor. Más es distinto.

Ningún ámbito de actividad humana ni sector industrial será inmune a la profunda reestructuración que está a punto de producirse a medida que los *big data* se abren paso en la sociedad, la política y los negocios. El hombre moldea sus herramientas. Y sus herramientas lo moldean a él.

Parte II. Distinto

La base del emprendimiento comercial es la información. Esto no ha cambiado. Era así para los mercaderes sumerios hace muchos siglos y también hace solo uno,

cuando Frederick Taylor realizó sus estudios sobre tiempos de ejecución y movimientos en las empresas de Estados Unidos.

Los detractores pueden pensar que el debate actual sobre los *big data* no es más que la prolongación del pasado, pero se equivocan igual que si dijeran que una tableta informática viene a ser lo mismo que una tablilla de piedra, o que la web es la continuación lógica de la paloma mensajera, o que el ábaco es algo parecido a una supercomputadora. Estas afirmaciones no estarían por completo descaminadas, pero sí lo bastante como para constituir una distracción inútil.

Lo importante de los *big data* es que nos permiten desarrollar cosas nuevas. Una de las maneras más prometedoras en las que se pueden usar los datos es el área llamada «aprendizaje automático». Es una rama de la inteligencia artificial, que a su vez lo es de la ciencia informática, pero con una dosis considerable de matemáticas. Explicada con sencillez, consiste en introducir una enorme cantidad de datos en un ordenador y hacer que este identifique patrones que los seres humanos no detectarían, o que tome decisiones basadas en probabilidades a una escala que las personas pueden hacer perfectamente, pero que las máquinas hasta hace poco no podían, y que quizá algún día obtengan a una escala inalcanzable para los humanos. Básicamente es una forma de lograr que un ordenador desempeñe funciones sin necesidad de enseñárselas explícitamente, consiguiendo que la máquina realice cálculos basados en cantidades masivas de información.

Lo importante de los *big data* es que nos permiten hacer cosas nuevas. Una de las maneras más prometedoras en la que se pueden usar los datos es el área llamada «aprendizaje automático»

Sus orígenes son bastante recientes. Aunque se concibió por primera vez en la década de 1950, la técnica no funcionaba bien al aplicarse al mundo real. Así que se pensó que era un fracaso. Pero en la última década se ha producido una revolución intelectual y técnica a medida que los investigadores han obtenido prometedores progresos al emplear el aprendizaje automático. Lo que faltaba antes eran datos en cantidades suficientes. Ahora que los hay, el método funciona. Hoy el aprendizaje automático es la base de todo, desde buscadores de internet, recomendaciones de productos *online*, traducción de lenguaje informático y reconocimiento de voz y muchas otras cosas más.

Para comprender lo que es el aprendizaje automático, resulta útil saber cómo surgió. En la década de 1950, un programador informático de IBM llamado Arthur Samuel programó un ordenador para que jugara al ajedrez. Pero no era un juego muy divertido. Samuel siempre ganaba, porque la máquina únicamente reconocía jugadas legales. Él sabía algo de estrategia, así que desarrolló un subprograma inteligente que, con cada movimiento, calculaba las probabilidades de que una configuración determinada del tablero condujera a ganar o a perder la partida.

Pero una partida entre hombre y máquina seguía sin funcionar bien; el sistema se encontraba en un estado demasiado embrionario. Entonces, Samuel dejó que

la máquina jugara contra sí misma. Al hacerlo, recopilaba nuevos datos. Al reunir más datos, la exactitud de sus predicciones mejoraba. Así que jugó contra la computadora y perdió. Una y otra vez. El hombre había creado una máquina que le superaba en habilidad en una tarea que él mismo le había enseñado.

De igual forma, ¿por qué tenemos coches que se conducen solos? ¿Acaso a la industria del *software* se le da mejor incluir todas las normas de circulación en un código? No. ¿Es por el aumento de la memoria de los ordenadores? Tampoco. ¿Procesadores más rápidos? No. ¿Algoritmos más inteligentes? De nuevo, no. ¿Chips más baratos? Tampoco. Todo esto ayudó, pero lo que de verdad hizo posible la innovación fue que los expertos en tecnología cambiaron la naturaleza del problema. Lo convirtieron en una cuestión de datos: en lugar de intentar enseñar al coche a conducir —algo difícil, porque el mundo es un lugar complejo—, el vehículo recoge todos los datos que lo rodean e intenta deducir él solo lo que tiene que hacer: que hay un semáforo, que está en rojo y no en verde y que esto significa que el coche debe detenerse. El vehículo puede tener que realizar hasta mil predicciones por segundo. El resultado es que se conduce solo. Más datos no significa solo más. Más datos supone que las cosas sean distintas.

La idea del aprendizaje automático ha llevado a una serie de descubrimientos inquietantes que parecen desafiar la primacía del ser humano como herramienta de interpretación del mundo. En un estudio del año 2011, investigadores de la Universidad de Stanford² introdujeron en un algoritmo de aprendizaje automático miles de muestras de células mamarias cancerosas, así como las tasas de supervivencia de las pacientes, y le pidieron al ordenador que identificara los signos que mejor predecían que una biopsia determinada resultara claramente cancerosa.

El ordenador produjo once indicadores que pronosticaban que una biopsia de células mamarias sería positiva. ¿Y cuál era el quid de la cuestión? Que la literatura médica solo había identificado ocho de ellos. Había tres que los patólogos no sabían que debían buscar.

De nuevo, los investigadores no le dijeron al ordenador lo que tenía que analizar. Se limitaron a suministrarle las muestras de células, sus características generales y datos sobre los índices de supervivencia de las pacientes (esta sobrevivió quince años al diagnóstico; esta otra murió once meses después). El ordenador encontró obviedades, pero también otras cuestiones no tan claras: firmas genéticas de la enfermedad que las personas no detectaban porque eran invisibles para el ojo humano. Pero un algoritmo sí las halló. El aprendizaje automático funciona porque se proporciona al ordenador muchos datos, más información de la que sería capaz de digerir un ojo humano en toda su vida, y más también de la que podría recordar de forma instantánea.

En este caso, el rendimiento del ordenador superó al de los humanos. Vio cosas que los especialistas no percibían. Y esto ha propiciado diagnósticos más precisos. Es más, debido a que es un ordenador, puede hacerlo a gran escala. Hasta aquí, el «más» de los *big data* no significa solo más de lo mismo; significa «mejor». Pero, ¿también «nuevo» y «distinto»? La respuesta es sí.

Pensemos un instante. Utilizando este método a gran escala podríamos ser capaces de interpretar biopsias diarias de una población al completo en lugar de una o varias veces en la vida de cada persona. Al hacerlo quizá detectaríamos el aspecto que presenta el cáncer en sus estadios más tempranos, de manera que fuera posible tratarlo con los procedimientos más sencillos, más efectivos y menos costosos. Con ello saldríamos ganando todos: los pacientes, la sociedad y los presupuestos para sanidad de los gobiernos.

¿Por qué es nuevo? Recordemos que el ordenador no solo aumentó la precisión del diagnóstico añadiendo nuevos indicadores. También logró un descubrimiento científico (en este caso, los tres rasgos distintivos obvios de cáncer antes desconocidos eran las relaciones entre células en un material celular llamado estroma y no solo rasgos dentro de las propias células).

El ordenador arrojó una conclusión que se les había escapado a los investigadores y que supone un avance para la comprensión humana.

¿Qué significa tener más datos? Manolis Kellis, investigador genético en el Broad Institute de Cambridge, en Massachusetts, nos ofrece un valioso ejemplo a modo de explicación. Tal y como señalaba un informe de la Casa Blanca de mayo de 2014 sobre los *big data*: «Disponer de un

La industria médica es un claro ejemplo de las repercusiones de los *big data*. Aunque en este ámbito hay grandes avances, las leyes restrictivas de privacidad obstaculizan los procesos

gran número de conjuntos de datos supone una ventaja crucial a la hora de identificar la variación genética significativa que determina una enfermedad. En esta investigación, una variante genética relacionada con la esquizofrenia no se constató al analizar 3.500 casos, se detectó débilmente en 10.000, pero se volvió relevante desde el punto de vista estadístico con 35.000 casos».³ Y como explicaba Kellis: «Hay un punto de inflexión en el que todo cambia».

La industria médica proporciona otro ejemplo contundente de las repercusiones que van a tener los *big data* en el sector. Los servicios sanitarios están repletos de ejemplos porque ya poseen gran cantidad de datos y, sin embargo, van un tanto atrasados a la hora de explotar su potencial. Así que, aunque en este ámbito se están produciendo importantes avances, la existencia de leyes restrictivas de privacidad está obstaculizando el proceso.

Pensemos en la manera de identificar una interacción adversa entre medicamentos; es decir, un caso en el que una persona toma dos fármacos que son efectivos y seguros por sí solos, pero que combinados tienen un efecto secundario peligroso. Con cientos de miles de medicamentos en el mercado es un problema difícil de abordar, puesto que es imposible examinar todos juntos. En 2013, Microsoft Research y varias universidades estadounidenses idearon un ingenioso método para reconocer estas sustancias: analizando consultas de búsqueda.⁴

Los investigadores elaboraron una lista de ochenta términos y expresiones asociados a síntomas de una enfermedad conocida, la hiperglucemia (por ejemplo, «nivel alto

de azúcar» o «visión borrosa»). Luego, observaron si las personas buscaban el fármaco paroxetina (un antidepresivo) y/u otro fármaco, pravastina (que reduce los niveles de colesterol). Después de analizar nada menos que ochenta y dos millones de búsquedas a lo largo de varios meses durante 2010, encontraron lo que estaban buscando.

Las consultas para solo los síntomas y ninguno de los fármacos eran extremadamente bajas, de menos del 1%; por tanto, ruido de fondo. Las personas que buscaban los síntomas y uno de los medicamentos suponían el 4%; para los síntomas y el otro fármaco, el 5%. Pero las que buscaban los síntomas y ambos fármacos ascendían a un sorprendente 10%. En otras palabras, había el doble de posibilidades de que las personas teclearan determinados síntomas médicos en la ventana de consultas del buscador si estaban buscando también los dos fármacos en lugar de solo uno.

Este descubrimiento es importante. Pero no es una pistola humeante. La policía no puede irrumpir en las casas de los directivos de los laboratorios y llevárselos presos. Es solo una correlación; no revela causalidad. No obstante, los resultados son significativos y tienen considerables consecuencias para los negocios y el valor corporativo. La interacción adversa entre estos dos fármacos no se conocía antes, no venía en el prospecto. Se descubrió analizando viejas búsquedas. Muchas, ochenta y dos millones.

El valor de estos datos es inmenso. Un paciente necesita saber esta información. Un médico quiere conocerla. A un asegurador médico le interesa especialmente. Y un regulador de fármacos la necesita. Quizá, alguien de Microsoft debería empezar a pensar en crear una división para negociar la cesión de datos como nueva fuente de beneficios y no limitarse a los ingresos que genera la publicidad que aparece con los resultados de búsquedas.

Este nuevo mundo de datos y la manera en que lo aprovechan las empresas choca con dos ámbitos de las políticas y las regulaciones públicas. El primero tiene que ver con los puestos de trabajo. Al principio, los líderes empresariales valoran la necesidad de incorporar nuevas clases de trabajadores a su plantilla... Es el auge del analista de datos o *data scientist*. Los asesores advierten gravemente de los peligros de una escasez de estos. Las universidades se apresuran a prepararse para la demanda existente. Pero esta forma de pensar revela poca visión de futuro. A medio y a largo plazo, los *big data* nos van a arrebatar nuestros puestos de trabajo. La tecnología va a traer consigo una oleada de desempleo estructural.

Esto se debe a que los *big data* y los algoritmos son la misma amenaza para los trabajadores no manuales y especializados en el siglo XXI que la que supuso la cadena de montaje para los obreros de los siglos XIX y XX. Entonces, el trabajo físico se consideraba una mercancía y las máquinas podían hacerlo mejor que las personas. En el futuro serán nuestras mentes las que demostrarán ser inferiores a las máquinas. Un estudio realizado por investigadores de la Universidad de Oxford⁵ predice que hasta el 47% del trabajo que se realiza hoy en Estados Unidos corre el riesgo de ser asumido por ordenadores.

Pensemos, por ejemplo, en el anatomopatólogo cuyo trabajo ya no es necesario porque un algoritmo de aprendizaje automático puede interpretar biopsias cancerosas con mayor precisión, velocidad y a un coste menor. Los anatomopatólogos suelen ser

titulados en Medicina. Se compran casas. Pagan impuestos. Votan. Entrenan al equipo de fútbol de sus hijos los fines de semana. En otras palabras, son participantes activos de la sociedad. Y, junto con una clase entera de profesionales como ellos, van a asistir a la transformación, o incluso quizá a la completa eliminación, de sus puestos de trabajo.

Un estudio realizado por la Universidad de Oxford predice que hasta el 47% del trabajo que se hace hoy en Estados Unidos corre el riesgo de ser asumido por ordenadores

La parte positiva es que los *big data* aportarán grandes beneficios a la sociedad. El peligro es que todos terminemos convertidos en profesores de yoga o camareros al servicio de un pequeño grupo de ingenieros informáticos millonarios. Nos gusta pensar que la tecnología genera empleo, incluso si esto sucede después de un periodo de trastocamiento transitorio. Desde luego que ocurrió así con la disrupción ocasionada por nuestra principal referencia, la revolución industrial, cuando las máquinas sustituyeron la mano de obra artesana. Las ciudades se llenaron de fábricas y los trabajadores agrícolas, pobres y sin educación, pudieron —una vez creadas las leyes laborales y el sistema educativo— mejorar su nivel de vida y disfrutar de movilidad social. Sin duda, fue un periodo de transformaciones traumáticas, pero que con el tiempo condujo a mejoras en la calidad de vida.

Sin embargo, esta extrapolación optimista pasa por alto el hecho de que algunos puestos de trabajo desaparecerán para no volver. Tal y como observó el economista y premio Nobel estadounidense Wassily Leontief, para los caballos la revolución industrial no fue precisamente una buena noticia.⁶ Es decir, que una vez que se introdujeron los tractores en las faenas agrícolas y los automóviles sustituyeron a los coches de caballos, básicamente estos animales dejaron de ser necesarios para la economía. Todavía se pueden apreciar las trazas de aquel cambio en las antiguas cocheras que abundan en el próspero West End londinense y que han sido reconvertidas en elegantes residencias.

La agitación que trajo consigo la revolución industrial generó revoluciones políticas que inspiraron nuevas teorías económicas y nuevos movimientos políticos como el marxismo. No hace falta ser un intelectual para predecir que surgirán nuevas filosofías políticas y nuevos movimientos sociales alrededor de los *big data*, los robots, los ordenadores e internet y sus efectos en la economía y en la democracia representativa. Los recientes debates sobre la desigualdad de rentas y los movimientos como Ocupa Wall Street parecen apuntar en esa dirección.

El segundo ámbito a legislar es la privacidad, que ya constituía un problema en la era de los «pequeños datos». Y lo será en la de los grandes. A primera vista puede que no parezca en esencia un problema distinto, sino el mismo a gran escala. Pero aquí también, más significa distinto. La naturaleza de proteger la información

personal cambia cuando la amenaza potencial para la privacidad no se produce una vez al día o una cada hora, sino mil por segundo. O cuando la recopilación de datos no se realiza por vías explícitas y activas, sino de manera invisible y pasiva, como consecuencia de otro servicio.

Así, por ejemplo, los sitios web en Europa están obligados a informar a quienes entran en ellos que utilizan *cookies* para identificar a los visitantes. Un requerimiento que, a simple vista, parece razonable. Pero ¿qué ocurre cuando cada punto de luz de un edificio identifica que hay una persona en la habitación por motivos de seguridad o protección? (como en el caso de un incendio, para que los bomberos sepan adónde tienen que acudir). Y el *software*, con un coste marginal de casi cero, es lo bastante avanzado como para identificar dónde están esas personas basándose en su silueta, su manera de andar o incluso también el pulso sanguíneo. Es difícil imaginar cómo las leyes clásicas sobre la privacidad gestionarían un mundo así; cómo podrá una persona que se sienta tratada de manera injusta tomar medidas, o siquiera ser consciente de la situación.

Y todavía hay más. Las leyes sobre la privacidad en todo el mundo se basan en el principio, recogido por las directrices sobre privacidad de la OCDE, de que una entidad se deshará de los datos una vez que los ha usado para el propósito con el que los recopiló. Pero la utilidad de los *big data* pasa precisamente por guardarlos para siempre, puesto que nunca se sabe qué valiosos fines se les podrá otorgar el día de mañana. Si Microsoft hubiera borrado sus antiguas búsquedas de 2010, no habría sido capaz de identificar la interacción adversa entre la paroxetina y la pravastina en 2013.

Por lo tanto, al igual que lo que define los *big data* es que más no es solo más, sino nuevo, mejor y distinto, los negocios modernos necesitarán reguladores que comprendan que las reglas que gobiernan los *big data* no pueden ser simplemente más de lo mismo. De hecho, las regulaciones actuales hacen poco por proteger la privacidad, así que continuar con medidas igualmente mediocres no parece tener demasiado sentido. En lugar de ello, el negocio de *big data* está pidiendo a gritos regulaciones nuevas, mejores y distintas.

Los *big data* cambiarán los negocios y los negocios la sociedad. La esperanza es que los beneficios superen las consecuencias negativas, pero no es más que eso, una esperanza. La realidad es que todo esto es muy nuevo y que a nosotros, como sociedad, no se nos da demasiado bien gestionar todos los datos que tenemos capacidad de recopilar. No hace tanto tiempo, en la Exposición Mundial de Chicago de 1893, se concedió la Medalla de Oro al inventor del armario archivador vertical, por aquel entonces una solución brillante al problema de almacenaje y localización de documentos en papel. Eran tiempos en los que el flujo de información sobrepasaba a las empresas, la «versión beta» de los *big data* en la vida corporativa actual.

Lo que está claro es que estos ejemplos pasados no nos sirven para predecir el futuro. La tecnología nos sorprende igual que a un hombre con un ábaco le asombraría un iPhone. Lo que es seguro es que más no será solo más. Será distinto.

El negocio de big data está pidiendo a gritos regulaciones nuevas, mejores y distintas

Notas

1

V. Mayer-Schönberger, *Delete: The Virtues of Forgetting in the Digital Age*, Princeton University Press, 2009.

2

A. H. Beck, *et al.* (con Daphne Koller), «Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated with Survival», en *Science Translational Medicine*, 3 (108), 2011 (<http://stm.sciencemag.org/content/3/108/108ra113.full.pdf>).

3

«Big Data: Seizing Opportunities, Preserving Values», Oficina del presidente de Estados Unidos, mayo de 2014 <http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf>

4

R. W. White, N. P. Tatonetti, N. H. Shah, R. B. Altman y E. Horvitz, «Web-Scale Pharmacovigilance: Listening to Signals from the Crowd», en *J Am Med Inform Assoc.*, 1 de mayo de 2013, 20 (3), pp. 404-408 <<http://www.ncbi.nlm.nih.gov/pubmed/23467469>>

5

C. B. Frey y M. A. Osborne, «The Future of Employment: How Susceptible Are Jobs to Computerisation?», documento de trabajo, Oxford University, 17 de septiembre de 2013 <http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf>

6

W. Leontief, «National Perspective: The Definition of Problems and Opportunities», en *The Long-Term Impact of Technology on Employment and Unemployment*, National Academy of Engineering, 1983, pp. 3-7 <http://books.google.com/books/about/The_Long_term_Impact_of_Technology_on_Em.html?id=hS0rAAAAYAAJ>

Canal OpenMind



Artículo

- Los big data y el futuro de los negocios

Conoce al autor

Kenneth Cukier

Artículos relacionados

- De la deconstrucción a los *big data*: cómo la tecnología está transformando las corporaciones
- *Big Data*: ¿a quién pertenece?
- Internet y los negocios

Lee el libro completo

- Reinventar la Empresa en la Era Digital

Otros libros



Compartir artículo

