OPINION ARTICLE

An active human role is essential in big data-led decisions and data-intensive science [version 1; peer review: 1 approved]

Mohamed L. Seghier

Department of Biomedical Engineering, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates

V1 First published: 08 Nov 2021, 10:1127
<https://doi.org/10.12688/f1000research.73876.1>
Latest published: 08 Nov 2021, 10:1127
<https://doi.org/10.12688/f1000research.73876.1>

Abstract

Big data is transforming many sectors, with far-reaching consequences to how decisions are made and how knowledge is produced and shared. In the current move toward more data-led decisions and data-intensive science, we aim here to examine three issues that are changing the way data are read and used. First, there is a shift toward paradigms that involve a large amount of data. In such paradigms, the creation of complex data-led models becomes tractable and appealing to generate predictions and explanations. This necessitates for instance a rethinking of Occam's razor principle in the context of knowledge discovery. Second, there is a growing erosion of the human role in decision making and knowledge discovery processes. Human users' involvement is decreasing at an alarming rate, with no say on how to read, process, and summarize data. This makes legal responsibility and accountability hard to define. Third, thanks to its increasing popularity, big data is gaining a seductive allure, where volume and complexity of big data can de facto confer more persuasion and significance to knowledge or decisions that result from big-data-based processes. These issues call for an active human role by creating opportunities to incorporate, in the most unbiased way, human expertise and prior knowledge in decision making and knowledge production. This also requires putting in place robust monitoring and appraisal mechanisms to ensure that relevant data is answering the right questions. As the proliferation of data continues to grow, we need to rethink the way we interact with data to serve human needs.

Keywords

Knowledge creation, big data, analytics, complexity, data-based decisions

Open Peer Review

Approval Status

1

version 1

08 Nov 2021

[view](#)

1. **Marta Ewa Kuc-Czarnecka** , Gdańsk
University of Technology, Gdańsk, Poland

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the [Research on Research, Policy & Culture](#) gateway.

Corresponding author: Mohamed L. Seghier (mseghier@gmail.com)

Author roles: Seghier ML: Conceptualization, Funding Acquisition, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was funded by Khalifa University of Science, Technology and Research.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Seghier ML. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Seghier ML. **An active human role is essential in big data-led decisions and data-intensive science** [version 1; peer review: 1 approved] F1000Research 2021, 10:1127 <https://doi.org/10.12688/f1000research.73876.1>

First published: 08 Nov 2021, 10:1127 <https://doi.org/10.12688/f1000research.73876.1>

Introduction

Big data describes voluminous data that are measured at different pace, scale, size and type. Big data can be produced anytime, anywhere and can come in structured, semi-structured or unstructured formats. **Figure 1** illustrates the main characteristics of big data (i.e. the 11 Vs) that make big data challenging and valuable at the same time,^{1,2} including volume, variety, velocity, visibility, variability, validity, veracity, value, volatility, vulnerability and versatility, though not all forms of big data possess all eleven characteristics.³ Many sectors have been transformed by the possibility to make accurate and useful decisions by harnessing big data through diverse effective strategies.⁴ For instance, in the health sector, sophisticated big data analytics can assist doctors to generate accurate and clinically useful individualized predictions for diagnostic or prognostic purposes, which ultimately translate into improved treatments, and more efficient and cost saving service delivery.^{5,6} In the education sector, the ability to process big data about students in real-time helps to understand and enhance learner's behavior and performance and to integrate data into the curriculum.^{7,8} As many sectors are competing to take full advantage of the big data revolution, with the potential to generate data-driven ingenious solutions to hard questions, digitization pace is increasing at faster rates thanks to the internet revolution (Internet of Things, 5G) and the widespread adoption of wearable smart devices. Many experts and initiatives around the world consider big data availability as a key driver to transform health and education sectors in the next decades (e.g.⁹⁻¹³), with far-reaching consequences to how knowledge is produced and shared.

For data specialists and policy makers, big data can help tackle problems in a reliable, accurate, unbiased, fast, and comprehensive way. Big data is adding many features to knowledge creation and decision-making processes (**Figure 2**). This 'added value' is completely reshaping and expanding these processes with the ability to: (1) affect events and systems while they are still unfolding, thanks to the possibility to make fast decisions based on a continuous flow of data; (2) generate an exhaustive and fine-grained picture about every aspect of a student's learning journey or a patient's condition in order to create a holistic decision making process; (3) discover patterns and relationships without a priori knowledge or hypotheses, hence minimizing human bias and framing; (4) account for as many variables and features as possible to generate reliable and useful decisions; (5) combine data with different types and from multiple modalities,

Figure 1. The main characteristics of big data (the 11 Vs).

Figure 2. The main added features, based on big data, to knowledge creation and decision-making processes.

including unstructured data, to continuously improve the accuracy of decisions; (6) increase mobility where decision making processes can take place anywhere, in the cloud, and on any device; and (7) maximize data sharing and collaboration between different decision makers with the possibility to combine data across diverse sectors and domains to generate in-depth and wide-ranging insights about people and events. These abilities are some of the most frequent arguments listed by decision makers for relying on big data to address a variety of questions in different sectors.^{14,15}

As the case for any technology, big data does not come free of challenges and limitations. Many challenges have been discussed extensively in the current literature regarding data capture, storage, searching, sharing, integration, transformation, analysis, visualization, consistency, completeness, scalability, timeliness, privacy, security, liability, accountability, governance and ownership.^{8,16,17} The aim of this article is to highlight another side of this big data revolution that warrants further discussion. This concerns the following issues: (1) the transformation of traditional research paradigms with emphasis on volume and quantity, (2) the erosion of the human role in decision making and knowledge discovery processes, and (3) the alluring nature of big data-based decisions. Below we discuss the implication of these issues on how big data is currently sought and treated.

Big data and good data

Rapid development of processor power and computer parallelization has now made it possible to study huge amounts of data with relative ease. The push for more data is not new as classic models of knowledge discovery emphasize the need for high statistical power, a measure that is positively associated with data size.¹⁸ As current analytics and machine approaches require copious amounts of training data, the increase in size and complexity of models' architecture will make the need for more data even bigger. Under the assumption that machines will be more accurate in extracting relevant and useful information with more and more data, quantity or volume is becoming the dominant feature in big data with sometimes little attention to data quality or relevance; for discussion see.^{19–21} Previous work has shown that data volume is not always a significant factor to drive innovation or improve performance.^{22,23} This is because the value of a data-driven decision cannot be better than the quality of data that is feeding the machines making that decision.²⁴ Poor or inaccurate decisions can be due to lack of data and/or lack of good data.²⁵ If a model or a process is fed poorly, data-driven features will be poor, and this will only yield poor decisions.²⁶

In this context, as the increasingly attractive solution to understand complex phenomena is to acquire more and more data, traditional inquiry research methods with well-controlled designs and well-selected samples may lose their appeal to researchers, funding agencies and policy makers.²⁷ It is important that collection of good data, even with a limited number of variables from samples that are not large, should not be deemed as a relic from the past. Many discoveries and breakthroughs made in the past, well before this era of big data, were the result of top-quality research with top-quality data that helped to generate accurate mechanistic accounts of many phenomena, mechanistic accounts that human users

were able to comprehend, model and harness for diverse applications.²⁸ However, what we learned from many solutions derived from big data is that the processes that yielded a solution remain opaque to human users as the solution itself is sometimes an intrinsic emerging property of big data. This gives very limited insights into the optimal mechanisms that human users can learn and follow to conquer similar problems. A new framework is needed to make sure that solutions derived from big data speak a similar language as solutions derived from prior knowledge by traditional ways of scientific inquiry.^{15,29}

Another important point concerns the creation of useful models, whether in educational or clinical settings, which is an important endeavor in science. It is apparent that availability of big data will have major ramifications on how models are created and compared. For instance, the design of complex models becomes more appealing as data is getting bigger. Although simple useful models are conventionally preferred (i.e. Occam's razor principle), complex data-led models may shift the way we look for explanations and predictions. Big data can extend the model space to search and assess more complex models with more variables and more features. The aphorism 'simple is beautiful' is gradually being replaced by 'complex is powerful'. Nevertheless, what we notice is that, while data is expanding in volume, this increase in model complexity is not necessarily translating into higher explicability or relevance, i.e. explained variation. Put another way, for a given phenomenon, the proportion of explained variance, as a proxy for the model's ability to explain that phenomenon, is not growing dramatically with the increase in data volume. This calls for a different way of designing and comparing models given data, as big data are not only answering questions but also generating more.

Big data with a human touch

The way big data is growing in its complexity, volume or type, makes it a product not crafted for human use or comprehension. How big data is typically created and collected illustrates that its purpose is mainly to feed sophisticated analytics, with little concern to how human users will read data.³⁰ The human aspect is gradually left aside in the processing of big data and in the making of big data-driven decisions, hence reducing human contribution to merely conveying the output of data processing tools. What's more concerning is that human users' ability to interpret and make sense of such big data-driven decisions is continuously eroded, with sometimes no say on how such analytics should read, process, summarize and present big data. When big data is submitted to an AI-powered tool for instance, the way it is cleaned, reduced and processed bear no attempt to account for what is relevant for the human end-users. The education sector provides an interesting example about the implications of this issue. In this sector, there is a growing interest in using big data and analytics to predict students' performance for diverse applications (for a recent systematic review, see³¹), with the ultimate goal to personalize learning and provide adequate and timely academic support.^{8,32} However, when students for instance are denied enrolment in a program or a track because an AI-powered warning system decided that they have an above-threshold risk of failing courses, this high-stake decision needs careful examination and opportunity for re-evaluation for a better protection of students' rights.³³ There is an even increasing risk that decision makers will hide behind the sophistication of such AI-powered tools, leaving no opportunity to challenge or question decisions because they have been made by intelligent machines on massive data, with the assumption that such decisions are, by design, error-free and unbiased. It is thus vital that opportunities for human involvement are encompassed in big data collection and processing. Processes that value the human input on top of big data-driven features should be supported.^{30,34} A proper dialogue needs to take place between policy makers and developers to create processes and tools that are accessible to all and that are in the service of the human decision makers.

We need big data with a human touch. Ongoing advances will increase big data's dual potential to either empower or isolate human users from decision making and knowledge creation processes. It is thus vital to gradually build synergies that can bring big data to their utmost purpose which is to serve human needs. Policy makers should put in place safeguards and regulations to monitor how big data are collected and treated to address a given question. Transparency should be guaranteed at all levels, in particular regarding the exact purpose of collected data and how it is going to be manipulated and processed.³⁵ When big data-driven decisions are made, the concept of legal responsibility and liability may become fuzzy. As analytics and machine learning algorithms are becoming increasingly impenetrable, decision makers are not always able to fully comprehend the breadth and consequences of a given data-driven decision. Hence, accountability needs to be clarified when wrong or inappropriate decisions or predictions are made with big data.³⁶ It is important to be able to trace back all processes and stages involved in the making of a decision. There must be integrated tools that analyze and interrogate the making process of a decision in case of an inquiry or an appeal.³⁷ Perhaps most importantly, the decision-making process must be freed from the opacity of big data and analysis methods, for example by scrutinizing the type of data selected or omitted in the process, the underlying assumptions behind the selected mathematical models, and the anticipated consequences of the decisions in the light of what is socially and legally acceptable.^{17,38,39}

The seductive allure of big data

When a presumably persuasive but irrelevant information is added to an argument, naïve or nonexpert users might be influenced by that information. For example, nonexperts may judge satisfying some bad explanations that are supported by

irrelevant information that has a persuasive power such as a neuroscience evidence or brain images.^{40,41} What this suggests is that, when people are nonexperts or cannot comprehend a process, they tend to be influenced (seduced) by other irrelevant peripheral features⁴²; for example, to be persuaded by peripheral aspects like an image of a brain, a complex equation, a massive number, or a photo of a famous person. Big data is no exception; nonexpert users and decision makers might be seduced by its size and complexity. Hence, there is a risk that decisions become more persuasive and significant just because the process used to generate them is based on big data. Decision makers might be tempted to give more authority (evidence) to decisions or information generated from big data. This might undermine appeals and grievance procedures if decisions and solutions are portrayed as the output of infallible machines on voluminous data.

It is thus important to dissociate core processes in big data-based decision making from peripheral aspects of that process. The way big data is visualized, presented and portrayed should not be made with the desire to obfuscate or complicate the information conveyed to users. When nonexperts are overwhelmed by the sheer complexity and the degree of details that result from a decision-making process on big data, they tend not to question such process. Data specialists should therefore devise user-friendly ways to make core processes as clear and as coherent as possible, such as how data is cleaned, reduced and what relevant features are selected. Additional data quality assurance processes must be put in place to ensure that big data is serving its main purpose in the most accurate and valid ways, using data sampling and profiling to minimize the risk for data quality degradation⁴³ and systematic processes to ensure data validity for other purposes (i.e. repurposed data).⁴⁴

This seductive nature of big data has implication on ethics. Results that emerge in a data-driven manner might not always be appropriate from an ethical point of view (e.g. results that stigmatize a group by gender, ethnicity or health condition), and participants from different groups might feel their data are being used or represented in ways they cannot fathom. This can create a divide between those who access and control big data and those who merely feature in them.^{45,46} Big data collection can be more invasive than necessary due to easy access,⁴⁷ and it is not unusual to see big data used to answer different questions than the ones originally consented for by participants.⁴⁸ Big data often contains much more information than is strictly necessary, hence guidelines and safeguards must be put in place to ensure fairness, equity and transparency. It is important that policy makers and ethics bodies uphold the same standards to big data given its longevity and intrusive nature. It is too easy to impress with big numbers, and it is a moral and legal obligation to communicate the right results with their very real limitations to society and policy makers.⁴⁹

Conclusion

Big data holds the potential for revolutionizing many sectors. We have seen the birth of many big data initiatives,⁵⁰ for instance UN's Global Pulse initiative, Europe's Data Saves Lives initiative, NIH's All of Us program, China's Cohort Consortium, and the UK Biobank. But big data is not the answer to all questions, regardless of how attractive and impressive big data is. As machines can process big data and discover new relationships between variables of any arbitrary shape, it is important to appraise the whole process by making sure that big data are answering the right questions. As we are moving toward a more data-intensive type of science, the objective must remain to understand than merely predict relationships. This necessitates a hybrid framework that incorporates insights from human experts in the data-led knowledge production. We are beholding how big data is gaining unprecedented authority in the decision making and knowledge production processes, while present standards (methodological, ethical and legal) are not able to keep pace with current growth in big data and their ramifications for many sectors. There is an urgent need for more critical reflection on how humans should interact with data and data-driven information.

Data availability

No data is associated with this article.

References

- Ristevski B, Chen M: **Big Data Analytics in Medicine and Healthcare.** *J. Integr. Bioinform.* 2018; **15**(3). 118940–118956. [Publisher Full Text](#)
- Sun Z, Strang K, Li R: **Big Data with Ten Big Characteristics.** *ICBDR 2018: Proceedings of the 2nd International Conference on Big Data Research.* 2018; p. 56–61.
- Kitchin R, McArdle J: **What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets.** *Big Data Soc.* 2016; **3**(1): 205395171663113. [Publisher Full Text](#)
- Al-Sai ZA, Abdullah R, Husin MH: **Critical Success Factors for Big Data: A Systematic Literature Review.** *IEEE Access.* 2020; **8**: 118940–118956. [Publisher Full Text](#)
- Wang Y, Kung L, Byrd TA: **Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations.** *Technol. Forecast. Soc. Chang.* 2018; **126**: 3–13. [Publisher Full Text](#)
- Hulsen T, et al.: **From Big Data to Precision Medicine.** *Front Med (Lausanne).* 2019; **6**: 34. [Publisher Full Text](#)
- Baig ML, Shuib L, Yadegaridehkordi E: **Big data in education: a state of the art, limitations, and future research directions.**

- Int. J. Educ. Technol. High. Educ.* 2020; **17** (in press).
[Publisher Full Text](#)
8. Luan H, et al.: **Challenges and Future Directions of Big Data and Artificial Intelligence in Education.** *Front. Psychol.* 2020; **11**: 580820.
[Publisher Full Text](#)
 9. Murdoch TB, Detsky AS: **The inevitable application of big data to health care.** *JAMA.* 2013; **309**(13): 1351–1352.
[PubMed Abstract](#) | [Publisher Full Text](#)
 10. Andreu-Perez J, et al.: **Big data for health.** *IEEE J. Biomed. Health Inform.* 2015; **19**(4): 1193–1208.
[PubMed Abstract](#) | [Publisher Full Text](#)
 11. Snyder M, Zhou W: **Big data and health.** *Lancet Digit Health.* 2019; **1**(6): e252–e254.
[PubMed Abstract](#) | [Publisher Full Text](#)
 12. Daniel BK: **Big Data and data science: A critical review of issues for educational research.** *Br. J. Educ. Technol.* 2019; **50**(1): 101–113.
[Publisher Full Text](#)
 13. Hasan MM, Popp J, Olah J: **Current landscape and influence of big data on finance.** *J. Big Data.* 2020; **7**: 21.
[Publisher Full Text](#)
 14. Kuch D, Kearnes M, Gulson K: **The promise of precision: datafication in medicine, agriculture and education.** *Policy Studies.* 2020; **41**(5): 527–546.
[Publisher Full Text](#)
 15. Kitchen R: **Big Data, new epistemologies and paradigm shifts.** *Big Data Soc.* 2014; **1**(1): 205395171452848–205395171452812.
[Publisher Full Text](#)
 16. Khan N, et al.: **Big Data: Survey, Technologies, Opportunities, and Challenges.** *Sci. World J.* 2014; **2014**: p. ID 1–18.
[Publisher Full Text](#)
 17. Sivarajah U, et al.: **Critical analysis of Big Data challenges and analytical methods.** *J. Bus. Res.* 2017; **70**: 263–286.
[Publisher Full Text](#)
 18. Jones SR, Carley S, Harrison M: **An introduction to power and sample size estimation.** *Emerg. Med. J.* 2003; **20**(5): 453–458.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 19. Dhindsa K, Bhandari M, Sonnadara RR: **What's holding up the big data revolution in healthcare?** *BMJ.* 2018; **363**: k5357.
[Publisher Full Text](#)
 20. Becker D, King TD, McMullen B: **Big data, big data quality problem.** 2015 *IEEE International Conference on Big Data (Big Data)*. CA, USA: Santa Clara; 2015; p. 2644–2653.
 21. Cai L, Zhu Y: **The Challenges of Data Quality and Data Quality Assessment in the Big Data Era.** *Data Sci. J.* 2015; **14**: 2.
[Publisher Full Text](#)
 22. Ghasemaghaei M, Calic G: **Assessing the impact of big data on firm innovation performance: Big data is not always better data.** *J. Bus. Res.* 2020; **108**: 147–162.
[Publisher Full Text](#)
 23. Ross JW, Beath CM, Quaadgras A: **You May Not Need Big Data After All.** *Harv. Bus. Rev.* 2013; **December 2013**.
 24. Yao H, et al.: **Learning with Small Data.** *KDD '20: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* 2020; p. 3539–3540.
 25. Diesner J: **Small decisions with big impact on data analytics.** *Big Data Soc.* 2015; **2**(2): 205395171561718.
[Publisher Full Text](#)
 26. Kristiansen TB: **Erroneous data and drug industry bias can impair machine learning algorithms.** *BMJ.* 2019; **367**: l6042.
[Publisher Full Text](#)
 27. Anderson C: **The End of Theory: The Data Deluge Makes the Scientific Method Obsolete.** *Wired.* 2008.
 28. Brush SG, Spencer JB, Osler MJ: **Physical science.** *Encyclopedia Britannica.* 2020.
 29. Leonelli S: **Scientific Research and Big Data.** *The Stanford Encyclopedia of Philosophy.* 2020. Summer 2020 Edition.
 30. Neff G: **Why Big Data Won't Cure Us.** *Big Data.* 2013; **1**(3): 117–123.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 31. Namoun A, Alshantqiti A: **Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review.** *Appl. Sci.* 2021; **11**: 237.
 32. Khan I, et al.: **An artificial intelligence approach to monitor student performance and devise preventive measures.** *Smart Learnin Environments.* 2021; **8**: 17.
[Publisher Full Text](#)
 33. Berendt B, Littlejohn A, Blakemore M: **AI in education: learner choice and fundamental rights.** *Learn. Media Technol.* 2020; **45**(3): 312–324.
[Publisher Full Text](#)
 34. Helzlsouer K, et al.: **Humanizing Big Data: Recognizing the Human Aspect of Big Data.** *Front. Oncol.* 2020; **10**: 186.
[Publisher Full Text](#)
 35. Kantarcioglu M, Ferrari E: **Research Challenges at the Intersection of Big Data, Security and Privacy.** *Front Big Data.* 2019; **2**: 1.
[Publisher Full Text](#)
 36. Kempeneer S: **A big data state of mind: Epistemological challenges to accountability and transparency in data-driven regulation.** *Gov. Inf. Q.* 2021; **38**(3): 101578.
[Publisher Full Text](#)
 37. Ulbricht L, von Grafenstein M: **Big data: big power shifts?** *Internet Policy Review.* 2016; **5**(1).
[Publisher Full Text](#)
 38. Ferretti A, et al.: **Ethics review of big data research: What should stay and what should be reformed?** *BMC Med. Ethics.* 2021; **22**(1): 51.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 39. Sethu SG, Nair R, Sadath L: **Big Data in Precision Medicine and its Legal Implications.** 2020 *IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*. India: RUPNAGAR; 2021.
 40. Weisberg DS, et al.: **The seductive allure of neuroscience explanations.** *J. Cogn. Neurosci.* 2008; **20**(3): 470–477.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 41. McCabe DP, Castel AD: **Seeing is believing: the effect of brain images on judgments of scientific reasoning.** *Cognition.* 2008; **107**(1): 343–352.
[PubMed Abstract](#) | [Publisher Full Text](#)
 42. Petty RE, Cacioppo JT: **The Elaboration Likelihood Model of Persuasion.** *Adv. Exp. Soc. Psychol.* 1986; **19**: 123–205.
[Publisher Full Text](#)
 43. Taleb I, et al.: **Big data quality framework: a holistic approach to continuous quality management.** *J Big Data.* 2021; **8**: 76.
[Publisher Full Text](#)
 44. Diaz-Garelli JF, et al.: **DataGauge: A Practical Process for Systematically Designing and Implementing Quality Assessments of Repurposed Clinical Data.** *EGEMS (Wash DC).* 2019; **7**(1): 32.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 45. Andrejevic MB: **The big data divide.** *Int. J. Commun.* 2014; **8**(1): 1673–1689.
 46. Metcalf J, Crawford K: **Where are human subjects in Big Data research? The emerging ethics divide.** *Big Data Soc.* 2016; **3**(1): 205395171665021.
[Publisher Full Text](#)
 47. Sula CA: **Research ethics in an age of big data.** *Bulletin of the Association for Information Science & Technology.* 2016; **42**(2): 17–21.
[Publisher Full Text](#)
 48. Xafis V, et al.: **An Ethics Framework for Big Data in Health and Research.** *Asian Bioethics Review.* 2019; **11**: 227–254.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 49. Hartung T: **Making Big Sense From Big Data.** *Front Big Data.* 2018; **1**: 5.
[Publisher Full Text](#)
 50. Hulsen T: **Sharing Is Caring-Data Sharing Initiatives in Healthcare.** *Int. J. Environ. Res. Public Health.* 2020; **17**(9)
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 24 March 2022

<https://doi.org/10.5256/f1000research.77562.r126888>

© 2022 Kuc-Czarnecka M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Marta Ewa Kuc-Czarnecka

Department of Statistics and Econometrics, Faculty of Management and Economics, Gdańsk University of Technology, Gdańsk, Poland

The article addresses significant problems of the impact of visible numbers (scoring, rankings) and invisible numbers (algorithms) used due to the growing popularity of big data. Important aspects such as overinterpretation of data, blurring of messages, misleading the recipient (intentionally or not), lack of transparency, and too much confidence in the number at the decision-making stage were noted. I believe that reasonable and gentle voices pointing out the problematic issues are vital. Especially in the era of admiration for measuring everything and everywhere, as well as over-modelling and numerification of reality. Perhaps the authors could also devote a bit of space to the issue of big data and fake news - on the one hand, big data should help in capturing them, and on the other hand, they are often used in disinformation or propaganda. I also recommend the authors to read:

- Saltelli, A., Dankel, D., Di Fiore, M., Holland, N. & Pigeon, M. *Science, the Endless Frontier of Regulatory Capture*. <https://papers.ssrn.com/abstract=3795058> (2021) doi:10.2139/ssrn.3795058.
- Saltelli, A. *et al.* Five ways to ensure that models serve society: a manifesto. *Nature* **582**, 482–484 (2020).
- Bruno, I., Didier, E. & Prévieux, J. *Statactivisme. Comment lutter avec des nombres*. (Zones, La Découverte, 2014).
- Saltelli, A. *et al.* *Why ethics of quantification is needed now*. *UCL Institute for Innovation and Public Purpose, Working Paper Series* (2021).
- Couldry, N. & Mejias, U. A. Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Telev. New Media* **20**, 336–349 (2019).

I have mentioned these articles because they refer to quantification in its broadest sense and its impact on decision-making, as well as the often questionable ethical use of numbers. Perhaps the

author, on the basis of these works, will be able to expand the scope of his research in the future.

I believe that the article in its current form meets the formal and substantive requirements, and therefore is suitable for indexing.

References

1. Saltelli A, Dankel D, Di Fiore M, Holland N, et al.: Science, the Endless Frontier of Regulatory Capture. *SSRN Electronic Journal*. 2021. [Publisher Full Text](#)
2. Saltelli A, Bammer G, Bruno I, Charters E, et al.: Five ways to ensure that models serve society: a manifesto. *Nature*. **582** (7813): 482-484 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Bruno I, Didier E, Pr: Comment lutter avec des nombres.
4. Saltelli A, Andreoni A, Drechsler W, Ghosh, Kattel R, et al.: Why ethics of quantification is needed now. *UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2021/05)*. [Reference Source](#)
5. Couldry N, Mejias U: Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television & New Media*. 2019; **20** (4): 336-349 [Publisher Full Text](#)

Is the topic of the opinion article discussed accurately in the context of the current literature?

Yes

Are all factual statements correct and adequately supported by citations?

Yes

Are arguments sufficiently supported by evidence from the published literature?

Yes

Are the conclusions drawn balanced and justified on the basis of the presented arguments?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: ethics of quantification, big data and ethics, composite indicators, taxonomy, spatial statistics, GIS.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research