Analyzing Codes of Ethics in Big Data Analytics: IBM, Splunk, and Databricks

Piper Macke

Columbian College of Arts and Sciences, George Washington University

DATS 2101: Ethical Life in a Digital World

Prof. Jamie Cohen-Cole

December 19, 2023

Introduction

In the professional field of big data analytics, companies rely on detailed, individualized codes of ethics to guide employee and client behavior in a way that is most conducive to the overall success of the business. In the realm of technology and artificial intelligence, these codes reflect the importance of ensuring proper data protection, workplace expectations, and confidentiality of sensitive information.

IBM, Splunk, and Databricks are three companies that are well-known in the big data world for their contributions to the industry as well as their continued success as innovators. They each outline the ethical and legal guidelines that all affiliates of the organization are expected to follow to keep the business running smoothly and without risk of internal or external failure.

Comparing Ethical Dimensions

Legal, Ethical, or Both?

Perhaps the most critical and explanatory dimension of each code of ethics is the legal and ethical grounds on which each code is founded. IBM, as the most established and longest-running business of the three, reasonably also sets forth the lengthiest code of ethics, or *Business Conduct Guidelines* (BCGs). While IBM's BCGs are carefully detailed, there is more emphasis put on employees acting in an honest and trustworthy manner as opposed to simply abiding by the law. Because IBM boasts its status and "one of the world's most ethical companies," it sensibly also holds all employees and partners to a higher standard than one that is contingent solely on fulfilling their legal obligation as a citizen (Business Conduct Guidelines, p. 6).

Contrasting IBM's ethics focused BCGs, Splunk's *Code of Business Conduct and Ethics* is aligned with a more legal framework. At Splunk, almost any ambiguity that may arise from the code is put in

the hands of the Splunk Legal Team, whereas concerns with the IBM BCGs are more likely to be resolved with a direct supervisor or manager. Databricks' *Global Code of Conduct* operates similarly to Splunk, requiring that reports of suspected violations of both laws and the code itself be handled by the Legal Compliance and Ethics Team at Databricks.

The variations in how these three codes are founded are naturally explained by observing each company's role in the big data industry. IBM has a significantly broader scope of involvement in both technology and consulting than Splunk or Databricks, each of which specialize almost exclusively in data analytics. A possible explanation as to why IBM's BCGs center more on ethics is due to the inevitable diversity in perspectives that is a byproduct of any historically renowned, globally established company. In Laura Stark's *The Science of Ethics: Deception, The Resilient Self, and the APA Code of Ethics*, she sheds light on how this phenomenon is not a new one, even in professions that would seem to hold a very strict set of shared standards. As Stark articulates, every moral compass points in a different direction, and each "by virtue of the distinctive environments in which they were trained and the unique life experiences they brought to bear on their work" (Stark, 2010, p. 339). A code of ethics that primarily encourages employees to act in the way they know is ethical creates a standard of behavior that is higher than any that could be explicitly defined by law.

Internal and External Compliance Measures

Given the foundational distinctions between each code of ethics, it is evident that each business also functions differently regarding its internal and external compliance measures. Internally, each company is in some capacity dedicated to confidentiality, legal compliance, and avoiding conflicts of interest. Considering that the "supreme currency" for modern businesses has shifted from dollars to data, ensuring employee and consumer protection at all levels is a common point of emphasis in big data analytics (Nissenbaum, 2011).

An interesting commonality between IBM and Splunk is that both companies state in their codes that intellectual property created by employees of the organization still belongs to the organization even if the employee departs. Conversely, Databricks does not set out strict guidelines prohibiting employees separated from the company from using their previously created IP. This is a glaring reflection of how these three companies function differently. Databricks is unique to the other two in that it was founded by creators that first championed the benefits of open-source computing with Apache Spark (Connell, 2022). Additionally, Databricks regularly contributes to other open-source projects, whereas this is not a pronounced function of either Splunk or IBM. Providing important historical context behind Databricks' support of open-source information is the 1998 Free Software Movement, known as being widely responsible for the modern perspectives and ethical justifications towards open-source software today. (Coleman & Golub, 2008).

Reference to Maintaining Reputation

Another means of compliance at both the internal and external levels for each of these three companies centers around public perception. Unsurprisingly, IBM, Splunk, and Databricks all demand that employees do not make statements on behalf of the organization unless they have been specially cleared to do so. Splunk warns especially about posting opinions on social media, and IBM strongly encourages employees not to comment publicly in almost any circumstance. However, where these organizations all align is how they direct employees to specialized teams within the company designed to deal with any media inquiries or employee violation of the communication policy. It is the collective existence of these teams that serves as an illustration of how the online sphere of society has become a place of extremely limited privacy and harsh judgment.

Individual privacy in the age of the modern internet is steadily becoming harder to accomplish, as data doubles continue to grow in value and the privilege of staying private remains a

freedom only some can fully achieve (Igo, 2018). This lack of privacy is essential to understanding why employees are so strongly encouraged not to make public statements. By enforcing these measures, big data analytics companies lend a small glimpse into their own knowledge of how data can easily be gathered and used maliciously. It goes to say that even companies with exceptional reputations are still not immune from the dangers of a breach of privacy.

Integrity of Books and Records

In big data analytics, one of the most critical aspects of ensuring safety and trust within an organization is creating and upholding standards of accurate recordkeeping. IBM, Splunk, and Databricks each emphasize how books and records are "valuable assets" to the company and must be treated as such (Business Conduct Guidelines, p. 25). Splunk warns particularly that business records regularly become public information, and for employees to be mindful of this in practice. Similarly, Databricks alludes again to upholding its positive reputation by mentioning how recordkeeping and never misrepresenting information are keys to an operational, responsible business.

These three companies feature in their recordkeeping some of the most integral parts of evidence-based governance and its reliance on quantitative and empirical knowledge (Merry, 2016). For example, Databricks asserts that any claims made on behalf of the company and its services must be "substantiated with reasonable supporting data," (Databricks, 2023). Splunk similarly expects all business records to be kept in "reasonable detail" to avoid any possible misinterpretation in the future (Splunk, 2022). Both of these measures are tributes to the necessity of proper data and record storage. Especially for companies in big data analytics, such guidelines act as stand-out features of any reputable business.

Reference to Government Agencies and Laws

Observing how IBM, Splunk, and Databricks each have varying approaches to governmental compliance is yet another example of how the companies differ according to what they specialize in. Because IBM is an internationally established corporation with offices in dozens of countries worldwide, the BCGs reference on multiple occasions to abide by laws of the country in which business is conducted. The BCGs caution employees to comply with legal regulations such as the U.S. Foreign Corrupt Practices Act, Brazil's Clean Company Act, and the UK Bribery Act. Databricks also includes in its code of ethics that employees are to adhere to U.S. and non-U.S. laws, especially those surrounding human rights and labor standards. Emphasizing that employees are bound both by the laws of the U.S. and any foreign country they may do business with is a telling facet of how all the codes heavily value respect as a quality of the organization.

Specific legislature mentioned in Splunk's guidelines includes the Federal Acquisition Regulation, whose purpose is to ensure the fairness and transparency of government purchases. A prominent aspect of Splunk's business operations includes the global trading of hardware, software, and other technology for monitoring and analyzing real-time and historical data. Technology with this purpose has long blurred the lines between what is legal and illegal, and there is no shortage of internet hackers that have used it both maliciously and with good purpose over the last several decades (Coleman, 2012). The ease in which this advanced technology can be exploited gives reason as to why Splunk's guidelines are crafted to ensure that any trading of valuable materials is approved by U.S. statutes.

Conclusion

After comprehensively analyzing the codes of ethics at IBM, Splunk, and Databricks, it is clear that there exist many common values between companies in the big data analytics field. The

fact that there are only mild variations at the most specific levels of the codes implies that there are underlying values and ethical standards that strongly guide the operation of each company.

The variations that do exist in the codes correlate most directly to the services and products the company offers. Because IBM, Splunk, and Databricks are all involved in areas of technology that are constantly improving in both efficiency and data capacity, these codes are a means to navigate areas of uncertainty before they arise.

Since they also stress the importance of employees acting ethically by their own intrinsic standards, the codes show their reliance on natural morality. Ethical behavior is defined in a way that is unique to an individual and their lived experience. While this may lead to variations in the way these individuals conduct business, many of their most central ethical values overlap. This is a testament to how, in many cases, cultures share the same fundamental principles for behavior. Implementing these similarities into how a company operates allows it to be reliant on each individual acting in the way they know is right. As a result, the codes of ethics companies truly abide by are not as much the ones written on paper as they are the intrinsic ethics that exist within each individual.

References

Coleman, E. G. (2012). Phreaks, hackers and trolls. The social media reader, 99-119.

Coleman, E. G., & Golub, A. (2008). Hacker practice: Moral genres and the cultural articulation of liberalism. Anthropological Theory, 8(3), 255-277.

Connell, C. (2022, January 14). Databricks - A history. Medium. https://medium.com/@chuck.connell.3/databricks-a-history-d8dd12fe9695

Databricks, Inc. (2023, January 30). Global Code of Conduct. Databricks. https://www.databricks.com/sites/default/files/2023-10/global-code-of-conduct-9-27-2023.pdf

Igo, S. E. (2018). The known citizen: A history of privacy in modern America. Harvard University Press.

International Business Machines Corporation. (2023). Business Conduct Guidelines. IBM. https://www.ibm.com/investor/att/pdf/IBM_Business_Conduct_Guidelines.pdf

Merry, S. (2016). The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking. Chicago: University of Chicago Press. https://doi.org/10.7208/9780226261317

Nissenbaum, H. (2011). A contextual approach to privacy online. Daedalus, 140(4), 32-48.

Splunk Inc. (2022, December 8). Code of Business Conduct and Ethics. Splunk. https://www.splunk.com/en_us/pdfs/legal/code-of-business-conduct-and-ethics.pdf

Stark, L. (2010). The science of ethics: Deception, the resilient self, and the APA code of ethics, 1966–1973. Journal of the History of the Behavioral Sciences, 46(4), 337-370.