

Michael Nerurkar, Christian Wadehul, Klaus Wiegerling:

Ethics of Big Data: Introduction

Big Data has been one of the trending ideas in the information technology business in the recent years. It has (or already had) its place in the Gartner Hype Cycle for Emerging Technologies¹ and the phrase is notorious for oscillating between being used as a buzz word and being a meaningful technological concept. There is some broader interest in Big Data, as the many attempts to give an answer to the question "What is Big Data?" indicate.² The idea that when speaking of Big Data we are dealing with more than just one of the current coins in marketing jargon is also supported by the fact that the term has been admitted into venerable dictionaries like the Oxford English Dictionary and Merriam-Webster. The widespread interest in the nature of Big Data, though, also hints that there lie some difficulties in giving a concise and informative definition of the concept. This, of course, is not a distinctive feature of Big Data technologies, but is the case with modern technologies in general, especially with information and communications technologies: They evolve rapidly and sometimes are subject to technological leaps or redefinitions that transcend concepts in use at a time (the computer becoming a computer in a network, the mobile phone becoming the smart phone, glasses becoming augmented reality glasses, the automobile becoming the driverless car etc.).

So what is Big Data?³ Big Data is very large volumes of data of various types, collected massively from heterogenous sources (sensors, cell phones, social networks etc.). Big Data is also the new technologies (algorithms/software and hardware) used to collect, store and process this data with high degrees of velocity. Technology forecasters and marketers envision Big Data to allow for entirely new kinds of applications and products to be developed in the field of information and communications. Big Data is marketed as a solution to many problems in informatics that could not be tackled before, or only with high costs and effort. Big Data is said to promise an increase in efficiency and productivity, to lower costs in industry and business, to enable new methods and knowledge in science, and to enhance control and regulation in personal life and in governance.

As regards those technologies as such it could seem that only technical disciplines would take an academic interest in them: It is a question for informatics whether envisioned Big Data applications actually are realizable, or whether they offer truly new approaches and solutions to classical problems. And it is a question for economics whether, for example, it would be economically rational to make use of Big Data technologies in business processes, or whether products built on them might find their markets. As regards ethics and the humanities it might seem not clear at first how Big Data technologies should actually raise any specifically *new* questions to think about. Are not problems of misuse and of harmful consequences of the use of technologies simply problems for a general ethics or an ethics of technology and lie in the responsibilities of individuals, public discourse and politics? Why, therefore, an "Ethics of Big Data"?

Outside of the technical literature the phrase "Big Data" (like, for example, its close associate "algorithm") is used in the media and a broader public to refer in a more or less technically informed way to "some technologies" that are more and more becoming part and shaping factor of our everyday life and that are thought to be potentially disruptive. Technology foresight scenarios give the impression that Big Data technologies can not only support or improve specific processes and activities but that they also enable new levels of regulation and control of individual, social, political und economical affairs and decisions. In this respect, "Big Data" would mean technologies that are socially contextualised and socially effective to an

¹ See the 2013, 2014 and 2015 versions of Gartner's Hype Cycle for Emerging Technologies: <http://www.gartner.com/newsroom/id/2575515>, <http://www.gartner.com/newsroom/id/2819918>, and <http://www.gartner.com/newsroom/id/3114217>.

² See for example <http://datascience.berkeley.edu/what-is-big-data/>.

³ This is, of course, also not meant to be a definition of the term "Big Data" since it clearly is only very vague (regarding the technologies in question) and is invoking terms ("large", "new", "high degree" etc.) that are meaningful only relative to a contemporary state of the art. We are only trying to sum up here what is circulating in the non technical literature and on the web about the nature of Big Data.

unusually high degree. A major effect (or benefit) of Big Data (as of any technology in general) might be to make things easier, more efficient, more rational, more economical etc. But the deployment and use of technologies currently referred to as "Big Data" could also cause a fundamental shift in our understanding of ourselves and our individual, social, cultural and political life. This invasive feature of Big Data should be subject to public, political and academic discourses. In this sense, "Big Data" as in "Ethics of Big Data" serves as an integrative concept. It does not express one single strict technological definition, but functions to integrate many different perspectives on a bigger phenomenon or paradigm shift that is technological in a narrower sense only at its core, but social and cultural in its impact. This throws up problems and questions in the various perspectives of disciplines like ethics, philosophy, epistemology, psychology, sociology, politics, jurisprudence or economics: Questions that are "ethical" in a wide sense, or "practical" in a classical sense: For example, autonomy and choice are concerned where supposedly highly accurate knowledge about persons "mined" from large sets of personal data leads to systems anticipating and pre-acting on our (presumed) desires, and "liberating" us from having to make our own decisions. Privacy is in question as regards how and with what amount of informed consent such individual data are collected and personality profiles are generated, stored, sold, made accessible to government agencies etc. Privacy is itself a necessary condition of autonomy since only when a private realm is guaranteed, and felt and known to be safe and free from external observation and intervention, one can engage in truly free processes of deliberation and decision making. Political freedom is at stake when government agencies practice data retention to profile citizens under the prospect of more security or more efficient regulation.

When talking about "data" one is led quickly to also talk about "information" and "knowledge", furthermore of "theory" and "science". In this regard, Big Data throws up not only ethical but epistemological and philosophical questions as well – to name just a few: What is data, ontologically? What is the relation between data, information, knowledge and theory, and how does one lead to the other? What sense is in speaking of data as "raw material" and of "mining" it? Big Data comes with claims to producing "better" knowledge, sometimes even with visions of making obsolete scientific models and theories, which clearly has epistemological and metaphysical implications since it purports to give us direct and "undistorted" access to reality. Those are claims that deserve thorough epistemological examination and critique. With Big Data we are given yet another label for self description of modern western type civilizations of the 21st century: Is it actually more appropriate to speak of "data societies" instead of "information societies" or of "knowledge societies"? – The contributions to this issue, however, are primarily dedicated to ethical/practical aspects of Big Data, though partly questions of theoretical/epistemological nature are also discussed.

Concerning the dominant optical metaphors within the Big Data discourse *Regine Buschauer* suggests a conceptual approach that allows clearer distinctions between Big Data as "visions" and as data technologies. Buschauer outlines three perspectives on the nexus between data and vision(s). Following Bruno Latour's counter-image of "oligoptica" Buschauer argues, more generally, in favor of a conceptual framework that understands Big Data as a sociotechnical infrastructure. Drawing on more recent studies, she discusses how such an approach allows to address social and ethical implications of present data technologies and practices in a more differentiated way.

In her contribution *Jessica Heesen* aims to explain, on one side, how a normative concept of the public sphere could be infiltrated by Big Data. Furthermore, she discusses how, on the other side, participative processes and common wealth can profit from a thorough use of Big Data analysis. Heesen introduces two important concepts: the numerical public (as a public that is constituted by machine-communication) and total politicisation (as a loss of negative freedom of expression).

Arne Manzeschke, *Galia Assadi* and *Willy Viehöver* consider Big Data to be a new form and instrument of biopolitics addressing both the categories of body and space. It is expected to fundamentally transform health care systems, domestic environments, and practices of self-observation and -reflection. Thus, the paper points out some problems and pitfalls as well as open questions that have emerged in the field of Ambient Assisted Living and that merit more attention in public and academic discourses.

Harald Weston asks whether predictive analytics enables businesses for purposes of lending money or extending credit, for example, to better assess characters of people beyond current objective measures of

credit scores and standard financial metrics. Can individual character be measured and predicted with predictive analytics? The pervasive data surveillance of people that goes with Big Data and predictive analytics is not only an invasion of privacy in general, but an impairment of the aspect of privacy called autonomy that will constrict and alter a person's choices and development of self.

In their article, *Andreas Kaminski* and *Philipp Richter* do not ask for possible losses of individual privacy and freedom but rather what Big Data visions mean for the concepts of self and human being. If "being oneself" according to Martin Heidegger is to live within a critical distance to predefined options, how can I know for sure then that I am "truly myself" and not just imitating behavior of others? Can Big Data help eliminate subjective distortions and illusions about our selves?

In his contribution, *Bruno Gransche* argues that the results of Big Data Analysis can be seen as today's oracle. The uncertainty of the modern, complex world can be felt as a threat and lead to a demand for better foreknowledge. Part of the promise of Big Data is to grant better foreknowledge while omitting problems of scientific theory and modeling. Gransche discusses that people would simply have to believe in the results of such data analysis which would make Big Data based outcomes a matter of faith.

We thank the authors for their contributions.

The Editors of this Volume:

Prof. Dr. Klaus Wieglerling / Dr. Michael Nerurkar / Christian Wadephul, M. A.:

- Institut für Systemanalyse und Technikfolgenabschätzung (ITAS), Karlsruher Institut für Technologie, Postfach 3640, 76021 Karlsruhe, Germany
- <https://www.itas.kit.edu>
- klaus.wieglerling@kit.edu / nerurkar@kit.edu / christian.wadephul@kit.edu
- +49-721-608-23839