

## Big Data From Social Media Perspective: A Case Study With Facebook

### Sosyal Medya Perspektifinden Büyük Veri: Facebook Vaka Çalışması

Sibel DİNÇ 

#### Abstract

Thanks to the development of internet science and the ability of computers to bring individuals, institutions, and communities together as a communication tool, social media platforms have gone beyond being used as a communication tool and have become perception management and manipulation tools of institutions. Using digital communication tools, user data was used to identify target users and create strategies. In the current age, the concept of data shows its importance frequently. The Economist's article "The world's most valuable resource is no longer oil, but data" shows that the concept of data has reached a striking point. The process of determining user profiles, behaviours and habits based on user information takes place thanks to cumulatively increasing huge data stacks. These massive stacks of data are called 'Big Data'. For Big Data, social media tools are used to obtain up-to-date, huge data stacks covering users from all walks of life. The role of big data in social media platforms is discussed in the study. The concepts of big data and social media and their development processes, the advantages and disadvantages of big data and data breaches will be discussed, and data breaches will be examined through the Facebook platform. The study aims to create a discussion point about data breaches and the limits of big data.

**Key Words:** Big Data, Digital Media, Data Breach, Facebook, Cambridge Analytica

#### Özet

Sosyal medya platformları internet biliminin gelişmesi ve bilgisayarların bir iletişim aracı olarak kişiler, kurumlar ve toplulukları bir araya getirebilme yetisi sayesinde sadece bir iletişim aracı olarak kullanılmasının ötesine geçerek kurumların algı yönetimi ve manipülasyon araçları haline gelmiştir. Dijital iletişim araçlarından yararlanılarak hedef kullanıcıların belirlenmesi ve stratejilerin oluşturulması için kullanıcı verileri kullanılmıştır. Bulduğumuz çağ itibarıyla veri kavramı her geçen gün önemini göstermektedir. The Economist'in "Dünyanın en değerli kaynağı artık petrol değil, veridir" isimli makalesi de veri kavramının çarpıcı bir noktaya ulaştığını gösteriyor. Kullanıcı bilgilerinden yola çıkarak kullanıcı profil, davranış ve alışkanlıklarının tespit edilmesi süreci ise kümülatif artan devasa veri yığınlarının olduğu bir sonuç ortaya çıkarmaktadır. Bu devasa veri yığınlarına 'Big Data' ismi verilmektedir. Big Data için güncel, her kesimden kullanıcıyı kapsayan ve devasa veri yığınlarını elde etmek için sosyal medya araçları kullanılmaktadır. Çalışmada büyük verinin sosyal medya platformlarındaki rolü tartışılmaktadır. Büyük veri ve sosyal medya kavramları ve bunların gelişim süreçleri, büyük verinin avantaj ve dezavantajları ile veri ihlallerinden bahsedilerek Facebook platformu aracılığıyla veri ihlalleri incelenecektir. Çalışma, veri ihlalleri ve büyük verinin sınırları hakkında bir tartışma noktası oluşturmayı amaçlamaktadır.

**Atıf için (how to cite):** Dinç, S., (2023). Big Data From Social Media Perspective: A Case Study With Facebook Fenerbahçe Üniversitesi Sosyal Bilimler Dergisi 2023;3(1),1-14  
DOI:10.58620/fbujoss.1310830

## 1.Introduction

As computer science develops the concept of data, which takes place in all areas of our lives, has become more visible. The Economist's (2017) article titled "The world's most valuable resource is no longer oil, but data" shows that the concept of data has reached a striking point. However, in this period when the internet and social media platforms entered our lives and gained momentum, the concept of data left its place for Big Data. The importance of big data lies in the huge amount of information it contains and the power that information can provide.

Another point that big data is so important is in the information that is desired to be transferred to the machines by the deep learning method, namely machine learning and artificial intelligence. The study conducted by Harvard Medical School (Prescott, 2016) reported that machine learning diagnoses increased to 99.5%, giving a much better result than human physicians in the latest findings. The most important source of this success in the field of health is the huge amount of information provided by big data. Because, thanks to the big data provided, it is possible to ensure that the machines can make the relevant determinations correctly, that is, the training of the machines.

With big data, analyses are made in many areas with huge data stacks. However, the largest data mass is on social media platforms. These platforms not only contain big data but also contain many codes about society and the individual. As Van Dijck (2014) mentioned, data from social media platforms are facilitators that reveal people's behaviour and mood. Analysing data from social media networks such as Google, Facebook, Twitter, LinkedIn, Foursquare and others, contains a huge mass of information that will shed light on the dynamics of society (Cukier and Mayer-Schoenberger, 2014).

However, despite its many advantages, big data also has disadvantages. The disadvantages of big data must be overcome for big data to have a lot of information about society and the individual through social media. Some of the ethical and legal aspects of collecting data from social media platforms are the most important issue currently discussed.

This article discusses the role of big data in social media platforms. In the second part, big data and social media concepts and their development processes, the advantages and disadvantages of big data, and data breaches will be mentioned. In the third part, the Facebook platform will be examined for data breaches as a case study. In the conclusion part, it will be concluded with a summary section about the issues mentioned. The article aims to create a point of discussion on data breaches and the limits of big data.

## 2.Related Studies in Literature

The concept of big data and its effects on society have an impact in every field. In the related studies of this part, an overview of the concept of big data was presented through the case study of the Cambridge Analytica Scandal with Facebook, one of the social media platforms, and the awareness and importance of the issue were mentioned by addressing the issues of data breach in terms of data privacy and its legal dimension.

Hewage et al. (2018), a comparative analysis was made based on big data techniques obtained from sixteen academic publications examining social media platforms and sites such as Facebook, Twitter, Amazon and Google between 2007 and 2015. It has been determined that the platforms examined in the related study also use big data in addition to their own business models. The study states that unlike other platforms, Facebook and Twitter are platforms that need big data, but they are also necessary to each other because this platform is isolated. While examining the big data requirements of the platforms examined in the study, similar and different aspects of each other are also explained.

In the study conducted by Fuller (2019), researcher draws attention to the concept of "data brokerage", while raising awareness for the acquisition, storage, dissemination and use of personal data for commercial purposes. The related study delves into the subject, bolstering current concerns about big data. With or without the consent of the users, when companies use this data for their own benefit, it also paves the way for the misuse of our personal data. It is known that there are measures and laws taken by the European Union and the United Kingdom on the relevant issue. The most well-known of these is the GDPR (General Data Protection Regulation), which came into force in 2018. However, the fact that these laws are in force on the basis of certain countries does not prevent the transfer of data to other countries' borders and their use in this age when it is easy to transfer data in digital media. Although this situation has come to light with the Facebook scandal, it does not indicate that certain measures have been taken, that the current discussions have ended and that there will be no potential discussions.

According to Panger (2016) focuses on the research design of the 2014 experiment conducted by Facebook by manipulating the content in the news sources and influencing the emotions and thoughts of the users. Based on this case, researcher offers implications for general validity problems, arguing that there are deficiencies in empirical research.

In the study conducted by Bronson (2015), researcher touches on how the concepts of small and big data are defined, and the concept of big data and the data taken for analysis has a structure that allows to identify current problems, identify trends and query data at the decision stage. In the related study, while addressing data systems, approaches to detecting and solving existing problems are included.

In the study by Henriksen (2019), while explaining that microtargeting techniques are a security threat to governmentality, researcher mentions that governments and companies use data obtained with similar techniques. The concept, which is compared with traditional statistical methods in the related study, explains the logic of big data analytics and the threat posed by microtargeting.

According to Metcalf (2016), an argument is made about the importance of the concept of data ethics and stopping the progress towards exemption of data science practitioners from current ethical regulations. The concept of data ethics is examined through existing examples, drawing attention to the variable, potentially harmful and controversial nature of data science research, which is exempt from the ethical framework it will bring in social sciences. Data subjectivity theory is presented for critical data studies in order to prevent the current and potential harms of data science in human-subject research.

The study by Jougleux (2022) mentions that the basis of the Facebook economy is a fundamental legal problem based on the exploitation of user personal data. Although the concept of data privacy is protected by legal agreements, the fact that these agreements are located only within the borders of certain countries does not give confidence that they cannot be taken out of the country. At the same time, it was mentioned in the related study that there are gaps in the relevant legislation for underage users, as another issue that needs to be discussed, causing the accountability of Facebook to be questioned.

### **3. Case Study As A Methodological Approach**

For this study, case study was chosen that is one of the qualitative analyse methods. According to Chmiliar (2012), the case study is a methodological approach that use systematically information and examine about topic of observing. According to Brown as well (2008), the case study is an important method because of gives a deeply view about cases and behaviours. As different from experimental studies the case studies aim to define categories of cases and behaviours to discover a specific case (Hancock & Algozzine, 2006). With Facebook case study chosen for this study is examine of big data concept and its reflections from social media perspective.

## **4. Big Data From Social Media Perspective**

### **4.1. History of Social Media**

The development of social media platforms occurred after the Internet has common. The internet, which was founded at the end of the 20th century, started with basic needs such as research and obtaining information. This Web 1.0 based internet provided only basic information. When the Web 2.0 based internet entered our lives, internet users were able to interact. With the opening of this period, the concept of social media entered our lives and platforms where people can exchange ideas and make friendships began to emerge. We are currently using Web 3.0 based internet. This leads to the

expansion of the services offered by social media platforms and the diversification of what can be done with the internet.

Considering the data provided by Statista (2020), the top 5 most popular social media platforms with active users worldwide in 2020 are respectively: Facebook, YouTube, WhatsApp, Facebook Messenger, and WeChat. 3 applications, including WhatsApp in the top 5, belong to the Facebook company. This shows that Facebook has not left the leadership with its continuous innovations and attacks in the age of social media.

#### **4.2. History of Big Data**

There are many definitions of Big Data. According to Zikopoulos et al. (2012), the term Big Data is a data stack that contains information that cannot be processed or analysed using traditional processes or tools. Microsoft defines big data as "an increasingly used term to describe the process of applying computing power to significantly large and often highly complex information sets" (Microsoft, 2013). In addition, big data is a term that describes the storage and analysis of complex data sets using techniques such as NoSQL, MapReduce and Machine Learning (Ward and Barker, 2013).

For a data set to enter the big data category, it must have some features. While these are 5V which are Volume, Velocity, Variety, Veracity and Value in the bases; Viscosity, Variability, Volatility, Viability and Validity features have been added over time (Gutta, 2020). The data set that provides the most basic features of Volume, Velocity, Variety, Veracity and Value is called big data and the analyses made and the programs used are selected accordingly.

Data sets accepted as big data are analysed with programs that have the capacity to process big data. One of them is Hadoop. It is Hadoop as a computing environment built on a distributed clustered file system designed specifically for very large-scale data processing (Zikopoulos et al., 2012). With Hadoop, data sets are stored as backups on more than one server, and this ensures that the data set is protected from backups in case of any problem or deletion in the data sets. Lack of backups of such large-sized data sets may cause great financial losses.

#### **4.3. Social Media and Big Data Relationship**

As social media platforms varied, they started to form enormous data stacks. At this point, the concept of big data started to use. However, the analysis of this massive data involves some analysis difficulties since the structure of the data is not structured. Zikopoulos et al. (2012) mention that many data sets taken from social media platforms consist of unstructured or semi-structured data. If a large part of the data is unstructured data, it takes a long time to bring this data to the analysis stage.

#### **4.4.How Advantageous is Big Data?**

Considering the advantages of big data, as we know from the science of statistics, the increase in our sample size enables the population to give the closest estimate to the truth. At this point, it gives the most accurate estimation possible with the large data stack it provides. Accurate analyses that can be made with big data will enable us to get real answers to our research questions.

However, large data sets and scattered data can cause prediction difficulties in the analysis process. On the other hand, the large data set may compensate for the clutter in the analysis process compared to the smaller data set (van Dijck, 2014). The size problem, which seems to be a disadvantage, makes the existing data scattered problem insignificant due to the increase in the size of the data set.

Another handicap of big data is time and cost issues. However, when a comparison is made in terms of cost, the reduction of data storage and processing costs and the fact that computer programs for big data analysis give better and faster results every day make it easier to work with big data.

Boyd and Crawford (2012) raised their concerns about the reliability and validity of the data in their study on Big Data. However, according to Mahrt and Scharrow (2013), Big Data can be safely reduced to medium-sized data with the sampling procedure and still give valid and reliable results.

Another disadvantage of big data and the one that is currently difficult to solve is data manipulation. The fact that users may intentionally shape their social media identities or the presence of false information about them also reduces the reliability of studies on this subject (Mazur, 2010). In addition to data manipulation, another issue open to discussion is the issue of data breaches. The data privacy of users is compromised with data obtained without permission.

#### **4.5. Data Breaches and Legal Aspects of Social Media**

Although social media data is mainly used for marketing and political purposes, it can also be used for security. Weerkamp and De Rijke (2012) cite that social media data creates an environment for law enforcement, such as the police and intelligence, to control major planned organizations. Similarly, The National Security Agency (van Dijck, 2014) reported that it prevented at least 50 terrorist attacks with the data collected from social media platforms. From a security perspective, it seems reasonable and logical to monitor data. On the other hand, we cannot be sure that data breaches will not occur, even for security reasons. When it is considered the civilian population, it is obvious that this situation cannot be controlled, and the data received without the knowledge of the person will cause data breaches. It is necessary for terms of data rights that a system that is so open to making subjective and unilateral decisions and that is far from auditing strictly adhere to certain legal rules.

When it comes to data breaches, not all events arise from user permissions. At the same time, companies must show similar sensitivity. Basically, big data research is based on the assumption that users indirectly allow their data to be collected and analysed by publishing them online (Mahrt and Scharrow, 2013). Both users, companies and researchers should pay attention to user permissions to prevent data breaches. As a matter of fact, awareness on this issue has begun to emerge. Providing the information that cookies will be used before entering a site and obtaining permission in this regard is an indication of awareness. At this point, the concept called cookie can be defined as a digital footprint of all the steps the user takes on the site. In addition to providing better service to the user, cookies can also be sold to other companies for marketing research without permission. For this reason, it is necessary to obtain our permission at every step to prevent data breaches.

#### **4.6. Correct Predictions Can Be Made with The Data, But?**

It was mentioned that the accuracy of the predictions made with big data will be high due to the large size of the data set. From a different perspective, considering the possibility of biased data, it is not possible to say that social media data cannot be manipulated, though on a huge scale. Algorithms used to extract data from social media platforms can be a guide for both those who create the algorithm and the users (van Dijck, 2014). Because human decisions and biases contain data and shape the design of algorithms (Benjamin, 2019a). If we look at the journey from the beginning to the end, namely the journey from data to artificial intelligence, it is obvious that the data must be processed objectively to be processed by machines. However, Caliskan et al. (2017) mention that machines can be a tool of discrimination based on gender, race, age or ethnicity. To give an example, as computer scientist Latanya Sweeney noticed on Google records, online search results observed that black names were associated with a much higher rate of arrest records than white names (Benjamin, 2019b). This sample data, which shows that social bias and discrimination will inevitably reflect on machine learning, shows that the correct results are a result of the fact that they contain correct and unbiased information.

Another handicap of the predictions of big data is that when we think of social media data, it only gives information about what users do, not why they do it (Mahrt and Scharrow, 2013). At this point, it is necessary to try to learn the sub meanings of the data. As in other disciplines such as psychology, education, or medicine, individual diagnoses and inferences require much greater precision to identify aggregate trends (Nabi and Oliver, 2009). Apart from the data obtained with this certainty, the meanings of these data should be known, that is, the message given should be read correctly.

#### **4.7. Most Popular Platform**

Computer and subsequent internet technologies have started to be available to the public since the end of the 20th century. The widespread use of the Internet has emerged way for big data, which we call huge data stacks. A substantial part of the data used for the analysis of big data is provided from



social media platforms. Thanks to the social media platforms that have gained velocity in the last decade, the platforms provide marketing opportunities for companies. The best known and most common of these platforms is Facebook. While Facebook had only a few thousand users in the year of its establishment in 2004, by 2020 it became the largest social network in the world with 2.7 billion monthly active users (Wallach, 2020). Looking at the data provided by Vincos (2020) in 2020, Facebook was measured as the leading social networking site in 151 of 167 countries. However, these data still show that China and Russia are not among the countries using Facebook.

There is a huge data stack on Facebook. Considering the 2020 data of 'every minute of the day' study calculated by Stata every year since 2012, Facebook users uploaded 147,000 photos and shared 150,000 messages during the year (Aran, 2020). This data is an example of how large a data set is on social media platforms.

#### **4.8. Are Social Media Platforms Free?**

Companies know very well how to easily get the processing of existing big data from people. Social media platforms, search engines and e-mail services do not charge any fees when they offer a service. Because even though no direct fee is charged, the data are worth gold in our age. With data, we can have a lot of information about people, societies and communities, countries, and so on. Big data studies also serve this purpose knowingly or not. If there is something that needs to be known at this point, it is us who are the products here if something is free (Worstell, 2012 cited in Coté, 2014). For example, we use the Facebook platform free of charge. However, our data is used for marketing or political purposes. Facebook does this by following all our digital footprints. Facebook stores where, when, on which device you are logged in, that is, your location, any information you like or think might be of interest to you, and the information of all applications connected to your account (Curran, 2018). Such power also provides a financial power to Facebook. In the Cambridge Analytica case, Facebook was accused of providing data to unauthorized third parties. Until this scandal, most users were unaware that their data was being used against them. The data we produce are moving on the orders of the capital and the state (Coté, 2014). This indicates that the right to speak of our personal data is derived from us and that we are always monitored.

#### **4.9. The Cambridge Analytica Scandal**

As a result of the years that have passed, Facebook completed mission as a social platform that only allows people to interact but has become a company where companies make a profit, deliver these ads to the target audience by advertising, and even personality profiles are removed from users' information and their political decisions are manipulated with this data. The best example of this situation is the Cambridge Analytica case, in which Facebook was involved.



Facebook's name has been involved in a scandal with its relationships with Cambridge Analytica, a political consulting firm focused on data mining and analytics. Using data from Facebook, Cambridge Analytica was detailed enough to allow users to create profiles with this data and to provide them with pinpoint ads and push them to new behavioural patterns (Clarke, 2020). Cambridge Analytica has collected personal information from users for research purposes. However, it not only used the information of the users who accepted the terms, but also the friends of those users. This data breach, along with the sale of data profits to other organizations, made their actions illegal and highly unethical. This can be summarized as a company that knows us better than we do, selling our data to other companies without our knowledge and approval. These companies could have the power to influence our decisions as well as making choices, apart from doing relatively small marketing efforts with our data.

Brexit and the American elections can be given as examples of concrete actions. In this process, with the support of Aleksandr Kogan, a researcher at Facebook Cambridge University, developed personality tests and transferred the collected data to Cambridge Analytica to use the data for his own interests (Clarke, 2020). "If we target enough persuasive people in the right places, these states will turn red instead of blue" by Kaiser, former business development manager of Cambridge Analytica, and "Blogs on every platform you can imagine until voters vote our candidate, that is, we see the world as we want, we bombarded it through websites, articles and videos" illustrates the extent of manipulation undertaken (Clarke, 2020). Moreover, the truth is, they got the result they wanted in the Brexit and American elections.

#### **4.10. Facebook's Power**

There may be some doubts as to how much the data Facebook obtains will define us. However, a study on this subject shows the essence of the work. Scientists at Cambridge and Stanford Universities analysed the data of more than 17,000 Facebook users who gave permission about their likes, with the help of an algorithm, through a personality survey. At the same time, a questionnaire was applied to the close circles of these users to compare them with the algorithm. The aim is to learn whether the algorithm or its immediate surroundings can define the user better. The results showed that the algorithm identifies a user's personality traits accurately with a person's 10 likes from taken colleagues, 70 likes from friends, 150 likes from family and 300 likes from itself (Quenqua, 2015).

#### **4.11. Technique Used**

Facebook is doing the extraction of all these personality characteristics with the help of big data. It needs a massive data stack to prototype users. At this point, machine learning techniques come into play. The information obtained with big data is taught to the algorithm through machine learning and underlying deep learning techniques. The more data the algorithm can be trained, the better it will reach the target for which it was trained. Once the data is collected, it can be analysed using machine

learning and compared to the browsing behaviour of millions of other users to make predictions with a single user's browsing history and create personalized experiences. This is called "Collaborative Filtering" and is often used for such tasks when creating personalized recommendations on shopping sites. Considering the marketing purpose, this may seem reasonable up to a certain point.

In fact, when the collection of data does not exceed certain limits, Personalized Experience is done correctly, personalization features are generally perceived as useful and pleasant for the user. For example, music recommendations on Spotify, movie recommendations on Netflix, product recommendations in an online store. At this point, however, companies like Facebook do not only take the data but sell their search histories to third parties by making data breaches.

This data collection is done with cookies used to collect consumer click data in personalization and advertising. Cookies record the digital footprint of each user on the site entered. At this point, click data is collected using cookies first, then machine learning is used to predict interests, and finally the purchase intention is determined.

#### **4.12. Legal Pressure**

With the case study given to obtain big data from social media, it is seen that there are still some ethical and legal problems. In terms of Facebook, it is seen that there are controls regarding data violations. Facebook is facing massive antitrust lawsuits filed by federal and state officials in the United States, and other countries are also increasing potential regulations (Wallach, 2020).

There now appears to be a greater awareness of data breaches and the necessary concrete legal sanctions. Facebook the existing pressures have prompted Facebook to take some initiatives. After the events, Facebook recruited thousands of people from its security team, invested in automated tools to detect badware content, and found and suspended accounts that aim to spread propaganda (Benjamin, 2019b). This situation shows that taking an action specific to Facebook is pleasing and there is promising news in terms of data breaches.

On the other hand, the larger the data, the more difficult it may be to prevent digital tracks from going beyond legal frameworks. When data are "large," unknown data points are more easily populated through prediction, imputation, and proxies, and as a result, staying confidential and hiding personal information cannot always be prevented (Williams et al., 2018). Because many different data traces are recorded digitally. This situation does not justify data breaches and the necessary arrangements in this regard should be made by those responsible.

## 5. Conclusion

This essay mentioned on the importance of big data in social media, Facebook review as a case study, and discussed the concept of big data with the dimension of data breaches. Big data has rapidly penetrated every aspect of our lives. In this context, data breaches have been presented to unauthorized third parties through social media platforms and have caused some legal problems.

Considering the disadvantages of big data, some of them are the clutter of the data set, the high cost of data analysis, the demanding of labour and time, the data set is open to manipulation by the user or the analyst, the validity-reliability problems of the analysis results and data breaches. However, some of these disadvantages can be overcome. The Central Limit Theorem, one of the most important theories of Statistical Science, states that increasing the size of the data set will increase the predictive power and decrease the margin of error. Data set disorganization and validity-reliability problems are minimized with a huge data set, and the margin of error is reduced to the lowest possible level. Developing computer programs for the analysis of Big Data sets, on the other hand, reduce the analysis costs and save time and effort. Even if the negative sides of Big Data have been eliminated so far, the fact that the data set is open to manipulation and data breaches is still important as the issues that big data cannot handle. Data processed without full user permissions from social media platforms is one of the most basic examples of data breaches. On the other hand, unawares manipulating computer programs created to analyse data sets by users and programmers is another important matter. How true the user information reflects? The machine learning, we use with the information provided by big data and the deep learning technologies underlying it can similarly cause programmers to manipulate the data without knowing it.

As a result, there are still some deficiencies that need to be overcome by the big data stacks obtained from Big Data and especially from social media platforms. The elimination of these deficiencies will make future studies in this field give more accurate estimates. For now, the most important concept to be considered is data violation. The awareness of the internet and social media users on these issues should be increased. Above all, the biggest responsibility is on social media platforms.

## References

- Aran, Ali. 2020. "Here's What Happens Every Minute on the Internet in 2020." Retrieved December 27, 2020 (<https://www.visualcapitalist.com/every-minute-internet-2020/>).
- Benjamin, Ruha. 2019a. "Assessing Risk, Automating Racism: A Health Care Algorithm Reflects Underlying Racial Bias in Society." *Science* 366(6464):421–22.
- Benjamin, Ruha. 2019b. *Race after Technology: Abolitionist Tools for the New Jim Code*. Medford: Polity Press.
- Boyd, Danah, and Kate Crawford. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information Communication and Society* 15(5):662–79.
- Bronson, N., Lento, T., & Wiener, J. L. 2015. "Open data challenges at Facebook." 2015 IEEE 31st International Conference on Data Engineering. Seoul, Korea (South) : IEEE.
- Brown, A. P. 2008. "A review of the literature on case study research." *Canadian Journal for New Scholars in Education*, 1-13.
- Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. 2017. "Semantics Derived Automatically from Language Corpora Contain Human-like Biases." *Science* 356(6334):183–86.
- Clarke, Laurie. 2020. "How the Cambridge Analytica Scandal Unravelling." Retrieved December 29, 2020 (<https://www.newstatesman.com/science-tech/social-media/2020/10/how-cambridge-analytica-scandal-unravelling>).
- Chmiliar, L. 2012. "Multiple-Case Designs. *Encyclopedia of Case Study Research*" (s. 582-583). içinde Thousand Oaks: SAGE Publications.
- Coté, Mark. 2014. "Data Motility: The Materiality of Big Social Data." *Cultural Studies Review* 20(1):121–49.
- Cukier, Kenneth, and Viktor Mayer-Schoenberger. 2014. *Datafication in Big Data*.
- Curran, Dylan. 2018. "Are You Ready? Here Is All the Data Facebook and Google Have on You." *The Guardian*. Retrieved December 29, 2020 (<https://www.theguardian.com/commentisfree/2018/mar/28/all-the-data-facebook-google-has-on-you-privacy>).
- Fuller, M. 2019. "Article Big data and the Facebook scandal: Issues and responses." *Theology*, 14-21.
- Gutta, Surya. 2020. "Data Science: The 5 V's of Big Data." Retrieved December 28, 2020 (<https://suryagutta.medium.com/the-5-vs-of-big-data-2758bfcc51d>).
- Hancock, R. D., & Algozzine, B. 2006. *Doing case study research*. New York: Teachers College Press.
- Henriksen, E. E. 2019. "Big data, microtargeting, and governmentality in cyber-times. The case of the Facebook-Cambridge Analytica data scandal." Oslo, Norway: University of Oslo.
- Hewage, T. N., Halgamuge, M. N., Syed, A., & Ekici, G. 2018. "Review: Big Data Techniques of Google, Amazon, Facebook and Twitter." *Journal of Communications*, 94-100.

- Jougleux , P. 2022. "Personal Data and Privacy Protection: Facebook and the Big Data Mountain." *Facebook and the (EU) Law: How the Social Network Reshaped the Legal Framework* (s. 13-92). içinde Springer.
- Mahrt, Merja, and Michael Scharkow. 2013. "The Value of Big Data in Digital Media Research." *Journal of Broadcasting and Electronic Media* 57(1):20–33.
- Mazur, Elizabeth. 2010. "Collecting Data from Social Networking Web Sites and Blogs." Pp. 77–90 in *Advanced methods for conducting online behavioral research*. American Psychological Association.
- Metcalf , J., & Crawford, K. 2016. "Where are human subjects in Big Data research? The emerging ethics divide." *Big Data & Society*, 1-14.
- Microsoft. 2013. "The Big Bang: How the Big Data Explosion Is Changing the World - Stories." Retrieved December 27, 2020 (<https://news.microsoft.com/2013/02/11/the-big-bang-how-the-big-data-explosion-is-changing-the-world/>).
- Nabi, Robin, and Mary Oliver. 2009. *The SAGE Handbook of Media Processes and Effects*. Los Angeles; London: SAGE.
- Panger, G. 2016. "Reassessing the Facebook experiment: critical thinking about the validity of Big Data research." *Information, Communication & Society*, 19(8), 1108-1126.
- Prescott, Bonnie. 2016. "Better Together: Artificial Intelligence Approach Improves Accuracy in Breast Cancer Diagnosis." Retrieved December 28, 2020 (<https://hms.harvard.edu/news/better-together>).
- Quenqua, Douglas. 2015. "Facebook Knows You Better Than Anyone Else." Retrieved (<https://www.nytimes.com/2015/01/20/science/facebook-knows-you-better-than-anyone-else.html>).
- Statista. 2020. "Most Popular Social Networks Worldwide as of October 2020, Ranked by Number of Active Users." Retrieved December 28, 2020 (<https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>).
- The Economist. 2017. "The World's Most Valuable Resource Is No Longer Oil, but Data." Retrieved December 28, 2020 (<https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>).
- van Dijck, José. 2014. "Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology." *Surveillance and Society* 12(2):197–208.
- Vincos. 2020. "World Map of Social Networks." Retrieved December 29, 2020 (<https://vincos.it/world-map-of-social-networks/>).
- Wallach, Omri. 2020. "Mapped: Facebook's Path to Social Network Domination (2008-2020)." Retrieved December 29, 2020 (<https://www.visualcapitalist.com/map-facebook-path-social-network-domination/>).
- Ward, Jonathan Stuart, and Adam Barker. 2013. "Undefined By Data: A Survey of Big Data Definitions."
- Weerkamp, W., and M. de Rijke. 2012. "Activity Prediction: A Twitter-Based Exploration" edited by Information and Language Processing Syst. *SIGIR 2012 Workshop on Time-Aware Information Access: #TAIA2012. Accepted Papers*.

Williams, Betsy Anne, Catherine F. Brooks, and Yotam Shmargad. 2018. "How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications." *Journal of Information Policy* 8:78–115.

Zikopoulos, Paul C., Chris Eaton, Dirk deRoos, Thomas Deutsch, and George Lapis. 2012. *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. edited by S. Sit. The McGraw-Hill Companies.