

Ethics for Big Data and Analytics

Daniel E. O'Leary

University of Southern California

© 2016

With Big Data and Analytics May Come Opportunities ...



Big Data has Led to Some Good News and Bad News - 1



STUART PALLEY FOR THE TIMES

GERI SIEGEL of Encino holds a solicitation for health insurance for her late husband, Martin, who died 14 years ago. He would have turned 65 this year.

Selling insurance to the deceased

DAVID LAZARUS

Being dead hasn't stopped Martin Siegel from being a solid insurance prospect.

Siegel, who died in 2002, would have turned 65 this year. As such, he's received solicitations from UnitedHealthcare, Kaiser Permanente and other insurers trying to sign him up for Medicare Advantage plans.

"It's just about every insurance company you can think of," said Geri Siegel, his widow.



"I receive his Social Security in my name, so the U.S. government obviously knows that he has been deceased for 14 years," the Encino resident said. "How could something like this happen?"

The answer to that question leads us into the murky world of "big data," in which people's personal information is bought, sold and shared by public and private entities.

One wrong data point can have a cascade effect throughout the cyberspace, creating erroneous files in potentially thousands of databases. I have some tips below.

Bad data is how a fraud
[See Lazarus, C4]

Big Data has Led to Some Good News and Bad News - 2

Medical privacy threatened by 'big data'

DAVID LAZARUS

The future of the U.S. healthcare system will be influenced to a large extent by a company that makes weapons of war.



"big data" — using advanced technology to predict people's future healthcare needs based on their interactions with doctors, hospitals and pharmacies, as well as information gleaned from other sources, such as social media.

Such systems, known as predictive analytics, aim to make healthcare more efficient and effective by opening the door to addressing medical issues before they become serious problems.

For example, you've complained to your doctor that you're having trouble losing weight. You're taking a cholesterol medication. You've posted on Facebook that you're feeling stressed because you've separated from your spouse or on LinkedIn that you're

[See Lazarus, C4]

Defense giant Northrop Grumman has signed a nearly \$92-million contract with the Centers for Medicare and Medicaid Services to build the second phase of a computer system that's currently focused on reducing fraud but down the road will play a greater role in anticipating beneficiaries' medical disorders.

It's the most prominent example of how public and private insurers are spending millions of dollars on



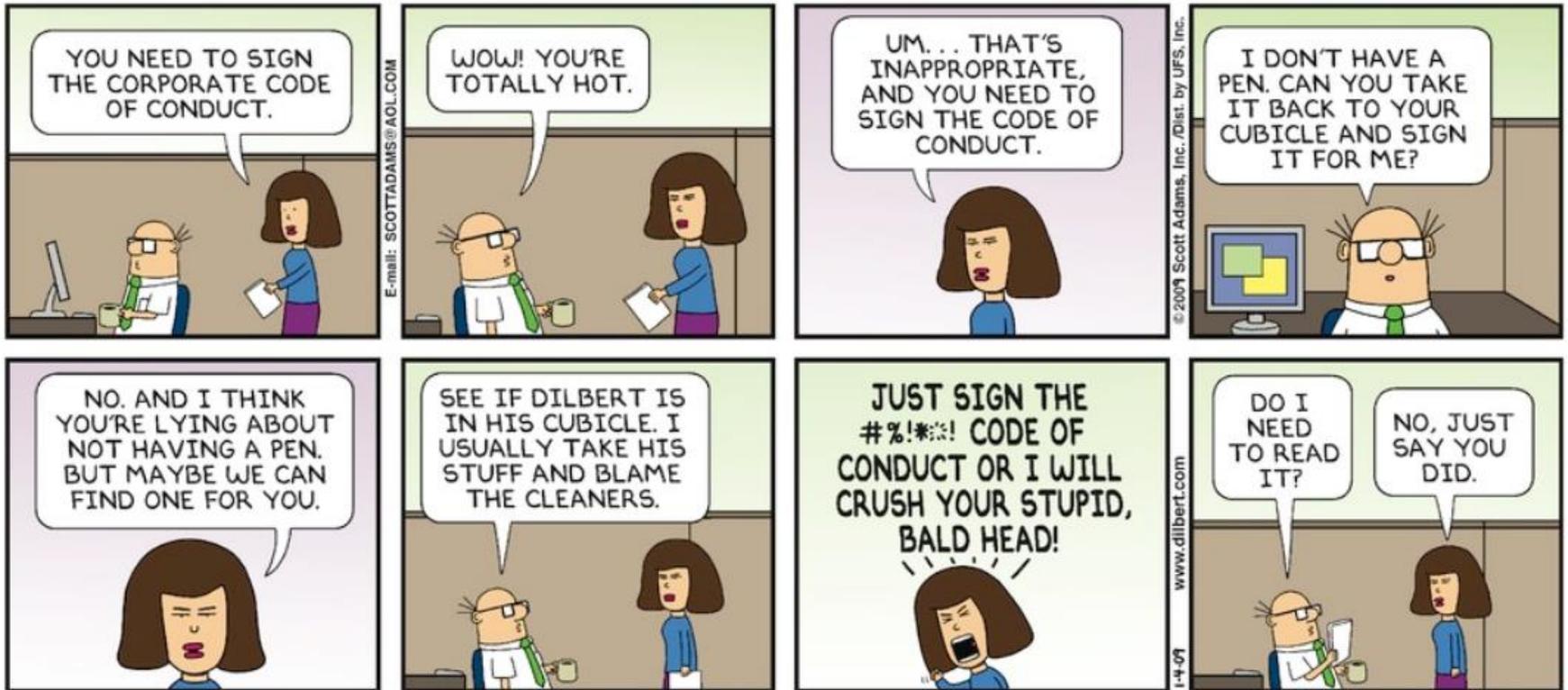
JAMIE GRILL/Getty Images/Tetra images RF

THE SHARING of healthcare data from various sources can help doctors anticipate patients' future needs, but such information comes at the expense of privacy.

Ethics in the World

- Ethics and computer ethics manifest themselves in the world as “codes of ethics” and “codes of conduct” developed by different organizations.
 - Companies build codes of conduct and ethics
 - Professional organizations build codes of conduct and ethics
 - Is anyone aware of such codes of conduct?

Codes of Conduct



Purposes of Codes of Ethics

- Thomas Wotruba and colleagues and others have suggested that such codes of ethics have at least three purposes. Whenever a group puts together a code of ethics, it indicates that the group is
 - concerned about ethics
 - transmitting the specific set of ethics to its group, and
 - ultimately affecting the group's behavior.
- *In addition*, codes of ethics provide a signal to those that interact with the relevant group as to what to expect of the group members.

Issues / Questions

- Is there a need for “big data” or “analytics” ethics or codes of ethics?
 - Is “regular” ethics enough?
 - Is “computer ethics” enough?
 - Does anyone already do big data / analytics ethics?
 - What would be a definition of big data / analytics ethics?

Computer Ethics

- Wikipedia defines computer ethics as “a part of practical philosophy (concerned with) ... how *computing professionals* should make decisions regarding professional and social conduct.”
- James Moor defined it as “the analysis of the nature and societal impact of computer technology and the corresponding formulation and justification of policies for the ethical use of such technology.”
- These definitions suggest a *strong tie between ethics and professional conduct* and an approach for influencing that conduct through *policies and rules*.
 - Determine what are those policies and rules

History of Computer Ethics

- Computer ethics has a history going back to the 1940s. Some researchers have argued that Norbert Wiener was among the first to suggest the notion of “computer ethics” (although he did not use the term “computer”).
- *“It has long been clear to me that the modern ultra-rapid computing machine was in principle an ideal central nervous system to an apparatus for automatic control... Long before Nagasaki and the public awareness of the atomic bomb, it had occurred to me that we were here in the presence of another social potentiality of unheard-of importance for good and for evil.”*

Two Points of View - Computer Ethics

- There is the equivalence of a debate in computer ethics regarding the role of computer ethics in the broader view of ethics.
- “Wiener–Maner–Górniak perspective”
 - Sees computer technology as ethically revolutionary, requiring human beings to re-examine the foundations of ethics and the very definition of a human life.
 - This perspective suggests that there is a need for a special branch of ethics for computer ethics.
- Johnson provides a more conservative perspective.
 - In that point of view, fundamental ethical theories will remain unaffected—that computer ethics issues are simply the same old ethics questions with a new twist— and consequently computer ethics as a distinct branch of applied philosophy will ultimately disappear.

Codes of Ethics and Conduct Related to Computers, Analytics and Big Data

Organization	Documentation
IEEE	Ethics and Member Conduct (www.ieee.org/about/ethics.html)
ACM	ACM Code of Ethics and Professional Conduct (www.acm.org/about-acm/acm-code-of-ethics-and-professional-conduct)
British Computer Society	Code of Conduct for BCS Members (www.bcs.org/upload/pdf/conduct.pdf)
Data Science Association	Data Science Code of Professional Conduct (www.datascienceassn.org/code-of-conduct.html)
INFORMS Certified Professional for the Analytics	Code of Ethics for Certified Analytics Professionals (www.informs.org/Sites/Certified-Analytics-Professional-Program/CAPs/CODE-OF-ETHICS)
American Association Statistical	Ethical Guidelines for Statistical Practice (http://biostat.mc.vanderbilt.edu/wiki/pub/Main/HeitmanSeminarMay08/ASAEthicalGuidelinesforStatisticalPractice.pdf)

Problems with Computer Codes of Ethics

- Which one(s) do I use?
- If I am a member of multiple societies, which dominates? Or, do I use all of them.
- How do I determine which “fits” best?
- Is a computer code of ethics enough? Do I need a specific big data or analytics (or both) code of ethics?

Case *for* Specific Big Data and Analytic Ethics (3 Primary Issues)

1. Computing ethics are about the computing artifact, while big data and analytics are about the data and what is done with it
2. The existence of specific codes of conduct for analytics and big data provide empirical evidence that they are different than computing ethics
3. The lack of specificity in computing or general ethics for big data and analytic issues, suggests a need for big data and analytic codes of ethics

Drill down ...

1. Computer Ethics vs. Big Data Analytics

“Computing Artifact vs. Data”

- Because there is controversy with computer ethics compared to general ethics frameworks, there can be a controversy over whether issues such as big data and analytics belong in computer ethics, or if they should be treated on their own.
- The initial focus of the need for computer ethics appears to have centered on the nature of the *computing artifact*.
 - As Wiener noted, “Cybernetics takes the view that the structure of the machine or of the organism is an index of the performance that may be expected from it.”

1. Computer Ethics vs. Big Data Analytics

“Computing Artifact vs. Data”

- However, the focus on big data is more concerned with *what is being processed*, the nature of what is being processed, the findings of analyzing the data and who the processing is being done for or by.
 - For example, big data has characteristics of volume, velocity, and variety, distinguishing it from other information being processed, such as transaction data. In addition, big data projects could be for individuals, organizations, or clients. As a result, issues such as data confidentiality and privacy can be a concern.
- Thus, computing ethics relates to the computer artifact, while big data relates to the data and analytics to the way that data is analyzed.
 - The artifacts are different

2. Analysis of Analytics and Big Data Codes of Conduct (1/2)

- The codes of ethics provide “empirical” evidence of the potential importance of a specific focus on big data ethics as compared to computer ethics or more general forms of ethics.
- First, the numbers of different codes of conduct for big data are one signal that there is something different about big data.
- Second, codes of conduct relating to big data come from multiple disciplines (not just computing): computing, statistics, operations research, and data science.
 - Big data and analytics appears to be multidisciplinary.
 - Some of those disciplines do not directly derive from computing.

2. Analysis of Analytics and Big Data Codes of Conduct (2/2)

- Third, in some cases the codes of conduct establish a vocabulary to ensure the appropriate communication of key concepts.
 - Perhaps the Data Science Association provides the most comprehensive vocabulary.
- Thus, the codes provide empirical evidence of difference

3. Lack of Specificity in Codes of Conduct

- Still another approach to ascertaining the extent to which big data ethics differ from other ethics frameworks is to apply existing general ethical frameworks or more specific computer ethics frameworks to big data ethics issues.
- As an example of using an existing general ethical framework to generate and facilitate analysis of ethical issues in big data, David Ross laid out seven basic axioms of right and wrong conduct.
- Two of those axioms potentially relate to analysis of big data:
 1. One ought to do what one can to improve the lot of others.
 2. One ought not to injure other people.

3. Lack of Specificity in Codes of Conduct

- It can be argued that many applications of big data are aimed at the first item (“improve the lot of others”).
- As an example, Jamie Cattell and colleagues argue that big data is being used to transform the United States healthcare system, improving pharmaceutical drug research with more timely, less constrained, and more effective analysis.
 - Big data allows analysis of both the main effects and side effects of drugs, facilitating greater innovation.
- As another example, big data can be used to facilitate smart cities, gathering sensor data from the Internet of Things (IoT) to facilitate improved city management

3. Lack of Specificity in Codes of Conduct

- Another approach is to apply a general computer framework to big data, such as the Ten Commandments of Computer Ethics. Of the 10, two appear potentially to apply directly to big data:
 3. Thou shalt not use a computer to harm other people.
 4. Thou shalt not use a computer to bear false witness.
- We can compare these principles to better understand the similarity of the two frameworks.
 - #3 and #4 could easily be made more general (“thou shalt not bear false witness”).
 - #1 is positive, whereas 3 and 4 are negative, suggesting that although general ethics frameworks may include positive rules, the more specific computer framework is largely negative, indicating what not to do.

3. Lack of Specificity in Codes of Conduct

- Both of these approaches illustrate that ethical frameworks can be applied to big data concerns. However, these ethical frameworks are focused on other settings, thus limiting their effectiveness for big data.
- These approaches illustrate that the application of such frameworks does not capture the full scope of ethical issues in big data.
- Perhaps the primary limitation is the lack of specificity that comes from applying an ethics framework that is more general than the use capabilities of a specific technology, such as big data.
 - As an example of greater specificity, the Data Science Association's code of conduct provides more ethical rules that directly draw on knowledge from the specific discipline

Specificity in Big Data and Analytic Codes of Ethics

- General codes of ethics and codes of conduct are not aimed at issues of concern with big data and analytics.



Definition of Big Data and Analytics Ethics (1/2)

- This discussion suggests that big data ethics differ from general ethics and computer ethics, as illustrated by
 - the differences between the artifacts,
 - the different emerging codes of ethics, and
 - the lack of specificity in existing computer or general ethical frameworks.
- Because of these differences, based on the previous research, I generated a potential parallel definition for big data ethics as “the analysis of the nature and societal impact of big data technology and the corresponding formulation and justification of policies for ethical use of big data.”
 - Similarly for analytics

Definition of Big Data and Analytics Ethics (2/2)

- Such a definition treats computing and big data as different technologies that require different sets of policies.
 - Unfortunately, with the development of a new technology, people and organizations do not fully understand what kinds of behavior to expect.
 - As a result, the rules and policies in place might not provide the appropriate guidance and control over behavior and might require greater specificity.
 - Codes of ethics can be developed to provide those guidelines.

Extension to Other Technologies

- Ultimately, this discussion is bigger than big data or analytics and can be generalized to a range of other types of technologies.
- For example, there is movement toward codes of ethics being designed around other technologies, such as the IoT (Internet of Things)
- Although many see the IoT as a source of issues associated with big data, it is likely that there will be important ethics specificity that can be generated for the IoT technology through its own code of ethics.

Codes of Ethics and Technology Life Cycles

- Furthermore, at one level, the existence of a code of ethics or conduct provides a signal as to where a technology is in its life cycle.
- Codes are developed, in part, to provide constraints on behavior.
 - Thus, development of codes of ethics indicates use of a technology and development of a set of rules to control that usage behavior.
- In general, the further along in the life cycle, the more likely the existence of one or more codes of ethics, and the more stable those codes are likely to be.
 - The further in the life cycle the more we know about what can go wrong, and what is needed to be constrained.

Summary

- Big data and analytics provide a setting for codes of ethics designed around the specific technologies:
 - *“The analysis of the nature and societal impact of big data technology and the corresponding formulation and justification of policies for ethical use of big data.”*

Questions?