

Big data, La revolución de los datos masivos

Viktor Mayer-Schönberger

Kenneth Cukier

Turner Noema, 2013 278 pp.¹

Análisis –más bien superficial- de las repercusiones del Big Data tanto en las metodologías de investigación como en la planificación de nuestras intervenciones sociales, tanto económicas como políticas.² Organizaremos el libro no siguiendo sus capítulos –es reiterativo- sino sus ideas fundamentales, ilustrados con pocos pero significativos ejemplos.

1º, el Big data ya está aquí...

El Big data es una realidad que afecta a numerosos sectores de la vida personal y social. Por ejemplo:

- *“Una forma de pensar en esta cuestión hoy en día –la que aplicamos en este libro– es la siguiente: los big data, los datos masivos, se refieren a cosas que se pueden hacer a gran escala, pero no a una escala inferior, para extraer nuevas percepciones o crear nuevas formas de valor, de tal forma que transforman los mercados, las organizaciones, las relaciones entre los ciudadanos y los gobiernos, etc. (145-148)”*
- *“Para apreciar hasta qué punto está ya en marcha la revolución de la información, considérense las tendencias que se manifiestan en todo el espectro de la sociedad. Nuestro universo digital está en expansión constante. Piénsese en la astronomía. Cuando el Sloan Digital Sky Survey arrancó en 2000, solo en las primeras semanas su telescopio de Nuevo México recopiló más datos de los que se habían acumulado en toda la historia de la astronomía. Para 2010, el archivo del proyecto estaba a rebosar, con unos colosales 140 terabytes de información. Sin embargo, un futuro sucesor, el Gran Telescopio Sinóptico de Investigación de Chile, cuya inauguración está prevista para 2016, acopiará esa cantidad de datos cada cinco días. Similares cantidades astronómicas las tenemos también más a mano. Cuando los científicos descifraron por primera vez el genoma humano en 2003, secuenciar los tres mil millones de pares de bases les exigió una década de trabajo intensivo. Hoy en día, diez años después, un solo laboratorio es capaz de secuenciar esa cantidad de ADN en un día. En el campo de las finanzas, en los mercados de valores de Estados Unidos, a diario cambian de manos siete mil millones de acciones, dos terceras partes de las cuales se negocian mediante algoritmos de ordenador basados en modelos matemáticos que procesan montañas de datos para predecir ganancias, al tiempo que intentan reducir los riesgos. Las compañías de internet se han visto particularmente abrumadas. Google procesa más de 24 petabytes de datos al día, un volumen que representa miles de veces la totalidad del material impreso que guarda la Biblioteca del Congreso de Estados Unidos. A Facebook, una empresa que no existía hace una década, se suben más de diez millones de fotos nuevas cada hora. Sus usuarios hacen clic en el botón de “me gusta” o insertan un comentario casi tres mil millones de veces diarias, dejando un rastro digital que la compañía explota para descubrir sus preferencias. Entretanto, los ochocientos millones de usuarios mensuales del servicio YouTube de Google suben más de una hora de vídeo cada segundo. El número de mensajes de Twitter aumenta alrededor de un 200 por 100 al año, y en 2012 se habían superado los cuatrocientos millones de tuits diarios. De las ciencias a la asistencia médica, de la banca a internet, los sectores pueden ser muy distintos, pero en conjunto cuentan una historia parecida: la cantidad de datos que hay en el mundo está creciendo deprisa, desbordando no solo nuestras máquinas, sino también nuestra propia imaginación. (160-181)”*

¹ Las citas se referencian no a la página impresa del libro, sino a la posición del Kindle de Amazon, que ya es decir...

² Yo/// Una vez más, se echa en falta formación filosófica relevante para el asunto. No hay remedio.

- En el año 2000, información analógica, 75%, digital, 25%. En el 2007, 7% analógica. En el 2013, de un total de 1200 exabytes, sólo es analógica un 2%...
- “Es más, debido al vasto tamaño de la información, muy a menudo las decisiones no las tomarán los seres humanos, sino las máquinas.”³ (346-347).

2º, Implicaciones metodológicas para el conocimiento científico.

Partiendo del viejo supuesto de que medir un fenómeno es lo mismo que entenderlo –“Medir es saber”, en palabras de Lord Kelvin, presupuesto que sufrió un severo revés con la mecánica cuántica, como reconoce el autor-, la medición es la palanca del progreso para la mentalidad Big data

1. La capacidad de almacenar *todos* los datos relevantes del conjunto de estudio hace innecesaria la epistemología de la muestra significativa –idea desarrollada en el capítulo II-;

Los estadísticos han demostrado que la precisión de la muestra mejora acusadamente con la aleatoriedad, no con el mayor tamaño de la muestra. En realidad, aunque pueda parecer sorprendente, una muestra aleatoria de 1.100 observaciones individuales sobre una pregunta binaria (sí o no, con aproximadamente las mismas probabilidades de darse) es notablemente representativa de toda la población. En 19 de cada 20 casos, presenta un margen de error inferior al 3 por 100, tanto si el tamaño de la población total es de cien mil como si es de cien millones. La razón resulta algo complicada de explicar en términos matemáticos, matemáticos, pero en resumen lo que ocurre es que, superado cierto punto, al principio, conforme las cifras van haciéndose mayores, la cantidad marginal de informaciones nuevas que se consigue de cada observación es cada vez menor. El hecho de que la aleatoriedad se impusiera al tamaño de la muestra supuso una revelación sorprendente. Allaná el camino para un nuevo enfoque de la recolección de información. Los datos que usan muestras aleatorias podían recopilarse a bajo coste y, sin embargo, extrapolarse para el conjunto con gran exactitud. Mayer-Schönberger, Viktor. Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle457-466). Turner. Edición de Kindle.

Principio que se utilizó en la elaboración de censos, negocios –calidad de las manufacturas, encuestas de consumidores-... pero tiene un talón de Aquilés: garantizar la aleatoriedad de la muestra pues:

Trae consigo una serie de debilidades inherentes. Su exactitud depende de que se haya garantizado la aleatoriedad al recopilar los datos de la muestra, pero el logro de esa aleatoriedad resulta peliagudo. Se producen sesgos sistemáticos en la forma de recopilar los datos que pueden hacer que los resultados extrapolados sean muy incorrectos. Las encuestas electorales efectuadas por teléfono fijo dan fe, por ejemplo, de algunos de estos problemas. La muestra está sesgada en contra de la gente que solo usa teléfonos móviles¹⁵ (que suelen ser más jóvenes y más progresistas), como ha señalado el estadístico Nate Silver. Esto se ha constatado en pronósticos electorales erróneos.(...) Y estos pequeños sesgos en la muestra global harán que los errores de los subgrupos sean más pronunciados. Por consiguiente, el muestreo deja de ser útil en cuanto se quiere ahondar más, para escrutar minuciosamente alguna subcategoría de datos que nos llame la atención. Lo que funciona en el nivel macro se viene abajo en el micro. (...) El muestreo requiere, además, una planificación y ejecución cuidadosas. Normalmente no se les puede “pedir” a los datos de la muestra cuestiones nuevas que no se hayan contemplado desde el principio. Así pues, aunque como atajo resulta útil, el coste de oportunidad es precisamente el de que, al final, solo es un atajo. Y siendo una muestra en lugar de un todo, el conjunto de datos carece de la extensibilidad o maleabilidad que serían necesarias para que los mismos datos pudieran ser analizados otra vez con un propósito enteramente distinto de aquel para el que fueron recopilados en origen.” Mayer-Schönberger, Viktor. Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle 478-504). Turner. Edición de Kindle.(La negrita es mía)

³ Y0// Retomar la advertencia nietzschiana: crearemos en Dios mientras creamos en la gramática. Tenemos un inmenso problema con la utilización del lenguaje a la hora de describir o referirnos a lo que las máquinas hacen.

Problemas que se resuelven cuando tenemos <<N = todo>> cuya relevancia está afectado sobre todo a las ciencias sociales... (los finales de partida de ajedrez con 6 piezas o menos ya están analizados en su totalidad, lo que permite que “los programas de ajedrez jueguen sin fallos las fases finales de las partidas de ajedrez, que son cruciales. Ningún ser humano podrá superar nunca al sistema.” (710 y ss.)

2. La enorme cantidad de datos manejables disminuye la importancia de la precisión-exactitud de los datos recogidos: “con menos errores de muestreo, podemos asumir más errores de medida” (275).

Lo relevante aquí es que la cantidad ingente de datos vuelve irrelevante las imprecisiones de medición de algunos de los datos –y eso a sabiendas de que, a más datos, más imprecisiones se integran en el sistema, que las cauteriza por vía cuantitativa-. El ejemplo es medir la temperatura de una viña en previsión de heladas: pocos sensores exigen mucha precisión; muchos, escasa...

Sorprendente la estrategia computacional de traducción: nada de lingüística + diccionario; sencillamente, correlaciones estadísticas entre las secuencias de los idiomas –que desconocen los estadísticos encargados del asunto...⁴. (715y ss.) El eslogan: “más es mejor que mejor”. La obsesión con la precisión es “un resabio de la era analógica privada de información.” (792)

Pero, ¿por qué hay imprecisión en la medición? La respuesta de los autores es maravillosa desde el punto de vista de la exhibición de su mentalidad:

“Merece la pena señalar que la confusión no es algo inherente a los datos masivos. Se trata, por el contrario, de una función de la imperfección de las herramientas que usamos para medir, registrar y analizar la información. Si la tecnología llegara a ser perfecta, el problema de la inexactitud desaparecería. Pero mientras siga siendo imperfecta, la confusión es una realidad práctica con la que tenemos que contar. Y es probable que nos acompañe durante mucho tiempo.” Mayer-Schönberger, Viktor. Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle 814-817). Turner. Edición de Kindle.⁵

Esta posibilidad de integrar metodológicamente la imprecisión, se vende como una virtud del sistema, con un argumento que es en sí mismo un homenaje inconsciente de Nietzsche:

“Por supuesto, puede que algunas etiquetas estén mal escritas, y esa clase de errores introduce imprecisión: no en los datos mismos, pero sí en cómo se organizan. Esto puede molestar a quienes estaban acostumbrados a la exactitud. Ahora bien, a cambio de cierta imprecisión en la forma de organizar nuestras colecciones de fotos, adquirimos un universo de etiquetas mucho más rico, y, por extensión, un acceso más profundo y amplio a nuestras fotos. Podemos combinar las etiquetas de búsqueda para filtrar las fotos de maneras que antes no eran posibles. La imprecisión inherente al etiquetado implica aceptar el desorden natural del mundo. Es un antídoto para sistemas más precisos que intentan imponer una esterilidad falaz sobre el tumulto de la vida real, fingiendo que todo cuanto hay bajo el sol puede disponerse en unas filas y columnas ordenadas. Hay más cosas en el cielo y en la tierra de las que sueña esa filosofía.” Mayer-Schönberger, Viktor. Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle 862-869). Turner. Edición de Kindle.⁶

Este desorden estructural es promovido y utilizado por los nuevos motores de búsqueda. Tanto es así que:

⁴ Tiene gracia el chiste mencionado: cada vez que echamos de la empresa a un lingüista, nuestro traductor mejora considerablemente.

⁵ Yo/// recordar el argumento contra la posibilidad de una medición perfecta... La vinculación entre racionalidad y medida viene de lejos –etimología logos y mente (de mens, mensurar, medir)-. Pero una cosa es medir, es decir, que toda realidad tiene, respectivamente, una dimensión, es dimensional, y otra distinta medir matemáticamente, lo que no siempre está claro por la sencilla razón de que la medida matemática exige, ya lo vieron los griegos, unidad, y no está clara la posibilidad de establecer la unidad de todo lo que de facto medimos. Puedo decir con sentido que el arte de Miguel Ángel supera a... pero no creo que sea factible decir que la proporción en que lo supera es de 1/3.

⁶ Yo/// No estaría de más preguntarle a los autores en qué punto de su exposición se cuele la imprecisión categorial que alaban.

“A cambio de vivir con el desorden, obtenemos servicios tremendamente valiosos que resultarían imposibles a esa escala y alcance con los métodos e instrumentos tradicionales. Según algunas estimaciones, solo el 5 por 100 de todos los datos digitales están “estructurados”, es decir, en una forma que encaja limpiamente en una base de datos tradicional. Si no se acepta la confusión, el 95 por 100 restante de datos sin estructurar, como páginas web y vídeos, permanecen en la oscuridad. Tolerando la imprecisión, abrimos una ventana a un universo de perspectivas por explotar.” Mayer-Schönberger, Viktor. Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle 946-951). Turner. Edición de Kindle.

“Y podremos tolerar lo borroso y lo ambiguo en áreas en las que solíamos exigir claridad y certeza, aun cuando se tratase de una claridad falsa y de una certeza imperfecta. Podremos aceptarlo, siempre y cuando obtengamos a cambio un sentido más completo de la realidad: es el equivalente de un cuadro impresionista, en el que cada pincelada resulta confusa si se la examina de cerca, pero basta con apartarse uno del cuadro para contemplar una imagen majestuosa.” Mayer-Schönberger, Viktor. Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle 962-965). Turner. Edición de Kindle.

3. La eficacia instrumental de las correlaciones estadísticamente significativas a la hora de orientar políticas concretas –en todos los órdenes, desde la economía a la sanidad-, **disminuye la importancia de descifrar relaciones causales con elevado grado de certeza...**: “Los datos masivos tratan del qué, no del porqué. No siempre necesitamos conocer la causa de un fenómeno; preferentemente, podemos dejar que los datos hablen por sí mismos.” (302-304).

Y esto, se mire como se mire, es una ventaja:

1º, por mucho que nuestro cerebro use la causalidad como principio de comprensión del mundo, no es nada fácil establecer correlaciones causales⁷:

En nuestra vida diaria pensamos tan a menudo en términos causales que podemos llegar a creer que la causalidad es fácil de demostrar. La verdad resulta mucho menos cómoda. A diferencia de las correlaciones, en las que el cálculo matemático es relativamente directo, no existe ninguna forma matemática obvia de “probar” la causalidad. Ni siquiera podemos expresar con facilidad las relaciones causales en forma de ecuaciones normales. De ahí que, incluso pensando despacio y detenidamente⁸, resulte difícil encontrar relaciones causales concluyentes. Como nuestras mentes están acostumbradas a la escasez de información, nos sentimos tentados a razonar con datos limitados, aun cuando muy a menudo están en juego demasiados factores para poder reducir simplemente un efecto a una causa particular. Mayer-Schönberger, Viktor. Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle 1282-1288). Turner. Edición de Kindle.

Por consiguiente, igual que sucede con las correlaciones, la causalidad rara vez –o ninguna– puede ser demostrada, solo mostrada con un elevado grado de probabilidad. Mayer-Schönberger, Viktor. Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle 1301-1302). Turner. Edición de Kindle.

Sin embargo, esto no sucede con:

“En cambio, los análisis no causales como las correlaciones suelen ser rápidos y baratos, y para ellos sí disponemos de los métodos matemáticos y estadísticos que permiten analizar las relaciones, y de las herramientas digitales para demostrar su solidez con confianza.” Mayer-Schönberger, Viktor. Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle 1307-1309). Turner. Edición de Kindle.

⁷ Yo!!! Lástima de más filosofía... (cómo resuenan Hume y Kant en todo esto)

⁸ Se refiere aquí a la distinción de Kahneman entre pensar rápido y despacio.

2º, el análisis de relaciones causales plantea problemas morales obvios, que no se dan en el análisis correlacional...⁹

“... los experimentos para confirmar conexiones causales a menudo no resultan prácticos o suscitan cuestiones éticas peliagudas. ¿Cómo podríamos realizar un experimento causal para identificar la causa de que determinados términos de búsqueda predigan mejor la gripe? Y, en el caso de la vacuna antirrábica, ¿podríamos someter a docenas, tal vez cientos de pacientes a una muerte dolorosa –como integrantes del “grupo de control” al que no se le administra la vacuna–, aun teniendo vacunas para ellos? E incluso cuando los experimentos resultan prácticos, no dejan de ser caros y requieren mucho tiempo.” Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition)* (Posición en Kindle1306-1307). Turner. Edición de Kindle.

3º, “... las correlaciones no solo son valiosas por sí mismas: también les indican el camino a las investigaciones causales. Al decirnos que dos cosas están potencialmente conectadas, nos permiten seguir investigando para ver si aparece alguna relación causal y, de ser el caso, por qué. Este mecanismo de filtración, barato y expeditivo, reduce el coste de hacer un análisis causal con experimentos controlados especialmente. A través de las correlaciones, podemos alcanzar un atisbo de las variables importantes que luego usamos en los experimentos para investigar la causalidad.” Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition)* (Posición en Kindle1309-1314). Turner. Edición de Kindle.

4º, además, las correlaciones evitan el sesgo metodológico que introducen las hipótesis (Yo/// o sea, más objetividad):

“En lugar del enfoque sustentado por hipótesis podemos emplear uno sustentado por datos. Puede que nuestros resultados sean menos sesgados y más precisos, y es casi seguro que los obtendremos mucho más deprisa. Las predicciones basadas en correlaciones son el corazón de los datos masivos. Los análisis de correlación se usan con tanta frecuencia hoy en día que, a veces, no valoramos bien el avance que han supuesto. Y sus usos no van a dejar de aumentar.” Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition)* (Posición en Kindle 1094-1097). Turner. Edición de Kindle.

Antes, cuando eran escasos los datos disponibles, tanto las investigaciones causales como los análisis de correlación empezaban con una hipótesis, que se sometía a prueba para desmentirla o confirmarla. Sin embargo, como ambos métodos requerían de una hipótesis de partida, los dos eran igualmente susceptibles al prejuicio y a la intuición errónea. Y además, los datos necesarios no solían estar disponibles. Hoy en día, con tantos datos a nuestro alrededor, y los que quedan por venir, las hipótesis ya no resultan cruciales para el análisis correlacional. Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition)* (Posición en Kindle1206-1210). Turner. Edición de Kindle.

5º. Su eficacia está fuera de toda duda, y es tan elevada que llevó al director general de la empresa FICO a decir: “Sabemos lo que usted va a hacer mañana”. Como muestra, la anécdota en Target:

“Las correlaciones incluso permitieron a la marca estimar la fecha de parto con un estrecho margen de error, de forma que podía enviar los cupones más adecuados a cada fase del embarazo. En su libro *The Power of Habit*, Duhigg cuenta la siguiente anécdota. Un día, un hombre furioso entró en tromba en una tienda Target de las afueras de Minnesota y exigió ver al director. —¡Mi hija ha recibido esto por correo! —gritó—. Todavía está en el instituto, ¿y ustedes se dedican a mandarle cupones para ropa de bebé y cunas? ¿Intentan animarla a que se quede embarazada? Cuando el director llamó al hombre unos días después para disculparse, se encontró con una voz más conciliadora. —He estado hablando con mi hija —dijo el hombre—. Resulta que en mi casa han tenido lugar ciertas actividades de las que yo no estaba del todo informado. Mi hija sale de cuentas en agosto. Soy yo el que les debe una disculpa.”

⁹ Yo/// La ingenuidad de tamaña afirmación se verá matizada por la última parte del libro en la que se analizan –también ingenuamente– los peligros del Big data.

Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle1142-1150)*. Turner. Edición de Kindle.

Es más: salvan vidas. Por ejemplo, los datos de monitorización de prematuros nos permiten desvelar correlaciones que anticipan infecciones antes de cualquier sintomatología clínica. Aunque el médico no lo vea, el ordenador sí.

“Puede que el algoritmo mismo no tome las decisiones, pero las máquinas hacen lo que saben hacer mejor, y ayudarán a los cuidadores a hacer lo que también ellos saben hacer mejor.” Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle1193-1194)*. Turner. Edición de Kindle.

Por supuesto, hay que tomar ciertas precauciones, pues incluso correlaciones fuertes no son “perfectas” y nuestros autores, ¡entre paréntesis!, nos recuerdan:

“(Dicho esto, conviene ser precavido: cuando el número de puntos de datos crece en orden de magnitud, también se aprecian más correlaciones espurias, fenómenos que parecen estar conectados aun cuando no sea así. Esto obliga a tener especial cuidado, como estamos empezando apenas a advertir).” Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle1067-1069)*. Turner. Edición de Kindle.

Problema: ¿supone todo esto el fin de la teoría en el conocimiento científico? Los autores mencionan el artículo de Chris Anderson en Wired en 2008. En síntesis:

*En 2008, Chris Anderson (53)¹⁰, director de la revista Wired, proclamó que “el diluvio de datos vuelve obsoleto el método científico”. En el artículo de portada, titulado “La era de los petabytes”, proclamó que representaba poco menos que “el final de las teorías”. El proceso tradicional del descubrimiento científico –el de la hipótesis puesta a prueba contra la realidad usando un modelo de causalidades subyacentes– está extinguiéndose, afirmaba Anderson, sustituido por el análisis estadístico de correlaciones puras carente de teoría. Para apoyar su argumento, Anderson describía cómo la física cuántica se ha convertido en un campo casi puramente teórico, porque los experimentos son demasiado caros, demasiado complejos y demasiado grandes para resultar viables. Hay teorías –sugirió– que nada tienen que ver ya con la realidad. Como ejemplos del nuevo método, se refirió al motor de búsqueda de Google y a la secuenciación genética. “Este es un mundo en el que las cantidades masivas de datos y las matemáticas aplicadas sustituyen a cualquier otra herramienta –escribió–. Con suficientes datos, las cifras hablan por sí mismas. Los petabytes nos permiten decir: ‘La correlación basta’”. Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle 1393-1403)*. Turner. Edición de Kindle.*

No hay que exagerar pues tanto las elecciones de datos, como la significatividad de las correlaciones se basan en teorías: “Está claro que la era de los datos masivos no carece de teorías: están por todas partes, con todo lo que ello implica.” (1421) Eso sí, la modificación en nuestra comprensión de la función de las teorías y las hipótesis en el conocimiento del mundo es poderosa y supone “un desafío para muchas instituciones” (1421)

4. La relevancia de las correlaciones significativas afecta directamente al conocimiento de los expertos, que son desplazados en beneficio de los estadísticos avezados...

“... el mayor impacto de los datos masivos consistirá en que las decisiones fundamentadas en datos probablemente están llamadas a incrementarse o a invalidar el juicio humano. En su libro Super

¹⁰Punto de vista de Anderson: Chris Anderson, “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete”, Wired, junio de 2008 (http://www.wired.com/science/discoveries/magazine/16-07/pb_theory/). Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle4497-4499)*. Turner. Edición de Kindle. Marcha atrás de Anderson: National Public Radio, “Search and Destroy”, 18 de julio de 2008 (<http://www.onthemedial.org/2008/jul/18/search-and-destroy/transcript/>). Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle4499-4502)*. Turner. Edición de Kindle.

Crunchers, Ian Ayers, economista y profesor de derecho en Yale, argumentaba que los análisis estadísticos fuerzan a las personas a pensar dos veces lo que les dice el instinto. A través de los datos masivos, esto se vuelve todavía más esencial. El experto en un área temática, el especialista sustantivo, perderá parte de su lustre en beneficio del estadístico y el analista de datos, quienes no se ven lastrados por las viejas formas de hacer las cosas y dejan que los datos hablen. Este nuevo marco se basará en unas correlaciones libres de juicios previos y prejuicios, de la misma forma que Maury no dio por buenas sin más las historias que los viejos capitanes le contaban acerca de un pasaje determinado mientras se tomaban una pinta en el pub, sino que confió en que los datos agregados revelaran verdades prácticas. Estamos asistiendo a la desaparición en muchas áreas de la influencia de los expertos temáticos. En el ámbito de los medios de comunicación, el contenido que se crea y publicita en páginas web como The Huffington Post, Gawker y Forbes se ve determinado por los datos, no solo por la opinión de los editores humanos. Los datos pueden revelar qué quiere leer la gente mejor que el instinto de los periodistas veteranos.” Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle2811-2821)*. Turner. Edición de Kindle.

“Seguramente, los expertos en campos determinados no desaparecerán por completo, pero su supremacía menguará. A partir de ahora, van a tener que compartir el podio con los geeks de los datos masivos, de la misma forma que la principesca causalidad tiene que convivir con la humilde correlación. Esto transforma nuestra manera de valorar el conocimiento, porque tendemos a pensar que las personas muy especializadas valen más que los generalistas: que la fortuna favorece a quien profundiza. Sin embargo, la especialización es como la exactitud: apropiada para un mundo escueto en datos, donde uno no tiene nunca información suficiente o correcta, y, por consiguiente, ha de apoyarse en la intuición y la experiencia para hallar el camino. En un mundo así, la experiencia desempeña un papel crucial, puesto que es la acumulación prolongada del conocimiento latente –conocimiento que no puede transmitirse con facilidad ni aprenderse en los libros, o quizá ni siquiera cara a su vida profesional. Las matemáticas y la estadística, tal vez con una pizca de programación y de ciencia de las redes, resultarán tan fundamentales para el puesto de trabajo moderno como lo eran hace un siglo los conocimientos básicos de aritmética, y la capacidad de leer y escribir antes aún. En el pasado, para ser un biólogo excelente uno necesitaba conocer a montones de biólogos. Eso no ha cambiado del todo. Sin embargo, hoy en día también importa la amplitud de los datos masivos, no solo el conocimiento exhaustivo de una especialidad. La solución de un problema de biología intrigante puede venir con la misma probabilidad por la asociación con un astrofísico que con un diseñador de visualización de datos. Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle 2838-2856)*. Turner. Edición de Kindle.

5. Datificación. Todo ellos es posible, claro está, por nuestra capacidad de datificación –cuya historia se recorre someramente en el cap. IV¹¹-. Algunas modestas ideas, y una loa apasionada:

La palabra latina data significa “dado”, en el sentido de “hecho”. Este término se convirtió en el título de una obra clásica de Euclides, en la que explica la geometría a partir de lo que se sabe, o se puede mostrar que se sabe. Hoy en día, por datos se entiende una descripción de algo que permite ser registrado, analizado y reorganizado. Mayer-Schönberger, Viktor. Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle1525-1527). Turner. Edición de Kindle.

“Datificar” un fenómeno es plasmarlo en un formato cuantificado para que pueda ser tabulado y analizado.” Mayer-Schönberger, Viktor. Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle1529-1530). Turner. Edición de Kindle.

“Ver el mundo como información, como océanos de datos que pueden explorarse cada vez más lejos y más hondo, nos ofrece un nuevo panorama de la realidad. Es una perspectiva mental que puede penetrar todas las áreas de la vida. Hoy formamos una sociedad aritmética porque presumimos que el mundo se puede comprender mediante los números y las matemáticas. Y damos por supuesto que el conocimiento se puede transmitir a través del tiempo y del espacio porque el concepto de la escritura está muy arraigado. Puede que el día de mañana las generaciones siguientes tengan una “conciencia de datos

¹¹ Y en la que ahí protagonistas principalísimos vinculados a la Iglesia Católica, mire por donde.

masivos”: la presunción de que hay un componente cuantitativo en todo cuanto hacemos, y de que los datos son indispensables para que la sociedad aprenda. La noción de transformar las innumerables dimensiones de la realidad en datos probablemente le parezca novedosa por ahora a la mayoría de la gente. Pero en el futuro, seguramente la trataremos como algo dado (lo cual, de forma agradable, nos retrotrae al origen mismo del término “dato”). Con el tiempo, puede que el impacto de la datificación deje pequeño el de los acueductos y los periódicos, rivalizando acaso con el de la imprenta e internet al facilitarnos las herramientas para cartografiar el mundo mediante datos. Por el momento, sin embargo, los usuarios más avanzados de la datificación se hallan en el mundo de los negocios...” Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle1921-1931)*. Turner. Edición de Kindle.

3. Repercusiones en el mundo de los negocios –pasamos-

4. Peligros.

Se citan tres:

1. La pérdida –“parálisis”- de la privacidad.

Cómo es obvio, los datos y los perfiles digitales que con ellos se crean amenazan la privacidad de las personas. Para salvaguardarla se ha recorrido a tres procedimientos que, sencillamente, no funcionan:

a) notificación y consentimiento previo:

“Este cambio socava el protagonismo que ahora tiene el individuo en las leyes de privacidad. Hoy en día, a este se le comunica en el momento de reunir la información qué datos se recogen, y con qué fin; tiene entonces la oportunidad de manifestar su consentimiento, antes de la recopilación. Aunque este concepto de “notificación y consentimiento” no es la única forma legal de recoger y procesar datos de carácter personal, según Fred Cate, experto en privacidad en la universidad de Indiana, ha sido transformado en la piedra angular de los principios de privacidad en todo el mundo.” (En la práctica, ha supuesto sobre todo unas alertas de privacidad de tamaño exagerado que rara vez lee alguien, y menos aún comprende; pero esa es otra historia). Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle3051-3057)*. Turner. Edición de Kindle.

b) exclusión voluntaria... que convierte a los excluidos en sospechosos –por el ejemplo de grabaciones de Google en la que propietarios de vivienda solicitaron ser excluidos... con lo que se propusieron como piezas más apetecibles...

c) anonimización:

“La aproximación técnica a la protección de la intimidad –la anonimización– tampoco funciona bien en muchos casos. La anonimización consiste en despojar los conjuntos de datos de todos los rasgos identificadores personales, como pueden ser nombre, dirección, número de tarjeta de crédito, fecha de nacimiento, o número de la seguridad social. Los datos resultantes pueden ser analizados y compartidos sin poner en peligro la privacidad de nadie. Pero esto solo funciona en un mundo escaso en información. Los datos masivos, al incrementar la cantidad y diversidad de la información, facilitan la reidentificación.” Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle3074-3079)*. Turner. Edición de Kindle.

2. La “propensión”.

Se refiere aquí a la tendencia a castigar conductas con criterios preventivos en base a correlaciones estadísticas presuntamente fiables. Los autores –en una reflexión de filosofía elemental, advierten-:

“Esto niega el concepto mismo de la presunción de inocencia, el principio básico de nuestro sistema legal y de nuestro sentido de lo que es justo. Y si hacemos responsable a la gente de unos actos futuros pronosticados, que puede que nunca lleven a cabo, también negamos la capacidad humana de realizar elecciones morales. Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle3234-3237)*. Turner. Edición de Kindle.”

3. La dictadura de los datos¹².

Se trata de confundir la realidad con lo que puede ser datificable-digitalizable¹³. Pero:

“Lo que somos capaces de recopilar y procesar siempre será una fracción minúscula de la información que existe en el mundo. Solo puede ser un simulacro de la realidad, como las sombras en la pared de la cueva de Platón. Como nunca podemos disponer de información perfecta, nuestras predicciones resultan inherentemente falibles. Tampoco significa que sean incorrectas, solo que siempre están incompletas. Esto no niega las percepciones que ofrecen los datos masivos, pero los pone en su sitio: el de una herramienta que no ofrece respuestas definitivas, solo algunas suficientes para ayudarnos por ahora, hasta que aparezcan métodos mejores y, por consiguiente, respuestas mejores. También sugiere que debemos usar esta herramienta con una generosa dosis de humildad... y de humanidad.” Mayer-Schönberger, Viktor. Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle 3910-3916). Turner. Edición de Kindle.

“Si los datos moldean las decisiones, ¿qué propósito queda para las personas, o para la intuición y el ir en contra de los hechos? Si todo el mundo apela a los datos y aprovecha las herramientas de datos masivos, quizá lo que se convierta en el punto central de diferenciación sea la imposibilidad de predecir: el elemento humano del instinto, la asunción de riesgos, el accidente y el error. De ser así, entonces surgirá una necesidad especial de hacerle sitio a lo humano: de reservar un espacio para la intuición, el sentido común y la buena suerte, para asegurar que no los eliminen los datos y las respuestas elaboradas por máquinas. Lo más grande de los seres humanos es precisamente lo que no revelan los algoritmos y los chips de silicio, aquello que no pueden revelar porque no puede ser capturado en forma de datos. No “lo que es”, sino “lo que no es”: el espacio vacío, las grietas de la acera, lo que aún no se ha dicho ni se ha pensado.” Mayer-Schönberger, Viktor. Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle 3886-3893). Turner. Edición de Kindle..

Para luchar contra estos peligros, proponen los autores:

1º, cómo tendencia fundamental, trasladar la responsabilidad de la gestión social de datos del individuo a las corporaciones privadas y públicas que los utilizan –entre otras cosas, porque el individuo jamás va a estar en disposición de conocer qué se registra y qué se hace con lo que registra-;

2º. Como medidas correspondientes: demandar transparencia, certificación y refutabilidad.

“El mismo principio debería aplicarse a las empresas privadas que adoptan decisiones importantes sobre las personas: contratarnos o despedirnos, ofrecernos un préstamo hipotecario o negarnos una tarjeta de crédito. Cuando estas decisiones se basen principalmente en predicciones a partir de datos masivos, recomendamos que se adopten determinadas salvaguardias. La primera es la transparencia: los datos y el algoritmo en que se fundamenta la predicción que afecta al individuo han de estar disponibles. La segunda es la certificación: una tercera parte experta ha de certificar que el algoritmo es correcto y válido para determinados usos sensibles. La tercera es la refutabilidad: hay que especificar formas concretas de que las personas puedan refutar una predicción sobre ellas. (Esto viene a ser análogo a la tradición, en el ámbito de la investigación científica, de revelar aquellos factores que pudieran debilitar

¹² De los datos, y gracias a los datos.: “Los registros civiles de los Países Bajos, conocidos por su exhaustividad, fueron utilizados por los invasores nazis para sus redadas de judíos. Los números de cinco dígitos tatuados en el antebrazo de los prisioneros internados en campos de concentración nazis correspondían inicialmente a números de tarjetas perforadas Hollerith de IBM; el procesamiento de los datos facilitó el asesinato a escala industrial.” Mayer-Schönberger, Viktor. Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle 3020-3023). Turner. Edición de Kindle.

¹³ Yo/// Lo que demanda una filosofía de la mens-ura. Valga, por lo que nos toca, la cita: “¿Parece que la educación se está yendo al traste? Foméntense los tests para evaluar el rendimiento, y penalícese a los profesores o escuelas que no den la talla. Que los tests sirvan para medir las aptitudes de los alumnos, la calidad de la enseñanza o las necesidades de una fuerza laboral moderna, creativa y adaptable es una cuestión aún pendiente; pero es una cuestión que los datos no admiten. ¿Se desea prevenir el terrorismo? Créense capas de listas de vigilancia y de personas a las que se prohíbe volar para vigilar policialmente los cielos. Pero resulta dudoso que esos conjuntos de datos ofrezcan la protección que prometen.” Mayer-Schönberger, Viktor. Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle 3314-3319). Turner. Edición de Kindle.

los hallazgos de un estudio). Lo más importante es que una garantía del albedrío humano nos protege de esa posible dictadura de los datos, la de atribuirles más sentido e importancia de la que merecen. Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle3505-3513)*. Turner. Edición de Kindle.”

Todo lo cual, por supuesto, pondrá de moda la profesión del futuro: el ALGORITMISTA interno y externo encargado de velar por nuestros derechos...

“Mediante estos escenarios podemos advertir el riesgo de que las predicciones basadas en datos masivos, y en los algoritmos y conjuntos de datos que tienen detrás, se conviertan en cajas negras que no nos ofrecen ninguna rendición de cuentas, trazabilidad o confianza. Para impedirlo, los datos masivos requerirán monitorización y transparencia, lo que a su vez hará necesarios nuevos tipos de expertos e instituciones, capaces de ayudarnos a escrutar las predicciones basadas en datos masivos y a que quienes se sientan perjudicados por ellos puedan solicitar reparación. Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle 3556-3560)*. Turner. Edición de Kindle.

Y alguna estrategia antimonopolio de datos masivos...

Para terminar:

“Los datos masivos nos ayudan a hacer mejor lo que ya hacemos, y nos permiten hacer cosas del todo nuevas. Sin embargo, no son una varita mágica. No van a traer la paz mundial, ni a erradicar la pobreza, ni a crear al próximo Picasso. Los datos masivos no pueden hacer un bebé, pero sí salvar a los prematuros. Con el tiempo, llegaremos a esperar que se apliquen en todas las facetas de la vida (y quizá nos sintamos ligeramente alarmados cuando falten en alguna), de la misma forma que esperamos que un médico pida una radiografía para descubrir los problemas que no podría advertir con un examen físico.” Mayer-Schönberger, Viktor. *Big data. La revolución de los datos masivos (Noema) (Spanish Edition) (Posición en Kindle 3842-3847)*. Turner. Edición de Kindle.

Lo que demanda una mentalidad nueva, una “mentalidad Big data”, que reconcilie las oportunidades del Big data con las mencionadas humanidad y humildad.

Amén.