



Extended Abstract

Big Data – Towards a new Techno-Determinism?

Stefan Strauß^{1,*}

¹ Institute of Technology Assessment (ITA) at the Austrian Academy of Sciences,
Strohgasse 45/5, AT-1030 Vienna

E-Mail: sstrauss(at)oeaw.ac.at

* Author to whom correspondence should be addressed; Tel.: +43(0)151581-6599

Accepted:

Big data promises a multitude of innovative options to enhance decision-making by employing algorithmic power to gather worthy information out of unstructured data sets. Exploiting petabytes of data is framed as remedy to deal with complexity and reduce uncertainty by paving the way for predictive analytics. However, the increasing complexity of big data analysis fed with increasing automation may trigger not merely uncertain but also unintended societal events.

Big data is often defined as “high-volume,-velocity, -variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making“. This definition refers to the Gartner Group (2001) and not least mirrors the strong role IT-marketing plays in the big data discourse as it puts emphasis on presenting big data as novel form of information processing that efficiently enriches decision-making. Less mystifying, (boyd/Crawford 2012) define big data as “a cultural, technological, and scholarly phenomenon” that rests on the interplay of technology, analysis and mythology. The latter addresses the “widespread belief that large data sets offer a higher form of intelligence and knowledge to generate insights previously impossible with the aura of truth, objectivity and accuracy“ (boyd/Crawford 2012).

This dimension of mythology is of particular interest in this contribution aiming at de-constructing some of the major claims of big data enthusiasm; such as a claim that the exploitation of large, messy data sets allows to win more insights in a natural/self-evident way as “[w]ith enough data, the numbers speak for themselves“ (Anderson 2008). In line with this delusive view is the perception that data quality decreases in importance and finding correlation is key to come to better decision making. Big data is closely linked to the trend of “datafication” (Cukier/Mayer-Schönberger 2013) aiming at gathering large amounts of every-day-life information to transform it into computerized, machine-

readable data. Behind the scenes of big data mystique and related trends there might be a new paradigm of data pragmatism on the rise as Boellstorff (2013) pointed out: „Algorithmic living is displacing artificial intelligence as the modality by which computing is seen to shape society: a paradigm of semantics, of understanding, is becoming a paradigm of pragmatics, of search”. If there is such a shift away from semantics then syntax might become more meaningful, especially for big data analysis. Together with an increase in automated decision-making big data then entails high risks of false positives and self-fulfilling prophecies, especially if correlation is mixed up with causation as the big data discourse suggests. This is *inter alia* visible in one of the seemingly “big” success stories, namely Google flu trends which was celebrated for its high accurate prediction of the prevalence of flu. However, as Lazer et al (2014) pointed out, in the end the prevalence of flu was overestimated in the 2012/13 and 2011/12 seasons by more than 50%. This and other examples underline the seductive power of big data to perceive it as novel tool to predict future events. If the results of predictive analytics are blindly trusted then their verification or falsification can become complicated. In particular then, if a predicted event triggers action to prevent this event. Together with developments towards predictive policing, aiming at identifying “likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions“ (Perry et al 2015), big data entails a number of serious challenges than can even strain cornerstones of democracy such as the presumption of innocence or the principle of proportionality. Threat scenarios referring to the movie “Minority report” might be overestimated. However, automated predictive analytics might increase the pressure to act and challenge to identify the red line between appropriate intervention and excessive pre-emption.

Big data algorithms (e.g. mapreduce) are most likely to be probability calculating pattern-recognition techniques. From a meta-perspective, big data might be understood as a pave maker for a new techno-determinism that is capable of re-shaping the future by transforming possibilities into probabilities. In this sense, big data might become a new source of political, economic and military power (Zwitter/Hadfield 2014). Implications range from sharpened views on realistic options for decision-making to constrained rooms of possibilities that impact privacy, informational self-determination and autonomy of the individual. Together with its “supportive relationship with surveillance“ (Lyon 2014) big data can reinforce a number of related threats, such as blurring boundaries between personal and non-personal information, de-anonymization and re-identification techniques (cf. Strauß/Nentwich 2013) and risks of surveillance such as profiling, social sorting and digital discrimination.

Big data represents a new source of networking power which (as every technology) can be boost or barrier to innovation in many respects. The “shady side” of winning new insights for decision-making may be new power asymmetries where a new data pragmatism celebrating quantity and probability curtails quality and innovation. To reduce the risks of big data, its likely reasonable reconsidering the thin line between overestimated expectations and underrepresented momentums of uncertainty that correlate with the big data discourse.

References and Notes

Anderson, C. (2008): in wired, http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

Boellstorff, T. (2013): Making big data, in theory. In: First Monday Vol. 18, No. 10.

boyd, D., Crawford, K. (2012): Critical Questions For Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." In: Information, Communication & Society, Vol. 15, No. 5

Gartner Group (2001): "3D Data Management: Controlling Data Volume, Velocity and Variety" <http://www.gartner.com/it-glossary/big-data/>

Cukier, K., Mayer-Schönberger, V., (2013): Big Data: A Revolution That Will Transform How We Live, Work and Think. Houghton Mifflin Harcourt,

W. L. Perry, B. McInnis, C. C. Price, S. C. Smith, J. S. Hollywood (2015): Predictive policing: The Role of Crime Forecasting in Law Enforcement Operations.

D. Lyon (2014): Surveillance, Snowden, and Big Data: Capacities, consequences, critique. In: Big data & society 2014 1-13, DOI: 10.1177/2053951714541861

Lazer, D. , Kennedy, R., King, G., Vespignani, A., (2014): The Parable of Google Flu: Traps in Big Data Analysis. In: Science Magazine, Vol. 343 no. 6176, pp. 1203-1205. DOI: 10.1126/science.1248506

Strauß, S., Nentwich, M. (2013):): Social network sites, privacy and the blurring boundary between public and private spaces, Science and Public Policy 40 (6), pp. 724-732

Zwitter, A. J., Hadfield, A. (2014): Governing big data. In: Politics and Governance, Vol. 2 Issue 1, pp. 1-2

© 2015 by the authors; licensee MDPI and ISIS. This abstract is distributed under the terms and conditions of the Creative Commons Attribution license.