

BIG DATA: ¿A QUIÉN SIRVE?

La era de la información es fecunda en la tarea de alumbrar terminología altamente seductora. *Big Data* no solo alude a una realidad en el tratamiento actual de la información, sino que posee además la capacidad de sugerir al lego la existencia de un nuevo saber al que debería prestar su atención y más frecuentemente su apoyo financiero. Pero, antes de entrar en detalles, quizá será mejor fijar algunos puntos acerca de qué puede entenderse por *Big Data*. Resulta curioso que el rasgo más relevante y en el que parecen coincidir todos los expertos sea presentar este nuevo dominio del saber por oposición a otros ya tradicionales. Así entendemos por *Big Data* la recogida, estudio y explotación de datos que por su magnitud, variedad o complejidad no pueden ser procesados por medios tradicionales. Es inevitable que esta definición recuerde en algo a la de la ya vieja *patafísica* dedicada, en este caso, al estudio de las excepciones que la ciencia tradicional no contempla. Quizá por eso, sea mejor considerar al *Big Data* como la recopilación y análisis de grandes cantidades de información del tipo más variado producida en la interacción de los usuarios, tanto civiles como institucionales a través de la Red.

El *Big Data* no se explica sin la Red y esto es algo que debemos tener en cuenta para entender su funcionamiento y aplicaciones, veámoslo con algún ejemplo. Supongamos que deseo poner en el mercado un producto orientado a los aficionados a los eventos deportivos, en particular el fútbol. Como es harto frecuente en la economía especulativa que viene imperando en los últimos años, mi objetivo no se centra en el producto, me es indiferente, solo deseo saber qué tipo de bien ha de ser, qué precio puedo pedir por él, en qué lugar debe ser distribuido, etc. Hasta no hace mucho, este tipo de trabajo recibía el nombre de *estudio de mercado* y se basaba tanto en el juicio experto de los analistas como en estudios estadísticos basados en pequeñas muestras representativas. Veamos cómo hace el *Big Data*. En este caso el estudio se orientaría no a obtener una muestra pequeña pero significativa, sino que intentaría ser lo más exhaustiva posible analizando toda la información que producen en la Red los usuarios de una cantidad considerable de eventos deportivos. A través del estudio de las correlaciones entre todo género de información producida en los mo-

mentos críticos, se obtendrían las conclusiones que servirían para adoptar una decisión de mercado u otra.

En el límite, imaginemos toda la información producida a nivel global en la Red, esto es, aquella que producimos en nuestras redes sociales, mediante dispositivos móviles, cuando adquirimos un producto, cuando efectuamos un trámite administrativo, cuando buscamos un dato o consultamos una información y simplemente sometámosla a una búsqueda intensiva de patrones y correlaciones. La sugerencia, porque todavía lo es, de los defensores de este tipo de estudios es que posiblemente hallaríamos conexiones funcionales entre comportamientos que solo pueden ser apreciados a esa escala y que no hubieran sido descubiertos simplemente postulando una hipótesis inicial, que muy difícilmente se nos habría llegado a ocurrir.

Esto lleva a otro punto que merece atención. Son muchos —me podría incluir a un nivel muy modesto— los analistas que son capaces de rastrear cantidades notables de información en ciertas redes sociales, Twitter sería el caso paradigmático, con el fin de obtener información valiosa acerca del comportamiento de sus usuarios. Las herramientas, como en el *Big Data*, son aplicaciones que extraen y procesan de forma automática cantidades de información que un agente humano no podría analizar, pero que tampoco requieren un tipo de maquinaria exclusiva. Estos análisis, que podríamos considerar como *Not-so-Big Data*, necesitan una hipótesis previa, un objetivo inicial y no se orientan a obtener la exhaustividad en el estudio de los casos relevantes, trabajan con datos previamente *estructurados*. Pero no son por eso menos interesantes.

Resulta sospechoso que la diferencia entre el *Big Data* y otros estudios de la información circulante en la Red sea a veces el tamaño y los recursos. El mensaje que indirectamente se manda no puede ser más claro: solo aquellos que disponen de los recursos exclusivos de que nosotros disponemos pueden obtener conclusiones de calidad; este no es un negocio al alcance de cualquiera, así que no confíes en aficionados. La cuestión es ahora averiguar qué hay de cierto en esto y para ello quizá convenga hacer notar que para desarrollar un proyecto de *Big Data* no solo hace falta una

gran maquinaria en términos de hardware y software, sino algo tan elemental como los propios datos, detalle en el cual rara vez se insiste. ¿Qué empresas, entidades o corporaciones poseen los datos que, no lo olvidemos, *producimos* como usuarios? La respuesta no es muy compleja por fortuna: aquellas que ofrecen búsquedas, como Google o Yahoo, las redes sociales, Facebook, Twitter o Whatsapp, las que dominan el sector del comercio electrónico como Ebay y Amazon y las centradas en el ámbito del entretenimiento, Youtube, Spotify entre otras. A estas se pueden añadir muchas otras de menor perfil y repartidas por casi todos los ámbitos de la actividad económica en el mundo desarrollado. Obsérvese que todas ellas son empresas privadas movidas, como es obvio, por un afán de lucro y cuyo principal capital es la información que producimos a diario y que queda a buen recaudo en sus bases de datos. Es cierto que para procesar tanta información no basta con un equipo medio, sino que se requiere la potencia de cálculo de grandes ordenadores que muy pocas entidades poseen. Por tanto, será

del comercio, entre las compañías que atesoran nuestros datos y aquellas otras, frecuentemente estatales, que poseen los recursos, de donde surja el *Big Data* tal y como se nos presenta en la actualidad.

¿Debemos temer esta suerte de conjunción entre los propietarios de los datos y los de la maquinaria de análisis? Esta cuestión, que en el fondo es la clave, no tiene, como todo lo importante, una respuesta fácil. Para los *neoluditas* y en general para un amplio colectivo que encaja con dificultad los cambios tecnológicos, la respuesta es obvia. Estaríamos no solo ante un *nuevo Leviatán*, sino ante el último y definitivo *Gran Leviatán*. En la línea de las mejores distopías tecnológicas nos veríamos avocados a un Estado omnipresente capaz de anticipar nuestros movimientos convirtiéndonos en títeres de una realidad alienada cuya trama quedaría finalmente fuera de nuestra vista. Las grandes corporaciones, en plena connivencia con los poderes públicos, crearían dependencias haciendo de nosotros poco más que autómatas perfectamente alineados con sus cadenas de producción. Esta respuesta me parece fácil; es más, me parece repetir en muchos de sus aspectos la reacción que provocó la propia Revolución Industrial en el mítico Ned Ludd y sus seguidores en los albores del siglo xix. ¡Quememos las fábricas y los telares! ¡Acabemos con el maquinismo que arruina la dignidad del hombre! Bien, quémense, que otros aún más grandes y atroces vendrán a ocupar su lugar. Cuando la Historia se fija una pauta con la intensidad que otorgó en su día a la Revolución Industrial, o en la actualidad a la propia Sociedad de la Información, hay muy poco que pueda hacerse por resistir el oleaje. Pero que quede claro, la alternativa tampoco es dejarse llevar.

Cuando me toca abordar este punto, cosa que ocurre con cierta frecuencia, siempre recuerdo el análisis que en su día se hizo de la *fábrica* como representante máximo de la deshumanización y alienación de las masas. El capitalista pensó que había alcanzado la solución perfecta, una forma de incorporar la materia prima humana como un elemento más a su maquinaria. Tenerlos a todos juntos para evitar la autonomía del hombre libre, aumentar sus horarios hasta la extenuación gracias a la concentración de los cuerpos en espacios cerrados y vigilados era el sueño de todo capitalista. Pero nunca tuvo en cuenta que los que antes estaban separados y no se conocían llegaron a convertirse, como en la guerra, en hermanos de sangre primero y más tarde, no mucho, en *cama-*



radar afiliados a sindicatos y partidos de clase. Toda acción genera una reacción y nunca es posible prever todas las consecuencias de una acción, entre ellas las indeseadas para los principales actores de las mismas.

Si extrapolamos este análisis al presente, veremos que la misma fuente de nuestros temores puede ser la forma en que nos protejamos de sus consecuencias más abyectas. Los datos que las grandes corporaciones y gobiernos precisan para sus proyectos de *Big Data* se generan en redes sociales que deben mantener su atractivo para un público pretendidamente variado y representativo. Si mi Red solo interesa a sectores muy definidos simplemente no resultará atractiva. Y ese pequeño margen que queda para el ciudadano digital concienciado puede valer su peso en oro en momentos críticos. No es una conjetura, sino un hecho. Nuestro país, en el que solemos dar poco valor a todo lo que tiene que ver con la innovación, ha sido protagonista de algunos de los movimientos sociales basados en el uso de las redes y las nuevas tecnologías más importantes de los últimos años, sobre todo por su capacidad para alterar la realidad política de manera efectiva. Este espacio para lo dispar es constitutivo de las redes y no puede ser eliminado de ellas sin hacer que el capital en que basan su poder, los datos, se devalúe definitivamente.

Por otra parte, también existen formas de defensa más activa, más cotidiana. Usando las similitudes, a veces fecundas, entre la realidad digital y material, podríamos decir que sin acumulación de capital no hay empresa, ni formación de plusvalía, auténtico motor de toda forma de capitalismo. El capital digital viene dado en este caso por los datos que los usuarios producimos a través de nuestro *trabajo*, digo bien, como usuarios de las más diversas redes. El trabajo realizado por estos usuarios requiere de su concurrencia activa en una plataforma, concurrencia sin la cual no se generará *plusvalía digital*, medida esta en términos del tiempo de uso de una Red. Para ello, es preciso que cada plataforma consiga dar con una fórmula de éxito capaz de atraer un número considerable de usuarios. La formación de capital digital es muy barata en términos de recursos —capital y maquinaria— materiales, pero requiere el despliegue de mucha habilidad para alcanzar el régimen de producción que permita la generación de plusvalía digital. Una Red que no cruce el rubicón del millón de usuarios registrados y activos apenas podría considerarse como una genuina *red social* de la era digital. Y esos valores no son sencillos de

alcanzar aunque sí de perder. Un paso en falso, una propaganda negativa creíble, una mala práctica o un sesgo muy reconocible puede hacer que una Red pierda rápidamente usuarios a favor de una alternativa eficiente. Pero tampoco hay necesidad de que los *señores de la red* lo hagan mal, basta con comportarnos de forma promiscua buscando alternativas eficientes que ralenticen, compliquen o vuelvan inestables los procesos de formación de capital digital. No toda búsqueda tiene por qué depender de Google, ni tenemos por qué centrar toda nuestra vida social en una sola Red, ni tampoco asociar toda nuestra identidad a un único perfil de usuario. Porque para impedir la formación de capital digital o moldearla a favor de un interés más general, no solo hay que resistirse a concurrir donde de hecho también concurre la mayoría, también cabe actuar sobre la fiabilidad o el valor de los datos.

En fechas recientes ha sido muy comentado el fiasco de Google en torno a las predicciones de la incidencia de epidemias de gripe en Estados Unidos. Supuestamente era fácil hacerse una idea muy precisa prestando atención a las consultas que los usuarios practicaban en su buscador y en las redes sociales del gigante, pero no fue así. Quizá aquellos que se encuentran afectados no tengan ganas de encender el ordenador salvo para ver una película relajante o imaginar unas vacaciones de ensueño... quizá sea por otro motivo, pero lo cierto es que falló. También podemos imaginar que no lo hizo y que, al mejor estilo *conspiranoico*, se nos hizo creer que sus predicciones fueron fallidas para apartar sus éxitos de nuestra mirada, pero no lo creo. Quizá sea debido a mi formación en el campo de la lógica matemática, pero he visto muchas veces como los intentos de manejar una totalidad de entidades del tipo que sea manteniéndola bajo estrictos criterios de control suele llevar a paradojas y limitaciones que ponen nuestras ambiciones en su sitio. Aún no se ha demostrado que el *Big Data* encierre en sí mismo un proyecto imposible, quizá no se haga nunca, pero tampoco creo que llegue a dominar nuestras mentes y voluntades a través de la acción de un *Gran Hermano* tecnológico. Como en otras ocasiones en la Historia de la humanidad eso es algo que solo ocurrirá si *We the People* así lo permitimos, y quiero creer que seremos lo suficientemente maduros y conscientes como para que así lo hagamos. Sirva lo que viene a continuación para que cada cual saque sus conclusiones.

Enrique Alonso