Pre-print of paper published in Künstliche Intelligenz 29(2) 2015, pp. 223-232. (C) Springer.
URL of this pre-print: http://people.cs.kuleuven.be/~bettina.berendt/Papers/berendt_buechler_rockwell_KUIN_2015.pdf .
The final publication is available at Springer via http://dx.doi.org/10.1007/s13218-015-0355-2 .

Künstliche Intelligenz manuscript No.
(will be inserted by the editor)

Is it research or is it spying?

Thinking-through ethics in Big Data AI and other knowledge sciences

Bettina Berendt · Marco Büchler · Geoffrey Rockwell

Received: date / Accepted: date

Abstract "How to be a knowledge scientist after the Snowden revelations?" is a question we all have to ask as it becomes clear that our work and our students could be involved in the building of an unprecedented surveillance society. In this essay, we argue that this affects all the knowledge sciences such as AI, computational linguistics and the digital humanities. Asking the question calls for dialogue within and across the disciplines. In this article, we will position ourselves with respect to typical stances towards the relationship between (computer) technology and its uses in a surveillance society, and we will look at what we can learn from other fields. We will propose ways of addressing the question in teaching and in research, and conclude with a call to action.

Keywords Big Data · Surveillance society · Privacy · Ethics · Interdisciplinary knowledge sciences

Bettina Berendt
Department of Computer Science, KU Leuven, Leuven, Belgium
Tel.: +32-16-32-7552
E-mail: bettina.berendt@cs.kuleuven.be

Marco Büchler
Göttingen Centre for Digital Humanities, Georg-August-University Göttingen, Göttingen, Germany
Tel.: +49-551-39-20479
E-mail: mbuechler@gcdh.de

Geoffrey Rockwell
Department of Philosophy, University of Alberta, Edmonton, Alberta Canada
Tel.: +1-780-492-6436
E-mail: grockwel@ualberta.ca

1 Introduction

"How to be a knowledge scientist after the Snowden revelations?" is a question we all have to ask as it becomes clear that our work and our students could be involved in the building of an unprecedented surveillance society.[1] We need to ask about our responsibilities to civil society and our students. We will argue that this asking can benefit from discussion across the knowledge sciences – those disciplines dealing with "Big Data" and its analysis, such as computer science, computational linguistics and the digital humanities. We will argue that in the first instance we have to ask about our own research and teaching practices and take responsibility for the ethical dimensions of what we do.

The present position paper appears in an Artificial Intelligence journal, but we explicitly want to point out that we describe our quest for an ethics of the knowledge sciences: It concerns all of us working with knowledge, specifically with knowledge about people, because of the link that the use of knowledge-processing and knowledge-producing technology forms between us. It concerns the majority of data miners, since most data scientists today (and therefore probably also many of the computer science students who study data mining today) deal, at least in parts of their work, with data about people [31].

[1] This essay comes out of working group that organized around this question at a Dagstuhl Seminar in July 2014 on Computational Humanities. We thank Chris Biemann, Joachim Scharloth, and Claire Warwick for the inspiring discussions.

Our argument also relates to a current debate about AI in general, manifested in the "Research Priorities for Robust and Beneficial Artificial Intelligence: an Open Letter"[2] published several weeks after the submission of the present paper and signed by a large number of prominent AI researchers. "Professional Ethics" are mentioned briefly, but only as a "short-term research priority" and in question form: "What role should computer scientists play in the law and ethics of AI development and use?" (p.3). We propose one answer to this question and argue that and why it should be a permanent rather than short-term priority.

Calls for professional ethics are often met by references to economic pressures. Ethical behaviour is easily viewed as a luxury that can be contemplated in academic environments but has to stand back in the realities of a graduate's job search and later professional actions. The present paper is not the place for contributing to this very broad question in any general terms. However, with regard to the forms of knowledge processing we discuss, there is reason for concern even if one cares about economics only: parts of the Snowden archives illustrate the amount of industrial espionage in intelligence agencies' surveillance activities (e.g. [17], see also the example in Section 5.1 below), and countries will have to protect themselves against this facet of surveillance in order to survive economically.[3] This is another motivation for making students aware of the consequences of data surveillance and focussing research and teaching on developing and deploying new tools and techniques. Investigating the relationships between "economic national sovereignty", "ethics and democracy", and "the student's own future paycheck" could be an interesting didactical challenge. However, a detailed investigation of these implications of the Snowden revelations lies beyond the scope of the present paper.

The remainder of the paper is structured as follows. We will first give some background to the Snowden revelations (Section 2) and position ourselves (Section 3) before outlining where we stand in contrast to typical stances towards the relationship between technology (especially computer science) and its uses in a surveillance society, and then looking at what we can learn from other disciplines (Section 4). We will propose ways of addressing these problems, in teaching as in research (Section 5), and close with a conclusion (Section 6).

[2] `http://futureoflife.org/misc/open_letter`

[3] This point was explored e.g. in the *Security and Privacy in a Post-Snowden World* November 2014 Symposium at KU Leuven (`http://eng.kuleuven.be/evenementen/arenbergsymposium-2014`).

2 Background to the Snowden revelations

Edward Snowden flew to Hong Kong in May of 2013 where he met with journalists including Glenn Greenwald and Laura Poitras, giving them access to a trove of thousands of classified documents about the large-scale data gathering activities of the NSA (National Security Agency of the USA) and partners. Since then journalists at media organizations such as The Guardian[4], Spiegel[5], The Washington Post[6], and The Intercept[7] have been working through the documents they have access to, redacting and reporting on what they reveal about the extraordinary signals intelligence (SIGINT) infrastructure developed.[8] Some of the revelations include:

- Phone companies such as Verizon were being forced to secretly share the phone records of millions of US customers with the NSA. This story was based on a leaked copy of a secret FISA court order and showed that the NSA was gathering phone metadata about calls within the USA and abroad. [9]
- The NSA and FBI were getting email content directly from technology media giants like Facebook, Google, Microsoft and Apple and mining it in a program code-named PRISM. [8]
- The United Kingdom's Government Communications Headquarters (GCHQ) was tapping into fibreoptic cables and siphoning off global communications that it shared with the NSA. [20]
- The NSA is gathering hundreds of millions of text messages a day including geolocation information and financial transactions where available. Analyst with the NSA and GCHQ can search the database of text content and metadata. [2]

[4] `theguardian.com`

[5] `spiegel.de`

[6] `washingtonpost.com`

[7] `firstlook.org/theintercept/`

[8] See [1] for a searchable archive and [10,29] for Greenwald's and Poitras' documentations.

In sum, the NSA and Five Eyes[9] partners have developed tools that let them gather and search the full content of messages of millions of citizens, and they have developed tools to map the network of people a citizen communicates with – activities many feel their governments shouldn't engage in, at least on the scale revealed. As Schneier [37] points out "the NSA has undermined a fundamental social contract" between government and its citizens. We assume democratic governments wouldn't eavesdrop on us. Now we know better, which raises the question "What should we do about it?"

These revelations also show that significant resources are going into the gathering of citizen communications, their storage and their mining. Since 2001 the intelligence budget for the United States alone has grown to $52.6 billion. [22] As a result there has been dramatic growth in the subcontracting of intelligence work to private companies. According to [39] some 70% of what is spent on intelligence by the US government goes to contracts. Wikileaks has documented the international reach of this new industry in their SpyFiles site[10]. It is not just US companies that are selling surveillance wares, but companies around the world are getting into the business and selling to governments not known for democratic institutions. Intelligence, especially signals intelligence has become a big business, and those of us who do research and teaching in related areas like data mining and knowledge representation need to acknowledge that our work could directly or indirectly be utilized for illegal or unacceptable purposes. We need to ask whether our research and teaching is contributing to unlawful government and commercial eavesdropping.

The Snowden revelations are not the root cause of our call for action, but they are a useful trigger: indeed, Internet surveillance has been going on and has been known to be going on for a long time, and even its extent and increasing breadth and depth have been surmised by many experts, cf. [33,39]. Questionable data uses, outright data abuses, and data leaks, by government and other actors such as big companies have been described and challenged for a long time[11]. Research into "Big Data" (formerly known as data mining, machine learning, knowledge discovery, etc.) has been increasing steadily also at computer science departments, business schools, and related institutions throughout the world, and research, business and government agencies have pushed for progress together (e.g., [28], pp. 56-63). At the same time techniques have also been explored in the social sciences and humanities for the study of large literary, historical, and philosophical corpora [24,14].

All these developments, however, were under the radar of most except for the occasional news story (e.g. [34]). These stories were treated as anomalies that didn't merit broad democratic discussion. This changed with the Snowden revelations, which (and this is not meant to be negative!) couldn't have been orchestrated better. The revelations have been staggered in "shock value", and coupled with exciting human drama – in short, made newsworthy. Snowden, Greenwald, Poitras and others at the core of the revelations have managed to provoke an intense public debate for an extended period of time, at least in certain countries like the USA and Germany. Notably absent in this debate are the disciplines, like ours, who benefit from increased investment in the knowledge sciences.

The public attention being given to these issues provides knowledge scientists with both a responsibility to engage and an opportunity to contribute.

3 What is our background?

When thinking about what our professions can and should do about social and ethical questions, we will unavoidably be affected by what we do and think already. In fact, we regarded this – more specifically, our coming from different disciplines – as a virtue of our working group. We start with the result of our reflections and discussion on positioning: how we perceive ourselves and the other members of the working group (idealised as a representative of a discipline) in terms of complementing concerns, experiences and skills/abilities (see e.g. [5] for more details on reflexivity and positioning).

The concerned **AI researcher** may feel responsible for a technology s/he has helped create and bring into the world, a technology that is now used for spying on innocent people, violating their privacy and other fundamental rights. Through his or her research practice and publications, the researcher has also brought a regard or disregard for individual data points, a care-

[9] The Five Eyes are the UK, the US, Australia, Canada and New Zealand; five Anglophone countries that have a history of collaborating on intelligence gathering.

[10] `http://wikileaks.org/spyfiles`

[11] for documentation, see `epic.org` or `europe-v-facebook.org`

fulness or sloppiness in data handling, an interpretation/use of statistics, and a vocabulary for talking about people and their relationships ("nodes and links", "predictability" of people), into the real world. S/he brings to the table abilities to explain technology, algorithms and the statistical methods behind them, and to identify their limitations, misinterpretations, and flaws in reasoning in applications. We call this thinking-through.

The concerned **Digital Humanities researcher** may realise that when analysing text corpora such as historical letters or census data, s/he is in a way doing something very similar (and with essentially the same methods and/or tools) as an intelligence agency that intercepts and analyses communications. After all, the social sciences and humanities are about society and the human. We are used to digging into people's personal lives, adapting their histories and arts to fit theoretical perspectives, and sometimes affecting lives by publishing results. Is there a fundamental difference between doing this to historical figures (who could have living relatives) and doing it to living citizens? Is there more to the difference than the historical role of the target or the scale of the eavesdropping? How can the researcher decide what is right and wrong here? S/he brings to the table abilities to closely analyse texts and other materials from different cultural viewpoints, to work in different languages, and to build up context for communications from traces. All of these paleographical, philological, linguistic and interpretative skills contribute to questioning the engineer's "solutionism" according to which there is a right and a wrong in social questions [26]. All of these skills make sense of the big data engineered for surveillance by thinking-through the technologies in the dual sense of thinking critically about them and thinking by using them.

The concerned **Ethics scholar**, in addition to the concerns and abilities s/he may have as one of these two (admittedly roughly sketched) groups, brings to the table the skills for teaching us how to have a structured discussion about an inherently emotionally laden problem with lots of unknowns and potentially no single satisfying solution. Thus, s/he can help other scientists not to fall into one of the common traps: voluntary ignorance ("I'm just a scientist, ethics is not my domain"), cynicism ("If I don't do [this piece or type of research], somebody else will anyway") resignation ("I can't change anything anyway"); well-intentioned but potentially futile or even abusable actionism ("I can tweak my algorithm so that it [doesn't do this one evil thing], so now I've contributed my share of goodness and can otherwise continue as before"). The Ethics scholar can also help us understand how thinking about the ethics of big data is always in a political and economic context.

Academics have the freedom to talk about ethics where others don't. We need to recognize that our students may not have a lot of choice in the jobs they take. For that matter, some academics don't even have the freedoms many of us take for granted.[12]

4 Where do we stand?

From the viewpoints sketched in the previous section, we will proceed to summarise what we perceive as predominant attitudes in the knowledge sciences (Section 4.1), why we think these are problematic (Section 4.2), and what fields we believe we can learn from (Section 4.3).

4.1 "It's just a tool": trying to not take a stand

"The surveillance that we display in our conferences, and discuss how to use, is available to any country in the world. [...] Do some countries use this technology to suppress political statements? Yes, I would say that's probably fair to say. But who are the vendors to say that the technology is not being used for good as well as for what you would consider not so good?" This quotes Jerry Lucas, President of ISS World, a trade fair for security technology.[13] This pattern of argument is pervasive among IT companies that produce surveillance technology, and only recently have the dual uses of such technology been acknowledged by their inclusion into the list of products regulated by the Wassenaar Arrangement[14] [16].

While – in the experience of the authors – similar arguments are often voiced by academic computer scientists too, it is more difficult to find them in quotable form. Reasons vary but certainly include the self-fulfilling prophecy that a computer scientist who considers her/his role that of "just doing science" for this very reason

[12] [19], `http://scholarsatrisk.nyu.edu/`
[13] `http://www.theguardian.com/technology/2011/nov/01/governments-hacking-techniques-surveillance`
[14] `http://www.wassenaar.org/controllists/`

will in general not publish or speak out in public on ethics that are "not their business". Instead, thoughts about the different uses of technology are generally left to authors outside informatics, as witnessed by the recent success of the book "Big Data" [23] and other, more profound, texts on the topic (e.g. [6,18]; see Section 4.3). It is good to see such comprehensive views, and we agree with them that data analytics regulation is also a political, legal and societal task. However, understanding "Big Data" also requires computational expertise, and computational design decisions have real effects that make it impossible to regard technology, and therefore research decisions, as neutral. We need to be careful that we don't agree to divisions of academic labor that haven't been thought through. We will illustrate this using the example of data mining and discrimination.

4.2 But whatever we do, we take a stand: the case of data mining and discrimination

Differentiation – making a distinction based on some features or attributes – is a fundamental characteristic of human cognition and behaviour. People apply differential treatment to other people, allowing some but not all to vote, applying certain laws to them, giving them jobs, and granting them loans – or denying them the privileges associated with these rights and decisions. Part of the social contract of any society is that certain attributes are accepted for differentiation, while others are not. Unacceptable differentiations are called "unlawful discrimination" or just "discrimination", with boundaries and the evaluation of the same differentiating descriptions and/or treatments of people being regarded as (legally and/or otherwise normatively) legitimate or not.

Differentiation is also the mainstay of statistical methods in general and data mining or "Big Data analytics" in particular, for example when "instances" (here: people) are clustered (persons A and B are similar to person C, but dissimilar to D, which in turn is similar to E and F) or classified (persons A and B are of this type / have this characteristic, while C-F are of that type). Of course, a formal method does not know about social contracts. Thus, the application of data mining may inadvertently lead to discrimination, for example when a pattern like "women/immigrants/... tend to default on their loan" in historical data is transformed into a decision rule "if the applicant is female/an immigrant/..., do not give them a loan".

To avoid such unwanted outcomes, one can modify algorithms such that they do not return differentiation based on certain attributes ("discrimination-aware data mining", founded by [30], "fairness-aware data mining", [15]; for an overview, see [4], for a discussion in relation to other disciplines' research on discrimination, see [36]). Sophisticated methods that also protect against harder-to-detect problems such as indirect discrimination (via seemingly innocuous but in fact correlated attributes) have been proposed. These methods are invaluable not only because they can block unwanted inferences and their application, but maybe even more so because they allow for a transparent, structured and quantifiable societal discussion about how much (e.g.) "prediction accuracy" one is willing to forgo in favour of how much "increased fairness".

Still, conceptual considerations show that this by itself cannot avoid unlawful discrimination, let alone discrimination in a wider, sociological sense, and user-study findings show that "sanitized patterns" do not necessarily mean "sanitized minds" in decision situations such as loan decisions, that the decision-making situation as a whole needs to be revisited [4]. It is at this point that many computer scientists, even if concerned about the implications of their science on society, give up: "Well, if data mining is inherently about differentiation/discrimination, what can we do?" (therefore, we won't do anything). This appears to be a form of the well-known "we know there are dual uses of the technology, but the technology itself is innocent" argument.

We believe that computer science can do better. Modifying algorithms is a valuable and much-needed starting point, but it is important that computer scientists are also aware of, and communicate, the limitations of this approach. In Sections 5.2 and 5.3, we will highlight some design choices where the data miner can and should exercise the responsibilities that their knowledge makes possible.

4.3 What can we learn from science and technology studies?

Literature, history, sociology and philosophy obviously have a lot to offer on the subject of the responsibilities of technologists such as us, more than can be covered here, but a few useful starting points include:

– There is a growing literature on the privacy, ethical and societal issues around Big Data. [23] contains an accessible chapter on "Risks", or one can look at Kitchin's [18] in-depth treatment in *The Data Revolution*. boyd and Crawford [6] propose a set of provocations for starting discussion. There is also the new journal *Big Data and Society*[15] that has good articles such as "Big Data ethics" [43] that introduce the subject.
– That new technologies can bring unexpected ethical challenges is not a new story. Novels like Mary Shelley's *Frankenstein* [38] and its many reinterpretations are just one way society reflects on technology [38]. The ways data mining are represented in science fiction and in the movies have an effect on public perceptions of our sciences which is why it is useful to engage with popular fictions. One reading of *Frankenstein* is that the pivotal crime is not the bringing to life of the monster (technology), but the abandoning of the technological son at the moment of conception. Much the same could be said about big data technologies – the danger doesn't lie in their development, but in their abandonment by researchers to other actors. These stories in circulation are also a way we can engage our students. Movies like *Minority Report* (2002), *The Lives of Others* (2006), and *Enemy of the State* (1998) can serve as a common text for discussing what has been done and what can be done.
– Many technologies, like the personal computer, are assumed to have beneficial democratizing effects as if technologies had politics. [21] Various claims are now being made for Big Data (e.g. [23]). While it seems naïve to suggest that technologies determine political structure, some technological systems do seem to be more compatible with certain political structures. Nuclear power, and even more so nuclear weapons, need a level of centralized control to be managed safely. Other "maker" technologies seem more compatible with distributed control. Our point is that we need to pay attention to both the politics of the technologies themselves and that of their contexts. [42] We need to pay attention in ways that leads to a dialogue that allows people to make choices without feeling railroaded by progress. This will take a level of interdisciplinary discussion between the technical and human disciplines. To understand the politics of a technology, we need both the technical and human sciences reading technologies together.
– An emerging field called Software Studies has taken software systems as artefacts for interdisciplinary study. To paraphrase Marshall McLuhan, if the software medium is the message, then we should study the medium of software. To that end the MIT Press has started a Software Studies Series[16]. Data scientists should consider working with digital humanists to study the tools of data analytics [43]. As described below in Section 5.1, studying the traces of Big Data software like the Olympia system shown in the Snowden-leaked CSEC slides is a way of engaging students in the issues, both the technical issues of what the system does and the ethical issues of whether such systems should be deployed by our governments.

5 What do we have to offer? Where can we go from here?

What then can researchers such as us do? What responsibilities do we have? This essay starts from a dialogical perspective – that none of us can tell another what they should do, but we can enter into dialogue and encourage others to carry the dialogue into their labs and classrooms. Dialogue is also recognition that we really don't know what constitutes ethical action as we don't know, for example, whether the surveillance system developed by the NSA and partners is effective. Dialogue is a way of thinking-through that is both means and end without being THE END.[17] We will describe some possible directions for us as teachers and as researchers. There are many other important roles, including as data providers, as mediators between science and the public, and as citizens; we do not discuss them here for reasons of space.

5.1 In teaching

One site for dialogue about the ethics of big data is the classroom. Those of us who teach courses about knowledge representation, data mining, text analysis

[15] `http://bds.sagepub.com/`

[16] `mitpress.mit.edu/books/series/software-studies`

[17] We can also think of "thinking-through" as a playful etymology of dia (between/through) and logos (thought).

and computational linguistics have an opportunity to introduce discussions about the ethical implications of knowledge technologies alongside teaching the very technologies in question. Here are some of the approaches we have had success with.

- Introduce academic readings that deal with ethical and social issues such as [6] "Critical Questions for Big Data" or [40] "Discrimination in Online Ads." Alternatively you can weave news readings into your course like "Why we need an algorithm ethic" [27] or The Guardian's interactive "The NSA Files: Decoded"[18].
- Create assignments where students review the projects of others. Berendt[19] has students who have taken a course on privacy and big data take on the role of "privacy consultants" for projects other students are doing. The "privacy consultants" engage with the project teams about privacy issues. The guidelines for this are derived from European Data Protection principles or the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data[20]. In addition, students are asked to consider technical choices such as anonymisation, distribution and encryption. The students are, among other things, learning about the business of privacy impact assessment and certification (see references in the teaching materials; see also e.g. Privacert[21]).
- Have the students perform a close reading of leaked materials like the CSEC slides leaked by Snowden and published by the Globe and Mail[22]. These slides purport to show how the Communications Security Establishment of Canada (the Canadian equivalent to the NSA) had been spying on the Brazilian Ministry of Mines and Energy along with screenshots of a "Network Knowledge Engine" called "Olympia" that can query the intelligence databases of other services (such as the NSA and GCHQ) and even automate targeting practices. Reading the slides encourages students to think about the value and interpretation of leaked information and how much one can infer from such materials. Reading the software screenshots engages the technical imagination – what can Olympia do? The value of these slides for a critical, digital-humanities-supported analysis was demonstrated by Rockwell and Sinclair [35].

No doubt there are many other ways of engaging students in dialogues from which we all learn. See for example Issue 14 of the International Review of Information Ethics on Teaching Information Ethics[23]. What is important is not to tell students how to be ethical, but to engage in dialogue in the public sphere, a fundamental component of participatory democracy. After all, we train students to not only be good at doing, but to be able to participate in the discussions that are part of decision making – local, national, and beyond.

5.2 As researchers applying methods and tools

Researchers can embed critical reflection into applying their methods and tools; what steps they take will be highly dependent on the specifics of these methods and tools. We therefore concentrate on one domain as an example, extending the discussion on data mining in Section 4.2.

Data miners understand the modelling decisions as well as the interpretations made in a machine learning / data mining application. We will illustrate this claim with the ontological commitments made by modelling choices, illustrating them with three popular attributes in current datasets and applications (for an extended example about classifier-learning, see [3]):

- Using an attribute claims that this characteristic exists. One example is the use of "race". Science has known for a long time that human races do not exist, and the term is deemed unacceptable throughout much of Europe. However, US-based datasets (where "race" is generally used to denote what Europeans would call "ethnicity") and their attribute names are often used as-is. Debates about the discrimination involved in describing individuals with certain words have gained ground in recent years – manifested, for example, in style guides of newspapers[24] – but they probably need to be made better

[18] http://www.theguardian.com/us-news/the-nsa-files
[19] http://people.cs.kuleuven.be/~bettina.berendt/teaching/2014-15-1stsemester/kaw/
[20] http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm
[21] privacert.com
[22] http://www.scribd.com/doc/188094600/CSEC-Presentation
[23] http://www.i-r-i-e.net/issue14.htm
[24] An example is the Guardian's and Observer's style guide, which does however use "race", judging by the example as referring to either ethnicity or nationality: http://www.theguardian.com/guardian-observer-style-guide-r

known in computer science. At the danger of alienating some readers: Using "race" as an attribute claims that "human races" exist, and this in itself is racist.

- Enumerating values that an attribute may take claims that these are the possibilities – and no others exist. A well-known recent example is the binary choice of "male" or "female" as gender enforced by states as well as private data collectors/processors and denounced as discriminatory for a long time[25]. Recently, some jurisdictions – and Facebook – have extended their lists of genders, cf. [13, 41].
- Defining what a term (attribute or value) means is a human activity, a "specification of a shared conceptualization" [11] and as such not objective. In many cases, the measurement (e.g., deciding whether an instance falls into one class or the other) is even more subjective. This will affect which instances (people) will be labelled as having a certain value on a certain attribute, and thereby the data mining model learned from the data, and thereby the predictions on new, unseen cases. Popular examples from the post-9/11 era include definitions of "terrorists" and "militants" [32, 12, 3].
- Modelling determines outcomes. The consequences of modelling choices for the results of data mining (or statistics in general) are familiar: Other relationships (such as between poverty and some outcome, when only "race" is modelled and also when both are modelled) may go undetected. Properties of subgroups (such as non-mainstream genders) get lost in the noise of inadequate binning. If enough training examples of "terrorist journalists" [12] are assembled, then more journalists will be classified accordingly in the future.
- Study design and data collection affect outcomes: The spectre looming over all of data mining's exploratory data analysis is of course the misinterpretation of correlation as causation.

Now a data miner may argue that he or she provides "only the table and neither the attributes nor values", while the "application partner" makes these ontological decisions. It would however be naïve to underestimate the normative power of factors that the computer scientist determines: the very format of the interface and its oftentimes existing constraints (such as the need to specify values of an attribute in advance), the examples given (in previous datasets, in help documents, in research papers, etc.), and the data scientist's own preconceptions during consultancy. The data are never neutral, they never "speak for themselves" [18]. The data miner can and should use their knowledge of method details to understand and highlight design choices and their implications.

Delegation of responsibility can occur not only vis-à-vis application partners. It can be tempting to delegate all ethical responsibility to the Ethics Board that has to approve a given research plan. However, such delegation in fact reinforces the "it's not my business" stance, and it harbours many other problems, including that no ethics board can understand all the method details of every scientific approach they are asked to assess. Ethics boards can be valuable partners in research, they are controversial for many reasons (e.g. [7, 6]), but they cannot absolve researchers of their responsibilities.

5.3 As researchers developing new methods and tools

Researchers can embed critical reflection also into the developments of new methods and tools, with the specifics again being highly domain-dependent. We will continue expanding on the data-mining example from Section 4.2.

The argument can be made that classifier-learning only learns how to (most accurately) replicate a principle that, through human decision making, has shaped prior data. This is true to the extent that the choice of which principle to apply for apportioning benefits and disadvantages is indeed external to the formal method itself. For example, whether a health-insurance system determines its premiums based on an individual's expected costs (which could be predicted with the help of a classifier learned from actuarial data) or on ability to pay (which can be fixed on the basis of taxable salary) is the decision of a country or at least insurance company as a whole – and if the expected-cost notion of social justice is chosen, then fine-grained interventions at the algorithmic level that avoid unlawful discrimination in premiums are indeed a valuable contribution.

However, just referring to decisions about principles, e.g. of justice, being made "outside the algorithm" is an unjustified oversimplification of how data-mining

[25] See for example the Yogyakarta Principles on the Application of International Human Rights Law in relation to Sexual Orientation and Gender Identity (2006). `http://www.yogyakartaprinciples.org/`

methods and technology work – it treats the innards of the method that is applied as a black box, when it is in fact his or her knowledge of this very black box that distinguishes the computer scientist / data miner and makes him or her a well-paid "added value" for the health insurance or other end users. Therefore, we argue that the refusal to engage with the consequences of all those design choices that to the outsider may remain a black box pushes responsibilities (a) away from the people who will make decisions that shape outcomes and who (should) understand what they are doing and (b) towards people who make such decisions, but do not have the technical expertise to gauge the consequences, and/or (c) who may have other goals than the truth.

5.4 What does what we say about others, say about us?

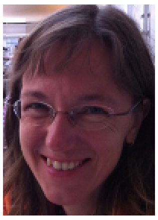
While at one time there were technical limitations to how much data could be collected and stored, at present this problem no longer exists. In fact, in systems such as Olympia (see Section 5.1 and the references there), information from different sources is brought together to form a wider 'common knowledge'. Text processing systems such as the Unstructured Information Management Architecture[26] provide very similar functionalities: for example, they can ingest heterogeneous plain texts and identify and link named entities throughout and across these sources. Both these systems can be easily extended to include large numbers of additional texts. With no technical boundary, the problem now is an ethical one. Even if it is for the sake of research, it is plain to see that tracking and recording every aspect of a person's life without their permission is not only intrusive, but also potentially dangerous. One need only think of being that person themselves to feel the intrusion. Researchers should ask themselves whether the same intrusion should be allowed on the lives of those who are now deceased and can no longer choose to share private aspects of their lives with science. There are currently many digitization efforts in Digital Humanities that are being carried out worldwide on diaries and private correspondence, which, though likely intrusive, are of incredible value to our understanding of history. It is, therefore, the responsibility of the researcher to decide which data and software-support it is ethically "right" to use and how. A good starting point for making this decision, is to ask oneself *what would I consider acceptable if I were the target of research*?

[26] `https://uima.apache.org/`

6 Conclusion

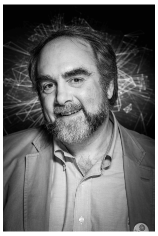
The discussion has shown that a multi-perspective view is needed for a critical theory of big data analyses, and that there is no single, and certainly no algorithmically computable, solution to it. Too many interests are involved, and these interests interact in too many ways. So we cannot hope to find a "decision tree" or any other schema for "solving" the question of how data can and should be analysed. Rather, we have to accept that this is a political problem, and that political problems require ongoing debate and have no clear-cut answers. We must refrain from "taking the politics out of politics" by having the "solutionism" mindset [26] and applying "algorithmic regulation" [25]. As knowledge scientists, we are in a unique position – and thus called upon – to use our in-depth understanding of knowledge modelling and processing to make potentials and limitations transparent.

Acknowledgements We thank the Flemish Agency for Innovation through Science and Technology (IWT), the Fonds Wetenschappelijk Onderzoek – Vlaanderen (FWO), the German Ministry of Education and Research (BMBF), and the Social Science and Humanities Research Council of Canada for support through the projects SPION (grant number 100048), Data Mining for Privacy in Social Networks (grant number G068611N), the early career research group eTRAP (No. 01UG1409), and the NovelTM project.



Bettina Berendt is a Professor of Computer Science in the Declarative Languages and Artificial Intelligence group at KU Leuven, Belgium. Her research interests include Web, text and social and semantic mining, privacy and anti-discrimination and how data mining can contribute to this, user issues and teaching of and for privacy. More information about Bettina Berendt can be found at `people.cs.kuleuven.be/~bettina.berendt`.

Marco Büchler leads a Digital Humanities Research Group at the Göttingen Center for Digital Humanities, Göttingen, Germany. His research includes Natural Language Processing on Big Humanities Data. Specifically, he works on Historical Text Re-use Detection and its transformation to business applications.



Geoffrey Rockwell is a Professor of Philosophy and Humanities Computing at the University of Alberta, Canada. He has published and presented papers in the area of philosophical dialogue, textual visualization and analysis, humanities computing, instructional technology, computer games and multimedia including a book, Defining Dialogue: From Socrates to the Internet. He is currently the Director of the Kule Institute for Advanced Study and a co-organizer of Replaying Japan, a conference on Japanese game culture.

References

1. American Civil Liberties Union: The NSA Archive. Searchable archive of Snowden documents (n.d.) `https://www.aclu.org/nsa-documents-search?`
2. Ball, J.: NSA collects millions of text messages daily in 'untargeted' global sweep (2014). `http://www.theguardian.com/world/2014/jan/16/nsa-collects-millions-text-messages-daily-untargeted-global-sweep`
3. Berendt, B.: Big Capta, Bad Science? On two recent books on "Big Data" and its revolutionary potential (2015). Extended version of book review to appear in Karakter. `http://people.cs.kuleuven.be/~bettina.berendt/Reviews/BigData.pdf`
4. Berendt, B., Preibusch, S.: Better Decision Support Through Exploratory Discrimination-aware Data Mining: Foundations and Empirical Evidence. Artif. Intell. Law **22**(2), 175-209 (2014).
5. Berger, R.: Now I see it, now I don't: researcher's position and reflexivity in qualitative research (2013). Qualitative Research, published online January 3, 2013, DOI 10.1177/1468794112468475
6. boyd, D., Crawford, K.: Critical Questions for Big Data (2012). Information, Communication & Society. 15(5): 662-679.
7. Gehring, P.: Fragliche Expertise. Zur Etablierung von Bioethik in Deutschland. [Questionable expertise. About the establishment of bioethics in Germany.] (2012). In M. Hagner (Ed.), Wissenschaft und Demokratie [Science and Democracy] (pp. 112-139). Berlin: Suhrkamp Verlag.
8. Gellman, B., Poitras, L.: U.S., British intelligence mining data from nine U.S. Internet companies in broad secret program (2013). The Washington Post. June 13, 2013. `http://www.washingtonpost.com/investigations/us-intelligence-mining-data-from-nine-us-internet-companies-in-broad-secret-program/2013/06/06/3a0c0da8-cebf-11e2-8845-d970ccb04497_story.html`
9. Greenwald, G.: NSA collecting phone records of millions of Verizon customers daily (2013). The Guardian. June 6, 2013. `http://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order`
10. Greenwald, G.: No Place to Hide: Edward Snowden, the NSA, and the U.S. Surveillance State (2014). New York: Metropolitan Books.
11. Gruber, T.R.: A translation approach to portable ontology specifications. Knowledge Acquisition **5**, 199–220 (1993)
12. Harrison, S.: Britain is treating journalists as terrorists – believe me, I know (2014). Guardian CommentIsFree. `http://www.theguardian.com/commentisfree/2014/mar/14/britain-journalists-terrorists-edward-snowden-nsa`
13. Heine, F.: M, F or Blank: 'Third Gender' Official in Germany from November (2013). Spiegel Online. `http://www.spiegel.de/international/germany/third-gender-option-to-become-available-on-german-birth-certificates-a-916940.html`
14. Jockers, M.L.: Macroanalysis: Digital methods and literary history (2013). Urbana, Illinois: University of Illinois Press.
15. Kamishima, T., Akaho, S., Asoh, H., Sakuma, J.: Fairness-aware classifier with prejudice remover regularizer (2012). In: ECML/PKDD (2), LNCS, vol 7524, pp. 35-50. Springer.
16. Kammerer, D.: 1A Trojaner. Geschäftsmodell Überwachung. [Prime Trojans. Business Model Surveillance] (2014). Le Monde Diplomatique (deutsche Ausgabe), 10.10.2014. `http://www.monde-diplomatique.de/pm/2014/10/10.mondeText1.artikel,a0009.idx,1`
17. Kirschbaum, E.: Snowden says NSA engages in industrial espionage: TV (2014). Reuters. 26 Jan 2014. `http://www.reuters.com/article/2014/01/26/us-security-snowden-germany-idUSBREA0P0DE20140126`
18. Kitchin, R.: The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. SAGE Publications Ltd (2014)
19. Lough, T., Samek, T.: Canadian University Social Software Guidelines and Academic Freedom: An Alarming

Labour Trend (2014). International Review of Information Ethics 21: 45-56. http://www.i-r-i-e.net/inhalt/021/IRIE-021-Lough-Samek.pdf
20. MacAskill, E., Borger, J., Hopkins, N., Davies, N., Ball., J.: GCHQ taps fibre-optic cables for secret access to world's communications (2013). The Guardian. June 21, 2013. http://www.theguardian.com/uk/2013/jun/21/gchq-cables-secret-world-communications-nsa
21. Markoff, J.: What the Dormouse Said: How the Sixties Counterculture Shaped the Personal Computer Industry. Penguin Publishing Group (2005).
22. Matthews, D.: America's Secret Intelligence Budget, in 11 (Nay 13) Charts. (2013). The Washington Post. August 29, 2013. http://www.washingtonpost.com/blogs/wonkblog/wp/2013/08/29/your-cheat-sheet-to-americas-secret-intelligence-budget/
23. Mayer-Schönberger, V., Cukier, K.: Big Data: A Revolution That Will Transform How We Live, Work, and Think (2013). New York: Houghton Mifflin Harcourt
24. Moretti, F.: Graphs, Maps, Trees: Abstract Models for Literary History (2007). London: Verso
25. Morozov, E.: The Real Privacy Problem (2013). MIT Technology Review. http://www.technologyreview.com/featuredstory/520426/the-real-privacy-problem/
26. Morozov, E.: To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems that Don't Exist (2013). Philadelphia, PA: Public Affairs.
27. Noller, S.: Why we need an algorithm ethic (2013). The Guardian. http://www.theguardian.com/media-network/media-network-blog/2013/jan/22/algorithm-ethic-mechanisms-control-security
28. O'Harrow Jr., R.: No Place to Hide (2005). New York: Free Press
29. Poitras, L. (Dir.): CITIZENFOUR (2014). Praxis Films. Film. https://citizenfourfilm.com/.
30. Pedreshi, D., Ruggieri, S., Turini, F.: Discrimination-aware data mining. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '08, pp. 560–568. ACM, New York, NY, USA (2008).
31. Piatetsky, G.: Data Science is mainly a Human Science (2014). KDnuggets. http://www.kdnuggets.com/2014/09/data-science-mainly-human-science.html
32. Pinsent Mason: Government definition of terrorism applies to firefighters (2002). http://www.out-law.com/page-3054
33. Priest, D., Arkin, W.: Top Secret America: The Rise of the New American Security State (2011). New York: Back Bay Books.
34. Risen, J., Lichtblau, E.: Bush Lets U.S. Spy on Callers Without Courts (2005). The New York Times. http://www.nytimes.com/2005/12/16/politics/16program.html?pagewanted=all&_r=0
35. Rockwell, G., Sinclair, S.: Watching out for the Olympians! Reading the CSEC Slides (2014). Paper submitted for review. An early draft can be found at http://theoreti.ca/?p=5057
36. Romei, A., Ruggieri, S.: A multidisciplinary survey on discrimination analysis. The Knowledge Engineering Review **29**(5), 582-638 (2014).
37. Schneier, B.: The US government has betrayed the internet. We need to take it back (2013). The Guardian. Sept. 5, 2013. http://www.theguardian.com/commentisfree/2013/sep/05/government-betrayed-internet-nsa-spying
38. Shelley M., W.: Frankenstein: or, The modern Prometheus (1888). London, George Routledge and Sons
39. Shorrock, T.: Spies for Hire: The Secret World of Intelligence Outsourcing (2008). New York, Simon & Schuster
40. Sweeney, L.: Discrimination in online ad delivery. Commun. ACM **56**(5), 44–54 (2013).
41. Weber, P.: Confused by All the New Facebook Genders? Here's What They Mean (2014). Slate. http://www.slate.com/blogs/lexicon_valley/2014/02/21/gender_facebook_now_has_56_categories_to_choose_from_including_cisgender.html
42. Winner, L.: Do Artifacts Have Politics? (1980). Daedalus, 109(1): 121-136.
43. Zwitter, A.: Big data ethics. Big Data & Society **1**(2) (2014).

All URLs were last accessed on Feb 23rd, 2015.