

Los problemas de la minería de datos

Tomado de "Post, Mine, Repeat, Social Media Data Mining Becomes Ordinary", Helen Kennedy, Palgrave Macmillan, UK, 2016.

1. Menos privacidad, más vigilancia

(Antes de empezar investigue quien es Edward Snowden)

En 2010, el fundador y CEO de Facebook, Mark Zuckerberg, anunció que en la era de las redes sociales, la privacidad ya no es "una norma social" (Johnson 2010)¹. A pesar de las críticas a las palabras de Zuckerberg, existe una aceptación generalizada de este punto de vista, ejemplificado, por ejemplo, en un artículo de una revista de Nueva York en 2007 que apodó a los jóvenes conocedores de las redes sociales como la generación "que dice todo". Esta generación, afirmó el artículo, piensan en sí mismos como teniendo una audiencia, archivan abiertamente los momentos de su adolescencia y tienen una piel más gruesa que las generaciones anteriores.

Varios comentaristas han impugnado este punto de vista, como Danah Boyd, cuya extensa investigación etnográfica sobre las actitudes de los adolescentes hacia las redes sociales se informa en su libro *It's Complicated: the social life of networked teens* (2014). Aquí, ella argumenta que la privacidad todavía es importante para los jóvenes de hoy. Anteriormente, en un discurso en 2010, Boyd utilizó la noción de "ser privado en público" para resaltar la compleja relación entre la publicidad y la privacidad en las redes sociales. Al comparar las redes sociales con los corredores, argumentó que si dos personas se topaban entre sí en el espacio público del corredor, uno podría decir algo privado al otro que no querría que se publicitara (boyd 2010). Sobre la base de esta metáfora, en su libro, Boyd hace una distinción entre ser público y estar en público. Los jóvenes podrían estar haciendo lo último, pero aún quieren privacidad, argumenta (boyd 2014).

Esto podría entenderse como "privacidad en contexto" o lo que la filósofa de medios Helen Nissenbaum (2009) ha denominado "integridad contextual". Nissenbaum argumenta que cuando las personas divulgan información personal en un contexto particular, tienen expectativas de lo que sucederá con sus datos en ese contexto. Ignorar esas expectativas es una violación de los derechos de datos de las personas. Lo que se necesita, entonces, es la integridad contextual, o respetar las expectativas de las personas sobre lo que razonablemente podría sucederle a sus datos dentro de un contexto dado. Contrariamente a lo que afirma Zuckerberg, estos argumentos sugieren que la privacidad sigue siendo, de hecho, una norma social.

Por supuesto, a las empresas de redes sociales les interesa ganar dinero vendiendo lo que los usuarios comparten en las plataformas de redes sociales para decirnos que ya

¹ Algunos argumentan que Zuckerberg hizo esto para justificar la controvertida decisión de su empresa de cambiar la configuración de privacidad predeterminada de las cuentas de sus usuarios y hacerlos más públicos de lo que habían sido anteriormente.

no nos importa la privacidad; de hecho, tales estrategias juegan un papel en la forma en que pensamos. Sin embargo, a pesar de los esfuerzos de Zuckerberg y otros para descartar la importancia de la privacidad en los entornos de las redes sociales, el concepto conserva la tracción, por lo que los ejemplos de 'invasión' de la privacidad de las redes sociales son recibidos por la preocupación del público, y los investigadores académicos buscan entender lo que llaman la "paradoja de la privacidad", o el hecho de que las prácticas de compartición de los usuarios de las redes sociales parecen contradecir sus preocupaciones expresadas sobre la privacidad. En relación con esta 'paradoja', algunos autores concluyen que, para los usuarios, existe una distinción entre la privacidad social (controlar qué personas dentro de sus redes tienen acceso a su información) y la privacidad institucional (la extracción de información personal por las plataformas de redes sociales y otras compañías comerciales) y que las preocupaciones de los usuarios de las redes sociales sobre el control de su información personal se relacionan con la primera, no con la segunda (Raynes-Goldie 2010; Young y Quan-Haase 2013).

Boyd (2014) proporciona una posible explicación para esto: ella argumenta que todos buscamos privacidad de aquellos que tienen poder sobre nosotros y, para muchos usuarios jóvenes de redes sociales, las personas que tienen poder sobre ellos están dentro de sus propios mundos sociales. Entonces, para algunos escritores, un problema con la minería de datos de redes sociales es que invade la privacidad personal. Otros han sugerido que este enfoque en la invasión de la privacidad personal no reconoce el problema estructural más significativo del aumento de la vigilancia que trae la minería de datos en las redes sociales. Netichailova (2012), por ejemplo, argumenta que un enfoque en las responsabilidades de los usuarios para tener más cuidado en el manejo de su privacidad personal ignora lo que ella llama 'aspectos sociales', con lo que se refiere a las fuerzas de estructuración político-económica dentro de las cuales la minería de datos toma lugar.

Al igual que la amenaza a la privacidad, la amenaza de una mayor vigilancia es una crítica importante a la minería de datos en las redes sociales, ya que circulan ejemplos de alto perfil de vigilancia en las redes sociales, siendo las revelaciones de Snowden las más visibles y espectaculares. Por lo tanto, las preocupaciones sobre menos privacidad y más vigilancia a veces se enfrentan entre sí, pero al mismo tiempo tienen mucho en común, y no menos importante su capacidad de llamar la atención del público.

En *Social Media as Surveillance* (2012), Trottier argumenta que la minería de datos en redes sociales es una nueva forma de vigilancia (ver también Fuchs 2014). Otros han argumentado que, junto con la vigilancia institucional y estatal, como se ejemplifica en las prácticas de minería de datos llevadas a cabo por gobiernos y actores comerciales, surgen nuevas formas de vigilancia social en las redes sociales. Esta es la frase que utiliza Marwick (2012) para describir la observación recíproca que tiene lugar en las plataformas de redes sociales. Asimismo, Albrechtslund (2008) describe las redes sociales en línea como vigilancia participativa; tales prácticas también se describen como vigilancia lateral o autovigilancia. Humphreys (2011) afirma que los participantes en su estudio de la aplicación de check-in Dodgeball se preocupan por estas formas de vigilancia social y lateral, no por la vigilancia institucional, haciéndose eco de la

distinción hecha en los estudios relacionados con la privacidad discutidos anteriormente.

Mark Andrejevic ha escrito extensamente sobre cómo las nuevas tecnologías digitales, incluidas las plataformas de redes sociales, amplían las prácticas de vigilancia (2004, 2007, 2011, 2013). Ya en 2004, argumentó que las tecnologías interactivas permiten que las actividades no supervisadas previamente estén sujetas a vigilancia y, por lo tanto, nuestras prácticas de comunicación digital se conviertan en parte de la sociedad de vigilancia. Más recientemente, al escribir sobre una gama de prácticas que incluyen formas de minería de datos de redes sociales como análisis de sentimientos y minería de opinión, pero también análisis de lenguaje corporal, neuromarketing y tecnología de drones, Andrejevic describe las diversas formas en que nuestras acciones están sujetas a un escrutinio vigilante, proporcionando una serie convincente de ejemplos para respaldar su punto (Andrejevic 2013).

Al escribir con Kelly Gates sobre la capacidad de las tecnologías de drones para capturar todo el tráfico de datos inalámbricos en un área, argumenta que en la era de big data, las definiciones de *vigilancia* cambian. Anteriormente entendida como "intencional, rutinario, sistemático y enfocado" (Murakami Wood et al. 2006, p. 4, citado en Andrejevic y Gates 2014, p. 189), ahora es mucho menos. Ahora, los drones capturan todos los datos sobre todos, y los almacenan para siempre, y la vigilancia es mucho más especulativa: los datos se capturan para una clasificación retrospectiva, en el caso de que puedan ser relevantes y de uso en el futuro. Esto es evidente en las palabras del Director de Tecnología de la CIA:

El valor de cualquier información solo se conoce cuando se puede conectar con otra cosa que llega en un momento futuro. [...] Dado que no puedes conectar puntos que no tienes, esto nos lleva a que fundamentalmente tratemos de recopilar todo y mantenerlo para siempre. (Sledge 2013, citado en Andrejevic y Gates 2014, p. 185).

Andrejevic y Gates señalan que existen serios desafíos de infraestructura para aquellos que recopilan todos los datos todo el tiempo, que se relacionan con el espacio de almacenamiento físico, así como con las herramientas y técnicas. Las plataformas de redes sociales como Facebook cuentan con la infraestructura que permite que las plataformas de comunicaciones sirvan como sistemas de vigilancia; en una reciente lista en línea de las diez bases de datos más grandes del mundo, figuran una serie de plataformas de redes sociales (Anonymizous 2012, citado en Andrejevic y Gates 2014, p. 189). Por lo tanto, las redes sociales están imbricadas en formas contemporáneas de vigilancia, como actores en prácticas de vigilancia, como infraestructuras que permiten la vigilancia y como bases de datos que albergan los conjuntos de datos que son vigilados.

Algunas de las prácticas reveladas por Snowden, y escritas por Andrejevic y otros, deberían preocuparnos, ya que plantean serias preguntas en relación con los derechos humanos, las libertades y la justicia social. Pero hay dos preguntas que deberíamos hacer en relación con las inquietudes en esta sección.

La primera pregunta es si tales preocupaciones se aplican igualmente a las formas contemporáneas de minería de datos, en las plataformas de redes sociales y en otros lugares, dada la diversidad actual de las prácticas de minería de datos. La segunda pregunta se relaciona con lo que se oculta a la vista cuando los ejemplos de alto perfil de invasión de privacidad y vigilancia total capturan tan fácilmente la imaginación del público. Ciertas historias reciben atención precisamente porque destacan las invasiones de la privacidad personal y la difusión de la vigilancia. El predominio de tales historias podría verse como un "efecto post-Snowden", ya que las revelaciones de Snowden pueden haber creado un espacio en el que es posible reconocer cómo la minería de datos contribuye a una mayor vigilancia y violación de la privacidad. Pero como estas historias ocupan el centro de atención, ¿qué otros aspectos de la minería de datos de redes sociales no son tan visibles? Una respuesta a esta pregunta se relaciona con cuestiones metodológicas sobre cómo la minería de datos clasifica, discrimina y estructura la vida social.

2. Discriminación y control

Varios investigadores han criticado las consecuencias discriminatorias de la minería de datos. Por ejemplo, Trow, ha llevado a cabo numerosos estudios en las industrias de publicidad digital para revelar cómo usan los datos para discriminar. En *The Daily You: cómo la nueva industria publicitaria está definiendo su identidad y su valor* (2012), Trow argumenta que la extracción de datos que llevan a cabo estas industrias conduce a la discriminación social. Esto sucede porque, a través de los procesos de análisis de datos, los "perfiles individuales" se convierten en "evaluaciones individuales" (2012, p. 6). Los cálculos del valor de marketing de cada persona se producen, en función del comportamiento y otras formas de seguimiento de datos, y cada individuo se clasifica como objetivo o desperdicio. Estos datos definen nuestra identidad y nuestro valor, argumenta Trow, determinando no solo lo que hacen las empresas de marketing, sino también cómo nos vemos a nosotros mismos y a los demás. Aquellos de nosotros que se consideran desperdicios recibimos opciones limitadas en los mensajes publicitarios que están dirigidos a nosotros y, según Trow, estos afectan nuestro sentido de identidad.

Beer y Burrows (2013) señalan las formas en que la discriminación de datos opera en otros aspectos de la cultura. Al escribir sobre las formas en que las tecnologías de consumo de música generan datos archivables sobre los hábitos de consumo (qué tipos de música se consumen, con qué frecuencia y durante cuánto tiempo, las ubicaciones geográficas de las prácticas de consumo), trazan un mapa de cómo estos datos juegan un papel en la constitución de hardware, software y comercio de consumo de música de diferentes maneras. Argumentan que los datos sobre las prácticas de escucha alimentan "la producción de sistemas geodemográficos nacionales a gran escala que a su vez proporcionan un análisis a nivel de código postal de los gustos y preferencias de las personas" (2013, p. 59). Como tal, los datos constituyen mucho más que cultura, y sirven también para dar forma a los regímenes de gobierno y control. En otro lugar, al escribir específicamente sobre el análisis de sentimientos, Hearn argumenta que la capacidad de la minería de datos para identificar sentimientos valiosos representa otro

mecanismo capitalista de extracción de valor. Describe a las personas que realizan este trabajo como "intermediarios de sentimientos", que "estructuran los sentimientos en beneficios para ellos y sus clientes" (2010, págs. 435–43). Del mismo modo, Andrejevic (2011) está preocupado por el papel desempeñado por el sentimiento de las redes sociales en la predicción y el control del afecto, que describe, citando a Massumi, como "una variable intrínseca del sistema capitalista tardío, tan infraestructural como una fábrica" (Massumi 2002 , p. 45, citado en Andrejevic 2011, p. 609).²

Estos escritores destacan las formas problemáticas en que las capacidades discriminatorias de la minería de datos se ponen en funcionamiento, a través de la vigilancia, la clasificación social y otras formas de control. Pero mientras que gran parte de este trabajo se centra en la discriminación intencional en la minería de datos, Barocas y Selbst (2014) llaman nuestra atención sobre las formas en que la minería de datos discrimina involuntariamente. La minería de datos, argumentan Barocas y Selbst, es discriminatoria por diseño. "El punto mismo de la minería de datos", escriben, "es proporcionar una base racional sobre la cual distinguir entre individuos" (2014, p. 6). En un documento que considera si la ley de los Estados Unidos está a la altura de legislar por la discriminación que ocurre en el contexto de la minería de datos, estos autores trazan las formas en que los distintos componentes técnicos de la minería de datos pueden conducir a la discriminación. Estos incluyen: el proceso de proponer definiciones; la selección de datos de entrenamiento; selección de características; el uso de poderes; y enmascaramiento, o la capacidad de los tomadores de decisiones con opiniones perjudiciales para hacer que la discriminación intencional parezca involuntaria. Por lo tanto, llaman la atención sobre los problemas con el funcionamiento de la minería de datos. Al destacar estos procesos no intencionales, Barocas y Selbst señalan cuestiones metodológicas y epistemológicas que han preocupado a otros comentaristas sobre los métodos de minería de datos.

3. Preocupaciones metodológicas

La minería de datos en las redes sociales es un método para tratar de comprender los patrones sociales y, como tal, plantea preocupaciones metodológicas. Estas preocupaciones se capturan hasta cierto punto en la idea de que los métodos de todo tipo, no solo la minería de datos de redes sociales, tienen una vida social (ver, por ejemplo, Law et al. 2011). Los métodos son del mundo social; nacen con propósitos y defensores. La encuesta por muestreo, escriben Law et al., ganó impulso en el Reino Unido en la década de 1960 como un medio para investigar la sociedad que omitió los puntos de vista de la élite y, como tal, fue apoyada o defendida por el gobierno de la época. Pero los métodos no solo están "formados por el mundo social en el que están ubicados" sino que también "ayudan a dar forma a ese mundo social" (Law et al. 2011, p. 2). Los métodos representan simultáneamente aspectos del mundo social y tienen efectos sociales, constituyendo las cosas que dicen representar (Law y Urry 2004). Al igual que todos los demás métodos de investigación, las prácticas de minería de datos

² Christian Fuchs is another writer who has highlighted how the discriminatory potential of data mining is captured in the interests of capital (for example, Fuchs 2011).

son "performativas": tienen efectos; ellos hacen diferencias; ellos promulgan realidades; y pueden ayudar a crear lo que también descubren (Law and Urry 2004, pp. 392-3). Tarleton Gillespie señala lo mismo sobre los algoritmos que subyacen a las técnicas de minería de datos, describiéndolos como "mecanismos socialmente contruidos y gestionados institucionalmente para asegurar la perspicacia pública: una nueva lógica del conocimiento" (2014, np). Por lo tanto, es importante pensar críticamente sobre los métodos", sobre lo que están haciendo los métodos y el estado de los datos que están haciendo" (Law et al. 2011, p. 7). Esto significa poner atención a los problemas, sobre la capacidad de la minería de datos de las redes sociales para proporcionar acceso a las opiniones, actitudes y comportamientos de las personas, y en su lugar ver los datos que se producen a través de estos métodos tal como están hechos y formados.

Las decisiones humanas influyen y dan forma al diseño, desarrollo, disposición e implementación de sistemas y artefactos tecnológicos, y el ensamblaje de la minería de datos con las redes sociales no es una excepción.

Gitelman y Jackson (2013) consideran este punto en su introducción a 'Raw Data' is an Oxymoron (2013), en el que contrarrestan los supuestos implícitos en el término 'datos en bruto' (de neutralidad y objetividad) con referencia a la afirmación de Bowker, parafraseada en el título del libro, de que 'los datos en bruto son tanto un oxímoron como una mala idea; por el contrario, los datos deben cocinarse con cuidado' (Bowker 2005, p. 184). Los datos no existen simplemente, argumentan Gitelman y Jackson (y muchos otros, como Boyd y Crawford 2012). Más bien, necesitan ser generados y, para ser generados, necesitan ser imaginados. Las afirmaciones epistemológicas sobre lo que puede lograr el análisis de big data representan una forma de imaginación. Bollier (2010) agrega a esto, argumentando que producir datos necesariamente involucra interpretación, y la interpretación, a su vez, está necesariamente sesgada por los filtros subjetivos que aplican los actores humanos individuales involucrados en su generación, como en los procesos descritos por Barocas y Selbst. (2014)

La forma en que se reconoce la "cocción" de los datos es tema de debate. Boellstorff (2013) argumenta que quienes trabajan con datos grandes o sociales entienden sus conjuntos de datos como representaciones limitadas del mundo, condicionadas por las teorías que enmarcan su generación. De manera similar, Havalais (2013) propone que "Nadie espera que las cantidades masivas de datos recopilados por el Gran Colisionador de Hadrones (Large Hadron Collider) o varios instrumentos astronómicos compartidos estén libres de la influencia del proceso por el cual fueron recopilados". Asimismo, Barnes señala que la suposición de que cualquier conocimiento científico, incluido el derivado de grandes datos sociales, es el producto de una "lógica universal incorpórea" (2013, p. 298) ha sido cuestionado durante mucho tiempo. Estos escritores sugieren que las limitaciones de los datos digitales y los métodos relacionados son ampliamente conocidos. Por otro lado, algunos críticos argumentan que se necesita una mayor conciencia de la configuración social de los datos a través de los métodos que se utilizan para generarlos, porque existe una 'creencia generalizada de que los grandes conjuntos de datos ofrecen una forma superior de inteligencia y conocimiento que puede generar ideas que antes eran imposibles, con el aura de verdad, objetividad y precisión' (Boyd y Crawford 2012, p. 663). En la introducción de un número especial de la revista First

Monday sobre big data, Helles y Bruhn Jensen (2013) están de acuerdo. Argumentan que la suposición de que los datos están 'disponibles y abundantes' existe entre los investigadores, así como en el debate público, y sin embargo, 'el proceso de generar los materiales que funcionan como datos a menudo sigue siendo opaco y ciertamente no está documentado en la investigación publicada' (2013).

Pero aunque los debates sobre la conformación y la creación de datos a través de la minería de datos se están llevando a cabo dentro de campos académicos como las ciencias sociales, que están acostumbrados a someter sus métodos a un escrutinio crítico, no puede decirse lo mismo de los actores en organizaciones de pequeña escala que comienzan a pensar en cómo la minería de datos en redes sociales podría ayudarlos a cumplir sus objetivos. Hay muchas razones para esto. Una es porque la "caja negra" (Pinch 1992) de los métodos de minería de datos y los procesos algorítmicos hace que sea difícil saber, y mostrar, cómo las herramientas y los procesos crean y dan forma a los datos. El concepto de "caja negra", resalta el hecho de que hay poca comprensión del funcionamiento interno de las tecnologías que constituyen nuestro mundo social.

Gillespie (2014) argumenta que este es un problema para varias formas de minería de datos de redes sociales, incluido el algoritmo de tendencia de Twitter y plataformas de medición de reputación digital como Klout, ya que los criterios por los cuales operan dichos sistemas son generalmente desconocidos (Gillespie 2014). Otra razón por la cual las formas en que la minería de datos construye datos puede no ser transparente para los actores "ordinarios" es que los datos se separan regularmente de sus condiciones de producción una vez "en la naturaleza". El ejemplo de *Target (25 productos cuya compra contribuyó a un "puntaje de predicción de embarazo")*, es un ejemplo. Aunque los analistas de datos de la compañía podrían haber sido conscientes de que solo identificaban la probabilidad de que la adolescente estuviera embarazada, una vez que sus datos se convirtieron en "números en la naturaleza", lo que antes eran posibilidades se convirtió en algo más concreto. Hablando sobre el movimiento de datos a través del espacio, Gerlitz (2015) cita a Espeland y Sauder (2007) quienes escriben: 'los números son fáciles de desalojar de los contextos locales y reinsertarse en contextos más remotos. Debido a que los números se descontextualizan tan a fondo, invitan a la recontextualización. La minería de datos en las redes sociales está motivada, produce y reproduce un' deseo de números 'que tiene varias consecuencias preocupantes, incluida su creación. Difícil crear un espacio para reflexionar críticamente sobre los métodos de minería de datos. Esta, entonces, es otra posible razón por la que, en el dominio público, rara vez se destaca la forma en que se crean y forman los datos. Más bien, su aparente objetividad se da por descontada. Por ahora, se da paso a un tema final en las críticas a la minería de datos: cuestiones de no representatividad, exclusión y acceso limitado a las habilidades, el conocimiento y la experiencia necesarios para participar en la minería de datos en las redes sociales y abrir su caja negra.

4. Problemas de acceso y desigualdad

Escribiendo en 2012 sobre el creciente interés en Big Data en las ciencias sociales, Boyd y Crawford insisten en la necesidad de abordar una serie de preguntas críticas sobre el

fenómeno de Big Data, incluido quién tiene acceso a él y las habilidades, conocimientos y experiencia necesarios para participar en la minería de datos. Por lo tanto, una de sus "seis provocaciones para Big Data (six provocations for big data)", como se llama su artículo, es que "el acceso limitado a Big Data crea nuevas divisiones digitales" (Boyd y Crawford 2012, p. 673). Señalan que, si bien gran parte del entusiasmo que rodea a Big Data proviene de la creencia de que es fácil y sencillo acceder, este no es el caso. Lev Manovich está de acuerdo, afirmando que 'solo las compañías de redes sociales tienen acceso a datos sociales realmente grandes. [...] Un antropólogo que trabaje para Facebook o un sociólogo que trabaje para Google tendrá acceso a datos que el resto de la comunidad académica no tendrá (Manovich 2011, p. 5).

Sin duda, las empresas comerciales de élite como Google, Facebook y Amazon tienen el mejor acceso a los datos, así como las mejores herramientas y métodos para darle sentido (Williamson 2014); algunas compañías incluso han ido tan lejos como para desalentar a los académicos de investigar las redes sociales. Algunas compañías restringen el acceso a los datos por completo, otras venden el acceso por una tarifa alta, algunas ofrecen pequeños conjuntos de datos a investigadores universitarios. Por lo tanto, aquellos con dinero o dentro de una empresa tienen acceso diferencial a los datos de las redes sociales comparado con aquellos que no tienen los recursos financieros necesarios para adquirirlos. Boyd y Crawford (2012) argumentan que, dada esta tendencia, pronto solo las universidades de nivel superior y con recursos suficientes podrán negociar el acceso a las fuentes de datos comerciales, y los estudiantes de esas universidades de élite y nivel superior probablemente sean los únicos invitados a trabajar dentro de las empresas de redes sociales, exacerbando los patrones existentes de desigualdad.

Manovich afirma que existen tres formas de relacionarse con Big Data: hay "aquellos que crean datos (tanto conscientemente como dejando huellas digitales), aquellos que tienen los medios para recopilarlos y aquellos que tienen experiencia para analizarlos" (2011, p. 10). En el caso de la minería de datos, como señalan Boyd y Crawford, quién se considera que tiene experiencia determina quién controla el proceso y el 'conocimiento' sobre el mundo social que resulta, conocimiento que a su vez reproduce el mundo social. "Usar APIs (interfaces de programación de aplicaciones), extraer y analizar grandes cantidades de datos es un conjunto de habilidades generalmente restringido a aquellos con antecedentes computacionales" (2012, p. 674), argumentan. Manovich confirma esto, afirmando que en 2011, "todos los proyectos de investigación verdaderamente a gran escala, que utilizan los datos con API (redes sociales) hasta ahora han sido realizados por investigadores en informática" (2011, p. 5). Él da algunos ejemplos de por qué este es el caso: por ejemplo, si desea extraer los libros digitalizados en Ngram Viewer de Google para analizar las frecuencias cambiantes en los temas de los libros a lo largo del tiempo, necesita experiencia en lingüística computacional y minería de textos para hacerlo.

Por lo tanto, el acceso a los datos, la minería de datos y las habilidades para hacerlo son desiguales y esto conduce a nuevas divisiones digitales. Boyd y Crawford advierten que existe el peligro de que surja una gran cantidad de datos ricos y de datos pobres en relación con el acceso a datos, herramientas, habilidades y experiencia. Pero, ¿por qué

importa ese acceso? Citan a Derrida para responder a esta pregunta. Derrida afirma que un criterio esencial por el cual se puede medir la "democratización efectiva" es "la participación y el acceso al archivo, su constitución y su interpretación" (Derrida 1996, p. 4, citado en boyd y Crawford 2012, p. xx). El acceso diferencial a los archivos de datos, la forma en que están constituidos y utilizados representan "desigualdades escritas en el sistema" (Boyd y Crawford 2012, p. xx). Las preocupaciones sobre las divisiones de datos no cuestionan las operaciones subyacentes de los datos. Más bien, llaman la atención sobre la distribución desigual de la minería de datos y las habilidades correspondientes, y buscan resaltar el carácter problemáticamente antidemocrático de tales desigualdades.

Las otras críticas a la extracción de datos sociales y de otro tipo se centran más explícitamente en las consecuencias discriminatorias de la extracción de datos y, como tales, señalan algunos de sus efectos preocupantes. Las críticas que se centran específicamente en las redes sociales sirven para contrarrestar los relatos más festivos de las posibilidades democráticas abiertas por las redes sociales y sus aclamadas posibilidades de participación (por ejemplo, en Jenkins 2008). Hacen el importante trabajo de hacer visible lo que es en gran medida invisible, es decir, el trabajo de monitorear, extraer y rastrear los datos que dejamos atrás a medida que avanzamos por las redes sociales. Pero no abordan la cuestión de si los seres humanos deben someterse a la lógica de las redes sociales y las bases de datos, o si es posible que la minería de datos sea compatible con los intereses de los actores, ciudadanos y públicos de pequeña escala.