Editor: **Daniel E. O'Leary,** University of Southern California, oleary@usc.edu

Ethics for Big Data and Analytics

Daniel E. O'Leary, *University of Southern California*

There are many definitions of "computer ethics." For example, Wikipedia defines it 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."[1] These definitions suggest a strong tie between ethics and professional conduct and an approach for influencing that conduct through policies and rules.

Computer ethics has a history going back to the 1940s. Some researchers (including Terrell Ward Bynam[2,3] and others) have argued that Norbert Wiener was among the first to suggest the notion of "computer ethics" (although he did not use the term "computer"). In particular, as Wiener noted[4]:

> 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.

Furthermore, computer ethics has a heritage building on that substrate, with a substantial literature examining key related issues. As a result, it probably is not surprising that there are multiple definitions, and some controversy, as to what constitutes computer ethics.

There is the equivalence of a debate regarding the role of computer ethics in the broader view of ethics.[5] One view, referred to as the Wiener–Maner–Górniak perspective,[6–9] "sees computer technology as ethically revolutionary, requiring human beings to re-examine the foundations of ethics and the very definition of a human life."[5] This perspective suggests that there is a need for a special branch of ethics for computer ethics. The contrasting point of view generated by Johnson provides a more conservative perspective.[10] 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."[5]

Purposes of Codes of Ethics and Codes of Conduct

Ethics and computer ethics manifest themselves in the world as "codes of ethics" and "codes of conduct" developed by different organizations. Table 1 offers sources for some of those codes of conduct, and additional codes of conduct and ethics are available elsewhere (see, for example, https://ethics.csc.ncsu.edu/basics/codes).

Thomas Wotruba and colleagues[11] and others (such as Fred Zacharias[12]) 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 who interact with the relevant group as to what to expect of the group members.

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

1541-1672/16/$33.00 © 2016 IEEE
Published by the IEEE Computer Society

Table 1. Codes of conduct and ethics related to 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 for the Certified Analytics Professional	Code of Ethics for Certified Analytics Professionals (www.informs.org/Sites/Certified-Analytics-Professional-Program/CAPs/CODE-OF-ETHICS)
American Statistical Association	Ethical Guidelines for Statistical Practice (http://biostat.mc.vanderbilt.edu/wiki/pub/Main/HeitmanSeminarMay08/ASAEthicalGuidelinesforStatisticalPractice.pdf)

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.

Computer Ethics versus Big Data and Analytics Ethics

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. For example, as Wiener noted,[6] "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." However, the focus on big data is more concerned with what is being processed, the nature of what is being processed 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.[1] 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.

Big Data Ethics and Codes of Ethics

Operationally, big data ethics are integrated into disciplines and user behavior using codes of conduct. Various aspects of big data are being claimed by a range of disciplines, including computing, statistics, and data sciences. Several potential sources of codes of conduct exist to help guide analysts in the investigation of big data projects. As Table 1 shows, an analysis of potentially relevant codes of conduct would lead to at least five codes that conceivably would be appropriate to follow in the analysis of a big data project. As a result, it is interesting to try to imagine which codes of conduct to follow in which situation for those doing big data. In addition, that count ignores other potential codes of conduct that might result because of the domain in which the analyst is operating (for example, finance or marketing), which would further complicate the choice of which code to follow. Finally, in addition to a wide range of codes of ethics, other artifacts also exist—for example, the Ten Commandments of Computer Ethics (http://cpsr.org/issues/ethics/cei).

Another approach to ascertaining the extent to which big data ethics differ from computer ethics is to examine the codes of conduct to determine the evolution of ethics and thoughts about ethics in those codes. This analysis results in several observations. 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: computing, statistics, operations research, and data science. These codes capture the notion that big data appears to be multidisciplinary. Furthermore, some of those disciplines do not directly derive from computing. Third, in some cases the codes of conduct establish a vocabulary to ensure the appropriate communication of key concepts. As an example, perhaps the Data Science Association provides the most comprehensive vocabulary. Thus, 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.

Application of Existing Ethical Frameworks

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 duties of right and wrong conduct.[13] Two of those duties potentially relate to analysis of big data:

- Duty 5: One ought to do what one can to improve the lot of others.
- Duty 4: One ought not to injure other people.

It can be argued that many applications of big data are aimed at Duty 5. 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.[14] 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.[15] Perhaps a key ethical principle concern for big data analysts is Duty 4. In particular, this would suggest that the use of big data should not be aimed at causing harm.

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:

- Commandment 1: Thou shalt not use a computer to harm other people.
- Commandment 5: Thou shalt not use a computer to bear false witness.

We can compare these principles to better understand the similarity of the two frameworks. For example, Duty 4 and Commandment 1 are quite similar in meaning. Commandment 5 could easily be made more general ("thou shalt not bear false witness"). Duty 5 is positive, whereas Commandments 1 and 5 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.

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 (see the "Data Science Association Code of Conduct Rules" sidebar). By focusing on data science, they can be more specific.

Big Data Ethics and Emerging Issues

This discussion suggests that big data ethics differ from 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." 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.

Ultimately, this discussion is bigger than big data 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 IoT (see Table 2). Although many see 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 IoT technology through its own code of ethics.

Data Science Association Code of Conduct Rules

- Rule 1: Terminology
- Rule 2: Competence
- Rule 3a: Abide by client decisions
- Rule 3b: No criminal or fraudulent activity
- Rule 4: Communication with clients
- Rule 5: Confidential information
- Rule 6: Conflicts of interest
- Rule 7: Duties to prospective clients
- Rule 8: Data science evidence, quality of data, and quality of evidence
- Rule 9: Misconduct
- Rule 9d: (Do not) engage in conduct that is prejudicial to methods of science
- Rule 9e: (Do not) misuse data science results to communicate a false reality or promote an illusion of understanding

Table 2. Internet of Things (IoT): movement toward codes of ethics.

Title of information source	Documentation
IOT Design Manifesto 1.0	http://iotmanifesto.org
IOT Trust Framework—Security, Privacy & Sustainability	http://otalliance.org/initiatives/internet-things
FTC Report: Privacy & Security in a Connected World	www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf

Ethical Interaction Effects

Although I have focused primarily on big data and ethics, similar comments could be made for analytics and its relationship with computer ethics. However, there also can be "ethical interaction effects" between big data and analytics (and computers). As an example, in many datasets, a "who" query is not likely to be an ethical or privacy issue. For example, when companies sell goods, they need to know who purchased them. However, in the case of 911 data, "who" likely is information of critical concern that deserves privacy and for which the release would not be ethical. Additional interaction effects might relate to the events analyzed or the thresholds used as part of a monitoring process. For example, too tight a threshold on an analytic could result in broad-based monitoring beyond ethical boundaries. Accordingly, both the data and the analytics need to be accounted for in codes of ethics and conduct, and there can be interaction effects.

So, what is the impact of this discussion on organizations in practice? First, codes of ethics provide a signal as to where a technology is in its life cycle that can be used to monitor technology development. Second, organizations likely have not captured the evolving nature of new technologies in their own codes of ethics. As a result, adapting to new technologies involves not only monitoring development in technologies, but also monitoring and implementing developments in new codes of ethics. Third, empirically, codes of ethics are largely negative, bounding behavior by what not to do. Unfortunately, such bounds can rarely fully anticipate what people might create or how they might behave until a given technology becomes more fully developed. Furthermore, such bounding largely ignores potential positive guidelines or suggestions. ■

References

1. J. Moor, "What Is Computer Ethics?" *Metaphilosophy*, vol. 16, no. 4, 1985, pp. 266–275.
2. T. Bynam, "Norbert Wiener's Vision: The Impact of 'the Automatic Age' on Our Moral Lives," 2005; www.cs.ox.ac.uk/activities/ieg/e-library/bynum.pdf.
3. T. Bynam, "Norbert Wiener's Foundation of Computer Ethics," 2011; http://rccs.southernct.edu/norbert-wieners-foundation-of-computer-ethics.
4. N. Wiener, *Cybernetics: Or Control and Communication in the Animal and Machine*, 2nd ed., 1961, MIT Press, pp. 26–27.
5. T. Bynam, "A Very Short History of Computer Ethics," *Newsletter on Philosophy and Computing*, Am. Philosophical Assoc., Summer 2000; www.cs.utexas.edu/~ear/cs349/Bynum_Short_History.html.
6. N. Wiener, *The Human Use of Human Beings*, Houghton Mifflin, 1954, 2nd ed., Doubleday Anchor, p. 57.
7. D.E. O'Leary, "'Big Data,' the 'Internet of Things,' and the 'Internet of Signs,'" *Intelligent Systems in Accounting, Finance and Management*, vol. 20, no. 1, 2013, pp. 53–65.
8. W. Maner, *Starter Kit in Computer Ethics*, Helvetia Press, 1980.
9. K. Gorniak-Kocikowska, "The Computer Revolution and the Problem of Global Ethics," *Science and Eng, Ethics*, vol. 2, no. 2, 1996, pp. 177–190.
10. D. Johnson, *Computer Ethics*, 2nd ed., Prentice-Hall, 1994.
11. T.R. Wotruba, L.B. Chonko, and T.W. Loe, "The Impact of Ethics Code Familiarity on Manager Behavior," *J. Business Ethics*, vol. 33, no. 1, 2001, pp. 59–69.
12. F.C. Zacharias, "Specificity in Professional Responsibility Codes: Theory, Practice, and the Paradigm of Prosecutorial Ethics," 69 *Notre Dame Law Rev.* 223, 1993.
13. W.D. Ross, *The Right and the Good*, Oxford Univ. Press, 1930.
14. J. Cattell, S. Chilukuri, and M. Levy, "How Big Data Can Revolutionize Pharmaceutical R&D," McKinsey, Apr. 2013; www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/how-big-data-can-revolutionize-pharmaceutical-r-and-d.
15. D. O'Leary, "Exploiting Big Data from Mobile Device Sensor-Based Apps: Challenges and Benefits," *MIS Quarterly Executive*, vol. 12, no. 4, 2013, pp. 179–187.

Daniel E. O'Leary is a professor in the Marshall School of Business at the University of Southern California. Contact him at oleary@usc.edu.