ERA Forum (2022) 23:209–219
https://doi.org/10.1007/s12027-022-00719-4

ARTICLE

Quality of data sets that feed AI and big data applications for law enforcement

Martyna Kusak[1,2]

Accepted: 26 September 2022 / Published online: 17 October 2022

Abstract
In the era of big data and artificial intelligence (AI), where aggregated data is used to learn about patterns and for decision-making, quality of input data seems to be of paramount importance. Poor data quality may lead not only to wrong outcomes, which will simply render the application useless, but more importantly to fundamental rights breaches and undermined trust in the public authorities using such applications. In law enforcement as in other sectors the question of how to ensure that data used for the development of big data and AI applications meet quality standards remains. This paper provides an overview of this topic, reporting selected issues stemming from big data, nonpersonal data and regulatory contexts. It concludes that the topic is still underexplored and sets areas for further research.

Keywords Law enforcement · Big data · Artificial intelligence · Data quality · Data protection · Directive 2016/680

1 Introduction

Law enforcement is increasingly being shaped by data analysis to support decision-making, detect criminals, analyze crime patterns, optimize resource deployment, prevent crime and even predict the time and place of future criminal events.[1] The use of such tools seems highly rational since law enforcement is a data-based activity involv-

[1] *Jansen* [21].

✉ M. Kusak
martyna.kusak@amu.edu.pl

1 Adam Mickiewicz University in Poznań, Poznań, Poland

2 Institute for International Research on Criminal Policy, Ghent University, Ghent, Belgium

ing the collection of large data sets (including personal data) relating to crime prevention, detection, and investigation. This data, processed by data mining techniques, can considerably modernize and make law enforcement authorities (LEAs) more efficient, correctly classifying objects that they have never seen with an accuracy exceeding that of humans and, thus, contribute to boosting the effectiveness of security.

While benefiting law enforcement, these new technologies have also increased the complexity of crime analyses, which require significant amounts of data, advanced statistical analyses, and often the use of computer algorithms, data mining, and machine learning.[2] In addition, big data and AI applications, in order to be developed, require vast amount of data. And the larger a given data set is, the easier it is to identify and map even the subtlest relations in the data. This, in turn, translates into more comprehensive and trustworthy results. Hence, access to data has proven itself a key ingredient in the AI landscape. Consequently, the EU has made significant efforts over the past years to improve accessibility and reusability of such data, using a number of avenues to do so, including opening up public sector information and publicly funded research results[3] and ensuring free flow of nonpersonal data.[4]

Besides ensuring data volume and availability, key challenges to successful uptake and implementation of data-based applications include ensuring the high quality of input data sets. Poor-quality data, such as incomplete or biased data, can lead to inaccurate, discriminative, or incorrect outcomes, a result that engineers colloquially call "garbage in, garbage out."[5] Various dimensions of the importance of the quality of data have been highlighted in a variety of documents produced in the EU legal field, including Ethics Guidelines for Trustworthy AI,[6] the White Paper on AI,[7] the report on the safety and liability implications of artificial intelligence, the Internet of Things and robotics,[8] A European strategy for data,[9] and in the European Union

[2] *O'Connor, Ng, Hill* [26], p. 1–3.

[3] Directive 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information [2019] OJ L 172/56.

[4] Regulation 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union [2018] OJ L 303/59.

[5] Which means that low-quality data lead to low-quality outcomes produced by algorithms, which in turn can lead to a violation of fundamental rights, European Union Agency for Fundamental Rights [14], p. 2.

[6] "The quality of the data sets used is paramount to the performance of AI systems. When data is gathered, it may contain socially constructed biases, inaccuracies, errors and mistakes. This needs to be addressed prior to training with any given data set. In addition, the integrity of the data must be ensured. Feeding malicious data into an AI system may change its behaviour, particularly with self-learning systems. Processes and data sets used must be tested and documented at each step such as planning, training, testing and deployment. This should also apply to AI systems that were not developed in-house but acquired elsewhere," Independent High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI [20], p. 17.

[7] "As with the risks to fundamental rights, these risks can be caused by flaws in the design of the AI technology, be related to problems with the availability and quality of data or to other problems stemming from machine learning", European Commission [6], p. 12.

[8] "The question arises if the Union product safety legislation should contain specific requirements addressing the risks to safety of faulty data at the design stage as well as mechanisms to ensure that quality of data is maintained throughout the use of the AI products and systems", European Commission [7], p. 9.

[9] "Data interoperability and quality, as well as their structure, authenticity and integrity are key for the exploitation of the data value, especially in the context of AI deployment", European Commission [8], p. 8.

Agency for Fundamental Rights focus paper *Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights*.[10]

In law enforcement, as in other sectors, the key element of ensuring quality and reliability of big data and AI apps is the quality of raw material. However, the negative effects of flawed data quality in this context extend far beyond the typical ramifications, since they may lead to wrong and biased decisions producing adverse legal or factual consequences for individuals,[11] such as detention, being a target of infiltration or a subject of investigation or other intrusive measures (e.g., a computer search). This, in turn, may cause negative consequences not only for citizens, such as discrimination and violation of fundamental rights and freedoms, but also for LEAs themselves, since relying on low-quality applications may be misleading and deceptive, and result in wrong decisions. This, consequently, may lead to a counterproductive fight against crime and undermine the legitimacy of LEAs, as well as their confidence in this technology.[12]

This article deals with the topic of data quality in the context of big data and AI applications for law enforcement. It starts by outlining selected issues relating to big data; the first subsection sketches the overall technical problem of data quality, which big data amplifies; and the second subsection focuses on law enforcement and data protection issues stemming from using big data for analytic purposes. The second section turns attention to data sets not containing personal data, pointing out that such data can also lead to fundamental rights violations and, thus, their processing also requires an assessment of their impact on fundamental rights. The third section outlines the already existing legal frameworks that enhance data quality (personal data protection), as well as the direct provision relating to data quality, foreseen in the proposed EU framework on AI.

2 Big data

2.1 Big data quality

Big data has evolved so quickly and in such a disorderly fashion that its universally accepted formal definition does not exist.[13] In a broader way big data is usually defined with the following properties associated with variety, volume, velocity, variability, complexity, and value. "Variety" means that big data has all kinds of data types consisting of raw, structured, semi-structured, and even unstructured data, which is difficult to be handled by the existing traditional analytic systems. "Volume" refers to a "big" amount of data, which is difficult to handle using the existing traditional systems. "Velocity" indicates that the data is constantly in motion and thus have to be dealt with in a timely manner, which is outside the scope of capacity for traditional

[10] "High quality data are essential for high quality algorithms", European Union Agency for Fundamental Rights [14], p. 1.

[11] *O'Connor, Ng, Hill* [25], p. 3.

[12] *Bennett Moses, Chan* [1], p. 818–819.

[13] *De Mauro, Greco, Grimaldi* [5], p. 5.

systems. "Variability" considers the inconsistencies of the data flow. "Complexity" refers to processes undertaken with the data coming from various sources, which requires connecting and correlating relationship, hierarchies and multiple data linkages in order to avoid having data that is out of control. "Value" refers to the economic value of the data, which impact companies and society. Usually the characteristics of big data come down to the "4Vs," which refer to the main characteristics of the data—volume, velocity, variety, and value—which altogether refer to specific technology and analytical methods to highlight the unique requirements needed to make use of such data, and underline its potential to create economic value for companies and society.[14]

Big data does not necessarily mean good-quality and error-free data, in fact, quite the contrary—often big data comes with "big noise," which means all the irrelevant information in a data set.[15] Thus, over time the tech-environment has developed various dimensions of data quality. At the outset of the research on this topic, data quality was understood as "fitness for use," and "data quality dimensions" were used as a set of data quality attributes that represent a single aspect or construct of data quality.[16] These dimensions are context-specific; for example, in the early 2000s, the healthcare, finance, and consumer product companies used the following dimensions (non-exhaustive list): accessibility (the extent to which data are available, or easily and quickly retrievable); believability (the extent to which data are regarded as true and credible); completeness (the extent to which data are not missing and are of sufficient breadth and depth for the task at hand); concise representation (the extent to which data are compactly presented); free-of-error (the extent to which data are correct and reliable); objectivity (the extent to which data are unbiased, unprejudiced, and impartial); relevancy (the extent to which data are applicable and helpful for the task at hand); and timeliness (the extent to which the data are sufficiently up to date for the task at hand).[17] An overall conclusion stemming from the various research is that already data quality is a multidimensional concept, difficult to characterise in precise definitions even in the case of well-structured data.[18] Big data poses even more questions pertaining to the applicability of existing data quality concepts, including the fundamental question on the actual importance of data quality for big data. In this regard, one school of thought argues that high data quality methods are essential for deriving higher level analytics, while another school of thought argues that data quality level will not be so important, as the volume of big data would be used to produce patterns and some amount of dirty data will not mask the analytic results that might be derived.[19] The latter standpoint seems to correspond more with business leaders who will always want more and more data storage, whereas the former one corresponds with the IT leaders who take technical aspects into consideration before storing all the data. If concluding that big data basically focuses on quality

[14] See the survey of existing definitions: *De Mauro, Greco, Grimaldi* [5], p. 6–7.

[15] *Waldherr, Maier, Miltner, Günther* [30].

[16] *Wang, Strong* [31].

[17] *Pipino, Lee, Wang* [27], p. 212.

[18] *Firmani, Mecella, Scannapieco et al.* [15], p. 19.

[19] *Juddoo* [22].

data storage rather than having very large irrelevant data (so that better results and conclusions can be drawn), further questions arise such as how to ensure which data are relevant, how much data would be enough for decision-making and whether the stored data are accurate and representative.[20] These questions are still a hot subject of research.[21]

The legal consequences of using poor-quality big data may be extremely serious, as pointed out by the FRA[22] and the European Parliament,[23] which lists among other things discrimination, privacy, and data protection breaches.

2.2 Data protection

Ensuring the quality of big data used in law enforcement is not the only challenge. Notwithstanding its high practical value, handling big data in everyday law enforcement activities is exposed to various legal and factual barriers. According to Europol, only 24% of the European countries (involved in the survey) use big data analytics as part of their work in, for instance, the identification of crime hotspots—even though 48% of them cooperate with academia and industry on big data and/or provide training. Various reasons justify the limited use of big data, ranging from the difficulty in seizing large amounts of data in a forensically sound manner, the time-consuming subsequent analysis of the data, lack of tool support, hardware and software costs (particularly data storage costs including backup solutions), to legal and privacy issues (such as how to protect personal data).[24]

The latter indeed prove difficult to comply with, as illustrated in the recent example of the so-called "Big Data Challenge[25]". This initiative aims to enable Europol to process large and complex data sets,[26] received as a contribution from Member States, for the purposes of strategic and operational analysis (personal data processing activities involved). According to the Proposal:

> Member States cannot detect such cross-border links through their own analysis of the large datasets at national level, as they lack the corresponding data on other crimes and criminals in other Member States. Moreover, some Member States might not always have the necessary IT tools, expertise and resources to analyse large and complex datasets. As regards the big data challenge for law

[20] *Katal, Wazid, Gouda* [23], p. 407.

[21] *Montero, Crespo, Piatini* [24].

[22] European Union Agency for Fundamental Rights [13], European Union Agency for Fundamental Rights [14].

[23] European Parliament [12].

[24] https://www.europol.europa.eu/iocta/2016/big-data.html accessed 1.2.2022.

[25] Proposal for a Regulation of the European Parliament and of the Council amending Regulation (EU) 2016/794, as regards Europol's cooperation with private parties, the processing of personal data by Europol in support of criminal investigations, and Europol's role on research and innovation, COM/2020/796 final, [2020].

[26] "Large data sets" are defined for the purpose of this decision as data sets, which because of the volume, the nature or the format of the data they contain, cannot be processed in the Europol Operational Network (OPS NET, which is the IT environment where Europol performed operational analysis) with regular tools, but require the use of specific tools and/or storage facilities, *ibidem*.

> enforcement, these problems can be tackled more effectively and efficiently at EU level than at national level, by assisting Member States in processing large and complex datasets to support their criminal investigations with cross-border leads. This would include techniques of digital forensics to identify the necessary information and detect links with crimes and criminals in other Member States.[27]

This proposal has drawn attention to the European Data Protection Supervisor (EDPS), who opened an own initiative inquiry on Europol's big data challenge. It raised concerns linked to the compliance with the Europol's data protection framework, in particular with the principles of purpose limitation, data minimization, data accuracy, storage limitation, with the impact of potential data breaches, location of storage, general management and information security.[28] The EDPS inquiry has shown, among other things, that it is not possible for Europol, receiving large data sets, to ascertain that all the information contained in them comply with data minimization and purpose-limitation principles. The volume of information is so big that its content is often unknown until the moment when the analyst extracts relevant entities for their input into the relevant database in OPS NET.[29] Moreover,

> the processing of data about individuals in an EU law enforcement database can have deep consequences on those involved. Without a proper implementation of the data minimisation principle and the specific safeguards contained in the Europol Regulation, data subjects run the risk of wrongfully being linked to a criminal activity across the EU, with all of the potential damage for their personal and family life, freedom of movement and occupation that this entails.[30]

The big data challenge was the subject of further consultations between Europol and the EDPS, which resulted in, *inter alia*, introducing a pre-analysis of personal data received with the sole purpose of determining whether such data fall into the categories of data subjects. This, however, has raised further concerns relating to maximum retention period for data sets lacking Data Subject Categorisation,[31] and the consultations in this regard are still ongoing.

The example of Europol's big data challenge perfectly illustrates how challenging it is for law enforcement to process big data in line with the data protection principles. The issues raised in the Europol context may also apply in the purely domestic context, not only for big data but for any data set concerning personal data. Unlawful or unfair processing of personal data may not only downgrade the data set quality (for example if the data are inaccurate or out of date), but also lead to fundamental rights violations, which the EDPS has clearly shown. It is worth noting that Member States in such context could take benefit from Art. 27 of the Law Enforcement Directive[32]

[27] *Ibidem*.

[28] European Data Protection Supervisor [10].

[29] European Data Protection Supervisor [10], Point 4.7.

[30] European Data Protection Supervisor [10], Point 4.10.

[31] European Data Protection Supervisor [11].

[32] Directive 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the pur-

(LED), which foresees data protection impact assessment in cases as processing big data for strategic and operational analysis, accompanied with prior consultations with the supervisory authority.

3 Quality of nonpersonal data

The risks of fundamental rights breaches are not exclusive for data sets consisting of personal data. As stated by the Committee of Ministers, human rights may still be negatively affected by the use of algorithmic systems operating on nonpersonal data, such as simulations, synthetic data, or generalized rules on procedures. In such cases both individuals and groups could be impacted, particularly when algorithmic systems are used for decision-making, adapting recommendations or shaping physical environments.[33] An example comes with UNICEF's paper on geospatial technologies,[34] revealing that geospatial data which identify trends in locations may lead to decision-making processes that affect individual cases. One of the reasons is that decision-making using geospatial data is frequently captured within big data, potentially resulting in unconscious discrimination against individuals in the absence of specific reflection on the individual case.

Another problem is that nonpersonal data may bring the feedback loop that negatively affects individuals. This issue was discussed in the context of PredPol, one of the predictive policing systems. PredPol operates under only three data points: past type of crime; place of crime; and time of crime. It uses no personal information about individuals or groups of individuals. However, the system does not take into account the fact that certain communities or districts are historically overpoliced, which results in the over policing of these areas which leads to disproportionate and unrepresentative crime statistics which continuously feed the same algorithm.[35] As a result, law enforcement following the recommendations of such a tool may generate deceptive statistics; and residents in over patrolled districts may find themselves subject to surveillance more often than others (bias, discrimination), while some districts might receive too little attention, leading to weaker security and poorer crime prevention. Therefore, the lack of any quality control for statistical data sets might lead to a variety of adverse human, social, and organizational effects.

4 Ensuring data quality through law

Specific legal criteria for "quality data" for a long time have been rather exceptional. An example of direct provision for quality data is Art. 8 of environmental information directive, which lists the requirements for environmental information as having

poses of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA (2016) OJ L 119/89.

[33] Committee of Ministers [4].

[34] *Berman, de La Rosa, Accone* [2], p. 17.

[35] *O'Donnell* [26], p. 561–562.

to be: up to date; accurate; and comparable.[36] Newer example is Commission implementing regulation 2021/2224 which envisages, inter alia, automated data control mechanisms, common data quality indicators and the minimum quality standards for storage of data.[37] In the literature, the role of data protection principles of data accuracy and completeness has been highlighted in the context of ensuring data quality, pointing out, however, the limits of data protection in this regard.[38] These conclusions should, however, be properly rethought in order to assess whether the EU directive 2016/680 has changed the state of play. Aside from the rules enhancing adequacy, up to date, completeness, timeliness, and review of the data, the directive also foresees the distinction (the nonexistent in the GDPR) between personal data and verification of their quality (Art. 7). It requires Member States to provide for personal data based on facts to be distinguished, as far as possible, from personal data based on personal assessments. As further explained in Recital 30 of the Directive, this provision aims at enhancing the principle of accuracy of data since the personal data are based on the subjective perception of natural persons and are not always verifiable. Moreover, para. 2 of Art. 7 of the LED requires Member States to take all reasonable steps to ensure that personal data that are inaccurate, incomplete, or no longer up to date are not transmitted or made available. To that end, each competent authority shall, as far as practicable, verify the quality of personal data before they are transmitted or made available. As far as possible, in all transmissions of personal data, necessary information enabling the receiving competent authority to assess the degree of accuracy, completeness, and reliability of personal data, and the extent to which they are up to date, shall be added.

A direct attempt to introduce quality criteria for data is Art. 10 of the proposal for Artificial Intelligence Act,[39] which introduces quality criteria and data governance rules for training, validation, and testing data sets used for high-risk AI systems, which is expected to ensure that the high-risk AI system performs as intended and safely and does not become the source of discrimination. To this end, it sets the data governance and management practices, for example: data collection; relevant data preparation processing operations, such as annotation, labelling, cleaning, enrichment and aggregation, a prior assessment of the availability, quantity and suitability of the data sets that are needed; the identification of any possible data gaps or shortcomings, and how those gaps and shortcomings can be addressed. Secondly, Art. 10.3 explicitly says that training, validation and testing data sets should be sufficiently relevant, representative, and free of errors and complete in view of the intended purpose of the system. Another feature enhancing data quality is their appropriate statistical properties, including in terms of the persons or groups of persons on which the high-risk AI system is intended to be used, including the features, characteristics or elements

[36] Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information and repealing Council Directive 90/313/EEC (2003) OJ L 41/26.

[37] Commission implementing regulation 2021/2224 of 16 November 2021 laying down the details of the automated data quality control mechanisms and procedures, the common data quality indicators and the minimum quality standards for storage of data, pursuant to Article 37(4) of the Regulation (EU) 2019/818 of the European Parliament and of the Council, (2021), OJ L 448/14.

[38] *Hoeren* [18]; *Hoeren* [19]; European Union Agency for Fundamental Rights [13].

[39] European Commission [9].

that are particular to the specific geographical, behavioural or functional setting or context within which such a system is planned to be used. Finally, in order to avoid discrimination that might result from the bias in AI systems, the providers should also be able to process special categories of personal data, as a matter of substantial public interest, so that they can ensure the bias monitoring, detection, and correction in relation to high-risk AI systems (Recital 44).

5 Conclusions and further research

While advantages of AI and big data in boosting the effectiveness of law enforcement and security may be numerous, they also pose various challenges, which have to be properly tackled to ensure acceptance of and trust in such applications by both LEAs and citizens. One of them is data quality, a complex and still underexplored issue, both from a technical and legal angle. The introduction of direct data quality requirements in the Artificial Intelligence Act proposal is, on the one hand, a move towards ensuring data quality through the EU law, but on the other it opens up a set of further questions, research, and policy steps. First, translating the legal quality requirements foreseen in Art. 10 of the proposal into IT data quality dimensions seems to be a challenging task in general, since quality models are still sought. The context of law enforcement will be even more demanding, given the high risk of fundamental rights violations and decreasing effectiveness in crime prevention and investigation, which usually require a specific approach.[40] Therefore, the quality dimensions will probably require field-oriented research which will also depend on the purpose of the use of the data (according to the "fit for purpose" rule) and will require conformity with the data protection principles, in particular purpose limitation and data minimization, as well as the rules' use of specific categories of data and their impact on inferences.[41] Secondly, it is still to be revealed to what extent personal data protection may be helpful in enhancing data quality for the purposes of developing big data and AI applications. While the ongoing discussion is mainly concerned with the accuracy of data in relation to individuals, the LED seems to offer much more. Given the high risk of privacy violations,[42] data protection could also be seen as a tool enhancing not only the rights of individuals, but also lawfulness, fairness, and accuracy of the AI and big data applications. The data quality topic also provokes the practical question of how to increase the quality of police data, both personal and nonpersonal (which should not be considered fully neutral) in everyday work,[43] what investments are necessary at the national and the EU levels to ensure interoperability between databases,[44] databases and historical data collected by various actors,[45] and to avoid

[40] Vide: data protection framework divided into the GDPR and the LED.

[41] *Wachter, B. Mittelstadt* [29].

[42] *Heilemann* [17], p. 54, 58–59.

[43] *Sanders, Condon* [28].

[44] *Jansen* [21], p. 2–3.

[45] *Blount* [3].

fragmentation across the big cities and rural areas, which deal with different crime and, consequently, different data sets.[46]

Funding The paper has been supported by the grant UMO-2019/33/B/HS5/02249 awarded by the National Science Centre, Poland.

Declarations

Competing Interests The authors declare no competing interests.

References

1. Bennett Moses, L., Chan, J.: Algorithmic prediction in policing: assumptions, evaluation, and accountability. Polic. Soc. **28**(7), 806–822 (2018). https://doi.org/10.1080/10439463.2016.1253695
2. Berman, G., de La Rosa, S., Accone, T.: Ethical considerations when using geospatial technologies for evidence generation. Office of Research – Innocenti Discussion Paper, DP-2018-02 (2018)
3. Blount, K.: Applying the presumption of innocence to policing with AI, artificial intelligence, big data and automated decision-making in criminal justice. Rev. Int. Droit Pénal **92**(1), 33–48 (2021)
4. Committee of Ministers: Recommendation CM/Rec (2020)1 of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems (2020)
5. De Mauro, A., Greco, M., Grimaldi, M.: What is big data? A consensual definition and a review of key research topics. In: Conference Paper at 4th International Conference on Integrated Information (2014). Available at https://www.researchgate.net/profile/Andrea-De-Mauro-2/publication/265775800_What_is_Big_Data_A_Consensual_Definition_and_a_Review_of_Key_Research_Topics/links/54e61d170cf277664ff2f0b4/What-is-Big-Data-A-Consensual-Definition-and-a-Review-of-Key-Research-Topics.pdf
6. European Commission: White Paper on Artificial Intelligence – A European approach to excellence and trust, Brussels, COM(2020) 65 final (2020)
7. European Commission: Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee Report on the safety and liability implications of artificial intelligence, the Internet of Things and robotics, COM/2020/64 final (2020)
8. European Commission: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Region, A European Strategy for Data, COM(2020) 66 final (2020)
9. European Commission: Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts, COM/2021/206 final (2021)
10. European Data Protection Supervisor: Decision on the own initiative inquiry on Europol's big data challenge, case 2019-0370 (2020). Available at https://edps.europa.eu/sites/default/files/publication/20-09-18_edps_decision_on_the_own_initiative_inquiry_on_europols_big_data_challenge_en.pdf
11. European Data Protection Supervisor: Decision on the retention by Europol of datasets lacking data subject categorisation, cases 2019-0370 & 2021-0699 (2021). Available at https://edps.europa.eu/system/files/2022-01/22-01-10-edps-decision-europol_en.pdf

[46] *Hardyns & Rummens* [16], p. 206.

12. European Parliament: Resolution of 14 March 2017 on fundamental rights implications of big data: Privacy, data protection, non-discrimination, security and law-enforcement (2016/2225(INI)) (2017)
13. European Union Agency for Fundamental Rights: #BigData: discrimination in data-supported decision making (2018). Available at https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-focus-big-data_en.pdf
14. European Union Agency for Fundamental Rights: Data quality and artificial intelligence: mitigating bias and error to protect fundamental rights (2019). Available at https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-data-quality-and-ai_en.pdf
15. Firmani, D., Mecella, M., Scannapieco, M., et al.: On the meaningfulness of "big data quality" (invited paper). Data Sci. Eng. **1** (2016). https://doi.org/10.1007/s41019-015-0004-7
16. Hardyns, W., Rummens, A.: Predictive policing as a new tool for law enforcement? Recent developments and challenges. Eur. J. Crim. Policy Res. **24**, 201–218 (2017)
17. Heilemann, J.: Click, collect and calculate: the growing importance of big data in predicting future criminal behaviour, artificial intelligence, big data and automated decision-making in criminal justice. Rev. Int. Droit Pénal **92**(1), 49–69 (2021)
18. Hoeren, T.: Big data and the legal framework for data quality. Int. J. Law Inf. Technol. **25**(1), 26–37 (2017)
19. Hoeren, T.: Big data and data quality. In: Hoeren, T., Kolany-Raiser, B. (eds.) Big Data in Context. Legal, Social and Technological Insights. Springer, Berlin (2018)
20. Independent High-Level Expert Group on Artificial Intelligence: Ethics Guidelines for Trustworthy AI (2019). https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8C1f-01aa75ed71a1
21. Jansen, F.: Working Paper: Data Driven Policing in the Context of Europe (2018). Available at https://datajusticeproject.net/wp-content/uploads/sites/30/2019/05/Report-Data-Driven-Policing-EU.pdf
22. Juddoo, S.: Overview of data quality challenges in the context of Big Data. IEEE, 1–9 (2015). https://doi.org/10.1109/CCCS.2015.7374131
23. Katal, A., Wazid, M., Goudar, R.: Big data: issues, challenges, tools and good practices. In: Procedures of the 2013 Sixth International Conference on Contemporary Computing, pp. 404–409. IEEE, Noida (2013)
24. Montero, O., Crespo, Y., Piatini, M.: Big data quality models: a systematic mapping study. In: Paiva, A.C.R., Cavalli, A.R., Ventura Martins, P., Pérez-Castillo, R. (eds.) Quality of Information and Communications Technology. QUATIC 2021. Communications in Computer and Information Science, vol. 1439. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-85347-1_30
25. O'Connor, C.D., Ng, J., Hill, D., Frederick, T.: Thinking about police data: analysts' perceptions of data quality in Canadian policing. The Police J., 1–20 (2021). https://doi.org/10.1177/0032258X211021461
26. O'Donnell, R.M.: Challenging racist predictive policing algorithms under the equal protection clause. N.Y. Univ. Law Rev. **94**, 544–580 (2019)
27. Pipino, L.L., Lee, Y.W., Wang, R.Y.: Data quality assessment. Commun. ACM **45**(4ve), 211–218 (2002)
28. Sanders, C., Condon, C.: Crime analysis and cognitive effects: the practice of policing through flows of data. Global Crime **18**(3), 237–255 (2017). https://doi.org/10.1080/17440572.2017.1323637
29. Wachter, S., Mittelstadt, B.: A right to reasonable inferences: re-thinking data protection law in the age of Big Data and AI. Columbia Bus. Law Rev. **2019**(2) (2019). https://ssrn.com/abstract=3248829
30. Waldherr, A., Maier, D., Miltner, P., Günther, E.: Big data, big noise: the challenge of finding issue networks on the web. Soc. Sci. Comput. Rev. **35**(4), 427–443 (2017). https://doi.org/10.1177/0894439316643050
31. Wang, R.Ym., Strong, D.M.: Beyond accuracy: what data quality means to data consumers. J. Manag. Inf. Syst. **12**(4), 5–33 (1996)