

Grau en Estadística

Títol: Reflexions sobre les potencialitats i els límits del Big Data

Autora: Laura Pereira Rueda

Directora: Mireia Fernández Ardèvol

Departament: Econometria, Estadística i Economia Aplicada

Convocatòria: 2a Convocatòria. 29 de juny de 2018



Resum

Les grans dades han esdevingut un concepte molt popular en el segle XXI, gràcies a les grans onades d'informació que disposem avui dia. El Big Data està transformant les nostres vides. Tot i això, encara no existeix una definició rigorosa del concepte. Aquest article ofereix una reflexió sobre les potencialitats i les limitacions de les grans dades, a partir d'una anàlisi de la bibliografia existent al voltant d'aquestes. A més, es presenten alguns dels elements clau en les grans dades com la dadificació, les correlacions o els algorismes, per tal d'ajudar a entendre el Big Data. En aquest article es fa especial menció a la pèrdua de privacitat i a la discriminació que poden estar vinculades a les grans dades. A partir d'aquest discurs us convido a reflexionar sobre el món de les grans dades. Un tema candent actualment però que encara s'està definint.

Paraules clau: grans dades, les Vs de les grans dades, dadificació, privacitat per disseny, gran germà, discriminació, aproximació crítica.

Reflections on the potential and limits of Big Data

Abstract

Big Data has become a very popular concept in the Twenty-First Century, thanks to the huge waves of information that we have today. Big Data is transforming our lives. However, there is not yet a rigorous definition of the concept. This article tries to offer a reflection on the potentialities and limitations of Big Data, from an analysis of the existing literature on the area. In addition, some key elements in the Big Data arena are discussed such as datafication, correlations or algorithms, in order to help to understand the Big Data. This paper, particularly focuses on the loss of privacy and discrimination that can be linked to Big Data. From this perspective, I invite you to reflect on the world of Big Data. A hot topic at the moment, but that is still being defined.

Keywords: Big Data, 3 Vs of Big Data, datafication, privacy by design, Big Brother, discrimination, critical approach.

Classificació AMS (*American Mathematical Society*)

G2P25: Applications to Social Science

Índex

INTRODUCCIÓ	5
JUSTIFICACIÓ	5
PROBLEMÀTICA DE RECERCA	5
OBJECTIUS	6
ESTRUCTURA	6
AGRAÏMENTS	7
METODOLOGIA	8
REFLEXIONS SOBRE LES POTENCIALITATS I LES LIMITACIONS DEL BIG DATA	10
EL MÓN DE LES GRANS DADES (BIG DATA)	11
LES DIMENSIONS DEL BIG DATA A PARTIR DE LES VS	13
DADIFICACIÓ	15
ALGORISMES	16
EL FINAL DE LA TEORIA	18
POTENCIALITATS I LIMITACIONS DE LES GRANS DADES	20
LES LLIBERTATS CIVILS ESTAN EN JOC?	24
LES GRANS DADES SÓN DISCRIMINATÒRIES?	29
CONCLUSIONS	32
BIBLIOGRAFIA	35
ANNEX	40
BIBLIOGRAFIA ANOTADA	40

Introducció

Aquest document constitueix la memòria del Treball Final del Grau d'Estadística, realitzat a la Universitat de Barcelona (UB) i la Universitat Politècnica de Catalunya (UPC). En aquest apartat es presenta la justificació, les hipòtesis, els objectius, l'estructura i els agraïments.

Justificació

El present treball proposa un article acadèmic sobre el Big Data. En els darrers anys el concepte Big Data ha esdevingut molt conegut, ja que ha estat treballat i desenvolupat en diverses disciplines. A més, actualment ens trobem en l'era digital o l'era de les grans dades on el Big Data n'és un dels eixos centrals. El motiu pel qual vaig escollir aquest Treball de Final de Grau sorgeix com a resposta d'una inquietud personal de buscar i conèixer més informació sobre el món de les grans dades, no únicament des d'un punt de vista analític i computacional, sinó també des del món acadèmic.

Com a estudiant del Grau d'Estadística, he pogut conèixer de manera pràctica i teòrica diverses tècniques d'anàlisi de grans dades com el *Data Mining*, *Machine Learning* o l'anàlisi multivariant, però mai ens hem centrat a dir què és el Big Data, com es caracteritza i en fer una anàlisi crítica d'ell.

En un futur m'agradaria treballar en el món acadèmic, poder continuar elaborant articles i seguir profunditzant i desenvolupant els meus coneixements en estadística, especialment en Big Data aplicat a les ciències socials.

Aquest projecte es desenvolupa amb la finalitat de conèixer i proporcionar, des d'un punt de vista crític, quines són les potencialitats i les limitacions de les grans dades. Crec que és un treball molt interessant per a qualsevol persona que vulgui dedicar-se al món del Big Data, ja que a partir de les lectures realitzades he fet una revisió de la seva història. I, tot això, m'ha permès conèixer, especialment, les seves debilitats, tot i que normalment el Big Data es destaca pels beneficis que ofereix.

Problemàtica de recerca

Atesa la naturalesa d'aquest Treball Final de Grau, és més pertinent parlar de problemàtica de recerca que d'hipòtesis. Plantejo quatre àmbits. En primer lloc, vull analitzar la definició del

concepte Big Data i el grau de consolidació d'aquest. En segon lloc, vull analitzar si és pertinent enfocar el seu estudi en termes de les potencialitats i les limitacions que caracteritzen el Big Data. En tercer lloc, sorgeix la qüestió de com és la privacitat i de si estem segurs en un món on cada vegada hi ha més dades. Finalment, em demano si les grans dades divideixen, és a dir, si el Big Data és discriminatori.

Objectius

L'objectiu principal que es persegueix amb aquest treball és conèixer les potencialitats i les limitacions de les grans dades, i oferir-ne una reflexió crítica i analítica a partir de l'elaboració d'un article acadèmic.

A més, donades les peculiaritats del meu Treball Final de Grau, s'han plantejat els següents objectius analítics específics:

- Proporcionar una visió analítica i crítica de les grans dades, en concret de les seves potencialitats i limitacions.
- Introduir conceptes sobre les grans dades, partint de si existeix una única definició per a Big Data.
- Conèixer les característiques o dimensions de les grans dades.
- Entendre quin és el paper d'alguns conceptes clau de les grans dades, en concret, els algorismes, les correlacions o la dadificació. Aquests termes ja fa molts anys que existeixen però, per què ara tenen tant de ressò?
- Analitzar la privacitat en l'era de les grans dades.
- Estudiar i analitzar si el Big Data divideix, és a dir, si és discriminatori.

Els objectius analítics s'acompanyen de dos objectius instrumentals:

- Aprendre a utilitzar i gestionar gestors bibliogràfics, en el meu cas Mendeley.
- Conèixer l'estil de citació APA, 6th *Edition*.

Estructura

L'element central d'aquest Treball Final de Grau és un article de caràcter teòric basat en una revisió bibliogràfica. Després d'aquesta secció introductòria es troba la metodologia, seguida de l'article pròpiament dit. Les conclusions tanquen el TFG i a l'Annex es pot trobar la bibliografia anotada, que ha servit de matèria primera per construir el cos del treball.

Agraïments

Vull mostrar el meu agraïment a la Doctora Mireia Fernández-Ardèvol, tutora del Treball Final de Grau, per la seva exigència, el seu compromís i la confiança d'acompanyar-me en la trajectòria del projecte.

A la meva família, per la seva paciència i suport aquests últims mesos.

I, finalment, a l'Adrià i l'Andrea pel seu assessorament i la seva visió crítica externa.

Metodologia

Per fer aquesta investigació he revisat la literatura relacionada amb les grans dades a partir de cerques bibliogràfiques. En primer lloc, vaig realitzar unes lectures introductòries, a partir de recomanacions de la tutora. Aquestes varen ser clau per a l'enfocament de la meua investigació. Les lectures van ser, per una banda, l'article de la Danah Boyd i la Kate Crawford (2012) *Critical questions for Big Data*, i, per l'altra, el llibre de Rob Kitchin (2014) *The Data Revolution*. A partir d'aquestes vaig poder entendre alguns conceptes bàsics de les grans dades i em van permetre conèixer algunes de les diferents aportacions que defineixen la problemàtica de la meua recerca. Així doncs, vaig poder començar una cerca autònoma d'una bibliografia més específica en diferents bases de dades, principalment a *Scopus* i a *Google Scholar*.

Per començar, vaig establir un filtre temporal de cerca comprès entre l'any 2013 i el 2018. A més, vaig definir unes paraules clau a l'hora de buscar en les diferents bases de dades: *Big Data*, *epistemological challenge*, *dataveillance*, *privacy*, *spurious correlation*, *discrimination*, *data deluge*, *data storage*, *hidden biases* i *datafication*. Les quals em van portar a conèixer altres termes com *vagueness*, *buzzword* o *privacy by design*. A mesura que anava trobant bibliografia especialitzada en l'àmbit de la meua recerca, en llegia la introducció i la conclusió per fer-me una idea del seu contingut. A la vegada, prenia la decisió de si aquests estudis eren rellevants o no per la construcció del meu article, i per tant, havia de posar especial atenció i llegir-los per complet.

Seguidament, a partir de la informació que anava trobant, buscava les referències i cites d'altres autors fins al punt de saturar la informació (Pacheco-Vega, 2016), és a dir, fins a trobar-me en el punt en el qual observava que moltes idees i cites es repetien regularment. I finalment, un cop vaig considerar que ja havia llegit prou sobre la temàtica de les grans dades, vaig elaborar un esquema amb un possible índex, on incloïa les cites dels autors que considerava més rellevants per estructurar el meu article.

A mesura que llegia els diferents estudis, elaborava la bibliografia anotada d'aquelles lectures que considerava més importants i interessants per al meu treball. La bibliografia anotada, que reproduïxo a l'Annex, ha estat el pilar principal a partir del qual he elaborat l'article, ja que en ella he inclòs resums i una anàlisi crítica de les lectures que he anat fent al llarg del treball.

A més, a part de fer una anàlisi d'articles científics i literatura especialitzada en les grans dades, m'he mantingut al corrent de l'actualitat d'aquestes, a partir de les notícies i notes de premsa. En aquest cas, he pogut trobar molta informació i exemples de temes que volia plantejar en

el meu article –fet que m’ha permès proporcionar una visió més propera de la situació actual de les grans dades.

Finalment, el gestor de referències i de cites bibliogràfiques escollit ha estat Mendeley, atesa la seva accessibilitat i la seva compatibilitat amb Word. Alhora, l’estil de citació triat ha estat APA (*American Psychological Association*) (2010), sisena edició.

Reflexions sobre les potencialitats i les limitacions del Big Data

Aquest article és un estudi recopilatori que engloba una investigació sobre treballs ja publicats al voltant de les grans dades. Avui dia, disposem d'una gran quantitat de recursos i eines per trobar informació, per això en el meu projecte vull proporcionar una reflexió crítica i analítica sobre la informació existent en vers al Big Data. Així doncs, en aquest estudi pretenc englobar i concentrar gran part de la bibliografia que el tracta –si més no, la més destacada– per tal de donar una visió concreta i completa sobre les potencialitats i les limitacions de les grans dades.

El segle XXI és també conegut com l'era de la informació. Gràcies als avanços en la tecnologia i la digitalització disposem cada vegada de més informació. En altres paraules, les persones cada dia estem creant i emmagatzemant grans quantitats de informació a partir d'activitats en la nostra vida quotidiana. Tot això desemboca en l'àmbit del Big Data –o les grans dades–.

El Big Data està canviant la manera de com entenem el món, la manera com vivim, ens gestionem, treballem i ens relacionem. Avui dia, les grans dades són una eina fonamental de moltes disciplines com la medicina, les ciències socials o l'economia. El Big Data és de gran aplicabilitat, ja que ha ajudat investigadors i acadèmics a descobrir fets que sense la tecnologia i les tècniques de Big Data podrien haver tardat anys a trobar. Concretament, les grans dades estan “de moda” perquè ofereixen respostes immediates i, per tant, ens permeten prendre decisions en temps real.

Encara que moltes empreses i organitzacions parlen de les grans dades, moltes no saben realment quina és la informació que han d'analitzar o com n'han de fer ús, és a dir, només arriben a conèixer-ne el concepte i els beneficis que comporten. Aquest motiu porta a molts acadèmics, científics i investigadors a analitzar què representen les grans dades i les seves característiques, des de les seves virtuts fins als seus inconvenients. Aquest és un dels motius principals de la meua investigació. Considero que el terme Big Data és àmpliament conegut però és un concepte molt vague, trobem moltes definicions però cap de concreta.

Molts cops, la gent no especialitzada en les grans dades, únicament coneix les característiques positives que les descriuen però no les seves limitacions, si és que coneixen el concepte del Big Data. Considero que com en qualsevol problema analític, és molt important conèixer la metodologia, saber quan es pot utilitzar i quines tècniques són les més adients en els seus estudis.

La metodologia emprada ha estat la d'una recerca bibliogràfica que m'ha permès elaborar una bibliografia anotada, a partir de la qual he elaborat el meu discurs.

Amb aquest article, doncs, pretenc descriure les grans dades i oferir quines són les principals potencialitats i limitacions a partir del que ja han dit els especialistes en Big Data. Per tant, la meva intenció és donar una visió global i concreta de la informació existent al voltant de les grans dades, per així ajudar i contribuir a la gent que vulgui treballar i continuar investigant en el món de les grans dades.

El món de les grans dades (Big Data)

En l'era digital, moment en el qual actualment ens trobem, un dels conceptes que més ressò ha tingut a la darreria d'aquesta època és el del "Big Data"¹. Probablement, no tothom sabrà què és realment el Big Data, i es demanarà per què és tan rellevant en el nostre dia a dia.

Tal com han assenyalat acadèmics com Floridi (2012), Jin et al. (2015) i Chandler (2015) no existeix una definició universal del Big Data. Diferents investigadors de les grans dades afirmen que el terme Big Data és pobre (Boyd & Crawford, 2012) i sostenen que és un terme que encara no està definit de manera rigorosa (Floridi, 2012; Mayer-Schönberger & Cukier, 2013).

El Big Data és un concepte que s'inclou i es desenvolupa en moltes disciplines, és per això que les grans dades han esdevingut un *buzzword*² o paraula de moda (Floridi, 2012). Acadèmics com Jin et al. (2015, p. 59) exposen que el "*Big data has rapidly developed into a hot topic that attracts extensive attention from academia, industry, and governments around the world*". És a dir, tot i no existir una definició universal per a les grans dades, és un tema candent en l'actualitat dels panorames tecnològic, social i econòmic.

Segons Diebold (2012), la història del terme Big Data, es remunta a mitjans de la dècada dels noranta. A més, argumenta que "les primeres referències acadèmiques significatives són probablement de Weiss i Indurkha (1998) en informàtica i de Diebold (2000) en estadística/econometria"(Diebold, 2012, p. 5).

¹ En aquest article es podrà trobar escrit Big Data o grans dades indistintament.

² El diccionari Oxford University Press, 2018 defineix aquest concepte de la següent manera: "[a] word or phrase, often an item of jargon, that is fashionable at a particular time or in a particular context".

<https://en.oxforddictionaries.com/definition/buzzword> (15/04/18).

Hi ha autors que relacionen el Big Data amb el volum de les dades. Acadèmics com Mauro, Greco i Grimaldi contextualitzen el Big Data de la següent manera:

the data sets and analytical techniques in applications that are so large and complex that they require advanced and unique data storage, management, analysis, and visualization technologies (2015, p. 102).

Pompeu Casanovas (2017, p. 103) considera que una definició corrent del Big Data podria ser la següent:

conjunts de dades extremadament grans que poden ser analitzats computacionalment per revelar patrons, tendències i associacions, en especial relacionats amb el comportament i interaccions humans.

Mayer-Schönberger & Cukier defineixen el Big Data destacant allò que podem fer amb les dades:

big data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments, and more (2013, p. 20).

Altres autors relacionen el Big Data amb les grans onades d'informació (Enriquez, 2012; Fan, Han, & Liu, 2013). De fet, la quantitat d'informació disponible és un dels trets més rellevants de l'era digital i, en conseqüència, de les grans dades. Un exemple de l'evolució d'aquest aspecte és que l'any 2003 només s'havien creat 5 exabytes d'informació, quantitat que ara és produïda tan sols en dos dies (Garden, 2012). Això, però, no és una novetat d'aquests darrers anys. Fa gairebé una dècada, IBM (2011) ja donava informació sobre l'afirmació anterior:

Cada dia creem 2,5 quintillons de bytes de dades, de manera que el 90% de les dades del món actual s'han creat en els últims dos anys. Aquestes dades procedeixen de tots els llocs: sensors utilitzats per recollir informació del clima, entrades [posts] en llocs de mitjans socials, imatges digitals, fotografies i vídeos, registres de transaccions comercials i senyals de GPS de telèfons mòbils, per citar unes poques referències. Aquestes dades són Big Data.

Es tracta d'una cita àmpliament reproduïda per diferents acadèmics, com per exemple, Calude & Logo (2017, p. 596) o Joyanes Aguilar (2014, p. 7). Ho podem entendre com l'explosió de grans volums de dades, on aquestes no paren de créixer i sembla que segueixen un model exponencial.

Boyd & Crawford, a partir de l'obra de Burkholder (1992), on es reflexiona sobre el gir computacional que es va produir en el pensament i en la investigació un cop van arribar els ordinadors a les llars, relacionen aquest gir amb el fenomen de les grans dades:

Big Data not only refers to very large data sets and the tools and procedures used to manipulate and analyze them, but also to a computational turn in thought and research (2012, p. 665).

De manera que aquestes han donat pas a un nou paradigma. Les grans dades comporten un canvi en totes les disciplines, des de la manera de fer ciència a la manera de pensar (Kitchin, 2014a). De forma que s'està redefinint la manera com vivim, treballem, pensem o guanyem diners. Així doncs, "les grans dades esdevenen un canvi cultural, tecnològic i acadèmic" (Boyd & Crawford, 2012, p. 662).

Les dimensions del Big Data a partir de les Vs

Tot i que no hi ha consens en la seva definició, diversos autors vinculen el terme Big Data al model de les tres V (3V o V³): Volum, Velocitat i Varietat (Chandler, 2015; Lycett, 2013; Turner, Schroeck, & Shockley, 2013).

La primera persona que va parlar sobre les 3V va ser Laney (2001). Va introduir-les com els principals reptes de la gestió de dades per a les empreses (D'Acquisto et al., 2015). Avui dia, aquest criteri s'utilitza per definir i descriure les grans dades.

A continuació, desenvolupo la definició dels tres conceptes a partir de D'Acquisto et al. (2015), Gandomi & Haider (2015) i Lycett (2013):

- **Volum:** és la característica o dimensió més popular i reconeguda del Big Data. De fet, molts investigadors l'utilitzen com a definició i tret principal de les grans dades. El volum es correspon amb grans quantitats de dades, en escala de terabytes, petabytes o zettabyte. S'ha de tenir present que pot ser que el que avui dia és considerat "gran" sigui relatiu, atès que aquests paràmetres varien en funció de factors com el temps i el tipus de dades. És a dir, pot ser que el que ara entenem com a gran, en un futur no ho sigui.
- **Varietat:** aquesta característica fa referència a l'heterogeneïtat estructural en el conjunt de les dades, és a dir, que les dades provenen de fonts molt diverses. Els tipus de dades es classifiquen en tres grans grups: estructurades (bases de dades

tradicionals), semiestructurades i no estructurades. En general, moltes de les dades que disposem en l'actualitat corresponen amb dades no estructurades.

- **Velocitat:** aquest tret posa especial atenció en la rapidesa amb la qual generem les dades i la velocitat amb la qual les grans dades són analitzades i utilitzades. Aquesta és una característica molt important i que dóna molt valor al Big Data. La velocitat obre la porta a tenir informació en temps real.

Autors com Schroeck et al. (2012), Gandomi & Haider (2015) o D'Acquisto et al. (2015) consideren una quarta i una cinquena dimensió, la veracitat i el valor.

- **Veracitat:** es pot considerar la quarta dimensió de les Vs. Aquesta característica representa el grau de confiança que hauríem de dipositar en les dades subjectes a ser analitzades. La veracitat determinarà la qualitat dels resultats. Aquesta dimensió ve donada, principalment, per la manca de fiabilitat que existeix, molts cops, en les grans dades. Aquests poden ser inexactes o incorrectes. Tot i això, en el Big Data, el que importa no són les dades individuals, sinó l'anàlisi del conjunt de dades.
- **Valor:** es pot considerar la cinquena dimensió de les Vs i una de les característiques més importants. Consisteix a convertir la informació en coneixement, i aquest en accions o en decisions. Dit d'una altra manera, utilitzar les dades de manera rentable i a la vegada eficient.

Com a curiositat, a les 5 Vs que defineixen el Big Data, Borne (2014) n'introdueix cinc de noves:

- **Variabilitat:** molt similar a la varietat. Aquest tret fa referència a diferents aspectes. Per una banda, a les desviacions que presenten les dades. I, per l'altra, la multitud de dades que provenen de diverses fonts i resulten de diversos tipus.
- **Validesa:** similar a la veracitat. Aquesta característica fa referència a com de precises i correctes són les dades.
- **Lloc ("Venue"):** Aquesta dimensió fa referència a la procedència de les dades, ja que aquestes estan distribuïdes en múltiples plataformes, de diferents sistemes, amb diferents requisits d'accés i format. El lloc on es troben distribuïdes les dades pot ser públic o privat.

- **Vocabulari:** Aquest tret identifica l'esquema, els models de dades, la semàntica, les ontologies, les taxonomies i altres metadades basades en contingut i context que descriuen l'estructura, la sintaxi, el contingut i la procedència de les dades.
- **Vaguetat:** Aquesta característica és conseqüència del fet que les dades ja no són primàries, són secundàries. Les dades són generades, normalment, per sistemes digitals, la qual cosa, a vegades pot generar confusió perquè podem arribar a no saber què volen ens volen dir.

Finalment, la idea de les Vs sembla que encara evoluciona, no està consolidada. De manera que encara hi ha propostes que amplien el model fins a 17Vs. L'any 2017, tres investigadors Arockia Panimalar, Varnekha Shree i Vensehia Kathrine van proposar les 17V i 1C: Afegint a les ja enumerades: Visualització, Virialitat, Viscositat, Volatilitat, Verbositat, Voluntariat, Versatilitat i Complexitat (2017).³

Dadificació

Actualment, produïm una gran quantitat d'informació gràcies a la petjada digital que genera la nostra vida quotidiana. Les dades ja no es generen únicament a partir de preguntes o propòsits específics, sinó que també són subproductes de les nostres activitats, i són emmagatzemades digitalment (Chandler, 2015). Aquest procés s'anomena dadificació.

La dadificació juga un paper molt important en l'era de les grans dades. És un fenomen que posa en format quantitatiu fets que ningú abans s'imaginava que poguessin arribar a tenir un valor mesurable (Mayer-Schönberger & Cukier, 2013). La dadificació té com a finalitat recopilar grans quantitats d'informació de la vida quotidiana per transformar-les en dades informatitzades i llegibles per màquines (Strauß, 2015). Per exemple, a partir d'aplicacions mòbils es poden controlar les passes, la son, la freqüència cardíaca o el cicle menstrual, d'una manera molt més simple que abans, gairebé sense assabentar-nos. De la mateixa manera, el fet de buscar roba, llibres o qualsevol altre producte en plataformes digitals està sent dadificat i, en conseqüència, es converteix en informació de valor comercial. Una bona mostra són plataformes com Amazon o Ebay que rastregen les nostres cerques per oferir-nos productes relacionats o similars, o suggerir-nos-en en funció de les nostres preferències.

Lycett (2013), basant-se en Normann (2001), utilitza els conceptes de desmaterialització, liquiditat i densitat per definir la dadificació. Lycett (2013) afirma que la desmaterialització és

³ Veure definicions a l'Annex (p. 73).

la capacitat de separar un bé o recurs i l'ús que se'n fa en el món físic. La liquiditat fa referència al fet que, un cop el recurs s'ha desmaterialitzat, la informació pot manipular-se i transferir-se fàcilment, agregant i desagregant dades d'una manera que abans era molt costosa. Finalment, la densitat es correspon amb la creació de valor. A més, Lycett (2013) reflexiona que en un món dadificat hauríem de veure les dades, primer des d'un punt de vista matemàtic, i després, les hauríem de posar en context.

El fet de transformar la informació en dades mesurables ens permet analitzar-la de manera més objectiva (Es & Schäfer, 2017), o això és el que s'espera (van Dijck, 2014). Igualment, van Dijck (2014) veu la dadificació com una oportunitat revolucionària per estudiar i comprendre millor el comportament humà, i així, captar la realitat complexa i dinàmica que ens envolta. Amb l'ajuda de les grans dades, ja no es veurà el món com una cadena de successos que expliquem com a fenòmens naturals o socials, sinó com un univers compost essencialment d'informació (Mayer-Schönberger & Cukier, 2013). Per tant, com exposa Casanovas (2017), vivim en un món informacional on som "prosumidors", és a dir, productors i consumidors d'informació.

La dadificació ens porta a pensar en la possibilitat d'un futur no gaire llunyà, on el món podrà arribar a ser controlat de manera gairebé automàtica (Casanovas, 2017; Crawford, Miltner, & Gray, 2014). En concret, gràcies al paper dels algorismes.

Algorismes

El terme algorisme, segons Miyazaki (2012) es remunta a l'Espanya del segle XII quan els manuscrits del matemàtic àrab Muḥammad ibn Mūsā al-Khwārizmī⁴ varen ser traduïts al llatí. Aquelles escriptures descrivien els mètodes de sumar, restar, multiplicar i dividir. A partir de llavors, algorisme significa: "*the specific step-by-step method of performing written elementary arithmetic*" (Kitchin, 2016, p. 16).

El llenguatge és una arma de poder que defineix el món i la manera com el veiem. Per a Finn (2017a), els algorismes són el nou llenguatge del segle XXI. Els defineix com una recepta, un conjunt d'instruccions i una seqüència de tasques per arribar a obtenir un determinat càlcul o resultat. En tant que instruccions, incorporen tècniques d'intel·ligència artificial amb l'objectiu d'optimitzar els processos de decisió.

⁴ Muḥammad ibn Mūsā al-Khwārizmī fou un matemàtic, geògraf i astròleg/astrònom creador dels termes àlgebra i algorisme. Font: https://ca.wikipedia.org/wiki/Muḥammad_ibn_Mussa_al-Khwarazm%C3%AD (10/05/18).

Avui dia, es creu que els humans són massa analítics, emocionals i esbiaixats. Aquestes actituds ens porten a ser lents en intel·ligència i càlcul (Uddin & Lee, 2016). Per aquest motiu apareixen els processadors, que poden realitzar milions d'operacions per segon, minimitzant l'error humà (Kitchin, 2016), en gran part, gràcies als algorismes. Malgrat tot, hi ha qui creu que sovint els algorismes es creen per determinats propòsits, lluny de ser neutrals i imparcials (Pasquale, 2015). Aquests identifiquen, ordenen, agrupen, classifiquen, simulen, visualitzen i regulen persones, processos i llocs (Kitchin, 2016). No hem d'oblidar que els algorismes estan dissenyats per éssers humans i deixen integrats els seus valors en els softwares que escriuen, mostrant ideologies o raons comercials (Pasquale, 2015; Uricchio, 2017; van Dijck, 2014). Per tant, entendre els algorismes únicament com una mera forma tècnica o objectiva de coneixement, ens porta a ignorar els seus aspectes tècnico-socials. Si aquest fos el cas, podríem caure en problemes de subjectivitat o bé de discriminació (Finn, 2017a; Kitchin, 2016).

A banda d'això, Pasquale (2015) relaciona els algorismes amb caixes negres perquè considera que el funcionament d'aquests és misteriós. Per evitar les complicacions que presenten aquests, s'haurien de complir els principis de transparència i responsabilitat. Tothom hauria de ser conscient de les conseqüències i el biaix que poden provocar. És important entendre l'accés i la correlació, la responsabilitat o conèixer la procedència de les dades amb les seves característiques. A més, s'haurien de sotmetre a una auditoria per garantir una validació i una comprovació on tot es documenti.

Així doncs, els reptes principals dels algorismes se centren en problemes d'accés, d'experiència, de responsabilitat i de transparència. Els problemes d'accés vénen donats perquè moltes empreses i organitzacions els consideren secret empresarial, és a dir, no els publiquen pel seu valor potencial. Encara que si tinguéssim accés a ells, tindríem el problema d'experiència perquè els algorismes es van retroalimentant d'ells mateixos, generant les seves pròpies millores, i poder arribar-los a controlar o entendre és un procediment complicat. Especialment, els algorismes han de ser el més transparent i responsables possible perquè, en cas contrari, podem arribar a obtenir conclusions i decisions basades en dades d'una qualitat desconeguda i en algorismes modelats de manera intel·ligible (Uricchio, 2017).

Els algorismes tenen un rol tan valuós perquè són capaços de proporcionar-nos nous coneixements i possibles solucions en el nostre dia a dia. Aquests estan arribant a representar una nova manera d'entendre la cultura. Segons Finn (2017a), es poden interpretar com una barreja entre el poder computacional i el coneixement humà.

Actualment, els algorismes ja dominen el mercat de valors, condueixen cotxes, escriuen o tradueixen (Finn, 2017a). Kitchin (2016), basant-se en les idees de Steiner (2012), creu que aviat els algorismes podran decidir amb qui parlem per telèfon, controlaran la música que arriba a la nostra ràdio o qui serà la nostra parella. Dyson afirma el següent: “Facebook defineix qui som, Amazon defineix què volem i Google defineix el que pensem” (2012, p. 308).

Finn (2017a) expressa que en l’actualitat tots som lectors i programadors d’algorismes en major o menor mesura. És per això que és important entendre i ser conscients de les seves potencialitats i limitacions.

El final de la teoria

No es pot parlar del Big Data sense anomenar el final de la teoria, exposat inicialment per Anderson (2008). Ell és el primer acadèmic a creure i afirmar que amb suficients dades els números poden arribar a parlar per ells mateixos. Altres autors com Kitchin (2014a) comparteixen algunes de les seves idees. Ambdós creuen que la ciència està basada en dades més que en el coneixement.

Anderson (2008) pretén trencar amb l’estadística tradicional. Creu que amb la revolució de les dades el mètode científic que coneixem avui dia quedarà obsolet. Així mateix, afirma que en el món de les grans dades hauríem de deixar de banda la causalitat per donar pas a la correlació. Aquesta idea és estudiada per diferents acadèmics, com per exemple Calude & Logo (2017), Mayer-Schönberger & Cukier (2013) o Crawford (2013). En les seves publicacions expliquen què està passant amb la correlació i la causalitat, però a diferència d’Anderson (2008), no creuen que les dades puguin arribar a parlar per elles mateixes, sinó que pensen que som nosaltres qui les fem parlar. Crawford (2013) exposa que les dades no poden ser objectives, ja que són creacions del disseny humà: som nosaltres qui donem veu als números, qui traiem conclusions i definim el seu significat a través de les nostres interpretacions. A més, es creu que pensar que el mètode científic pot arribar a quedar obsolet és un error perquè les grans dades es reforcen en la teoria. Gross⁵ presenta la seva opinió de la següent manera:

In practice, the theory and the data reinforce each other. It’s not a question of data correlations versus theory. The use of data for correlations allows one to test theories and refine them (Citat a: Bollier, 2010, p. 7).

⁵ Patrick Gross és el president de *The Lovell Group*, una empresa d’assessorament i inversió en negocis i tecnologia. Font: <http://csreports.aspeninstitute.org/Roundtable-on-Information-Technology/2013/participants/details/51/patrick-gross> (10/05/18).

Segons Kitchin (2014a) i González-Bailon (2013), dins del Big Data, com ja s'ha pogut demostrar, existeixen dues postures: una empírica on es creu que les dades poden parlar per elles mateixes sense la necessitat de la teoria i una altra, de teòrica, que es basa en l'abducció⁶, la inducció⁷ i la deducció⁸.

Calude & Logo (2017) reflexionen que si bé el Big Data es basa en correlacions, no podem abandonar la teoria. A més, opinen que en segons quines disciplines no està ben vist fer un ús acrític de les correlacions. Per tant, les grans dades s'haurien de basar en una teoria sòlida (Fong, 2017). González-Bailon (2013) opina que la teoria ha d'existir per formular hipòtesis creïbles sobre allò que les dades exposen. Llavors, en la mateixa línia de Gross (2010), aquests acadèmics consideren que el que cal fer és millorar el mètode científic.

Un objectiu de les grans dades és treure valuoses correlacions sense conèixer-ne les causes subjacents. A vegades, establir causalitats provoca biaix, ja que podem arribar a treure conclusions que realment no existeixen, a partir dels prejudicis dels nostres coneixements i creences (Mayer-Schönberger & Cukier, 2013). Les correlacions són molt útils en els estudis del Big Data, gràcies al seu poder aproximatiu, ja que ens permeten capturar el present i predir el futur (Mayer-Schönberger & Cukier, 2013). Tot i això, resulta important destacar que les grans dades impliquen alts riscos de fracàs i de subjectivitat quan la correlació es barreja amb la causalitat (Strauß, 2015). Per tant, ens hem d'allunyar al màxim de la causalitat per deixar pas a la correlació. És a dir, hem de buscar ser més inductius i realistes, deixant de banda les suposicions causals, abstractes i a vegades reduccionistes (Chandler, 2015). Amb altres paraules, amb el Big Data no ens interessa saber per què passen les coses (causalitat), ens interessa saber què està passant o què ha passat (correlació). A més, la causalitat no és fàcilment demostrable, en canvi les correlacions tenen una matemàtica al darrere (Mayer-Schönberger & Cukier, 2013).

⁶ Abducció: Sil·logisme en què la premissa major és evident, però la menor, i per tant la conclusió, és solament probable. Font: <https://dlc.iec.cat/results.asp?txtentrada=abducci%F3&Submit2=Cerca+directa+al+diccionari> (19/05/18).

⁷ Inducció: Raonament pel qual hom remunta de la part al tot, del particular al general, de l'efecte a la causa, dels fets a la llei que suposen. Font: <https://dlc.iec.cat/results.asp?txtentrada=inducci%F3&Submit2=Cerca+directa+al+diccionari> (19/05/18).

⁸ Deducció: En lògica, raonament en què una conclusió se segueix necessàriament d'una o més premisses. Font: <https://dlc.iec.cat/results.asp?txtentrada=deducci%F3&Submit2=Cerca+directa+al+diccionari> (19/05/18).

Malgrat el gran poder de les correlacions, en el Big Data hi poden haver correlacions espúries i ens poden induir a caure en fal·làcies (Calude & Longo, 2017). En les grans dades, la correlació espúria fa referència a les variables no correlacionades que es troben falsament associades, a causa del gran volum de dades que se'n disposen (Gandomi & Haider, 2015).

Finalment, m'agradaria fer menció a Hal Varian, economista principal de Google, que ens convida a reflexionar sobre el paper de la teoria:

Theory is what allows you to extrapolate outside the observed domain. When you have a theory, you don't want to test it by just looking at the data that went into it. You want to make some new prediction that's implied by the theory. If your prediction is validated, that gives you some confidence in the theory. There's this old line, 'Why does deduction work? Well, because you can prove it works. Why does induction work? Well, it's always worked in the past' (Bollier, 2010, p. 5).

Per tancar aquest apartat crec que és important considerar que el Big Data no demostra el final de la teoria perquè les inferències de grans dades haurien d'estar fonamentades en conceptes estadístics o matemàtics.

Potencialitats i limitacions de les grans dades

Si les grans dades han esdevingut tan importants en la nostra societat, no és només pel gran volum d'informació que disposem en temps real. És per la quantitat de beneficis i oportunitats que proporcionen. Tanmateix, no tot és positiu, també tenen els seus reptes i vulnerabilitats. En aquest apartat es parlarà de les potencialitats i limitacions del Big Data i s'explicarà que no tot és blanc o negre.

Les grans dades es veuen actualment com una nova oportunitat d'anàlisi que pot oferir avenços significatius en diferents aspectes de la nostra vida diària, incloent-hi la salut, l'oci, el medi ambient, el treball o l'economia (D'Acquisto et al., 2015). Així mateix, el Big Data ofereix la possibilitat d'una intel·ligència artificial, detallada i en temps real de molts aspectes de la nostra vida (Kitchin & Lauriault, 2014).

Les grans dades creen especial valor en l'economia mundial. Segons Tene & Polonetsky, "les dades han esdevingut la força motriu que hi ha darrere de gairebé totes les interaccions entre individus, empreses i govern" (2013, p. 30). Altres com Jin et al. (2015) creuen que les grans dades poden esdevenir "el petroli" del segle XXI. Així mateix Lohr (2012) compara les dades amb actius econòmics, com poden ser la moneda o l'or. D'aquesta manera, les grans dades

ofereixen solucions a reptes empresarials, transformen el coneixement, les empreses, les organitzacions i inclús la mateixa societat (Turner et al., 2013).

El Big Data crea oportunitats i són la base de nous avantatges competitius (Barocas, 2014; Eastin, Brinson, Doorey, & Wilcox, 2016; Letouzé, 2012). Les empreses i organitzacions veuen com les grans dades proporcionen la capacitat per comprendre i predir millor el comportament dels clients i així poden satisfer les seves necessitats de manera òptima (Schroeck et al., 2012).

Actualment, gràcies, entre d'altres, a les xarxes socials, les grans dades són capaces de respondre a moltes preguntes de recerca en ciències socials (Felt, 2016). Gràcies a aquests estudis podem conèixer amb més profunditat el comportament humà. Per tant, com exposen Helles & Bruhn-Jensen (2013), el Big Data és un recurs més per entendre com les persones es comuniquen i interactuen socialment. Alhora, les grans dades permeten descobrir patrons de població i heterogeneïtats que no era possible de descobrir a escales més petites (Fan et al., 2013).

Segons Power (2016), les grans dades ofereixen grans beneficis i oportunitats a les societats però aquest progrés ha d'anar acompanyat d'una reflexió: quin ús i quina anàlisi volem fer del Big Data per tal de poder oferir una diferència positiva en la vida de les persones, ja que és cert que ens pot ajudar a prendre millors decisions i extreure beneficis.

Per contra, les grans dades no només ofereixen beneficis. S'ha de tenir present que en el Big Data participen qüestions ètiques, metodològiques i polítiques. A vegades, amb les grans dades, existeix una àrea grisa entre allò que és legal i allò que és correcte (Schroeck et al., 2012).

Així mateix, és important conèixer les vulnerabilitats del Big Data, ja que existeix el perill que aquest, en un futur, es converteixi en el Big Brother⁹. Amb les grans dades existeix la por de caure en un estat totalitari, on es violi la privacitat de les persones o se'n faci un ús inadequat. Hu (2013), una investigadora biomètrica¹⁰, estudia els perills de l'ús de les grans dades, com la instal·lació i normalització de la ciberseguretat en la nostra vida quotidiana. Un bon exemple és el cas de la Xina on el seu govern té accés il·limitat a les dades personals dels seus ciutadans (Jones, 2016).

⁹ El concepte Big Brother neix en la famosa novel·la distòpica "1984" de George Orwell.

¹⁰ La biometria és la branca de la biologia que estudia els fenòmens quantitius en els éssers vius. Definició Biometria segons el diccionari de la llengua catalana, <https://www.iec.cat/activitats/entrada.asp> (07/01/18).

De fet, actualment, ja s'estan recopilant i analitzant grans quantitats de dades biomètriques. Són molts els que estudien les limitacions de les grans dades, no per refutar-les sinó per reforçar-les. Segons Kitchin (2014b), les limitacions que més preocupen a la societat són la vigilància massiva, la classificació social i la corrupció.

Com a conclusió a les idees d'aquests estudis, es podria dir que l'ús del Big Data comporta el risc de perdre part de la nostra llibertat. Enriquez (2012) compara les dades amb els tatuatges. Explica que deixem petites marques, algunes més visibles que d'altres, de qui som, amb qui estem i què ens agrada. Per ell aquest fet és com tatuar-se electrònicament. Amb aquesta analogia també es pot entendre que les dades ens acompanyaran tota la vida i que seran difícilment eliminades.

S'ha de tenir en compte que autors com Lesk (2013) pensen que la perdurabilitat de les dades personals té impactes socials que poden durar dècades, ja que Internet mai oblida. Alhora Power (2016) reflexiona que vivim en un món digitalitzat i interconnectat on difícilment les dades poden ser anònimes. Ell reflexiona que potser no és raonable esperar que el que la gent digui o faci sigui anònim o confidencial en la societat actual, valora que el problema principal és veure quin ús es fa de les nostres dades i en quines mans cauen.

Als crítics del Big Data els preocupa que les grans dades puguin ser mal utilitzades i que se n'abusi d'elles. Un exemple seria el d'organitzacions o governs que tinguessin habilitats per manipular als consumidors. Els experts en dades temen que els possibles abusos de les dades puguin posar en perill la intimitat personal, les llibertats civils i les llibertats dels consumidors (Bollier, 2010). Les investigadores Boyd & Crawford ens conviden a reflexionar sobre les grans dades i exposen que "només perquè siguin accessibles no significa que siguin ètiques" (2012, p. 671). Tanmateix, declaren que és molt difícil obtenir el consentiment de totes les persones que publiquen contingut a Internet. De manera que el fet que el contingut sigui accessible al públic no significa que sigui destinat a ser consumit per qualsevol, per tant les seves accions poden no ser ètiques.

En general, existeixen molts problemes relacionats amb l'ètica de la recopilació i l'anàlisi de les dades en línia. Per altra banda, es creu que les dades reflecteixen amb precisió el món social però hi ha bretxes significatives, amb poca o cap informació provinent de comunitats particulars, és un problema de cobertura (Crawford, 2013). Dit d'una altra manera, en les grans dades existeixen problemes de representativitat que poden venir donats per situacions econòmiques, geogràfiques o laborals (Chandler, 2015). De manera paral·lela, Graham (2012) reflexiona que no podem veure les grans dades com un mirall socioeconòmic de la societat

perquè si no estem creant bretxes digitals i fent generalitzacions que no tenen per què ser certes. Per a ell és necessari contextualitzar les dades i segmentar-les en petits subconjunts de dades per així no perdre pel camí idees o patrons de la societat. Tal com afirma Strauß (2015), les grans dades fora de context poden perdre el seu significat.

En les grans dades no solament trobem desafiaments ètics sinó que hi ha altres reptes. Entre ells trobem: la gran quantitat de dades i l'heterogeneïtat de les mateixes. Aquestes ens introdueixen a reptes computacionals i estadístics, on apareixen l'acumulació de soroll, la correlació espúria o errors de mesura (Fan et al., 2013; Lycett, 2013). Jin et al. (2015) escriuen que la complexitat inherent de les dades fa que la seva percepció, representació, comprensió i càlcul siguin molt complicats i provoquin una complexitat computacional. En conseqüència, un dels reptes més grans existent en les grans dades és l'adquisició de les habilitats necessàries per a fer-les operatives, des del punt de vista d'anàlisi i gestió (Turner et al., 2013). Habilitats i tècniques que poden ser analítiques, informàtiques o polítiques.

En l'actualitat, la disponibilitat de grans dades ha dividit l'opinió pública (Gonzalez-Bailon, 2013). Per una banda, tenim els escèptics que qüestionen l'ús legítim de les dades en referència a la privacitat o aspectes ètics. I per l'altra banda, els entusiastes que se centren en l'impacte transformador de tenir més informació que mai. En funció del seu posicionament, els acadèmics i investigadors de grans dades veuen el seu impacte d'una manera més o menys positiva.

En definitiva, les grans dades se centren en l'obtenció i l'optimització de la informació. Aquest és el seu poder i al mateix temps, el seu major problema (D'Acquisto et al., 2015). L'opinió que es pugui tenir de les grans dades depèn en gran part de l'ús que en fem d'elles, per tant, l'únic que es pot demanar és exigir un ús responsable de la informació (Gonzalez-Bailon, 2013; Lesk, 2013), on l'única preocupació sigui poder garantir una diferència positiva en la vida de les persones. El Big Data reformula la manera com vivim, treballem i pensem. És per això que en aquesta nova etapa hem de potenciar les capacitats humanes com la creativitat, la intuïció o l'ambició de coneixement, atès que l'enginy humà és la principal font de progrés (Mayer-Schönberger & Cukier, 2013). És molt important evitar empresonar-nos en probabilitats, encara que les prediccions siguin molt bones, ens hem de protegir de la dependència excessiva de les dades a partir dels nostres propis recursos.

Per poder avançar amb les grans dades encara és necessari fer-se preguntes crítiques sobre què signifiquen les dades, qui hi té accés o què volem fer amb elles (Boyd & Crawford, 2012). El gran repte per a Tene & Polonetsky (2013) seria aconseguir un equilibri entre els riscos de la privacitat i els grans beneficis que ens ofereixen les grans dades.

Un cop presentades les potencialitats i les limitacions del Big Data, faré una anàlisi més profunda de la privacitat i la discriminació existent al voltant de les grans dades. Tractaré alguns punts com les conseqüències de la privacitat en l'era digital i els problemes a l'hora de generalitzar.

Les llibertats civils estan en joc?

L'any 1949 George Orwell va publicar la novel·la *1984*. Power (2016) explica que aquesta novel·la de ficció distòpica presenta un món sense llibertat d'expressió, on tots són sempre vigilats pel seu líder, el Gran Germà o "Big Brother". Els ciutadans no tenen llibertats de cap tipus, ni tan sols la possibilitat de pensar en privat. Aquest llibre introdueix el concepte del Gran Germà, un líder sempre present, que tot ho pot veure i controlar a través de monitors. Aquesta obra va ser escrita fa gairebé setanta anys i són diversos els acadèmics que creuen que cada dia estem més a prop del Big Brother, com per exemple Power (2016), Lesk (2013) o Mai (2016). Power (2016), reflexiona que el Gran Germà és una advertència metafòrica sobre les conseqüències del govern si utilitza les tecnologies digitals per mantenir el poder i el control de la gent.

En l'actualitat, estem constantment generant dades a partir de les nostres activitats. En l'era del Big Data existeix el perill d'una gestió de la ciberseguretat que no respecti la nostra privacitat, ja que, cada vegada més, s'estan recopilant i analitzant grans quantitats de dades biomètriques (Hu, 2013). Aquesta tasca les realitzen empreses privades com poden ser Facebook, Twitter o Amazon (Lesk, 2013). D'aquesta manera, les empreses i els governs poden arribar a saber més de nosaltres mateixos del que realment ens agradaria (Garden, 2012).

L'any 2013, Edward Snowden, un ex-contratista tècnic i ex-treballador de la CIA, que treballà per una empresa contractada per l'Agència de Seguretat Nacional (NSA) dels Estats Units, va revelar i transmetre a la premsa informació classificada sobre el programa PRISM ("*Planning Tool for Resource Integration, Synchronization, and Management*") (Anthony, Stablein, & Carian, 2015). Aquest és un programa que compren una vigilància massiva de la Internet arreu del món (Cole, 2014). PRISM està creat amb l'ànim de recopilar tot el contingut dels correus electrònics, registrar trucades de telèfon o registrar ubicacions.

A partir d'aquell moment, es va generar un debat públic sobre les preocupacions en vers a la privacitat, la vigilància massiva dels ciutadans i l'abús governamental en l'ús de les grans dades. Buttarelli (2017), en el seu article, exposa que la privacitat ens hauria de preocupar a tothom perquè aquesta és un dret humà fonamental. Tothom necessita una part del seu espai

íntim. La privacitat és un requisit previ per gaudir d'altres llibertats molt lluitades com la llibertat d'expressió i la no discriminació per motius de sexe, raça, orientació sexual, creences polítiques o religioses (Buttarelli, 2017). Letouzé (2012) creu que la privacitat es pot entendre com el dret de les persones a controlar la divulgació de la informació que es relaciona amb elles. A més, expressa que en una societat sense privacitat, seguretat, diversitat o innovació, les nostres llibertats bàsiques estarien en perill.

Hu (2013) reflexiona que la gran majoria de les persones no són conscients de les conseqüències d'una cibervigilància. Aquesta opinió és compartida per Es & Schäfer (2017), que afirmen que els usuaris desconeixen fins a quin punt les seves dades arriben a ser recollides i analitzades. A més, Kitchin (2014b) critica que la societat no llegeix les polítiques de privacitat. Els motius que exposa són per mandra o bé perquè no les entenen, atès que moltes vegades aquestes polítiques de privacitat són incomprensibles. La gent les accepta diàriament com a condició necessària per participar en l'economia en línia (Turow, 2007).

Mai (2016) considera que en l'era de les grans dades, necessitem pensar diferent sobre la privacitat. Creu que s'ha de redefinir el concepte perquè troba que la manera com entenem la privacitat pot estar obsoleta. Kitchin (2014b) , també, afirma que la manera que tenim d'entendre la privacitat està canviant tant a escala social com legal. D'Acquisto et al. argumenten que "no poden haver grans dades sense privacitat" (2015, p. 17). Per a ells, si les grans dades no respecten els principis de privacitat, aquestes no podran satisfer les necessitats dels individus, és a dir, aquests no estaran adequadament protegits. Per tant, els reptes de la privacitat han de ser vistos com a oportunitats, que poden generar confiança pel benefici tant dels usuaris com per la indústria de les dades (D'Acquisto et al., 2015). Per d'Acquisto et al. (2015) els principals reptes de la privacitat són la falta de control i transparència, la reutilització de les dades, la re-identificació de les dades, és a dir, la pèrdua de l'anonimat, el perill de prendre decisions de manera automàtica o la dificultat per revisar els controls de protecció de dades.

Avui dia, la informació personal és vista com una mercaderia, un negoci que es compra i es ven (Mai, 2016). Fins al moment, la privacitat anava molt lligada al consentiment. En l'era de les grans dades, com expressa Mai (2016), el marc de la privacitat hauria d'estar menys centrat en el consentiment individual, especialment en el moment de la recopilació, i més centrat a assumir la responsabilitat d'ús. Per tant, amb les grans dades es produeix un canvi. Es passa de la privacitat per consentiment a la privacitat per responsabilitat (Mayer-Schönberger & Cukier, 2013). És un canvi molt necessari per governar les dades de manera eficaç. Mayer-Schönberger & Cukier (2013) consideren que també s'ha de redefinir el concepte de justícia, ja que amb el Big Data es pot arribar a jutjar a un individu no per allò que ha fet, sinó per allò

que s'ha predit que faria. Ells creuen que s'ha de continuar jutjant a les persones per la seva responsabilitat individual i el seu comportament real.

D'Acquisto et al. (2015) creu que el tipus de privacitat que pot ser òptima en l'era de les grans dades és la privacitat per disseny (PbD), concepte introduït per Ann Cavoukian en els anys 90 (Cavoukian, 2009). Cavoukian & Jonas (2012) defensen que la privacitat per disseny promou l'opinió de què el futur de la privacitat no pot garantir-se únicament mitjançant l'acompliment de la legislació i els marcs normatius; més aviat, la garantia de la privacitat hauria de convertir-se en la manera de funcionament predeterminat de les empreses i organitzacions. Segons Kitchin (2014b), basant-se en Cavoukian (2009), la privacitat per disseny hauria de ser un sistema que busca explícita i activament assegurar la privacitat dels usuaris. L'objectiu de la PbD és garantir la protecció de privacitat i obtenir el control personal sobre la pròpia informació i, per les organitzacions, obtenir una avantatge competitiu sostenible (Cavoukian & Jonas, 2012). Tot això es pot aconseguir a partir de la pràctica dels set principis fonamentals de la PbD, definits per Cavoukian (2009). Els set principis són manifestacions del mateix objectiu i s'enllacen i dialoguen entre ells:

1. Protecció preventiva i proactiva, no reactiva.
2. Privacitat "per defecte".
3. Privacitat incorporada en el disseny.
4. Privacitat de funcionalitat completa, és a dir, suma positiva. Ara ja no és privacitat vs. seguretat, ja que ambdues són possibles.
5. Protecció durant tot el cicle vital, "end to end".
6. Garantir visibilitat i transparència.
7. Respecte per la privacitat de l'usuari.

Segons Cavoukian (2009), en un món en constant canvi tecnològic com el nostre, el dret a la privacitat és més important que mai. És per això que ens hem de mantenir alerta en la protecció de la privacitat, ja que és la base de la nostra llibertat.

Si fem una comparativa entre les lleis que regulen les grans dades a Europa, Estats Units o la Xina, trobem grans diferències, especialment entre Europa i el país asiàtic. Segons Irazabal (2018) el reglament que seguim a Europa garanteix majors drets en els usuaris, com per exemple, dret a la limitació del tractament de dades, dret a la portabilitat, augment en la transparència, dret a la informació, major dret d'indemnització i d'interposició de denúncies i, en especial, el dret a l'oblit. En canvi, la Xina és un país amb pocs controls sobre el seu poder

(Jones, 2016). La Xina és el primer país on s'està implementat un estat totalitarista digital (Jones, 2016).

M'agradaria fer una especial menció al Reglament Europeu de Protecció de Dades (GPRD) que va entrar en vigor, a la Unió Europea, el passat 25 de maig de 2018. A partir d'aquest moment, hem de donar el nostre consentiment perquè les empreses puguin utilitzar les nostres dades. Així, doncs, les empreses i organitzacions estaran obligades a informar-te de quines dades estan utilitzant, com les estan tractant, per què i qui és la persona responsable de les mateixes (Jiménez Garcia, 2018).

El problema rau en el fet que avui dia, a mesura que la gent es va familiaritzant amb les noves tecnologies, les pors respecte a la privacitat i la seguretat desapareixen i això és realment perillós (Anthony et al., 2015). Però, com bé indica Wenning (2010), si la gent fos més conscient de la realitat que ens envolta, no divulgaria tant les seves idees, gustos, preferències o ubicacions. A més, remarca que no ens hem de preocupar per la privacitat únicament quan tenim un problema, ja que la privacitat és o hauria de ser un requisit per poder gaudir de la nostra llibertat (Buttarelli, 2017).

L'objectiu seria trobar un equilibri entre l'ús de les grans dades i la protecció de la privacitat de les dades personals (D'Acquisto et al., 2015). Hem de pensar que en el món del Big Data molts cops guanyem seguretat a canvi de la nostra privacitat, però Kitchin (2014b) critica que això no hauria de ser així, ja que la privacitat és un dret humà bàsic. Aquesta opinió és compartida per Casanovas (2017). Finalment, no hem d'oblidar que els drets individuals han d'estar per sobre de qualsevol benefici (Tene & Polonetsky, 2013).

Com a curiositat, m'agradaria destacar dos esdeveniments recents. Primer, parlaré de l'escàndol de la xarxa més gran del món, Facebook, amb Cambridge Analytica. I en segon lloc d'Alexa, el majordom d'Amazon.

En primer lloc, Cambridge Analytica era una consultora política britànica especialitzada en la recopilació i l'anàlisi de dades per a la creació de campanyes publicitàries i polítiques («Cambridge Analytica», s.d.). Fa relativament pocs mesos, el 17 de març de 2018, *The New York Times*, *The Guardian* i *The Observer* van denunciar que l'empresa estava explotant informació personal dels usuaris de Facebook. La consultora està acusada d'haver obtingut informació de milions d'usuaris, atemptant contra les polítiques d'ús de la xarxa social i d'haver utilitzat aquestes dades per crear anuncis polítics durant les eleccions presidencials de 2016 als Estats Units.

El que va passar és que Facebook va crear una aplicació en forma de qüestionari, “*thisismydigitallife*”, on es recopilava informació per predir la personalitat i la psicologia de les persones. Però l’aplicació no només agafava les dades de les persones que participaven en els qüestionaris, sinó que també dels seus contactes de Facebook. Aquesta informació després era reportada a Cambridge Analytica. Tot això es va utilitzar per anuncis polítics específics en la campanya del referèndum del Brexit, així com per l’equip de Trump en les eleccions nord-americanes de 2016.

Fa relativament poc, Snowden¹¹ va escriure una piulada molt crítica respecte a les xarxes socials. Ell exposa que les xarxes socials, com per exemple Facebook, són companyies de vigilància rebatejades sota el nom de xarxes socials. No hem d’oblidar el que s’acostuma a dir: si un producte és gratis, és perquè el producte som nosaltres.

El segon exemple també es troba en relació amb la violació de la privacitat. A continuació, us presento a Amazon Echo, conegut com a Alexa («Amazon Echo», s.d.). Alexa és el nom de l’assistent intel·ligent d’Amazon. És un altaveu amb funcions, respon a preguntes, reproduïx música, apaga i encén les llums i sap quan has de sortir de casa per agafar amb temps el transport públic. Es pot dir que connectada a Internet, Alexa ho pot arribar a saber gairebé tot.

El dispositiu, per defecte, escolta les converses que es van produïnt a l’interior d’una casa, buscant la paraula arrel, Alexa, per poder reaccionar. Per tant, escolta converses privades.

Tot i que l’invent ha estat creat i dissenyat per ajudar en les tasques quotidianes, des de la seva presentació han sorgit diversos fets que ens porten a preocupar-nos per la nostra privacitat. Per exemple, el maig de 2018, *The New York Times* publicava una notícia on una parella de Portland havia estat gravada mentre mantenia una conversa sobre terres de parquet i la seva conversa havia estat enviada per missatge a un company de feina del marit (Chokshi, 2018). Realment, el problema és alarmant, encara que en aquesta situació és “anecdòtic”, però posa de manifest el difícil equilibri existent entre la convivència i la privacitat que aporten les màquines intel·ligents. L’any 2016, ja va haver-hi un debat al respecte quan es va demanar a Amazon que col·laborés en una investigació d’un assassinat que disposava d’un assistent Alexa, tot i que Amazon es va negar a aportar tota la informació i les dades en benefici de la privacitat dels seus clients (Cordón, 2016).

¹¹ Edward Snowden [Snowden]. (2018, març, 17). Businesses that make money by collecting and selling detailed records of private lives were once plainly described as «surveillance companies. » Their rebranding as «social media» is the most successful deception since the Department of war became the Department of Defense. [tweet]. Recuperat 11 juny 2018, de <https://twitter.com/Snowden/status/975147858096742405>

Ens hauríem de demanar si, realment, les autoritats poden accedir a les nostres dades privades en cas de necessitat. Aquestes situacions haurien d'estar regulades perquè la jurisprudència és pràcticament inexistente i està poc preparada per lidiar amb un món on cada vegada són més escoltats per dispositius intel·ligents.

Les grans dades són discriminatòries?

Avui dia hem dotat a les màquines de gran part del poder humà. Ara, les màquines i processadors tenen sistemes automatitzats de decisió i de classificació de forma que poden afectar en la nostra vida (Eubanks, 2018).

Tot i que les grans dades ofereixen molts beneficis i oportunitats, s'ha de tenir en compte que presenten qüestions ètiques, metodològiques i polítiques que convé resoldre. Leurs & Shepherd (2017) reflexionen que en els estudis de Big Data s'acostumen a fer generalitzacions. Consideren que el "nosaltres", en les grans dades, acostuma a ser indeterminat i, per tant, el Big Data discrimina intrínsecament els subjectes ja marginats per la societat (Leurs & Shepherd, 2017).

De la mateixa manera, Crawford (2013) i Graham (2012) critiquen que la gent veu les dades com un mirall socio-econòmic-polític. Crawford (2013) exposa que això no pot ser veritat perquè hi ha carències significatives d'informació en algunes comunitats. Fa referència a comunitats pobres que, per exemple, no tenen tants dispositius electrònics com en zones més riques, o bé la gent gran que, en general, és menys donada a l'ús de dispositius electrònics, si més no, a una escala molt menor de la joventut. Tot això està generant bretxes digitals (Crawford, 2013; Graham, 2012). Barocas (2014) comparteix la mateixa opinió, creu que no tots els ciutadans es veuen representats en l'anàlisi de les grans dades. Ell ho atribueix a un biaix de selecció o de discriminació. Això vol dir que de manera intencionada o no, en els estudis de grans dades poden haver-hi col·lectius omesos per prejudicis de gènere, raça, edat, religió o situació econòmica.

Als Estats Units, Eubanks (2018) argumenta que els grups marginats, sigui per situacions econòmiques, religioses o racials, s'enfronten a nivells més alts de recopilació de dades quan volen accedir a beneficis públics o travessar fronteres. Aquests grups marginats són assenyalats per una política pública sancionadora i una vigilància més intensa.

Floridi (2012), en el seu article *Big data and their epistemological challenge*, reflexiona que el veritable problema epistemològic en el Big Data són els patrons petits. Personalment, entenc

els patrons petits com aquells grups de la societat que poden no estar ben representats en moltes investigacions de Big Data. Floridi (2012) exposa que l'objectiu de les grans dades és detectar on es troben els patrons petits per explotar-los millor, i així crear riquesa i avançar en el coneixement. Alhora, argumenta que els patrons petits només poden ser significatius si s'agreguen adequadament. A més, reforça la idea que quan la informació no està activa o és inexistent, també és significativa. Floridi (2012) per consolidar les seves idees fa menció a Sherlock Holmes¹², que resol un misteri a partir del silenci d'un gos que hauria d'haver bordat. Floridi, en relació a l'obra de Conan Doyle i les grans dades comenta el següent: *"If big data are not "barking" when they should, something is going on, as the financial watchdogs (should) know"* (2012, p. 437).

A més, Helles & Bruhn-Jensen (2013) reflexionen sobre com s'elaboren les dades. Ells creuen que és un procés poc documentat on apareixen múltiples agents socials. Des del seu punt de vista, aquest procediment s'hauria de millorar i ser el més transparent i responsable possible, ja que si no ningú no ens assegura que el que obtenim és informació esbiaixada i, llavors, podem caure en problemes de discriminació o classificació social discriminatòria (Kitchin, 2014b), a vegades sense ser-ne conscients.

En l'era de les grans dades, per tant, podem arribar a ser discriminats o etiquetats en aspectes que podem o no estar d'acord, a partir del paper de les classificacions algorísmiques (Kitchin, 2014b). Realment, com no tenim gaire control sobre les nostres dades, aquesta informació ens pot perseguir durant molt de temps. A més, hem o hauríem de ser conscients que les nostres dades poden ser utilitzades en usos secundaris, dels quals no som coneixedors. Un bon exemple d'això podria ser el que està passant a la Xina. Recordem que l'estat asiàtic és el primer país on s'està implementant un estat totalitari digital. Segons Jones (2016), a la Xina són experts en fer llistes. Allà es publiquen, per exemple noms, fotos o números de telèfon de morosos, assassins o delinqüents, fent que aquests siguin fàcilment identificables. Aquestes llistes les fan perquè les institucions, organitzacions i governs facin sancions a partir d'elles, com restringint la compra d'un bitllet d'avió o la compra o venda d'una casa. El problema rau, com bé explica Jones (2016), en el fet que avui dia, encara existeixen moltes limitacions en les grans dades, així que en aquelles llistes poden haver-hi errors i, per tant, es podria caure en discriminació a partir d'aquestes.

Mayer-Schönberger & Cukier (2013) destaquen que un dels riscos més rellevants de les grans dades és trobar-nos en una possible dictadura de dades. Exposen que les grans dades es basen en correlacions, per tant, és una eina inadequada per a jutjar la causalitat i, que en cas de

¹² Conan Doyle, A. (1892). *The Adventure of Silver Blaze*. The Strand Magazine.

jutjar en funció de la causalitat i no de la correlació, podríem caure en problemes de discriminació. Eubanks (2018) critica que la presa de decisions automatitzades destrueix la seguretat social i intensifica la discriminació. Per tant, és important tenir present el paper dels algorismes. En l'era de les grans dades els prejudicis dels algorismes poden ser molt més perillosos i transcendents del que ens imaginem (O'Neil, 2016). En aquest sentit, O'Neil (2016) considera que aquests prejudicis poden arribar a distorsionar la societat i la vida de les persones.

Finalment, m'agradaria exposar un exemple que va passar al 2015, en la plataforma Google Photos. Fa gairebé tres anys, Google va determinar que les cares d'unes persones de color eren goril·les i les va catalogar com a tal, fet que va provocar un escàndol (Simonite, 2018). En realitat, aquest fet amaga un error en els seus algorismes. Aquests algorismes són d'aprenentatge automàtic, és a dir, es van retroalimentant amb els criteris que han estat dissenyats. Per tant, en aquest cas ens trobàvem davant d'un algorisme racista, un algorisme que confonia goril·les amb persones. Segurament perquè la base de dades d'aprenentatge no incloïa un volum suficient de fotografies de persones no-blanques, en concret de persones de pell negra.

Conclusions

El segle XXI és també conegut com l'era de la informació i la digitalització. Tant la digitalització com la informació han estat dos factors decisius en l'explosió del Big Data. Actualment, vivim en un món on les dades es reproduïxen, s'emmagatzemen i s'analitzen de manera gairebé automàtica i en temps real. Les grans dades han donat pas a un nou paradigma. Com afirmen Boyd & Crawford "el Big Data esdevé un canvi cultural, tecnològic i acadèmic"(2012, p. 662). Les grans dades són revolucionàries i ens ofereixen noves oportunitats en totes les disciplines, des de les ciències socials a la medicina.

Avui dia, no existeix una definició rigorosa per al concepte Big Data, tot i que actualment, ha esdevingut una paraula de moda. Diversos autors i investigadors lliguen el concepte de les grans dades al model de les 3 Vs – Volum, Varietat i Velocitat–, concepte inicialment introduït per Laney (2001). El Volum fa referència a la gran quantitat de dades, en escala de terabytes, petabytes o zettabytes. La Varietat és la característica o dimensió que defineix l'heterogeneïtat de les dades. I, la Velocitat. Manifesta la rapidesa amb la qual generem i analitzem les dades. Són diversos els acadèmics que creuen que n'existeixen d'altres. Actualment, hi ha experts que consideren que aquest model té més dimensions. Així podem parlar de les 5 Vs (les ja citades més Veracitat i Valor), de les 10 Vs (les cinc anteriors més Variabilitat, Valides, *Venue*, Vocabulari i Vaguetat) i, fins i tot, de les 17 Vs i 1 C (afegint a les ja comentades: Visualització, Viralitat, Viscositat, Volatilitat, Verbositat, Voluntariat, Versatilitat i Complexitat).

Al voltant de les grans dades existeixen alguns conceptes claus com poden ser el de dadificació, algorisme o correlació. La dadificació té com a finalitat posar en format quantitatiu informació de la nostra vida quotidiana per tal de poder recollir-la i analitzar-la ràpidament. Els algorismes esdevenen el nou llenguatge del segle XXI, segons Finn (2017b). Gràcies a ells, es poden obtenir avantatges competitius i es poden prendre decisions minimitzant l'error humà. Però, no hem d'oblidar que poden ser caixes negres – és a dir, algorismes de difícil explicació, ja que no es coneix el seu funcionament intern–. A més, en aquesta nova era informacional, es pretén deixar de banda la causalitat, és a dir, oblidar per què passen les coses, per donar pas a la correlació, és a dir, explicar què passa o què ha passat. Per aquest motiu la correlació és tan important, tot i que també es poden trobar correlacions espúries.

Des del meu punt de vista, en les grans dades no existeixen ni blancs ni negres, es tracta d'una fina línia grisa. En altres paraules, és molt difícil catalogar el Big Data en potencialitats i limitacions, ja que molts cops les seves potencialitats esdevenen les seves principals limitacions, i viceversa. Destaco alguns dels grans reptes que he trobat en els estudis de les

grans dades com poden ser la gestió de la privacitat, l'ús ètic de les dades o la discriminació i classificació automatitzada de les persones.

En aquest sentit, en l'article queda palès que una gran quantitat de terminologia i conceptes encara estan evolucionant, i n' existeixen d'altres que penso que s'haurien de redefinir, com és el cas la privacitat o la justícia. Es creu que el tipus de privacitat idoni és el de la privacitat per disseny. Aquest tipus de privacitat es basa en un tipus de privacitat preventiva i proactiva, "per defecte", incorporada en el disseny, de funcionalitat completa – és a dir, de suma positiva–, que garanteixi protecció en tot el procés del producte i el principi de transparència, i, finalment, que respecti la privacitat de l'usuari.

A més, trobem que les grans dades divideixen. Els algorismes estan dissenyats per éssers humans i en ells podem trobar valors i prejudicis dels seus creadors, mostrant ideologies o bé raons comercials. Els algorismes són capaços d'identificar, classificar, segmentar i ometre. No tothom està representat en els estudis de Big Data. Seria interessant analitzar qui és el "nosaltres" en els estudis de Big Data, ja que no tothom proporciona informació de la mateixa manera.

Els resultats que destaquen de la recerca que he dut a terme són quatre. En primer lloc, molts investigadors defineixen les grans dades a partir de les Vs però aquestes no solament defineixen el Big Data, sinó que també suposen alguns dels seus principals reptes. Algunes de les característiques més conegudes del Big Data són la gran quantitat de dades (Volum) i l'heterogeneïtat de les mateixes (Varietat). Doncs, justament aquestes dues característiques, com ve assenyalen Fan et al. (2013) i Lycett (2013), introdueixen reptes computacionals i estadístics, on apareix el soroll o la correlació espúria, que fan difícil les seves anàlisis i el seu tractament. En segon lloc, s'exposa que una de les principals potencialitats del Big Data és que a partir de la dadificació podem entendre i comprendre millor el comportament humà. Tanmateix, no tothom queda representat en els estudis de les grans dades. Per tant, estem discriminant i no s'estan exposant veritats absolutes. En tercer lloc, realment, el fet de crear una diferència positiva en la vida de les persones o en el funcionament de les empreses, organitzacions i governs depèn de l'ús i de la gestió de les grans dades. I, finalment, també ens podríem fer aquestes preguntes que encara manquen de resposta precisa: Què passa amb totes les persones que viuen a "l'ombra de les grans dades"?, les correlacions són suficients per interpretar els resultats? O és necessària una mica de subjectivitat humana per donar una història o una interpretació el més real possible?, és a dir, és necessària la teoria?

En la revisió a la bibliografia existent al voltant de les grans dades, detecto que el Big Data és un concepte relativament nou i que molts cops els aspectes més innovadors o més curiosos i

interessants es troben en articles webs, més que en llibres o articles. De la bibliografia llegida m'ha cridat l'atenció que molts acadèmics i investigadors de Big Data exposen que la discriminació en les grans dades és un fet, però sempre parlen d'exemples basats en el nivell econòmic o la nacionalitat de les persones. Potser caldria preguntar-se què passa amb una gran majoria de la gent d'edat avançada que no utilitza gairebé dispositius electrònics, i si n'usen, ho fan d'una manera molt desigual als joves. Per tant, pot ser que si seguís estudiant i informant-me sobre les grans dades, orientés futures investigacions cap aquest tema.

Finalment, com a curiositat, Dalton (2016) exposa que l'estadística tradicional va ser vista com una ciència moral i política, i, no va ser fins al segle XX, als anys noranta, que va ser considerada com una ciència exacta, que és el que podria estar passant amb el Big Data. Aviat, per tant, caldria esperar que el Big Data esdevingui una disciplina reconeguda i consolidada.

Bibliografia

- Amazon Echo. (s.d.). Recuperat 3 juny 2018, de https://en.wikipedia.org/wiki/Amazon_Echo
- American Psychological Association. (2010). *Publication Manual of the American Psychological Association, Sixth Edition* (Sixth Edit). Washington, DC.
- Anderson, C. (2008). *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*. *Wired Magazine*. Recuperat de <https://www.wired.com/2008/06/pb-theory/>
- Anthony, D., Stablein, T., & Carian, E. K. (2015). Big Brother in the Information Age: Concerns about Government Information Gathering over Time. *IEEE Security and Privacy*, 13(4), 12-19. <https://doi.org/10.1109/MSP.2015.70>
- Barocas, S. (2014). Data Mining and Discourse on Discrimination.
- Bollier, D. (2010). *The Promise and Peril of Big Data*. *The Aspen Institute*. Washington, DC. Recuperat de <http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/InfoTech09.pdf>
- Borne, K. (2014). Top 10 Big Data Challenges – A Serious Look at 10 Big Data V's. Recuperat de <https://mapr.com/blog/top-10-big-data-challenges-serious-look-10-big-data-vs/>
- Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662-679. <https://doi.org/10.1080/1369118X.2012.678878>
- Burkholder, L. (1992). *Philosophy and the Computer*. Boulder, San Francisco and Oxford: Westview Press.
- Buttarelli, G. (2017). Privacy matters: updating human rights for the digital society. *Health and Technology*, 4(5), 325-328. <https://doi.org/10.1007/s12553-017-0198-y>
- Calude, C. S., & Longo, G. (2017). The Deluge of Spurious Correlations in Big Data. *Foundations of Science*, 22(3), 595-612. <https://doi.org/10.1007/s10699-016-9489-4>
- Cambridge Analytica. (s.d.). Recuperat 3 juny 2018, de https://es.wikipedia.org/wiki/Cambridge_Analytica
- Casanovas, P. (2017). *Sub lege pugnamus*. Barcelona: Universitat de Barcelona. Publicacions.
- Cavoukian, A. (2009). *Privacy by Design: The 7 Foundational Princ. Information and Privacy Commissioner of Ontario*.
- Cavoukian, A., & Jonas, J. (2012). Privacy by Design in the Age of Big Data. *Information and Privacy Commissioner of Ontario*, 14.
- Chandler, D. (2015). A world without causation: Big data and the coming of age of posthumanism. *Millennium: Journal of International Studies*, 43(3), 833-851. <https://doi.org/10.1177/0305829815576817>
- Chokshi, N. (2018). Is Alexa Listening? Amazon Echo Sent Out Recording of Couple's Conversation. Recuperat 3 juny 2018, de <https://www.nytimes.com/2018/05/25/business/amazon-alexa-conversation-shared-echo.html>
- Cole, D. (2014). We Kill People Based on Metadata. Recuperat 6 gener 2018, de <http://www.nybooks.com/daily/2014/05/10/we-kill-people-based-metadata/>
- Cordón, C. (2016). Surge De Nuevo La Polémica, EEUU Pide Datos De Amazon Echo Para Resolver Un Asesinato.
- Crawford, K. (2013). The Hidden Biases in Big Data. *HBR Blog Network*, 9-10. Recuperat de <https://hbr.org/2013/04/the-hidden-biases-in-big-data>
- Crawford, K., Miltner, K., & Gray, M. L. (2014). Critiquing big data: Politics, ethics, epistemology. *International Journal of Communication*, 8(1), 1663-1672. Recuperat de

- <http://ijoc.org/index.php/ijoc/article/view/2167/1164>
- D'Acquisto, G., Domingo-Ferrer, J., Kikiras, P., Torra, V., Montjoye, Y.-A. de, & Bourka, A. (2015). *Privacy by design in big data: An overview of privacy enhancing technologies in the era of big data analytics*. ENISA. <https://doi.org/10.2824/641480>
- Dalton, C. M., Taylor, L., & Thatcher, J. (2016). Critical Data Studies: A dialog on data and space. *Big Data & Society*, 3(1), 205395171664834. <https://doi.org/10.1177/2053951716648346>
- De Mauro, A., Greco, M., & Grimaldi, M. (2015). What is big data? A consensual definition and a review of key research topics. En *AIP Conference Proceedings* (Vol. 1644, p. 97-104). <https://doi.org/10.1063/1.4907823>
- Diebold, F. X. (2000). «Big Data» Dynamic Factor Models for Macroeconomic Measurement and Forecasting. En M. Dewatripont, L. P. Hausen, & S. J. Turnovsky (Ed.), *Advances in Economics and Econometrics, Eighth World Congress of the Econometric Society* (p. 115-122). Seattle. Recuperat de <https://www.sas.upenn.edu/~fdiebold/papers/paper40/temp-wc.PDF>
- Diebold, F. X. (2012). A Personal Perspective on the Origin(s) and Development of «Big Data»: The Phenomenon, the Term, and the Discipline, Second Version. *SSRN Electronic Journal*, 5. <https://doi.org/10.2139/ssrn.2202843>
- Dyson, G. (2012). *Turing's Cathedral: The Origins of the Digital Universe*. New York, NY: Pantheon Books.
- Eastin, M. S., Brinson, N. H., Doorey, A., & Wilcox, G. (2016). Living in a big data world: Predicting mobile commerce activity through privacy concerns. *Computers in Human Behavior*, 58, 214-220. <https://doi.org/10.1016/j.chb.2015.12.050>
- Enriquez, J. (2012). Reflections in a Digital Mirror. En R. Smolan & J. Erwit (Ed.), *The Human Face of Big Data* (p. 18-21).
- Es, K. van, & Schäfer, M. T. (2017). Introduction New Brave World. En M. T. Schäfer & K. van Es (Ed.), *The Datafied Society. Studying Culture through Data* (p. 13-23). Amsterdam, The Netherlands: Amsterdam University Press.
- Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (Firts Edit). New York, NY: St. Martin's Press.
- Fan, J., Han, F., & Liu, H. (2013). Challenges of Big Data Analysis. *National Science Review*, 1(2), 293-314. <https://doi.org/10.1093/nsr/nwt032>
- Felt, M. (2016). Social media and the social sciences: How researchers employ Big Data analytics. *Big Data & Society*, 3(1), 1-15. <https://doi.org/10.1177/2053951716645828>
- Finn, E. (2017a). *What Algorithms Want? Imagination in the Age of Computing*. Information, Communication & Society. Cambridge, MA: MIT Press. <https://doi.org/10.1080/1369118X.2017.1409785>
- Finn, E. (2017b). What Is an Algorithm? En *What Algorithms Want? Imagination in the Age of Computing* (p. 15-56). Cambridge, MA: MIT Press.
- Floridi, L. (2012). Big Data and Their Epistemological Challenge. *Philosophy & Technology*, 25(4), 435-437. <https://doi.org/10.1007/s13347-012-0093-4>
- Fong, W. M. (2017). Bursting the big data bubble. *Significance*, 14(3), 20-23. <https://doi.org/10.1111/j.1740-9713.2017.01035.x>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144. <https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Garden, D. (2012). Introduction. En R. Smolan & J. Erwit (Ed.), *The Human Face of Big Data*

- (p. 14-16).
- Gonzalez-Bailon, S. (2013). Social science in the era of big data. *Policy and Internet*, 5(2), 147-160. <https://doi.org/10.1002/1944-2866.POI328>
- Graham, M. (2012). Big data and the end of the theory. *the guardian*, 1-4. Recuperat de <https://www.theguardian.com/news/datablog/2012/mar/09/big-data-theory>
- Helles, R., & Bruhn-Jensen, K. (2013). Introduction to the special issue «Making data - Big data and beyond». *First Monday*, 18(10), 1-12.
- Hu, M. (2013). Biometric ID Cybersurveillance. *Indiana Law Journal*, 88(4), 1475-1558.
- IBM. (2011). What Is Big Data? Recuperat de <https://www-01.ibm.com/software/sg/data/bigdata/>
- Irazabal, E. (2018). LAS FRONTERAS REGULATORIAS DEL BIG DATA : CRECE EL PODER DE LA UNIÓN. Recuperat de <https://blogs.deusto.es/bigdata/las-fronteras-regulatorias-del-big-data-crece-el-poder-de-la-union-europea/>
- Jiménez Garcia, S. (2018). *Protección de Datos de Carácter Personal. Agencia Estatal Boletín Oficial del Estado*. Recuperat de <http://www.boe.es/buscar/pdf/1999/BOE-A-1999-23750-consolidado.pdf>
- Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and Challenges of Big Data Research. *Big Data Research*, 2(2), 59-64. <https://doi.org/10.1016/j.bdr.2015.01.006>
- Jones, S. (2016). China invents the digital totalitarian state. *The Economist*, 1-8. Recuperat de <http://www.economist.com/news/briefing/21711902-worrying-implications-its-social-credit-project-china-invents-digital-totalitarian>
- Joyanes Aguilar, L. (2014). ¿Qué es el Big Data? En *Big Data: Análisis de grandes volúmenes de datos en organizaciones* (MARCOMBO, p. 22). México: Alfaomega.
- Kitchin, R. (2014a). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 1-12. <https://doi.org/10.1177/2053951714528481>
- Kitchin, R. (2014b). *The data revolution*. Sage. <https://doi.org/10.1017/CBO9781107415324.004>
- Kitchin, R. (2016). Thinking critically about and researching algorithms. *Routledge*, 20(1), 14-29. <https://doi.org/10.1080/1369118X.2016.1154087>
- Kitchin, R., & Lauriault, T. P. (2014). Towards critical data studies : Charting and unpacking data assemblages and their work. *Geoweb and Big Data*, 1-19. Recuperat de http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112
- Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. *Application Delivery Strategies*, 949(February 2001), 4. <https://doi.org/10.1016/j.infsof.2008.09.005>
- Lesk, M. (2013). Big Data, Big Brother, Big Money. *Security & Privacy, IEEE*, 11(4), 85-89. <https://doi.org/10.1109/MSP.2013.81>
- Letouzé, E. (2012). Big Data for Development: Challenges & Opportunities. *Global Pulse*, (May), 1-47. Recuperat de <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>
- Leurs, K., & Shepherd, T. (2017). Datafication & Discrimination. En M. T. Schäfer & K. van Es (Ed.), *The Datafied Society. Studying Culture through Data* (p. 211-232). Amsterdam, The Netherlands: Amsterdam University Press.
- Lohr, S. (2012). The Age of Big Data. *The New York Times*, 1-5. <https://doi.org/10.1126/science.1243089>
- Lycett, M. (2013). «Datafication»: Making sense of (big) data in a complex world. *European*

- Journal of Information Systems*, 22(4), 381-386. <https://doi.org/10.1057/ejis.2013.10>
- Mai, J.-E. (2016). Big data privacy: The datafication of personal information. *The Information Society*, 32(3), 192-199. <https://doi.org/10.1080/01972243.2016.1153010>
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: a revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt.
- Miyazaki, S. (2012). Algorhythmic: Understanding micro-temporality in computational cultures. *Computational Culture*, (2). Recuperat de <http://computationalculture.net/algorhythmic-understanding-micro-temporality-in-computational-cultures/>
- Normann, R. (2001). *Reframing Business: When the Map Changes the Landscape*.
- O'Neil, C. (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy* (First edit). New York: Crown.
- Pacheco-Vega, R. (2016). How to do a literature review: Citation tracing, concept saturation and results' mind-mapping. Recuperat 15 juny 2018, de <http://www.raulpacheco.org/2016/06/how-to-do-a-literature-review-citation-tracing-concept-saturation-and-results-mind-mapping/>
- Panimalar, A., Shree, V., & Kathrine, V. (2017). The 17 V's Of Big Data. *International Research Journal of Engineering and Technology (IRJET)*, 4(09), 329-333.
- Pasquale, F. (2015). The Black Box Society. <https://doi.org/10.4159/harvard.9780674736061>
- Power, D. J. (2016). "Big Brother" can watch us. *Journal of Decision Systems*, 25(January), 578-588. <https://doi.org/10.1080/12460125.2016.1187420>
- Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). Analytics: The real-world use of big data: How innovative enterprises extract value from uncertain data. *IBM Institute for Business Value*, 1-20. Recuperat de https://www.ibm.com/smarterplanet/global/files/se__sv_se__intelligence__Analytics_-_The_real-world_use_of_big_data.pdf
- Simonite, T. (2018). When it comes to gorillas, google photos remains blind.
- Steiner, C. (2012). *Automate this: How algorithms took over our markets, our jobs, and the world*. New York, NY: Portfolio Penguin.
- Strauß, S. (2015). Datafication and the seductive power of uncertainty -A critical exploration of big data enthusiasm. *Information (Switzerland)*, 6(4), 836-847. <https://doi.org/10.3390/info6040836>
- Tene, O., & Polonetsky, J. (2013). Privacy and Big Data: Making Ends Meet. *Stanford Law Review Online*, 66, 25.
- Turner, D., Schroeck, M., & Shockley, R. (2013). Analytics: The real-world use of big data in financial services How innovative banking and financial markets organizations. *IBM Global Business Services*, 16. Recuperat de https://www-935.ibm.com/services/multimedia/Analytics_The_real_world_use_of_big_data_in_Financial_services_Mai_2013.pdf
- Turow, J. (2007). Niche Envy: Marketing Discrimination in the Digital Age. *International Journal of Communication*, 1, 2006-2007.
- Uddin, M. F., & Lee, J. (2016). We Are What We Generate - Understanding Ourselves Through Our Data. En D. C.H. (Ed.), *Procedia Computer Science* (Vol. 95, p. 335-344). Elsevier B.V. <https://doi.org/10.1016/j.procs.2016.09.343>
- Uricchio, W. (2017). Data, culture and the Ambivalence of Algorithms. En M. T. Schäfer & K. van Es (Ed.), *The Datafied Society. Studying Culture through Data* (p. 125-138). Amsterdam, The Netherlands: Amsterdam University Press.

- van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance and Society*, 12(2), 197-208.
- Weiss, S., & Indrukya, N. (1998). *Predictive Data Mining: A practical guide*. Morgan Kaufmann Publishers, Inc.
- Wenning, R. (2010). *Privacy - Why on earth should we care? W3C Workshop on Privacy for Advanced Web APIs* (Vol. 12). London. Recuperat de <https://www.w3.org/2010/api-privacy-ws/papers/privacy-ws-37.html>

Annex

Bibliografia anotada

- Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine*, 16(07), 1-10. Recuperat de <https://www.wired.com/2008/06/pb-theory/>

L'autor exposa que estem vivint en l'era del Petabyte. En un món on les dades han de ser interpretades primer de manera matemàtica i després posades en context.

Anderson, l'autor de l'article, és qui afirma per primera vegada que amb suficients dades, els números poden parlar per si mateixos. Proclama el final de la teoria. Creu que la revolució de les dades farà que el mètode científic esdevingui obsolet.

Finalment, Anderson afirma que hem de deixar de banda la causalitat per donar pas a la correlació. No importa per què passen les coses, ja que ara es poden quantificar i mesurar.

És un autor que presenta una postura optimista i revolucionària enfront de l'era del Big Data. Pretén trencar amb l'estadística tradicional.

- Anthony, D., Stablein, T., & Carian, E. K. (2015). Big Brother in the Information Age: Concerns about Government Information Gathering over Time. *IEEE Security and Privacy*, 13(4), 12-19. <https://doi.org/10.1109/MSP.2015.70>

Els autors de l'article exposen com a partir de les declaracions que va fer Snowden sobre la NSA (Agència de Seguretat Nacional) es va generar un debat públic sobre les preocupacions de vigilància massiva, la privacitat, etc. Els autors afirmen que aquestes preocupacions vénen d'anys enrere.

Aquest article exposa l'evolució de la privacitat al llarg del temps als Estats Units. Al llarg de l'argumentació es veu una crítica sobre la privacitat dels EEUU. Els autors exposen que les lleis de privacitat no s'han mantingut al mateix ritme que les noves tecnologies. La qual cosa provoca que hi hagi menys privacitat.

Com a conclusió, els autors argumenten que avui dia existeix el problema que a mesura que la gent es va familiaritzant amb les noves tecnologies les pors i les preocupacions desapareixen i això és el que esdevé, realment, perillós.

- Barocas, S. (2014). Data Mining and Discourse on Discrimination.

A partir de les dades els analistes de Big Data poden saber el nostre sexe, la nostra raça, la nostra religió, la nostra orientació sexual, etc. Per tant, les dades poden crear avantatges competitiu o bé poden fer-nos molt de mal.

En aquest article, l'autor se centra en les possibles discriminacions que poden existir en les dades. Barocas exposa que no sempre la discriminació és intencionada. Ell destaca la bretxa digital com una de les grans discriminacions que existeix. Exposa que no tots els ciutadans es veuen representats en les anàlisis de Big Data, ho atribueix a un biaix de selecció¹³.

Per altra banda, destaca la correlació. Critica que la correlació no sempre és objectiva, és a dir, a vegades els analistes poden fer inferències incorrectes, especialment en les cerques en línia.

- Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662-679.
<https://doi.org/10.1080/1369118X.2012.678878>

L'article examina sis provocacions de caràcter cultural, tecnològic i acadèmic per tal d'obrir debats sobre el Big Data.

Actualment, la definició de Big Data ja no només depèn del nombre de dades sinó de la capacitat de cerca, agregació i referències creuades de grans conjunts de dades. El Big Data està canviant la definició del coneixement. A més, l'article critica que les afirmacions d'objectivitat porten a l'error, que les grans bases de dades no són sempre les millors, i que si es treu fora de context, el Big Data perd el seu significat. Així mateix, fa una crítica al fet que les dades siguin accessibles, ja que això no implica que sigui un assumpte ètic. Finalment, l'article exposa que l'accés limitat al Big Data està creant una nova bretxa digital.

¹³ El biaix mostral s'entén com una distorsió que es introduïda degut a la forma com es selecciona la mostra. Existeixen diferents tipus de biaixos mostrals: d'espai, d'informació, de participació, d'estudi, etc.

Aquest article és útil per la meua recerca, ja que parla de diferents conceptes que engloben al Big Data, i, personalment, opino que l'autora aconseguirà que el públic reflexioni sobre aquest fet.

- Bunz, M. (2017). The Need for a Dialogue with Technology. En M. T. Schäfer & K. van Es (Ed.), *The Datafied Society. Studying Culture through Data* (p. 249-255). Amsterdam, The Netherlands: Amsterdam University Press.

Aquest capítol manifesta la necessitat d'interactuar amb la tecnologia perquè si no esdevindrà una caixa negra per a la societat.

El Big Data i la tecnologia aporten a la societat grans beneficis però alhora, tots dos, són causants de molts altres perills com la vigilància massiva, el rastreig, etc. La tecnologia no és una solució per als problemes polítics, però a vegades pot arribar a ser la causant.

- Buttarelli, G. (2017). Privacy matters: updating human rights for the digital society. *Health and Technology*, 4(5), 325-328. <https://doi.org/10.1007/s12553-017-0198-y>

En aquest article, Buttarelli, exposa que la privacitat ens ha de preocupar a tots, especialment en l'era digital. L'autor argumenta que la privacitat hauria de ser un requisit per tal de poder gaudir altres llibertats i no patir discriminacions per sexe, raça, orientació sexual, creences, etc.

En l'article es fa un repàs de la privacitat al llarg de la història. Es creu que és necessari un canvi en la reformulació de què és la privacitat i com i quan s'ha d'aplicar, ja que exposa que la privacitat ha de ser per a tothom i per igual, i sobretot, real i present.

- Calude, C. S., & Longo, G. (2017). The Deluge of Spurious Correlations in Big Data. *Foundations of Science*, 22(3), 595-612. <https://doi.org/10.1007/s10699-016-9489-4>

Els autors argumenten que no hem de creure que les dades poden parlar per si mateixes. Afirmen que el fet de pensar que en l'era de les grans dades el mètode científic queda obsolet és un error.

Al llarg de l'argumentació exposen els autors que en l'era del Big Data es parlarà de correlació no de causalitat. Amb les grans dades no ens interessa saber "per què" passen les coses (causalitat), ens interessa saber "què" està passant o ha passat (correlació). A més, expliquen

que és cert que hem de donar pas a la correlació i “oblidar” la causalitat, però hem de ser conscients que en les grans bases de dades hi poden haver correlacions espúries i ens poden induir a caure en fal·làcies.

Finalment, els autors afirmen que tot i que ens basem en correlacions no podem deixar de banda la teoria, en depèn quines disciplines no està ni estaria ben vist fer un ús acrític de les correlacions. Que a partir de l'era de les grans dades el que hem de fer és millorar el mètode científic.

- Casanovas, P. (2017). *Sub lege pugnamus*. Barcelona: Universitat de Barcelona. Publicacions.

És un llibre de filosofia redactat per Pompeu Casanovas. En el llibre s'explica la transformació que ha patit el món a partir de les grans dades i la Internet de les coses, sempre des d'un punt de vista filosòfic, polític i legislatiu.

L'autor argumenta que en el món en el qual vivim hem de poder trobar un equilibri entre la llibertat i les polítiques de seguretat. Alhora, assumeix que no sempre som conscients dels resultats que ens ofereixen les dades.

Per altra banda exposa que vivim en un món informacional on, com ell afirma, som “prosumidors”, és a dir, productors i consumidors d'informació. Destaca que la nostra societat, en un futur no molt llunyà, serà controlada de manera gaire bé automàtica.

A més, a partir de les seves línies de reflexió veus que no creu que les dades parlin per si soles, sinó que nosaltres les fem parlar.

Finalment, exposa que en el món on vivim, la seguretat i la privacitat són o haurien de ser un bé bàsic per a tots els ciutadans. Alhora, defensa que la societat ha canviat molt i que, per tant, la llei no ha de continuar sent entesa com es feia fins al moment.

- Cavoukian, A., & Jonas, J. (2012). *Privacy by Design in the Age of Big Data*. *Information and Privacy Commissioner of Ontario*, 14.

Els autors argumenten que les grans dades tenen la capacitat de generar un enorme valor en la societat, malgrat això en el seu article no presenten els beneficis del Big Data, sinó que se centren en la privacitat, especialment en la privacitat per disseny.

Cavoukian i Jonas (2012) reflexionen que en el passat, la privacitat era vista com un bé personal, més que com un bé social. Alhora, els autors argumenten que, actualment, els canvis tecnològics poden perjudicar la privacitat, per exemple, les anàlisis més potents poden fer possible la re-identificació. Per tant, ells reflexionen que seria òptim que la privacitat s'integrés per defecte, durant l'arquitectura, el disseny i la construcció dels processos. Tot això dóna pas a la privacitat per disseny (PbD), concepte introduït per Cavoukian (2009)¹⁴.

La PbD emfatitza el respecte per la privacitat de l'usuari i la necessitat d'integrar la privacitat com una condició per defecte, però preserva el compromís amb la funcionalitat en una estratègia "win-win" o suma positiva. La privacitat per disseny es basa en 7 principis fundacionals:

1. Proactiva, no reactiva; preventiva no reparadora, és a dir, la privacitat per disseny ve abans dels fets, no després.
2. La privacitat "per defecte". L'objectiu de la PbD és oferir el màxim grau de privacitat garantint la protecció automàtica de les dades personals en qualsevol sistema informàtic o pràctica empresarial. Amb aquest principi, l'usuari no necessita fer res, la privacitat es manté intacta, està incorporada en el sistema, per defecte.
3. La privacitat incorporada en el disseny.
4. Funcionalitat completa, suma positiva, no suma 0. Busca satisfer tots els interessos i objectius de manera que tots surtin guanyant. La PbD evita les dicotomies falses com la privacitat vs. seguretat, demostrant que és possible tenir ambdues.
5. Seguretat d'extrem a extrem: protecció total del cicle de vida del producte.
6. Visibilitat i transparència, tant pels usuaris com pels proveïdors.
7. Respecte per la privacitat de l'usuari.

Aquests principis s'han d'introduir des del principi, durant la planificació, el disseny i els procediments operatius.

- Chandler, D. (2015). A world without causation: Big data and the coming of age of posthumanism. *Millennium: Journal of International Studies*, 43(3), 833-851.
<https://doi.org/10.1177/0305829815576817>

Al llarg de l'article s'estudien els supòsits epistemològics i ontològics en què es basen les grans dades. Els supòsits epistemològics fan referència a la teoria del coneixement, mentre que els

¹⁴ Cavoukian, A. (2009). Privacy by Design: The 7 Foundational Princ. Information and Privacy Commissioner of Ontario.

ontològics fan referència a la part de la metafísica que tracta de l'ésser en general¹⁵. L'autor pretén fer-nos reflexionar sobre les noves formes de govern i de producció de coneixement i/o informació.

L'autor, com molts altres, argumenta que no existeix una única definició per explicar què és el Big Data però creu que la definició sempre ha d'anar lligada a les 3 Vs: Volum, Veracitat i Varietat¹⁶.

Avui dia, les dades no es generen a través de cap pregunta o hipòtesi, són un efecte secundari de les activitats de la nostra vida quotidiana. Tot això comporta beneficis com prevencions enfront d'esdeveniments futurs, respostes ràpides, etc. Per altra banda, les grans dades ens poden portar a la discriminació dels grups més petits o amb menys possibilitats (econòmiques, geogràfiques, laborals, entre d'altres.).

Segons l'autor, les grans dades busquen ser més inductives que les dades tradicionals. Volen "obtenir" la realitat, a vegades omesa, pels supòsits causals. És a dir, la correlació ha de reemplaçar la causalitat. També, fa èmfasi en el fet que les grans dades no permeten canviar la nostra situació actual, únicament ens ajuda a adaptar-nos a elles i a ser-ne conscients.

Alhora, l'autor critica que si el govern continua governant de manera tradicional, és a dir, si no és capaç d'adaptar-se a les noves tecnologies i fer un bon ús de les dades, recopilar grans quantitats de dades no té sentit.

- Cole, D. (2014). We Kill People Based on Metadata. Recuperat 6 gener 2018, de <http://www.nybooks.com/daily/2014/05/10/we-kill-people-based-metadata/>

En aquest article, l'autor argumenta la importància de la privacitat. En molts moments fa referència a la privacitat que existeix en els Estats Units. Argumenta com la NSA (Agència de Seguretat Nacional) investiga a tots els ciutadans, especialment a aquells que són estrangers.

S'exposa que la NSA va crear PRISM (Planning Tool for Resource Integration, Synchronization, and Management) amb l'ànim de recopilar tot el contingut dels correus electrònics, registrar trucades de telèfon, ubicació, etc. Estem vivint en una societat de vigilància massiva.

¹⁵ Definició d'epistemologia i ontologia segons el diccionari Institut d'Estudis Catalans (DIEC) <https://www.iec.cat/activitats/entrada.asp> (09/05/18)

¹⁶ Veure Gadomi i Haider (2015) per a una definició de les 3V.

Òbviament, l'autor posa de manifest que no totes les conclusions que es treuen són negatives però fa una crítica a què hem d'estar alerta, ja que les nostres llibertats civils estan en joc.

- Crawford, K. (2013). The Hidden Biases in Big Data. *HBR Blog Network*, 9-10.
Recuperat de <https://hbr.org/2013/04/the-hidden-biases-in-big-data>

En aquest article es pot apreciar la discriminació existent en el món del Big Data. L'autora critica que el Big Data no sempre reflecteix la veritat objectiva. També exposa que les dades són creacions del disseny humà, per tant no tenen per què ser objectives.

Kate Crawford en el seu assaig fa una reflexió molt interessant per la meua investigació. Ella reflexiona que se suposa que les dades reflecteixen amb exactitud el món social, però exposa que hi ha carències significatives en algunes comunitats. Fa referència a comunitats pobres que, per exemple, no tenen tants dispositius electrònics com en zones més riques, o bé, la gent gran que en general és menys donada a l'ús de dispositius electrònics. Tot això està generant bretxes digitals.

- Crawford, K., Miltner, K., & Gray, M. L. (2014). Critiquing big data: Politics, ethics, epistemology. *International Journal of Communication*, 8(1), 1663-1672. Recuperat de <http://ijoc.org/index.php/ijoc/article/view/2167/1164>

L'objectiu d'aquest article, com exposen els autors, no és destruir el fenomen del Big Data, és millorar la forma en què es realitza la investigació social i cultural. Ho fan mitjançant una crítica teòrica i pràctica sobre l'emmagatzematge de les dades, l'ètica d'ús, la discriminació, la falta d'accés, etc. En aquestes "crítiques" s'exposen arguments i positius per crear una opinió al lector.

Com a curiositat, Cornelius Puschmann i Jean Burgess descobreixen que les grans dades es poden explicar de dues maneres. Una com un recurs a consumir i una altra com una força natural a controlar.

Finalment, l'article exposa que hauríem de començar un nou escepticisme sobre la validesa a priori del Big Data com a representació naturalitzada del món per així poder descobrir les seves limitacions i potencialitats.

- D'Acquisto, G., Domingo-Ferrer, J., Kikiras, P., Torra, V., Montjoye, Y.-A. de, & Bourka, A. (2015). *Privacy by design in big data: An overview of privacy enhancing technologies in the era of big data analytics*. ENISA. <https://doi.org/10.2824/641480>

Aquest article estudia la privacitat en el marc de les grans dades. L'objectiu seria trobar l'equilibri perfecte entre l'ús de les grans dades i la protecció de la privacitat de les dades personals. Els autors exposen que no és una tasca senzilla.

Es presenten els principals riscos de la privacitat, que són els següents: la falta de control i transparència, la reutilització de les dades, la re-identificació de les dades, és a dir, la pèrdua de l'anonimat, el perill de prendre decisions de manera automàtica, i, finalment, la dificultat per revisar els controls de protecció de dades.

Els autors exposen que el Big Data i la privacitat poden arribar a ser molt contradictoris, però que no creuen que pugui existir el Big Data sense privacitat. Expliquen que si les grans dades no necessitessin privacitat, les dades que avui dia es comparen amb el petroli, tindrien un valor molt menor.

Finalment, els autors presenten un tipus de privacitat que pensen que pot ser òptima per l'era de les grans dades. La privacitat per disseny. Ho defineixen com un procés en el qual intervenen components tecnològics i organitzatius que apliquen els principis de privacitat i de protecció de dades.

- Dalton, C. M., Taylor, L., & Thatcher, J. (2016). Critical Data Studies: A dialog on data and space. *Big Data & Society*, 3(1), 205395171664834. <https://doi.org/10.1177/2053951716648346>

L'article es basa en una entrevista on els autors discuteixen els reptes, les idees, les responsabilitats i les possibilitats dels estudis del Big Data. Aquests estudis es fan des d'un punt de vista reflexiu i crític, i en un nivell teòric i pràctic. Una de les conclusions més rellevants que he pogut extreure de l'article és que el Big Data divideix a diferents nivells: segons la riquesa, l'accés a les tecnologies, al coneixement, a l'anàlisi, etc. Així mateix, en l'article queda reflectit que el Big Data no pot ser entès clarament si es treu fora de context. A més, s'afirma que la no resposta també és una resposta i que els límits de les persones molts cops van lligats als límits de les dades. També, relaciona la mitologia del Big Data amb el seu valor comercial.

En l'àmbit comercial és molt útil el Big Data però també s'han de tenir en compte les limitacions i els riscos que suposa per a la presa de decisions.

Com a curiositat, recalcar que l'estadística va ser vista com una ciència moral i política, no va ser fins als noranta que va ser considerada com una ciència exacta, que és el que pot estar passant amb el Big Data i per això hem de descobrir les seves limitacions i potencialitats per fer-lo més ric.

Finalment, exposo una petita crítica en relació a l'article, ja que sempre fa referència a l'estudi de les dades, en general, però no es detalla que el que realment s'està fent són crítiques en relació al fenomen del Big Data.

- De Mauro, A., Greco, M., & Grimaldi, M. (2015). What is big data? A consensual definition and a review of key research topics. En *AIP Conference Proceedings* (Vol. 1644, p. 97-104). <https://doi.org/10.1063/1.4907823>

Els autors de l'article afirmen que avui dia la paraula "Big Data" ha esdevingut un "buzzword" o paraula de moda, tot i que reflexionen que el concepte és relativament jove. En el seu article analitzen les grans dades a partir de quatre temes que ells creuen que són els termes que van associats al Big Data: Informació, Tecnologia, Mètodes i Impacte.

Els autors destaquen que ser conscients de les limitacions de les grans dades és un recurs fonamental per a les organitzacions que volen impulsar la presa de decisions basades en dades. A més, afirmen que les grans dades ofereixen avantatges competitius a les organitzacions.

De Mauro et al. creuen que el nucli del concepte de les grans dades pot ser expressat per:

- Les 3V, concepte introduït per Laney (2001)¹⁷ per descriure les característiques de la Informació.
- Una tecnologia específica i mètodes d'anàlisi concrets, per aclarir els requisits estrictament necessaris per fer ús d'aquesta informació.
- La transformació en coneixement, i conseqüentment, creació de valor econòmic, com la principal forma en què el Big Data està impactant a les empreses i a la societat.

¹⁷ Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. *Application Delivery Strategies*, 949(February 2001), 4. <https://doi.org/10.1016/j.infsof.2008.09.005>

Finalment, a partir dels trets anteriors proposen la següent definició: *“Big Data represents the Information assets characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value.”*(p.103)

- Diebold, F. X. (2012). A Personal Perspective on the Origin(s) and Development of «Big Data»: The Phenomenon, the Term, and the Discipline, Second Version. *SSRN Electronic Journal*, 5. <https://doi.org/10.2139/ssrn.2202843>

Per Diebold les grans dades es troben en el punt de mira de la ciència i dels negocis. L'autor discuteix el terme Big Data per una banda, des del punt de vista del fenomen de les grans dades, en el sentit de les grans onades d'informació que disposem, la velocitat, la informació en temps real, etc. I, per l'altra, des del punt de vista del concepte. On exposa, com altres autors inclosos en aquest document, que no existeix una única definició per les grans dades i que hem d'anar redefinint el concepte.

També, remarca que el Big data, en l'actualitat, no és únicament un fenomen i un concepte. L'autor expressa que les grans dades estan esdevenint una nova disciplina.

Finalment, l'autor exposa que el terme “Big Data” es va començar a fer servir a mitjans de la dècada dels noranta. A més, Diebold en la seva conclusió exposa el següent:

“The first significant academic references are arguably Weiss and Indurkha (1998) in computer science and Diebold (2000) in statistics/econometrics. An unpublished 2001 research note by Douglas Laney at Gartner enriched the concept significantly. Hence the term “Big Data” appears reasonably attributed to Massey, Weiss and Indurkha, Diebold, and Laney. Big Data the phenomenon continues unabated, and Big Data the discipline is emerging.” (2012, p. 5)

- Domínguez, E. R. (2016). Big Data y Black Mirror : China siempre estuvo cerca. Recuperat de <http://dominguezprost.com/bigdatayblackmirror/>

En aquest blog es compara la xarxa social Instagram amb un capítol de la sèrie *Black Mirror*. En el capítol, es narra la història d'un aplicatiu que a mesura que la gent es va creuant amb altres persones es van donant punts en funció de les seves funcions, vestuari, gustos, etc. Representa que les persones amb més punts són aquelles que es podrien considerar com a “socialment respectats o estimats”. A més, representa que en depenent dels punts que reben les persones poden tenir descomptes, accedir a certs viatges, etc.

En aquest blog, l'autor ens vol fer reflexionar exposant que a la Xina es vol portar a terme una cosa semblant a partir de l'any 2020 (BBC)¹⁸. Es planteja el naixement d'un sistema de crèdits socials, on combinarien la informació dels ciutadans en diferents eixos de la vida quotidiana: treball, trajectòria educativa, comportament com a consumidor, antecedents, ... Per tant, a partir de diferents bases de dades i d'acord amb diferents ponderacions l'estat podria puntuar als ciutadans d'acord a unes bones o dolentes conductes. L'autor argumenta que en un rànquing social, podríem viure en un món enganyós i superficial.

Finalment, l'autor expressa que el fet de tenir més informació no sempre és sinònim de millores, transparència i benestar. Ho deixa palès amb l'exemple de la Xina.

- Eastin, M. S., Brinson, N. H., Doorey, A., & Wilcox, G. (2016). Living in a big data world: Predicting mobile commerce activity through privacy concerns. *Computers in Human Behavior*, 58, 214-220. <https://doi.org/10.1016/j.chb.2015.12.050>

Aquest article exposa que la recopilació de les dades mòbil cada vegada han esdevingut més interessants per als acadèmics i professionals. A més, en l'article s'argumenten les preocupacions de la privacitat dels usuaris dels mòbils com pot ser la violació de la privacitat, control, ús secundari no autoritzat, accés inadequat, rastreig d'ubicació, etc.

Tot i que també s'esmenta que pot arribar a ser un gran avantatge competitiu.

Aquest article m'ha semblat important per la meua recerca, ja que parla del Big Data a partir dels mòbils. Personalment, opino que és molt interessant, ja que avui en dia tots o gairebé tots som usuaris d'ells.

- Enriquez, J. (2012). Reflections in a Digital Mirror. En R. Smolan & J. Erwitte (Ed.), *The Human Face of Big Data* (p. 18-21).

L'autor exposa que ens trobem en les primeres etapes de la revolució del Big Data i ja s'estan transformant totes les disciplines. Cada vegada hi ha més i més dades.

En un moment donat, l'autor compara les grans dades amb els tatuatges. Exposar que el fenomen del Big Data és com si ens tatuéssim electrònicament, amb totes les seves conseqüències. Personalment, m'agrada la comparativa de les conseqüències de les grans

¹⁸ Font: <http://www.bbc.com/news/world-asia-china-34592186> (07/06/18).

dades amb el fet, cada dia més habitual, de portar tatuatges. És una manera molt objectiva de descriure que totes aquelles dades no estructurades i estructurades que produïm diàriament estaran sempre al nostre costat i mai podran ser eliminades. És una manera d'identificar patrons, persones, etc.

Finalment, critica que la revolució de les dades és relativament nova i que hem d'aprendre a gestionar les dades, ja que ens estem quedant sense espai on emmagatzemar les dades.

- Es, K. van, & Schäfer, M. T. (2017). Introduction New Brave World. En M. T. Schäfer & K. van Es (Ed.), *The Datafied Society. Studying Culture through Data* (p. 13-23). Amsterdam, The Netherlands: Amsterdam University Press.

En aquest capítol se'ns presenta el Big Data com el motor d'un procés tecnològic i social que ens està portant a la revolució de les dades. Està transformant la manera com vivim, treballem, pensem, guanyem diners, etc. Les dades s'han convertit en objectes ontològics i epistemològics d'investigació.

En l'era de les grans dades s'espera que l'anàlisi de les dades aportí una manera objectiva de captar la realitat complexa i dinàmica que ens envolta. En l'àmbit acadèmic, ha estat constantment criticada. Així com Rob Kitchin, en el seu llibre *The data revolution* (2014) exposa afirmacions que sosté l'empirisme del Big Data. Algunes d'elles són que el Big Data pot capturar la totalitat, no hi ha necessitat de teoria perquè els números parlen per si mateixos, etc.

És necessari que facin transparents el paper dels algorismes.

Aquest llibre és molt interessant per la meua recerca, no només la Introducció, ja que ens convida a fer reflexions sobre els reptes que estan sorgint en la societat a causa del Big Data.

- Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (Firts Edit). New York, NY: St. Martin's Press.

Virginia Eubanks és una persona que lluita per la justícia social en l'era de la informació. En el seu llibre, explica a partir d'històries d'altra gent i de la seva pròpia experiència que és la desigualtat automatitzada. En el llibre ella argumenta com afecten la mineria de dades, els algorismes a la gent de classe treballadora i als pobres en els Estats Units. Eubanks introdueix el concepte de casa digital dels pobres.

A partir dels seus relats, afirma que tot el fenomen de les grans dades afecta a tothom però que de manera més invasiva a la gent pobre, ja que argumenta que les eines digitals amb les quals interactuem, empitjoren els prejudicis, destrueixen la democràcia i violen els drets humans.

A partir del seu llibre ens convida a reflexionar sobre una veritat a vegades amagada. Ella relata exactament el que passa amb la gent que viu a l'ombra de les grans dades.

- Fan, J., Han, F., & Liu, H. (2013). Challenges of Big Data Analysis. *National Science Review*, 1(2), 293-314. <https://doi.org/10.1093/nsr/nwt032>

Els autors presenten els reptes principals de les grans dades. Argumenten que els principals reptes que té el Big Data són de volum, de cost computacional i estadístics. Això fa que sigui necessari un nou paradigma. Cal estudiar i investigar nous mètodes estadístics i computacionals per tal de poder superar els reptes plantejats pel Big Data.

Els principals reptes als quals ha de fer front el Big Data des d'un punt de vista de volum, computacional i estadístic són els següents: les grans dades tenen problemes de soroll, presenten correlacions espúries, homogeneïtat incidental, etc. A més, com les grans dades són dades agregades de múltiples fonts generen reptes d'heterogeneïtat, dades no estructurades, biaixos, etc.

Alhora, els autors exposen que les grans dades ben utilitzades poden presentar infinitat de beneficis.

- Felt, M. (2016). Social media and the social sciences: How researchers employ Big Data analytics. *Big Data & Society*, 3(1), 1-15. <https://doi.org/10.1177/2053951716645828>

Avui dia les xarxes socials amb la col·laboració de les ciències socials estan ajudant a comprendre la interacció humana i, així en conseqüència, millorar la condició humana.

Els mitjans de comunicació social, o més coneguts com a xarxes socials, són una font de dades bastant rica. Són capaços de respondre a moltes preguntes de recerca en ciències socials.

En aquest àmbit, com en qualsevol altre, és necessària una anàlisi crítica de les dades. En aquest cas, s'han d'utilitzar mètodes digitals d'acord a la dinàmica de la plataforma.

En les investigacions de les ciències socials es considera realment important enriquir l'analítica de les dades amb metodologies tradicional per a proporcionar una descripció més completa.

- Finn, E. (2017). *What Algorithms Want? Imagination in the Age of Computing. Information, Communication & Society*. Cambridge, MA: MIT Press.
<https://doi.org/10.1080/1369118X.2017.1409785>
- **Cap. I: “What Is an Algorithm?” (p. 15-56)**

En aquest capítol l'autor pretén contextualitzar als algorismes. Els algorismes són el nou llenguatge del segle XXI, per això és necessari entendre'ls per tal de donar-los sentit.

L'autor relaciona els algorismes amb la teologia. Els compara amb la figura de Déu, ja que estan per totes bandes. Tot i que comenta que els algorismes, la ciència, ens aparta de la religió.

Els algorismes ens proporcionen coneixements i solucions del nostre dia a dia. Alhora, argumenta que només creure en el poder dels algorismes ens pot portar a la discriminació.

Els algorismes són una nova manera d'entendre la cultura. Es pot observar com una barreja entre el poder computacional i els humans. Ofereix un nou paradigma cibernètic.

L'autor expressa que en l'actualitat tots som lectors i programadors d'algorismes, en major o menor mesura. Per això considera que és vital entendre'ls i ser conscients que estem vivint amb ells. Tot i que només he considerat interessant el primer capítol, en la resta de capítols trobem exemples d'algorismes relacionats amb Siri, Netflix, Uber, etc.

- Floridi, L. (2012). Big data and their epistemological challenge. *Philosophy and Technology*, 25(4), 435-437. <https://doi.org/10.1007/s13347-012-0093-4>

Floridi argumenta que el fenomen del Big Data està de moda, tot i que no està gens clar que significa exactament. Ell proposa diferents definicions, tot i que critica que les definicions que existeixen són poc rigoroses. Alhora, fa èmfasi a què la definició no es pot basar únicament en el terme “gran”.

Planteja que existeixen dues confusions pel que fa a l'epistemologia. Una en relació al problema de què cada dia hi ha més i més dades; i una altra en relació a la solució d'intentar reduir les dades, per tal que puguin ser gestionades. Argumenta que és absurd queixar-se de cada dia són més rics en dades que això no pot ser el problema, i que, per tant, s'ha de reconsiderar la solució.

Finalment, s'exposa que el veritable problema epistemològic en el Big Data són els patrons petits, que argumenta que només poden ser significatius si s'agreguen adequadament.

Pel que fa a la meua recerca, he arribat a la mateixa conclusió que l'autor. S'han d'elaborar millors tècniques per veure els petits patrons de dades, i és necessària una millor epistemologia per filtrar aquells que són valuosos, ja que avui dia disposem de moltes dades i moltes poden no tenir valor.

- Fong, W. M. (2017). Bursting the big data bubble. *Significance*, 14(3), 20-23.
<https://doi.org/10.1111/j.1740-9713.2017.01035.x>

En l'article s'exposa que el Big Data té limitacions i que s'han de tenir presents, ja que si no obtenim correlacions espúries, prediccions errònies i per què no rendiments negatius.

En general, el text fa referència constantment a les rendibilitats borsàries. Exposa que aquest món de la borsa està ple de soroll, com moltes altres àrees on s'apliquen les grans dades. Fong argumenta que les grans dades haurien d'anar acompanyades d'una teoria sòlida.

L'autor fa una crítica, exposa que les activitats de cerca per Internet només podrien estar lleugerament correlacionades amb els esdeveniments del món real.

- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144.
<https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2014.10.007>

Aquest article pretén definir el Big Data a partir de les seves característiques úniques. Considera que existeixen dues clares definicions del Big Data, una sobre "què és" i unes altres sobre "què fa". Com altres autors, afirma que la grandària de les dades no és més que un criteri del Big Data. El Big Data es pot definir a partir de les 3 Vs. Volum, per la magnitud de les

dades, Varietat, per l'heterogeneïtat de les bases de dades, i Velocitat, per la rapidesa en la qual es generen les dades i per la velocitat a la qual han de ser analitzades i utilitzades.

Argumenta, que algunes de les característiques de les grans dades com l'heterogeneïtat, el soroll i la grandària de les dades requereix el desenvolupament d'algorismes eficients que puguin evitar grans paranys en les dades, com la correlació espúria (molt habitual quan hi ha moltes dades).

- Garden, D. (2012). Introduction. En R. Smolan & J. Erwitte (Ed.), *The Human Face of Big Data* (p. 14-16).

El Big Data és un terme que descriu l'acumulació i l'anàlisi de molta informació. Dan Garden afirma que el Big Data és molt més que "grans" dades, és la capacitat d'extreure significat de totes elles.

En molts àmbits les grans dades són vistes com a beneficis, com és el cas de la medicina. Però no ens podem enganyar, vénen acompanyades de perills, com és el cas de la vigilància massiva.

No som conscients però les empreses i el govern poden saber més de nosaltres mateixos del que realment ens agradaria.

Dan Garden comença la introducció del llibre fent referència a edificis de vidre on se'ns pot vigilar massivament; i conclou la introducció exposant que amb l'anàlisi dels nostres patrons de comportament podrien arribar a saber com actuaríem, i per tant, inclús els nostres caps podrien ser transparents. És una clara pèrdua de llibertat, i, és per això que s'han de definir uns drets.

- Gonzalez-Bailon, S. (2013). Social science in the era of big data. *Policy and Internet*, 5(2), 147-160. <https://doi.org/10.1002/1944-2866.POI328>

L'article té un objectiu doble: per una banda, posa de manifest que no s'ha d'abandonar la teoria, ja que s'han de formular històries creïbles sobre allò que les dades exposen. Per l'altra, permet als investigadors de les ciències socials reexaminar velles preguntes a partir de les noves tecnologies. Els aspectes fins ara comentats han portat nous reptes en l'era del Big Data. Com els investigadors de les ciències socials ja no poden investigar per si sols, és necessari treballar sobre un llenguatge comú, en definitiva, s'ha de fer compatible l'anàlisi analítica amb la teòrica. A més, des del punt de vista de l'autor existeixen dues postures vers el Big Data:

una postura escèptica i una altra entusiasta. Finalment, també posa de manifest les conseqüències de la manera com s'aleatoritzen: la filtració i anàlisi de les dades també poden induir a un biaix i, per tant, és el motiu pel qual l'autor creu que les dades no poden parlar per si mateixes.

- Graham, M. (2012). Big data and the end of the theory. *the guardian*, 1-4. Recuperat de <https://www.theguardian.com/news/datablog/2012/mar/09/big-data-theory>

Graham exposa que la primera persona a afirmar el fi de la teoria va ser Anderson (2008)¹⁹. Ell, al llarg del seu article, argumenta que la teoria ha de ser més que necessària. Exposa que el big data pot donar respostes, però no totes.

Explica que generalment les dades vénen donades de plataformes digitals o xarxes socials. Fa una crítica a les bases de dades, ja que són increïblement esbiaixades.

Finalment, Mark Graham acaba dient que si confiem que el Big Data és un mirall socioeconòmic/polític ens estem equivocant. Per fer-ho bé s'han de segmentar les dades, si no estem creant bretxes digitals i fent generalitzacions que no tenen per què ser certes.

A partir d'aquest article, arribo a la conclusió que és necessari contextualitzar les dades i segmentar-les en petits subconjunts de dades per així no perdre pel camí idees o patrons de la societat.

- Helles, R., & Bruhn-Jensen, K. (2013). Introduction to the special issue «Making data - Big data and beyond». *First Monday*, 18(10), 1-12.

L'article examina com s'elaboren les dades. I arriba a la conclusió que és un procés poc documentat on participen múltiples agents socials. I, tot això s'ha de solucionar, ja que si no podem obtenir informació esbiaixada i podem caure en la discriminació. Per tant, fa èmfasi al fet que és necessari establir unes competències i un diàleg sobre la gestió i l'anàlisi de les dades d'Internet i altres mitjans digitals.

El Big Data, al cap i a la fi, és un recurs més per entendre com les persones es comuniquen i interactuen socialment, però s'ha de tenir en compte que presenta qüestions ètiques, metodològiques i polítiques.

¹⁹ Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Recuperat de <https://www.wired.com/2008/06/pb-theory/>

- Himelfarb, S. (2014). Can Big Data Stop Wars Before They Happen? Recuperat de <http://foreignpolicy.com/2014/04/25/can-big-data-stop-wars-before-they-happen/>

L'autor argumenta que davant de la gran quantitat de dades socials que tenim en els nostres dies podem saber i tenir més accés al qual la gent pensa i sent. Ell creu que a partir d'això ho podem aprofitar per a predir i prevenir esdeveniments com podria ser una guerra. Tot i això, remarca que és molt necessari entendre com analitzar les dades i que ens volen dir aquestes. Destaca que s'han d'ajuntar professionals tecnològics, analítics de grans dades i sociòlegs per mirar de millorar aquest possible benefici. Creu que en un futur podem estar en el camí correcte però que en l'actualitat hem de continuar treballant.

- Hu, M. (2013). Biometric ID Cybersurveillance. *Indiana Law Journal*, 88(4), 1475-1558.

Aquest article fa referència a les dades biomètriques. Per posar-nos en context, la biometria²⁰ és la branca de la biologia que estudia els fenòmens quantitius en els éssers vius.

En l'era de les grans dades, existeix el "perill" d'instal·lar i normalitzar la ciberseguretat en la nostra vida quotidiana, ja que s'estan recopilant i analitzant grans quantitats de dades biomètriques. Tot això pot provocar una vigilància burocràtica.

Els autors exposen que aquestes dades comencen a estar a disposició del govern a partir dels rebuts de les targetes de crèdit, registres de telèfon, etc.

Els autors afirmen que la gran majoria de les persones no són conscients de les conseqüències d'una cibervigilància, podem arribar a perdre les nostres llibertats civils. Per altra banda, els autors afirmen que en determinades disciplines tenen molts beneficis. El que seria ideal és trobar un equilibri entre els beneficis i els riscos.

²⁰ Definició Biometria segons el diccionari de la llengua catalana, <https://www.iec.cat/activitats/entrada.asp> (07/01/18).

- Irazabal, E. (2018). LAS FRONTERAS REGULATORIAS DEL BIG DATA : CRECE EL PODER DE LA UNIÓN. Recuperat de <https://blogs.deusto.es/bigdata/las-fronteras-regulatorias-del-big-data-crece-el-poder-de-la-union-europea/>

En aquest article es posa en relleu les polítiques de les grans dades a la Unió Europea. En l'article se'ns explica que el Reglament que seguim a Europa garanteix majors drets en els usuaris, com per exemple, dret a la limitació del tractament de dades, dret a la portabilitat, augment en la transparència, dret a la informació, major dret d'indemnització i d'interposició de denúncies i, en especial, el dret a l'oblit²¹.

La Unió Europea pretén que els ciutadans tinguin més control sobre les dades. L'autora de l'article exposa que la UE es posiciona com a protectora de les dades i limita la intervenció en la vida de les persones pel poder polític.

- Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and Challenges of Big Data Research. *Big Data Research*, 2(2), 59-64. <https://doi.org/10.1016/j.bdr.2015.01.006>

L'article posa de manifest que el Big Data avui dia està present en totes les disciplines. En aquest article es classifiquen les grans dades en dues categories: les dades del món físic (experiments, ciència, ...) i les dades de la societat humana (xarxes socials, Internet, ...).

S'exposa que en la investigació i aplicació del Big Data, la indústria està per davant de l'acadèmia. Tot i que, l'aparició del Big Data ha donat pas a un nou paradigma d'investigació. Implica un canvi en la manera de fer ciència i en la manera de pensar.

Argumenta que els reptes reals del Big Data se centren al voltant de les seves característiques úniques, les quals defineixen les grans dades, és a dir, la varietat, la velocitat i la veracitat. La complexitat inherent de les dades fa que la seva percepció, representació, comprensió i càlcul siguin molt complicats i provoqui una complexitat computacional.

En aquest article podem trobar els grans reptes de les grans dades però també els grans beneficis que ens poden i ens estan arribant a portar. En un punt de l'article s'argumenta que el Big Data podria arribar a ser el nou petroli del segle XXI.

²¹ El dret a l'oblit es va reconèixer per primera vegada a la Unió Europea el 13 de maig de 2014.

- Jones, S. (2016). China invents the digital totalitarian state. *The Economist*, 1-8. Recuperat de <http://www.economist.com/news/briefing/21711902-worrying-implications-its-social-credit-project-china-invents-digital-totalitarian>

L'article exposa la situació del Big Data a la Xina, un país amb pocs controls sobre el seu poder. La Xina és el primer país on s'està implementat un estat totalitari digital. A la Xina, el govern té accés il·limitat de les dades personals dels seus ciutadans. Com a curiositat, en funció de les dades personals que produeixen els treballadors tenen unes recompenses administratives, com rebre diners, una millor assegurança...

Resulta que són experts en fer llistes. A partir de les grans dades estan fent una llista dels delinqüents del país. Resulta que és una llista bastant extensa, ja que allà apareixen morosos, assassins, etc. Aquestes llistes les fan perquè les institucions, organitzacions i governs realitzant sancions a partir d'elles com no poder comprar bitllets d'avió, no poder comprar o vendre una casa, etc. El problema rau en el fet que avui dia encara existeixen moltes limitacions en les grans dades, per tant, en aquelles llistes poden haver-hi errors i molts cops es cau en la discriminació.

Finalment, aquest article m'ha cridat molt l'atenció perquè mentre ho he pogut relacionar amb una sèrie de Netflix, que es diu Black Mirror. Sincerament, tot i que sóc conscient que hi ha capítols d'aquesta sèrie que poden esdevenir reals en un futur pròxim, em sorprèn que en alguns indrets puguin haver-hi casos "reals".

- Joyanes Aguilar, L. (2014). ¿Qué es el Big Data? En *Big Data: Análisis de grandes volúmenes de datos en organizaciones* (MARCOMBO, p. 22). México: Alfaomega.

L'autor d'aquest llibre pretén donar resposta a què és el Big Data, tot i que afirma que no existeix una única definició de les grans dades. En el seu relat, l'autor destaca diferents definicions, de diverses empreses i organitzacions. Joyanes creu que "existeixen diferents aspectes on gairebé totes les definicions estan d'acord i amb conceptes consistents per capturar l'essència de les grans dades: creixement exponencial de la creació de grans volums de dades, origen o fonts de dades i la necessitat de la seva captura, emmagatzematge i anàlisi per a aconseguir el benefici més gran per a organitzacions i empreses juntament amb les oportunitats que ofereixen i els riscos de la seva no adopció" (2014, p. 2).

Joyanes exposa que les grans dades es poden definir a partir de les seves característiques: Volum, Velocitat, Varietat, Veracitat i Valor, és a dir, mitjançant les 5V.

- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 1-12. <https://doi.org/10.1177/2053951714528481>

L'article exposa com la disponibilitat de grans dades, juntament amb l'anàlisi de les dades, està definint noves estratègies epistemològiques a través de les diferents disciplines. A més, l'article fa una petita referència a la fi de la teoria, argumenta que la ciència està basada en dades més que en el coneixement. Tot això està provocant canvis en moltes disciplines. El Big Data intenta donar sentit a la cultura, la història, l'economia i la societat. En l'article s'exposa que el Big Data està redefinint la investigació i, alhora, ofereix una reflexió crítica sobre la revolució de les dades.

En l'article s'exposen diferents característiques inherents del Big Data. Les grans dades són enormes en volum, d'alta velocitat computacional, exhaustives, de gran abast, d'una naturalesa racional, flexibles i escalables.

Les grans dades permeten respondre preguntes de noves maneres i això està donant un nou sentit a la ciència, s'està definint un nou paradigma. Dintre de la ciència existeixen dues postures. Una empírica on es creu que les dades poden parlar per si mateixes sense la necessitat de la teoria i, una altra que creu en l'abducció, la inducció i la deducció (basat en les dades).

En l'article s'argumenta que les grans dades no substituiran els estudis de petites dades o "small data", ja que aquests són millors per a l'anàlisi de preguntes més concretes.

Finalment, s'exposa que l'àrea de les ciències socials, actualment, té un dèficit de competències per analitzar i donar sentit a les dades. L'objectiu hauria de ser crear una ciència reflexiva impulsada per les dades.

- Kitchin, R. (2014). *The data revolution*. Sage. <https://doi.org/10.1017/CBO9781107415324.004>

El llibre examina el canvi que s'està produint amb la revolució de les dades, gràcies a les Tecnologies de la Informació i la Comunicació (TIC) i la quantitat d'aparells electrònics dels quals disposem avui dia. Alhora, avalua com està afectant els diferents conceptes de negoci,

de mercat, de govern i dels ciutadans des d'un punt de vista crític, analític, ètic, polític i legal.

Aquest llibre toca molts temes relacionats amb el Big Data i tot allò que l'envolta. Resulta d'interès per la meua investigació, ja que no es pot entendre el concepte de "revolució de dades" només a partir del Big Data.

A continuació, faré menció als capítols que m'han semblat més rellevants per a la meua investigació.

- **Cap. 7: "The Governmental and Business Rationale for Big Data" (p. 152-166)**

La revolució de les grans dades és present en totes les disciplines, ja que ofereix una nova forma d'entendre i gestionar tots els aspectes de la vida humana. Tot i això, destaca que vivim en una societat de vigilància massiva i n'hem de ser conscients.

Kitchin argumenta que el que realment ofereixen les grans dades és la possibilitat d'una intel·ligència exhaustiva, detallada i en temps de reals de molts aspectes de la nostra vida. A més, assegura que el motor principal de les grans dades no és tecnològic sinó financer.

En el capítol s'observa com el govern i els negocis i/o organitzacions són uns dels principals generadors i consumidors de dades.

Finalment, conclou el capítol exposant que no ens hem d'oblidar de les conseqüències negatives com pot ser la vigilància massiva, la classificació social, la corrupció, etc. I ho acompanya amb la següent afirmació de Salmon (2014): "*The more field is run by a system, the more that system creates incentives for everyone (employees, customers, competitors) to change their behaviour in perverse ways – providing more of whatever the system is designed to measure and produce, whether that actually creates any value or not.* (p. 165)", que he trobat que era molt encertada.

- **Cap. 10: "Ethical, Political, Social and Legal Concerns" (p. 206-223)**

Una de les conclusions que es pot treure un cop llegit el capítol és que els ciutadans guanyem seguretat a canvi de la nostra privacitat, però Kitchin exposa que això no hauria de ser així, ja que la privacitat és un dret humà bàsic.

L'autor argumenta que la manera que tenim d'entendre la privacitat està canviant tant a escala social com legal. Alhora, critica que la gent no llegeix les polítiques de privacitat, bé

perquè no les entenen, per mandra, etc. Argumenta que vivim en una societat de desconeixement i molts cops ens genera desconfiança.

Per altra banda, afirma que les dades poden arribar a ser molt discriminatòries o ens poden etiquetar en aspectes que podem o no estar d'acord. Com no tenim gaire control sobre les dades aquesta informació ens pot perseguir durant molt de temps. A més, explica que hem de ser conscients que les nostres dades poden ser utilitzades en usos secundaris, dels quals no som gens conscients.

Finalment, exposa que al ritme que estan creixent les dades no es té prou temps per disminuir els problemes de seguretat, que és més aviat el contrari, es multipliquen.

- **Cap. 11: “Making sense of the Data Revolution” (p. 225-232)**

Kitchin comença el capítol plantejant-se qüestions sobre components ontològics, epistemològics, ideològics i metodològics. Es demana què són les dades? Quina és la seva naturalesa? En quin marc estan catalogades les dades?

Afirma que les dades són tant socials com materials, ja que no són independents de les nostres idees, coneixements, persones, etc.

Conclou el capítol de la mateixa manera que el comença. Es planteja moltes qüestions envers el Big Data, possiblement per fer-nos reflexionar i per ser capaços de donar-li sentit després d'haver llegit el llibre.

- Kitchin, R. (2016). Thinking critically about and researching algorithms. *Routledge*, 20(1), 14-29. <https://doi.org/10.1080/1369118X.2016.1154087>

En aquest article Kitchin reflexiona sobre la postura dels algorismes. L'autor creu que els algorismes, la tecnologia i les grans dades estan canviant la manera com treballem, vivim i pensem.

En primer lloc, Kitchin es demana per què estudiem els algorismes i ell explica que ho fem perquè ens permeten analitzar moltíssimes operacions en segons i sense biaixos humans, a més reduïm en costos. Malgrat això, per ell no tot són beneficis. Ens demana si realment som conscients del poder que exerceixen els algorismes sobre nosaltres. Kitchin exposa que els algorismes són construïts a base de prova i error.

En l'article Kitchin presenta un pensament crític en vers els algorismes. Una de les grans crítiques és que no els hem de veure com a veritats absolutes, ja que vénen carregats d'un alt component humà, social i ideològic. Explica que els algorismes han de ser racionals i contextualitzats.

A més a més, s'exposa que els algorismes poden arribar a ser discriminatoris. Argumenta que els algorismes poden moure el món i si no ens som coneixedors de les seves potencialitats i limitacions poden arribar a tenir enormes conseqüències.

Altres grans reptes que presenten els algorismes poden ser que són vistos com a caixes negres, que són molt heterogenis (en el sentit de diferents dades, lleis, estàndards, etc.), han de ser constantment revisats i actualitzats, etc.

- Lesk, M. (2013). Big Data, Big Brother, Big Money. *Security & Privacy, IEEE, 11(4)*, 85-89. <https://doi.org/10.1109/MSP.2013.81>

Lesk exposa que en l'era de les grans dades hem de vigilar en no caure en les mans d'un govern controlador, on puguin controlar tot el que la gent diu o pensa.

Les dades i la recopilació de dades, avui dia, estan en mans d'empreses privades com poden ser Facebook o bé Amazon.

Argumenta que les grans dades ofereixen molts beneficis per a la nostra societat, per això són de gran valor per a les empreses. Però, també recorda que hi ha perill d'obtenir correlacions espúries, crear desigualtats entre la societat, etc. Realment, l'opinió que es pugui tenir sobre el Big Data depèn en gran part de l'ús que en fan.

L'autor afirma que tot i que les grans dades ens ofereixen grans beneficis, cada dia estem més a prop del Big Brother. I un cop arribem a aquest estat, serà molt difícil fugir d'ell, ja que eliminar les dades d'Internet és gairebé impossible.

- Letouzé, E. (2012). Big Data for Development: Challenges & Opportunities. *Global Pulse*, (May), 1-47. Recuperat de <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>

En aquest article, com bé indica el nom d'ell, se'ns presenten els reptes i les oportunitats que ens ofereixen les grans dades. Letouzé afirma que el Big Data permet prendre millors decisions

gràcies a les dades, ens permet analitzar la informació en temps real, ens ofereix una nova forma de fer negoci, ens aporta un avantatge competitiu, etc. Les grans dades ofereixen i garanteixen noves oportunitats en diferents sectors, com pot ser el món de l'economia i les finances, la salut i les ciències socials.

L'autor exposa que en l'actualitat estem vivint en la revolució de les dades o "data deluge", és a dir, vivim envoltats de grans quantitats de dades. Alhora, Letouzé, afirma que la revolució de les dades s'està desenvolupant en tot el món de diferents maneres i a velocitats molt diferents. La qual cosa, ens provoca una bretxa digital enorme.

Pel que fa als reptes de les grans dades. L'autor exposa que una de les principals preocupacions que provoquen repercuteix en la privacitat. Leotuzé creu que la privacitat es pot entendre com el dret de les persones a controlar la divulgació de la informació relacionada amb elles. L'autor diu que una ciutat sense privacitat, seguretat, diversitat, pluralisme, innovació, les nostres llibertats bàsiques estan en perill.

Un altre gran repte de les grans dades ve donat per qui té accés i què fan amb les nostres dades. Finalment, l'autor exposa que altres reptes venen donats per l'anàlisi de les grans dades i el cost computacional.

- Leurs, K., & Shepherd, T. (2017). Datafication & Discrimination. En M. T. Schäfer & K. van Es (Ed.), *The Datafied Society. Studying Culture through Data* (p. 211-232). Amsterdam, The Netherlands: Amsterdam University Press.

Aquest capítol pretén suggerir una intervenció epistemològica des d'una postura crítica i antiopressiva, que busca reintegrar a les persones dintre de la societat, ja que en els estudis de Big Data existeix molta discriminació.

En els estudis de Big Data s'acostuma a fer generalitzacions. El "nosaltres" acostuma a ser indeterminat. Ens ho hauríem de preguntar des del punt de vista del privilegi i del poder.

Les grans dades discriminen intrínsecament als subjectes ja marginats per la societat.

És necessari descobrir les operacions discriminatòries inherents a les dades per a poder parlar d'un "nosaltres".

- Lohr, S. (2012). The Age of Big Data. *The New York Times*, 1-5.
<https://doi.org/10.1126/science.1243089>

En aquest article del New York Times les dades són vistes com un nou actiu econòmic, com ho és la moneda o bé l'or. L'autor veu el Big Data com una oportunitat per entendre el món i prendre decisions. S'ha de tenir present que la gran majoria de les dades que formen el big data són dades no estructurades, és a dir, imatges, vídeos, sons, etc.

Exposa que el Big Data està transformant l'estudi de diferents disciplines.

Alhora, exposa els perills que s'estan presentant en l'era de les grans dades. Tot i els grans beneficis, també hi ha perills. El Big Data es basa en algorismes matemàtics que poden induir a descobriments falsos, biaixos, tripijocs estadístics, etc.

- Lycett, M. (2013). «Datafication»: Making sense of (big) data in a complex world. *European Journal of Information Systems*, 22(4), 381-386.
<https://doi.org/10.1057/ejis.2013.10>

Aquest autor exposa que les grans dades es poden definir a partir de les 3Vs: volum, velocitat i varietat.

Introdueix el concepte de la dadificació. El defineix a través dels conceptes de "desmaterialització", "liquiditat" i "densitat". L'autor entén la desmaterialització com la capacitat de separar un bé/recurs i el seu ús del món físic. La liquiditat com un cop el bé o recurs és desmaterialitzat, la informació pot manipular-se i transferir-se fàcilment, agregant i desagregant dades de manera que abans eren molt costoses o difícils. I, finalment, la densitat és la creació de valor. Tot i que destaca que la densitat no sorgeix únicament de les dades.

L'autor exposa que un món dadificat hauríem de veure les dades matemàticament primer, i després posar-les en context.

- Mai, J.-E. (2016). Big data privacy: The datafication of personal information. *The Information Society*, 32(3), 192-199.
<https://doi.org/10.1080/01972243.2016.1153010>

En l'era de les grans dades existeix un gran repte a l'hora de definir els termes clau que les componen. És per això que s'han de redefinir molts conceptes, un d'ells la privacitat. Ens convé

un canvi de pensament. En el segle XXI és gairebé impossible no generar dades a partir de les nostres activitats quotidianes.

Avui dia, argumenta l'autor, la informació personal és una mercaderia, un negoci, que es compra i es ven. Fins al moment la privacitat anava molt lligada al consentiment.

En aquest article s'exposa la privacitat de la informació i els reptes que ens trobem com pot ser la definició, l'enfocament en peces individuals (és a dir, la privacitat que els individus i/o usuaris poden controlar), la subjectivitat, el fet de poder diferenciar el que és públic d'allò que és privat i, finalment, el seu valor.

La privacitat de la informació emfatitza amb la responsabilitat de l'individu i se centra en la capacitat de controlar la seva informació.

Tots aquests canvis vénen donats per les grans dades i per la dadificació. Tot això està redissenyant moltes àrees de coneixement.

- Markham, A., & Buchanan, E. (2017). Research Ethics in context. Decision making in Digital Research. En M. T. Schäfer & K. van Es (Ed.), *The Datafied Society. Studying Culture through Data* (p. 201-210). Amsterdam, The Netherlands: Amsterdam University Press.

L'objectiu d'aquest capítol és fer arribar a tots els lectors una crítica i/o reflexió sobre l'ètica en el context de la revolució de grans dades.

En el capítol s'exposa que els usuaris desconeixen fins a quin punt les seves dades arriben a ser recollides i utilitzades. Es critica que l'ètica en la revolució de les dades és molt subjectiva, ja que com i quan s'aplica el consentiment?

S'argumenta que com la informació digital en algun moment involucra als éssers humans de manera individual, és molt important tenir definits uns principis. A més, els problemes ètics es poden produir en tots els processos de la investigació.

La conclusió més important que s'ha de treure d'aquest capítol és que els drets individuals, humans, han de ser més importants que els beneficis.

- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: a revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt.

Aquest llibre neix en el context de què Internet ho ha canviat tot i ha aparegut el Big Data... Amb el Big Data es poden quantificar coses que mai abans s'havien fet i/o pensat, aquest concepte rep el nom de dadificació que és diferent de la digitalització, ja que aquest últim només és a escala computacional, mentre que l'altre ho engloba tot.

El Big Data ajuda a respondre el "què" però no el "per què".

En aquesta nova era, s'han produït tres grans canvis: es passa de dades petites a grans dades, s'accepta el desordre en les dades, a causa de l'heterogeneïtat de les dades, i es deixa de banda la causalitat per la correlació.

És important conèixer les vulnerabilitats del Big Data, ja que existeix el perill que aquest es converteixi en el Big Brother. Pel que fa a les dades, podem aprendre com millorar la nostra vida però sempre amb sentit comú. El Big Data reformula la manera com vivim, treballem i pensem. Tot i això, s'ha de tenir present que en un món de Big Data també s'ha de potenciar les capacitats humanes com la creativitat, la intuïció o l'ambició de coneixement, atès que l'enginy humà és la principal font del progrés. Finalment, cal dir que no s'ha de valorar només el poder del Big Data, sinó també les seves limitacions.

Aquest llibre toca molts temes de gran interès per a la meva investigació. Es pot considerar com un text seminal. Alhora, es pot tenir present que és un llibre de 2013, i pot ser que algunes definicions que fins al moment no eren gaire rigoroses, avui dia ho siguin una mica més.

A continuació exposaré els capítols que he considerat més rellevants per a la meva investigació:

- **Cap. I: "Now" (p. 7-41)**

El Big Data està molt present en diferents àrees. Actualment, encara no està definit de manera rigorosa. En el llibre queda definit de la següent manera: "*big data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments, and more (p.20)*". A més, les grans dades estan generant nous reptes en les diferents disciplines.

S'argumenta que el Big Data pretén eliminar la causalitat. El big data no té per què explicar el "per què" passa alguna cosa, sinó que exposa "que" passa alguna cosa.

Finalment, en aquest capítol se'ns introdueix el concepte "*Datafication*", és a dir, afegir un valor quantitatiu a tot, inclús a tot allò que mai ens haguéssim pensat que podia ser mesurable.

- **Cap. II: "*More*" (p. 42-67)**

Les grans dades es caracteritzen per analitzar grans quantitats de dades, per privilegiar el desordre de les dades, en lloc de l'exactitud, i, per la correlació en lloc de la causalitat.

A la revolució de les dades es poden observar canvis a diferents velocitats. És ben sabut que cada dia disposem de més dades i més informació, que les eines han canviat però la metodologia roman igual.

Finalment, convé tenir present que abans el mostratge donava molta informació. Avui dia és millor tenir totes o quasi totes les dades, ja que garanteixen més llibertat per explotar i analitzar les dades.

- **Cap. III: "*Messy*" (p. 68-101)**

Amb el Big Data s'accepta el caos i el desordre, ja que queda compensat per la dimensió de les dades. Aquest caos és difícil d'acceptar pels analistes. Amb el Big Data s'ha d'aprofitar del desordre en lloc de tractar d'eliminar-lo. Netejar una base de dades implica perdre informació. Treballar amb dades completes i desordenades ens ajuda més a apropar-nos a la realitat.

Els models simples amb grans quantitats de dades són, en general, millors que els models elaborats amb menys dades. Tot i que, en molts casos pot ser molt important treballar amb petites dades. Tot depèn de l'objecte d'estudi.

- **Cap. IV: "*Correlation*" (p. 102-146)**

Un objectiu del Big Data és treure valuoses correlacions/associacions sense conèixer les causes subjacents. Saber el "què" i no el "per què". Amb això es pretén abandonar la causalitat (establir causalitats ens provoca biaix, ja que podem arribar a treure conclusions que realment

no existeixen, a partir dels prejudicis dels nostres coneixements i creences). Si tot és causat per alguna cosa, la lògica dicta que no som lliures de decidir res.

La causalitat no és fàcilment demostrada, en canvi les correlacions tenen una matemàtica al darrere.

Tot això no demostra el final de la teoria, ja que el Big Data està fomentat per conceptes estadístics, informàtics, matemàtics, etc.

- **Cap. V: “Datafication” (p. 147-195)**

“Datafication”: Fenomen de posar en format quantitatiu a coses que ningú s’imaginava que poguessin arribar a tenir un valor quantitatiu. Representa un enriquiment essencial en la comprensió humana. Amb l’ajuda del Big Data, ja no es veurà el món com una cadena de successos que expliquem com a fenòmens naturals o socials, sinó com a un univers compost essencialment d’informació.

- **Cap. VI: “Value” (p. 196-244)**

La informació és el que els economistes anomenen un bé “no competitiu”. L’ús que una persona fa d’ella no impedeix l’ús a una altra. La informació no es desgasta per l’ús.

Les dades no són sempre útils i re-utilitzables. El repte és saber quines dades ja no són útils.

Hem d’aprendre de les dades.

“Data exhaust”: Dades esgotades, pot arribar a ser un gran avantatge competitiu per a les empreses.

- **Cap. VII: “Implications” (p. 245-297)**

Avui dia existeixen tres tipus d’empreses diferents de Big Data. Es diferencien pel valor que ofereixen. Poden oferir dades (“*data holder*”), habilitats (“*data specialist*”) o idees (“*Big-data mindset*”). Els dos primers tipus són els que més atenció reben actualment. Tot i que els últims són els que tenen més valor.

En un futur, el que tindrà més valor seran les dades per si mateixes. L’impacte més gran serà que les decisions basades en dades estan preparades per augmentar o invalidar els jutges

humans. Convé tenir present que l'experiència (el coneixement latent) és el que ens permet prendre les decisions més intel·ligents. A més, aprofitar les dades no és sempre garantia d'èxit, però mostra allò que és possible.

- **Cap. VIII: "Risks" (p. 298-338)**

Si Internet amenaçava la privacitat, el Big Data encara ho fa més. És el costat fosc del Big Data, tot i que no és l'únic. Estem sota una vigilància constant.

Els riscos més rellevants són la violació de la privacitat, la tendència a jutjar a partir de les prediccions i caure en una possible dictadura de dades.

En l'article s'exposa que anonimitzar les dades ja no és útil, ja que ara es disposa de moltes dades i és possible arribar a perdre l'anonimat. A més, argumenta que amb el "*profiling*" es poden tenir grans inconvenients, ja que si no s'utilitza bé pot conduir a la discriminació de determinats grups i a la culpa per causalitat, cosa que pretén erradicar el Big Data. Les grans dades es basen en correlacions, per tant, és una eina inadequada per jutjar la causalitat.

L'objectiu de les grans dades és ser menys discriminatori i ser més individualitzat.

S'ha d'anar molt alerta amb el Big Data, ja que pot arribar a amenaçar la nostra moralitat individual i arribar a empresonar-nos en probabilitats. Per tant, encara que les prediccions siguin molt bones ens hem de protegir de la dependència excessiva de les dades.

- **Cap. IX: "Control" (p. 339-365)**

El Big Data està transformant molts aspectes de les nostres vides. El canvi és molt ràpid. És per això que les grans dades plantegen un canvi de paradigma.

En l'article s'exposa que el marc de privacitat del Big Data hauria d'estar menys centrat en el consentiment individual, en el moment de la recopilació, i més centrat a assumir la responsabilitat d'ús. Per tant, es produeix un canvi. Abans era privacitat per consentiment i ara passa a ser privacitat per responsabilitat. És un canvi molt necessari per governar de manera eficaç les dades.

En aquesta nova era, s'ha de redefinir el concepte de justícia, ja que amb el Big Data es pot arribar a jutjar a un individu no per allò que ha fet si no per allò que s'ha predit que faria. S'ha

de continuar jutjant a les persones per la seva responsabilitat individual i el seu comportament real.

Segons l'autor, les grans estratègies per controlar el Big Data són: garantir proteccions de privacitat, fer que tothom participi en el Big Data i garantir la responsabilitat i la transparència dels algorismes.

Finalment, exposa que podem permetre que les grans dades es desenvolupin més enllà de la capacitat humana.

- **Cap. X: "Next" (p. 366-392)**

A partir del Big Data velles creences estan sent qüestionades avui dia. Els autors exposen que els reptes que generen les grans dades poden no tenir respostes fixes.

Encara que el Big Data és beneficiós, amb totes les seves característiques, no s'ha de deixar de banda les petites dades. La tipologia de les dades depèn d'allò que es vol estudiar.

S'ha d'anar molt alerta, ja que podem arribar a ser presoners de les nostres accions.

Finalment, ens podem demanar què passa amb els nostres trets més humans com la creativitat, la intuïció, l'ambició intel·lectual, etc. Doncs, realment, haurien de ser els trets més fomentats, ja que el nostre enginy és la nostra font de progrés.

L'espurna de la invenció es converteix en el que les dades no diuen.

- Menichelli, F. (2015). Review: The Data Revolution. *Surveillance & Society*, 13(2), 319-321. Recuperat de <https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/Kitchin/5458>

Aquest article és una "review" del llibre de Kitchin (2014)²², *The Data Revolution*.

L'article afirma que estem vivint una revolució de dades que està canviant totes les disciplines. Alhora implica canvis en la manera com es produeixen, s'administren, s'emmagatzemen i

²² Kitchin, R. (2014). *The data revolution*. Sage. <https://doi.org/10.1017/CBO9781107415324.004>

s'analitzen les dades. Es fa una crítica de què són les dades i què estan fent en aquest món a partir del llibre de Kitchin.

Es posa de manifest que les dades no són independents de les idees, tècniques, sistemes, persones, etc.

La revolució de les dades, que no únicament és produïda pel Big Data. La revolució de les dades ve donada per les grans dades, les dades obertes, les infraestructures de les dades i, òbviament, per les seves conseqüències. Aquest moviment sembla que únicament és positiu i ofereix beneficis per a tothom però s'ha d'anar alerta perquè alhora és una amenaça per a la privacitat, poden aguditzar les divisions socials, entre altres perills.

- O'Neil, C. (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy* (First edition). New York: Crown.

Aquest és un dels articles més crítics que he llegit fins al moment. Exposa que a part de viure en l'era del Big Data, vivim en l'era de l'algorisme. Les decisions es prenen a partir de models matemàtics que poden ser erronis o no estar regulats. Tot això ens porta al costat obscur del Big Data, com pot ser la discriminació.

L'autor exposa com els prejudicis escrits en algorismes poden arribar a distorsionar la societat i la vida de les persones. Reflexiona que no entén com la gent assumeix que el Big Data és sempre el millor quan realment existeix una gran falta de precisió.

Finalment, O'Neil exposa que s'han d'observar les dades, ja que si únicament utilitzem dades històriques, l'únic que acabarem fent és repetir la història.

Aquest article permet al lector una gran reflexió. Aquesta ha estat la meva: sembla que en l'era de les grans dades els negocis juguen un paper molt important, fins al punt de buscar beneficis en els estudis de grans dades en lloc de veritats objectives de gran interès per a tota la societat.

- Panimalar, A., Shree, V., & Kathrine, V. (2017). The 17 V's Of Big Data. *International Research Journal of Engineering and Technology (IRJET)*, 4(09), 329-333.

En aquest article, els autors exposen que les grans dades estan creixent de manera exponencial, ja que aquestes són creades per tots i per tot, dispositius mòbils, servidors webs, xarxes socials, etc.

Afirmen que les grans dades són una idea i concepte “relativament nou”, i té diverses definicions d’investigadors, organitzacions i individus. Els autors argumenten que Laney (2001) va ser el primer a catalogar les dimensions de les grans dades, a partir de les 3V: Volum, velocitat i varietat. Després, amb el temps van sorgir les 4V, afegint Veracitat. Seguidament van sorgir les 5V, afegint Valor. L’any 2014 sorgeixen les 10 Vs de la mà de Borne (2014)²³, afegint Variabilitat, Validesa, Lloc, Vocabulari i Vaguetat.

A continuació, ells llisten les 14V, tot i que no queda clar si és una proposta seva o es basen en les aportacions d’algú altre. Ells afirmen que tota investigació en grans dades hauria de girar entorn d’aquestes característiques i/o dimensions (14 Vs i 1 C) amb la finalitat de gestionar i utilitzar amb eficàcia i eficiència les grans dades. Aquestes són:

- Volum: Grandària de les dades.
- Velocitat: la rapidesa de les dades.
- Valor: la importància de les dades.
- Varietat: tipus de dades (estructurades, no estructurades, semi-estructurades).
- Veracitat: la qualitat de les dades.
- Validesa: garantir l’autenticació de les dades
- Volatilitat: ens indica durant quant de temps les dades són útils.
- Visualització: la manera com es representen les dades.
- Viralitat: Velocitat de difusió de les dades.
- Viscositat: fa referència a la demora de l’esdeveniment, és a dir, a la diferència de temps entre l’esdeveniment passat i l’esdeveniment que s’està descrivint.
- Variabilitat: Diferenciació entre les dades que són soroll i aquelles que són importants i rellevants.
- Lloc (“Venue”): fa referència a la procedència de les dades, ja que aquestes estan distribuïdes en múltiples plataformes, de diferents sistemes, amb diferents requisits d’accés i format.
- Vocabulari: descriu l’esquema, els models de dades, la semàntica, les ontologies, les taxonomies i altres metadades basades en contingut i context que descriuen l’estructura, la sintaxi, el contingut i la procedència de les dades
- Vaguetat: posa de manifest la confusió sobre el significat de les grans dades.

²³ Borne, K. (2014). Top 10 Big Data Challenges – A Serious Look at 10 Big Data V’s.

- Complexitat: la correlació de les dades. Les dades provenen de diferents fonts i és necessari calcular els canvis, siguin grans o petits, per poder tenir la informació real ràpidament.

Finalment, ells consideren que s'haurien d'afegir 3V més, fent d'aquestes 17, per facilitar una gestió simple i efectiva de les grans dades. Les V que afegeixen són:

- Verbotat: es defineix com la redundància de la informació disponible en diferents fonts.
 - Voluntariat: és un dels trets que es defineix com la plena disponibilitat de grans dades per ser utilitzat d'acord amb el context.
 - Versatilitat: es defineix com la capacitat de les grans dades per ser suficientment flexibles com per ser utilitzats de forma diferent en diversos contextos.
- Power, D. J. (2016). "Big Brother" can watch us. *Journal of Decision Systems*, 25(January), 578-588. <https://doi.org/10.1080/12460125.2016.1187420>

George Orwell va ser el creador del concepte "*Big Brother*", una societat amb un estat totalitari. Avui dia, el concepte *Big Brother* pot esdevenir real, podem arribar a viure en una societat deshumanitzada. Estem vivint en un món digital interconnectat.

La societat d'avui dia està preocupada per la privacitat, per la vigilància massiva i per l'abús governamental de les dades, entre altres coses. En el supòsit que algun dia caiguem en la societat definida pel Big Brother serà molt difícil tornar enrere.

Avui dia, les dades ja difícilment són anònimes i molta gent no n'és conscient.

Com molts altres articles, Power exposa que les grans dades ofereixen grans beneficis a les societats però demana que ens plantejem quin ús i quina anàlisi volem fer del Big Data per tal de poder oferir una diferència positiva en la vida de les persones, ja que és cert que ens pot ajudar a prendre millors decisions i extreure beneficis.

Alhora, afirma que aquest procés ja no farà marxa enrere sinó que s'accelerará de manera molt més ràpida. Per això exposa que haurem de ser nosaltres qui vetllem per la nostra privacitat i la nostra llibertat personal.

Finalment, com a curiositat, m'agradaria remarcar que avui dia ja existeixen assegurances que ofereixen tarifes més baixes als conductors que es posen un dispositiu al cotxe per tal de poder controlar el seu comportament.

- Rhoen, M. (2017). Rear view mirror, crystal ball: Predictions for the future of data protection law based on the history of environmental protection law. *Computer Law & Security Review*, 33(5), 603-617. <https://doi.org/10.1016/j.clsr.2017.05.010>

En aquest article s'exposa que la llei de protecció de dades és una resposta al desenvolupament tecnològic. Rhoen argumenta que en la Unió Europea existeix el Reglament General de Protecció de Dades (GDPR) per fer front als riscos del tractament de dades personals. Critica que la protecció de dades no hauria d'anar únicament associada a la privacitat individual, tot i que afirma que és el més visible. L'autor indica que també s'ha de protegir la privacitat col·lectiva, a gran escala, ja que el fet de tenir una societat lliure i democràtica pot ser igual o més important que la privacitat individual.

L'autor exposa que eliminar les dades és una tasca difícil, ja que les dades queden emmagatzemades en moltes bases de dades, i amb la proliferació de la dadificació cada dia es registren moltes més dades. Aquests factors faciliten la vigilància, la classificació social, etc. i és cada cop més difícil de controlar, per això ha d'haver-hi un redisseny de la legislació.

Alguns investigadors han considerat la dadificació com un problema de contaminació en l'era digital.

- Schulz, G. (2016). The Human Face of Big Data. *Science*, 351(6274), 673-673. <https://doi.org/10.1126/science.aaf3194>

Segons Schulz, l'objectiu del llibre és iniciar una conversa i/o reflexió sobre un conjunt de tecnologies que estan sorgint emergents que tenen el potencial de tenir més impacte en la societat que en Internet.

El Big Data no es tracta d'un fenomen únicament de números, ni a la programació, ni a la ciència. Són moltes més coses. El Big Data podria considerar-se el conjunt d'eines més poderós que la raça humana hagi tingut mai. Ens ofereix una nova forma de veure les coses. Però haurà de superar reptes generalitzats.

L'autor se centra en l'ètica de les grans dades, exposa que no s'ha de permetre que el Big Data esdevingui el nou Big Brother.

Les grans dades no haurien de ser exclusives, haurien de ser per tothom.

- Sep, B., & Collider, L. H. (2013). Tape rescues big data. Recuperat de <https://www.economist.com/blogs/babbage/2013/09/information-storage>

Aquest és un article curiós que explica la necessitat que tenen les dades d'estar emmagatzemades. S'exposen els avantatges de les cintes com a mètode d'emmagatzematge de dades, tot i que es planteja una jerarquia en funció del tipus de dades.

Segons els autors existeixen tres tipus de dades: "hot data" que fan referència a les dades d'ús immediat han d'estar guardades a la memòria flash; "lukewarm data" fan referència a les dades que s'utilitzen de manera freqüent però no immediatament, han de ser guardades a discs durs; i, finalment, les "cold data" són les dades que han de ser emmagatzemades a llarg termini han de ser guardades a les cintes.

Segons diferents estudis, després d'un parell de mesos, en general, el 90% de les dades d'una empresa es refreden, és a dir, esdevenen "cold data".

Realment, no sé si serà un material que utilitzaré per a l'elaboració del meu article però és força curiós.

- Strauß, S. (2015). Datafication and the seductive power of uncertainty -A critical exploration of big data enthusiasm. *Information (Switzerland)*, 6(4), 836-847. <https://doi.org/10.3390/info6040836>

L'autor expressa que les grans dades tenen un poder seductor, ja que permeten millorar la presa de decisions a partir dels algorismes. Gràcies a això, a partir del Big Data es pot millorar el benestar social, es pot aprofitar informació digital per tal de crear nous models de negoci, ens pot ajudar a comprendre esdeveniments passats i saber què està passant, etc. Tot i els beneficis significatius de les grans dades, l'autor expressa que les grans dades també tenen les seves limitacions, com poden ser les correlacions espúries que ens poden portar a predir fal·làcies.

Strauß expressa que una de les principals limitacions del Big Data és la correcta interpretació. La seva analítica predictiva té límits naturals. Les dades per tal que siguin vàlides han d'estar en algun context, perquè si no, l'únic fet vàlid és la seva existència.

- Tene, O., & Polonetsky, J. (2013). Privacy and Big Data: Making Ends Meet. *Stanford Law Review Online*, 66, 25.

El Big Data ha generat grans oportunitats i beneficis en sectors com l'economia, la seguretat i la salut. Però tots aquests beneficis es veuen alterats per les preocupacions de la privacitat i la protecció de dades, la llibertat d'expressió, etc.

El Big Data té un gran repte implícit, que seria aconseguir un equilibri entre els riscos de la privacitat i els grans beneficis que ens ofereix. Realment, hem de pensar que els drets individuals han d'estar per sobre de qualsevol benefici.

L'autor exposa que a partir del big data obtenim molts beneficis però realment no tots són el millor, a vegades simplement són interessants.

A partir de la lectura d'aquest article observo que encara no hi ha una política definida per poder regular les grans dades. Em dóna la sensació que hi ha idees però no estan definides encara de manera rigorosa.

- Turner, D., Schroeck, M., & Shockley, R. (2013). Analytics: The real-world use of big data in financial services How innovative banking and financial markets organizations. *IBM Global Business Services*, 16. Recuperat de https://www-935.ibm.com/services/multimedia/Analytics_The_real_world_use_of_big_data_in_Financial_services_Mai_2013.pdf

L'article es basa en una investigació feta sobre una enquesta a professionals i acadèmics del món de les grans dades per tal de treure unes conclusions si més no, indiquen els autors, objectives.

Una gran majoria va coincidir que gran part de la confusió en el món de les grans dades ja ve donada per la mateixa definició. Argumenten que per tenir una bona definició de Big Data on quedi reflectit exactament el què és han de ser presents les 3 Vs: Volum, Varietat i Velocitat (i a poder ser Veracitat).

Els autors exposen que la gran majoria dels enquestats coincideix en el fet que les grans dades ofereixen solucions a reptes empresarials, transformen el coneixement, les empreses, les organitzacions, inclús la mateixa societat.

Per altra banda, també posa de manifest els problemes que apareixen amb el Big Data com pot ser la seguretat i la privacitat de les dades. Realment, són vistos com els grans reptes d'aquesta nova era. Ja que exposa que aquests reptes estan evolucionant i poden variar moltíssim d'un país a un altre, perquè cada un té les seves pròpies lleis.

Finalment, l'autor conclou que el repte més gran es converteix en l'adquisició de les habilitats necessàries per a fer operatives les grans dades, habilitats tècniques, analítiques, polítiques, etc.

- Uddin, M. F., & Lee, J. (2016). We Are What We Generate - Understanding Ourselves Through Our Data. En D. C.H. (Ed.), *Procedia Computer Science* (Vol. 95, p. 335-344). Elsevier B.V. <https://doi.org/10.1016/j.procs.2016.09.343>

L'article exposa que avui dia, gràcies a Internet i l'era digital, estem produint grans quantitats de dades. Exposa que tots som generadors, consumidors i recol·lectors de dades. Aquestes dades estan sent utilitzades per grans companyies amb l'objectiu de crear anuncis personalitzats i tàctiques de màrqueting, és a dir, les dades s'estan convertint en un gran negoci.

Els autors exposen models i algorismes matemàtics que analitzen les personalitats humanes a partir d'allò que fem. Els autors consideren que els humans són massa analítics, emocionals i esbiaixats, tot això segons el seu punt de vista ens porta a ser lents en intel·ligència i càlcul. L'article es qüestiona diferents reptes com que avui dia s'analitzen grans quantitats de dades però som prou "madurs" en tecnologies i algorismes com perquè no ens portin a l'error? Què passa amb la privacitat de les dades? I què passaria si deixem de produir-ne? També, remarca que s'han de tenir presents les interpretacions manipulables.

- Uricchio, W. (2017). Data, culture and the Ambivalence of Algorithms. En M. T. Schäfer & K. van Es (Ed.), *The Datafied Society. Studying Culture through Data* (p. 125-138). Amsterdam, The Netherlands: Amsterdam University Press.

En aquest capítol es defineix l'algorisme com la formula, procés o conjunt de regles expressades en notació algebraica. L'autor critica que existeix certa ambigüitat amb el terme, ja que la seva definició varia en funció de la disciplina que el defineix.

A vegades, els algorismes poden ser perillosos perquè no sempre demostren veritats absolutes, objectivitats, etc. S'ha de tenir present que en els algorismes es poden trobar prejudicis profundament humans.

Els problemes principals dels algorismes se centren en problemes d'accés i experiència. Per a moltes institucions i/o empreses acostumen a ser secrets molt ben guardats.

Per tant, els algorismes han de ser el més transparent possible perquè si no obtindrem conclusions i/o decisions basades en dades d'una qualitat desconeguda i en algorismes modelats de maneres desconegudes.

- van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance and Society*, 12(2), 197-208.

Aquest article pretén destruir els fonaments ideològics de la dadificació. Segons van Dijck la dadificació es basa en afirmacions ontològiques i epistemològiques problemàtiques.

El concepte dadisme implica convicció. La gent confia informació personal a les xarxes socials i aquestes la poden arribar a compartir amb les agències d'intel·ligència. Tot i que és de gran sabuda, la gent ho fa. El dadisme exposa una creença en l'objectivitat de la quantificació i en el potencial de rastrejar tot tipus de comportament humà i social a través de les dades en línia.

La dadificació hauria d'estar definida de manera més rigorosa. És una oportunitat revolucionària d'investigació per a estudiar la conducta humana.

Tot i això, s'ha de tenir present que les prediccions poden estar subjectes a manipulacions. Els algorismes poden estar calculats mostrant preferències, ideologies, raons comercials, etc. A més, les dades poden ser d'una qualitat desconeguda.

El concepte "*dataveillance*" o vigilància de les dades implica el seguiment continu de les dades amb finalitats predeterminades.

- Wenning, R. (2010). *Privacy - Why on earth should we care? W3C Workshop on Privacy for Advanced Web APIs* (Vol. 12). London. Recuperat de <https://www.w3.org/2010/api-privacy-ws/papers/privacy-ws-37.html>

S'exposa que hem d'entendre què és la privacitat i quin és el seu objectiu, i no ens hem de preocupar únicament d'ella quan tenim algun problema. Alhora, l'autor descriu que la pèrdua de privacitat posa en perill la democràcia, ja que actuem en funció de normes. Es demana si, realment, podríem viure en una societat sense privacitat.

Destaca que per treballar el tema de la privacitat, hem de tenir molt present el tema de la igualtat.

Finalment, conclou que la millor manera de viure la privacitat és que se'ns assegurí certa transparència. Critica que si la gent fos més conscient de la realitat, no divulgaria tant les seves dades. Per tant, els serveis haurien d'informar a l'usuari en tot moment de què faran amb les seves dades.

Un cop llegit l'article deixa la sensació que l'autor vol fer-nos pensar i/o reflexionar sobre qui és el responsable de la privacitat i de la seva gestió.