

## Technological Policing: Big data vs real data

Sónia M. A. Morgado

ICPOL – Centro de Investigação do Instituto Superior de Ciências Policiais e Segurança  
Interna; Lisboa

CIEQV – Life Quality Research Center; Portugal

 <https://orcid.org/0000-0002-7036-5174>

Sérgio Felgueiras

ICPOL – Centro de Investigação do Instituto Superior de Ciências Policiais e Segurança  
Interna; Lisboa

CAPP – Centro de Administração e Políticas Públicas; Lisboa

 <https://orcid.org/0000-0001-6885-7439>

DOI: <https://doi.org/10.57776/hkcb-br21>

**Resumo:** Big Data assume um papel preponderante na intervenção nas mais diversas áreas, à qual não é alheia a área policial. O presente artigo teórico procura de abrir caminhos para a compreensão do fenómeno Big Data na intervenção policial a nível micro e macro, enquanto ferramenta inserida na tecnologia policial. Por outro lado, apresentar o conceito agregado à data real, e o seu distanciamento na intervenção policial. A metodologia utilizada para o efeito foi a de análise bibliográfica, em que se apresentam critérios de inclusão e exclusão, por forma, a consolidar a análise em questão. Na perspetiva dos autores salienta-se que enquanto instrumento da tecnologia policial, Big Data providencia a informação necessária, em termos de volume e quantidade, tendo em conta, os diferentes níveis de intervenção policial, e é um recurso que incrementa a sua eficiência na manutenção da ordem pública.

**Palavras-Chave:** Big Data; Legitimidade; Polícia; Real Data; TechPol

**Abstract:** Big Data assumes a preponderant role in the intervention in the most diverse areas of intervention, to which the police area is no stranger. This theoretical article seeks to open ways to understand the Big Data phenomenon in police intervention at micro and macro levels, as a tool inserted in police technology. On the other hand, to present the concept aggregated to real data, and its distancing in police intervention. The methodology used for this purpose was literature analysis, in which inclusion and exclusion criteria are presented to consolidate the analysis in question. From the authors' perspective, it is pointed out that as an instrument of police technology, Big Data provides the necessary information, in terms of volume and quantity, taking into account the different levels of police intervention, and is a resource that increases efficiency in the maintenance of public order.

**Keywords:** Big Data, Law enforcement, Legitimacy, Real Data, TechPol

**Resumen:** El Big Data assume un papel preponderante en la intervención en los más diversos ámbitos de intervención, al que no es ajeno el área policial. Este artículo teórico pretende abrir caminos para entender el fenómeno del Big Data en la intervención policial a nivel micro y macro, como herramienta inserta en la tecnología policial. Por otro lado, presentar el concepto agregado a los datos reales, y su distanciamiento en la intervención policial. La metodología utilizada para ello fue el análisis bibliográfico, en el que se presentan los criterios de inclusión y exclusión para

consolidar el análisis en cuestión. Desde la perspectiva de los autores, se señala que como instrumento de la tecnología policial, el Big Data proporciona la información necesaria, en volumen y cantidad, teniendo en cuenta los diferentes niveles de intervención policial, y es un recurso que aumenta la eficiencia en el mantenimiento del orden público.

**Palabras-Clave:** Big Data, Legitimidad; Policía, Real Data, TechPol

Submetido em: 17/01/2022

Aceite em: 3/04/2022

## **I. Introduction**

A dataveillance society is prevailing in the contemporary world. Thus, Big Data is the way of having normative categorization of behaviours. This means knowability, legibility, and the ability to reveal the utmost data in policing using predictive models.

The analytic system unveils the prediction by transforming data into behavioural patterns, profiles, and predictive policing (Morgado & Felgueiras, 2021, p. 218).

Globalization, as an integration process (Marx, 1977, Mackinder, 1943, Saint-Simon, 1814), inevitable (Friedman, 2007, Stiglitz, 2003) and irreversible process (Morgado, 2013), and the concerns around the subject “unfolds a set of thoughts that expresses the need to be supported and up-hold the decision-making process” (Morgado & Felgueiras, 2021, p. 217). Alongside globalization, which allows opening all gates, overcoming barriers and exploit the full potential of economic (Morgado et al., 2020), personal, and digital efficiency, Big Data plays an evident role in this process. The central core for Big Data is the result of having a myriad of elements at the disposal of companies, organisations, and institutions.

A literature review concerning the concept was executed to understand the subject comprehensively. The literature search considered the most widespread and complete databases, such as, google scholar, EBSCO, and Scopus. The inclusion criteria for the analysis were the following: i) complete articles published in journals with peer review; ii) complete articles published in Proceedings with peer review; iii) documents which were subjected to peer evaluation (for example, book chapters); iv) conference papers with peer review; v) written in English, or French; and, vi) title contained the terms Big Data, law enforcement, twenty first century, globalization, policing, artificial intelligence, surveillance, reorganisation, information society, peace, privacy, data. To be able to maintain scientific integrity, some exclusion criteria were considered: i) documents only presented the abstract in English or French, but the paper was not; ii) after the reading, the article or document added no value for the analysis or was not related to the prime topic.; iii) documents that presented no quality or deterred a poor understanding of the concept; iv) working papers, master thesis, doctoral dissertations, posters, abstracts and posters.

This article is a theoretical study, and the structure obeys an introduction, state of the art, perspectives (guidelines) and conclusion (practical or theoretical implications).

## II. State of the Art

In the academic landscape, it is difficult to find consensual proposals for the definition of the term Big Data. Mauro et al. (2016a, p. 122) state that "in fact, the term 'Big Data' itself has been used with several and inconsistent acceptations and lacks a formal definition". The difficulty sometimes to find definitions that can, on the one hand, limit a particular reality, and on the other hand, differentiate or confer an identity to that reality, often does not appear as a task of easy execution. For this reason, many scholars present partial proposals of the Big Data phenomenon, depending on approaches that are sometimes either distorted or emphasise certain features.

The multiplicity of proposals has led "the absence of a consensual definition of Big Data often brought scholars to adopt 'implicit' definitions through anecdotes, success stories, characteristics, technological features, trends or its impact on society, firms and business processes" (Mauro et al., 2016a, p. 126).

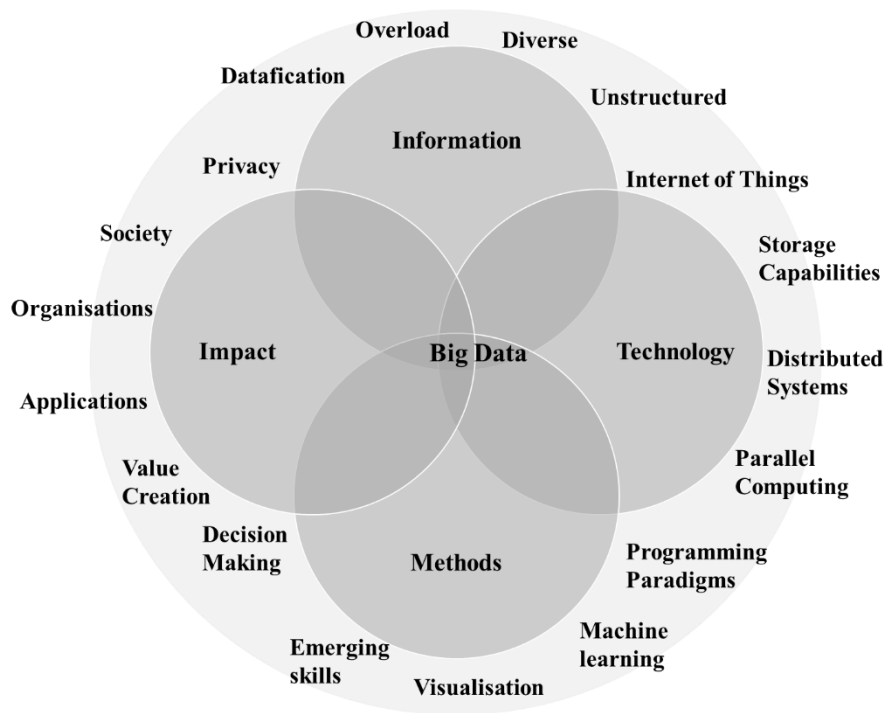
These different approaches have undoubtedly contributed to conceptual diversity, causing considerable anarchy in the state of the art (Mauro et al., 2016a, 2016b).

In fact, in a work developed by Mauro et al. (2016a) the analysis of the different proposals of definition of Big Data allowed the identification of four analysers that correspond to the focus placed on the description of the phenomenon. This analysis facilitates the understanding of the phenomenon and its dimensions, thus allowing introducing greater rationality in the construction and discussion of the Big Data concept. The four analysing categories proposed by Mauro et al. (2016) are the attributes of the data, the technological needs, exceed the thresholds and the social impact. In the study presented by Mauro et al. (2016) the propositions presented in each category were examined to verify which themes associated with Big Data (the information, the technology, the method, and the impact) were present in the concept proposal.

Figure 1 schematizes the significant themes and concepts associated with Big Data to facilitate understanding. No concept proposal presented all the Big Data themes. However, seven proposals presented three themes, five definitions identified two themes and, finally, three concepts admitted one theme.

## Figure 1

*Big Data themes and related topics in existing literature*



*Note:* This figure demonstrates the themes associated to Big Data provided by literatures. Adapted from “A formal definition of Big Data based on its essential features” by A. de Mauro, M. Greco, & M. Grimaldi, *Library Review*, 65(3), p. 126 (<https://doi.org/10.1108/LR-06-2015-0061>). Copyright 2016 by the Emerald Insight.

Mauro et al. (2016a) allowed the presentation of a concept that transversely includes all topics related to Big Data. "Big Data is the Information asset characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value" (p. 127). This definition is sufficiently comprehensive and allows rebalancing the factors implicit in the Big Data phenomenon.

This asset of information is also mentioned by Ferguson (2017, p.8), because he considers that ““Big Data” is a short-hand term for the collection and analysis of large data sets with the goal to reveal hidden patterns or insights”. Large data sets are also presented as the common element for the concept presented by Maras and Wandt (2019). For these authors “Big data is a term used to describe extremely large data sets that can be analysed to reveal patterns, trends and associations, especially relating to human behaviour and interactions” (p. 161).

The result of the changes occurred in the interaction of people with digital and online technologies, urged the need of systematic advances in computer sciences in treating and working and subsequently analysis of “huge troves of digital is commonly referred to as Big Data” (Mc Daniel & Pease, 2021, p. 29). Is this analysis of the mass digital data, data creates value for every organization or institution.

The value of Big Data is added with the element of “modality of surveillance”. The widespread of adoption of big data analytics as a “means to increase efficiency, improve prediction and reduce bias”, is recurrent in an eclectic array of organisations on all sectors, from finance, healthcare, education to law enforcement (Christin, 2016). In fact, “the turn to Big Data brought about a feverish desire to track, capture, compute, and analyze endless customer data points to gain “insights” about customer preferences and behaviors” (Bridges, 2021, p. 4).

It is accepted that the tool Big Data analytics enables augmentations of pre-existent surveillance methods and consequently transformations in surveillance activities (Braynes, 2017).

Converging in the fundamentals of Maslow hierarchy of needs, as stated by Morgado and Felgueiras (2021), the beginning of the process passes by the data, that is the physiological need of Big Data, and the ending is of the self-actualization of the pyramid is the value.

Because the lack of consensus, Big Data is still an ambiguous term varying across fields and institutional frameworks (Brayne, 2017). However, the ambivalent construct, the Big Data for the purpose of this paper, is the one purposed by Brayne (2017) Big Data “is a data environment characterized by four features: it is vast, fast, disparate, and digital” (p. 980).

### **III. Perspectives/Guidelines**

The potentiality of Big Data is embraced by the complexity of the dynamics revolving the concept of Big Data. In fact, the main elements of Information, Impact, Technology and Methods, develop the need and the overachieving concepts of Internet of Things, Machine Learning, M2M, and wrapped by Artificial Intelligence (AI).

The algorithms are the main element for determining and processing data to provide quality with the overload, diverse and unstructured level of data on the Internet of Things. Because the technology allowed the increase of storage and the rapid distribution of information, the machine learning systems powered by the set of instructions were developed by the programming paradigm. Humans designed the coding necessary for enabling computer systems to translate data into conclusions, information, or outputs (Kaye, 2018). The contradiction of this process lies over the biases that humans can consciously or unconsciously put in the analysis because of their background, education, values, socioeconomic class, and other elements (Noble, 2018).

It is a fact that “machine-driven decision-making is everywhere. Decisions based on machine learning algorithms are supplementing or replacing human decision-making in vastly different aspects of society, including consumer finance, employment, housing, healthcare, and sentencing, among others” (Selbst, 2018, p. 113). The technological process for police decision-making should be the subject of a deep reflection considering the ethical and deontological dimension associated with the police function, the protection of citizens’ fundamental rights, the transparency of data processing, privacy and the quality of policing. The scrutiny on the use of technologies must be implemented to minimize the negative impacts that derive from the use of Big Data in policing. “Regarding predictive policing specifically, society lacks basic knowledge and transparency about both the technology’s efficacy and its effects on vulnerable populations” (Selbst, 2018, p. 118).

We are well aware that the use of new technological tools is a certainty for developing the current policing models. “Though predictive methods such as crime mapping and offender profiling are not new, predictive policing is something different, a creature of the world of Big Data” (Selbst, 2018, p. 113). The use of Big Data for police purposes is a contemporary reality. As Selbst argues, “one particularly important area of rapid adoption is predictive policing, a popular and growing method for police departments to prevent or solve crimes” (Selbst, 2018, p. 113). The potential technological transformations to be introduced in the policing models do not mean that there will be an increase in effectiveness, efficiency and quality in the way people are watched over and ensure their safety. The emergence of new surveillance practices raises questions about the legitimacy of the new digital panopticon. Undoubtedly, the use of Big Data and associated technology opens new possibilities for the operationalization of mass

surveillance. As Brayne states (2017, p. 977), “in the past decade, two major structural developments intersected: the proliferation of surveillance in everyday life and the rise of “big data””, namely, with the massive deployment of sensors throughout our cities and the ability to collect, analyze and use knowledge produced by Big Data tools that have allowed the design of police surveillance activities based on the Big Data system. However helpful Big Data is to police surveillance raises the question “whether and how the adoption of big data analytics transforms police surveillance practices” (Brayne, 2017, p. 978).

The massification of the use of Big Data in various sectors of activity is related “to increase efficiency, improve prediction, and reduce bias” (Christin, 2016, cit in Brayne, 2017, p. 979). Despite this aspiration, some problems emerge from the extensive use of Big Data in decision-making processes. For many, the use of Big Data in policing represents the miracle solution to solve significant security problems; for others, Big Data is not the Holy Grail of policing because it raises a set of ethical, deontological, fundamental rights protection, privacy, legitimacy, public recognition and, most likely, the systematic reproduction of biases. These are some of the issues raised by the emergence of new surveillance practices.

Suppose the preparation of decisions is based on the history of the information collected. In that case, we may be reproducing past errors and biases, as Barocas and Selbst (2016, p. 674) state, “data mining can reproduce existing patterns of discrimination, inherit the prejudice of prior decision-makers, or simply reflect the widespread biases that persist in society”. On the other hand, police discretion raises some problems in how police officers may approach a suspect or a criminal; police officers “make many choices every day regarding who is a suspect and who appears to be a criminal” (Selbst, 2018, p. 119).

Even though of a risky biased analysis, the visualization and the emerging skills of the method provided by the machine learning, and consequently by the Artificial Intelligence (AI), results in what its “called algorithmic decision making” (Ashraf, 2020, p. 165).

The main concern in what kind of information is necessary for the different levels of command in law enforcement - micro and macro level – is also an element of ponderation for decision makers. The elementary question is: Does the Police officer in operation level has the need to perform his duty to access to Big Data?



The decision-making process, and so-called algorithmic decision making, must be endorsed of coding for operational level, to provide the settings and the knowledge that can improve a better understanding of the situation in hand. While the macro or strategic level decision-makers are meant to work and decide, having Big Data, that gives them a holistic positioning of the problem. While the macro or strategic level decision-makers are meant to work and decide, having Big Data, that gives them a holistic positioning of the problem.

The impact of AI is deciding what are your preferences based on the choices that the user makes (e.g., Spotify suggestions of playlist), managing powers self-driving cars, and above all assisting law enforcement in deciding where and when to patrol and helps pilots fly (Raso et al., 2018), as envisioned in some of futuristic universes of cinematography (Minority Report; AI; Matrix).

The impact of Big Data, and AI intervention in society is yet to be measured, namely in the human rights besides freedom of expression and privacy (Raso et al. 2018).

Even though the widespread application, law enforcement, medical, sports, education, communication, and easy acceptance by the users, there are some concerns about the impact on the freedom of expression and privacy. The question is when AI is too much for the privacy of the human, and its speech, when the algorithms are created to be more moderate, exclude comments that do not convey to the instated.

The organization of society with AI is core element of surveillance in every kind of regime. To keep the mainframe of positive embedment, the word smart, smart cities, smart policing (Feldstein, 2019), is used for the approach to the citizens and to the users of technologies.

Keeping and strengthening security of countries, banks, roads, people, the development of facial recognition, digital passports, closed-circuit television, crime analysis, normalises the AI-based surveillance, has an unintended effect (Ashraf, 2020). Even though, AI is used to identify potential terrorists' threats, gather open-source intelligence, enhance medical research and academic investigation, not always Big Data.

Not all data can be treated as measurable and legible as such. Therefore, the algorithm of prediction can be flowable because deconstruction depends on the biases of the coders. On the other hand, people are not techno-socio-economic objects that can provide certainty about their behaviours.

The rise of AI has been unprecedented, widespread on the ground (Brayne & Christin, 2020). Predictive software targeting victims and offenders and predicting the next crime will occur is recurrent (Brayne 2017; Ferguson 2017). In fact, in law enforcement, Big Data is used “for both person-based and place-based predictive identification, in addition to risk management, crime analysis, and investigations” (Brayne & Christin, 2020, p. 3). Predictive models in policing and courts are justified as objective and rational, avoiding discretionary judgments.

Through the digital crime mapping, it's possible to identify in real-time hotspots, create pattern-matching results from massive police databases "filled with hitherto unimaginable volumes of data; routine form-filling and data entry processes are being assisted by automated predictive text software; and facial and vehicle recognition cameras are locating suspects as they move" (McDaniel & Peasy, 2021, p. 2).

Nonetheless, managerial surveillance or this dataveillance society is controversial and can be affected by data obfuscation.

Authors such as Bridges (2021) consider that in this world of dataveillance, the algorithmic that allows the prediction of behaviours is isentropic, fugitive and queer. These three characteristics may carry problems in terms of legitimacy and data protection. It has been proven that the concept is blemished prone to jaundiced cyphering that can lead to political abuse, preservation of some interests by social engineering and over-surveillance the marginalised (Bridges, 2021, Sanders & Sheptycki 2017). For this reason, the notion of being rigorous, systematized, mathematical and neutral logic is apparent (Sanders & Sheptycki 2017).

Arguments about the counter effects of surveillance are pointed out Bernal (2016): i) its harmful; ii) damages security and, iii) impacts on human rights (right to respect for private and family life; freedom of thought, conscience and religion; freedom of expression; the right to a fair trial; freedom of assembly and association; prohibition of discrimination). Sanders and Sheptycki (2017) also present this argumentation, stating that few can manage the algorithms, creating mismanagement.

However, more recent studies (Leiendecker, 2021) highlight that legitimacy of law enforcement is not compromised for the potential loss described by the authors (Bridges, 2021; Bernal, 2016).

Hopefully, the guidelines provided in this analysis will allow a sensible, science-based approach to the discussion/conclusion.

#### **IV. Discussion/Conclusion**

It is common sense that predictive analytics, artificial intelligence, and automated decision systems are uprising and the leading public governance technology.

Adjusting society to this unrelenting and irreversible process is necessary because a dataveillance or perpetual surveillance will prevail, cutting out some of them until now well-grounded fundamentals of society described in the human rights.

Big Data will be consistent with real data, making it more manageable to intervene and to provide insights of satisficing data to micro and macro level in the decision-making process.

Taking advantage of the "uberization" of security control in a panoptic society, intelligence-led policing is the leverage to crime-fighting and increasing the perception of security of society.

The more effective surveillance is based on an extensive and profound analysis of information that allows deterring or eliminating some problems because it can detect anomalies, irregularities and prevents potential criminal behaviour. It's a way of being more agile and disruptive to face criminal activity.

The overwhelming speed of the use of Big Data in the daily life of companies, government's institutions is a process that is growing because it has achieved its maturity. For this reason, the need to reconcilable Big Data to overcome the problems concerning discrimination, segregation, and privacy is the element to dismantle the negative externalities presented by the massive use of this instrument, without causing any disturbance in its core objective that is to prevent, deter, thwart, and mitigate crime, in an efficient way. Other drawbacks may be considered, such as ethical and deontological issues, and does not convey with the protection of fundamental human rights. The core of Big Data is the storage of information that is not only from individuals with disruptive behaviour but from every citizen, whether he/she is a felon, potentially one or simply a bystander, that obeys the law.

Alongside the bias on which the algorithm is based, there is one more important question. That question concerns how the information is processed when it enters the blue box of Big Data. Issues like unstructured Big Data or the historical decision-making process supported in the analytics of predictive systems can result in misleading decisions, even though Big Data is actionable intelligence.

Nonetheless, the fundamental issue is that Big Data helps predict and monitor crime locations, criminal identification and allows cases closure. This is a result of improving the effectiveness of police activity because they have access to a wider display of integrated, actionable intelligence that helps to solve crimes.

### References

- Ashraf, C. (2020). Artificial intelligence and the rights to assembly and association. *Journal of Cyber Policy*, 5(2), 163-179. <https://doi.org/10.1080/23738871.2020.1778760>
- Bernal, P. (2016) Data gathering, surveillance and human rights: recasting the debate. *Journal of Cyber Policy*, 1(2), 243-264. <https://doi.org/10.1080/23738871.2016.1228990>
- Barocas, S., & Selbst, A. (2016). Big Data's disparate impact, *Californian Law Review*, 104(3), 671-732. <http://dx.doi.org/10.15779/Z38BG31>
- Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review*, 82(5), 977-1008. <https://doi.org/10.1177/0003122417725865>
- Brayne, S., & Christin, A. (2020). Technologies of crime prediction: The reception of algorithms in policing and criminal courts. *Social Problems*, 0, 1-17. <https://doi.org/10.1093/socpro/spaa004>
- Bridges, L. E. (2021). Digital failure: Unbecoming the “good” data subject through entropic, fugitive, and queer data. *Big Data & Society*, 8(1), 2053951720977882. <https://doi.org/10.1177/2053951720977882>
- Bridges, L. E. (2021). Digital failure: Unbecoming the “good” data subject through entropic, fugitive, and queer data. *Big Data & Society*, (ifirst), 1-17. <https://doi.org/10.1177/2053951720977882>
- Christin, A. (2016). From daguerreotypes to algorithms: Machines, expertise, and three forms of objectivity. *ACM Computers & Society*, 46(1), 27–32

- Christin, D. (2016). Privacy in mobile participatory sensing: Current trends and future challenges. *Journal of Systems and Software*, 116, 57-68. <https://doi.org/10.1016/j.jss.2015.03.067>
- De Mauro, A., Greco, M., & Grimaldi, M. (2016a). A formal definition of Big Data based on its essential features. *Library Review*, 65(3), 122-135. <https://doi.org/10.1108/LR-06-2015-0061>
- De Mauro, A., Greco, M., & Grimaldi, M. (2016b). Beyond data scientists: A review of Big Data skills and job families. *Proceedings of 11th International Forum of Knowledge Asset Dynamics: FKAD 2016. Towards a new architecture of knowledge: Big Data, culture and creativity* (pp.1844-1857). IFKAD.
- Ferguson, A. G. (2017). *The rise of big data policing*. New York University Press.
- Friedman, T. L. (2007). *The world is flat 3.0: A brief history of the Twenty-first century* Picador.
- Kaye, D. (2018). *Promotion and protection of the right to freedom of opinion and expression*. A/ 73/348. United Nations.
- Mackinder, H. J. (1943). The round world and the winning of the peace. *Foreign Affairs*, 21(4), 595-605.
- Maras, M. H., & Wandt, A. S. (2019). Enabling mass surveillance: data aggregation in the age of big data and the Internet of Things. *Journal of Cyber Policy*, 4(2), 160-177. <https://doi.org/10.1080/23738871.2019.1590437>
- Marx, K. (1977). *Capital: A critique of political economy*. Random House.
- McDaniel, J. L. M., & Pease, K. G. (2021). Introduction. In J. L. M. McDaniel & K. G. Pease (Eds.). *Predictive Policing and Artificial Intelligence* (pp. 1-38). Routledge-Taylor & Francis Group.
- McDaniel, J. L., & Pease, K. (2021). Policing, AI and choice architecture. In J. L. McDaniel & K. Pease (Eds.) *Predictive Policing and Artificial Intelligence* (1st ed.) (pp. 79-110). Routledge.
- Morgado S.M.A., Felgueiras S. (2021) Big Data in Policing: Profiling, Patterns, and Out of the Box Thinking. In Á. Rocha, H. Adeli, G. Dzemyda, F. Moreira, A. M. Ramalho Correia (Eds), *Trends and Applications in Information Systems and Technologies. WorldCIST 2021. Advances in Intelligent Systems and Computing*, vol 1365 (pp. 217-226) Springer, Cham. [https://doi.org/10.1007/978-3-030-72657-7\\_21](https://doi.org/10.1007/978-3-030-72657-7_21)

- Morgado, S. (2013). Going global: health organisations and networking – Information society and social media. In M. Mokryš, Š. Badura, & A. Lieskovský (Eds). *Proceedings in Scientific Conference 2013* (pp-47-51). EDIS – Publishing Institution of the University of Silina, Slovakia.
- Noble, S. (2018). *Algorithms of oppression: How search engines reinforce racism*. (1st ed.). NYU Press.
- Raso, F. A., Hilligoss, H., Krishnamurthy, V., Bavitz, C., & Kim, L. (2018). *Artificial intelligence & human rights: Opportunities & risks*. Berkman Klein Center for Internet & Society at Harvard University.
- Saint-Simon, C-H. d. R. (1814). *De la reórganizati3n de la soci3té europ3enne*. Delaunay Libraire.
- Selbst, A. (2018). *Disparate impact in Big Data policing*, *Georgia Law Review*, 52, 109-195.
- Stiglitz, J. E. (2003). *Globalization and its discontents*. W. W. Norton & Company, Inc..
- Leiendecker, L. J. (2021). *Big Data policing and its legitimacy*. University of twente.