Ethics in Digital Data Gathering

Carlos Ocasio Rodriguez

Charleston Southern University

CSCI 325-401: Object-Oriented Programming

Prof. Henderson

October 23, 2020

Abstract

In the era of Big Data, it is now possible to capture and store vast amounts of data about users, and their routine online browsing meant for the delivery of custom-made services using predictive algorithms, as witnessed in the fields including government, online commerce, research, and politics (Saqr, 2017). With the constantly expanding capabilities of big data, so does the ethical dilemmas associated with current privacy self-management policies. Privacy self-management through notifications and optional agreement to terms gives the user the autonomy in regulating access to personal information by consenting or denying access. However, in reality, data continues to be collected, used, analyzed, and sometimes abused and shared with third parties. The users do not always read or fully understand consent. Those users who fail to accept may not get certain services, may choose an inappropriate alternative, or sometimes lack alternatives. These conditions make an agreement to terms mandatory, and hence users have no control over consent. This paper explores data gathering techniques and the importance of obtaining consent from users in order to maintain proper ethical policies in a business environment.

Keywords: Privacy, ethics, information technology

Ethics in Digital Data Gathering

With the rapid development in the field of data science, the speed has superseded the legal and ethical values related to the collection, storage, and analysis of user data. In computer science, some well-established guidelines and organizations govern the ethical conduct of Information Technology (IT) professionals. These Include the ACM (Association for Computing Machinery) and the IEEE (Institute of Electrical and Electronics Engineers) code of Ethics. However, some uses of personal data may sometimes challenge these guidelines.

Nonetheless, the bulk of data generated by users on the internet remains accessible, obtainable, and can be hacked, provided one gains entry to databases. Since some users have a sense of ownership/privacy of the data they generate, the usage of their data without consent may generate conflict (Jurkiewicz, 2018). Accordingly, many social media companies and government agencies rarely admit their continuous monitoring and storage of personal data, including content, demographics, physical and mental health, locations, search history, purchase history, among others. Therefore, whenever users access websites, applications, or social media forums, they unknowingly allow these companies to collect, use, and sell their data without limitations.

Notably, the aspect of informed consent is important in data collection. The idea of informed consent outlines that consent should be sought before data collection and that users should have a sound understanding of the implications of sharing their personal data and the possible consequences. This is outlined in the ACM Ethical principles of avoiding harm, respecting privacy, and honoring confidentiality (ACM Code of Ethics and Professional Conduct, 1998). The Fifth IEEE Code of ethics also requires that IT professionals “improve the

understanding by individuals and society the capabilities and societal implications of conventional and emerging technologies including intelligent systems." (Pugh, 2009). The modern technologies, however, contravenes these provisions due to automatic capture or lack of informed consent by users. The modern computer algorithms record everything attributed to the possible future usefulness of data and the cheap nature of recording the data.

Additionally, the low cost of gathering vast amounts of data implies that users cast long data shadows captured in daily activities, including credit cards, travels, phone calls, and interactions with intelligent personal assistants. Such data can reveal individual activities, social circles, personal interests, and their beliefs (Hand, 2018). The ethical dilemma for IT personnel comes with the fact that while such personal data is valuable to track terrorists and criminals, but may also fall into the wrong hands, including authoritarian governments for control of its law-abiding citizens through blackmail and embarrassments. Furthermore, data can be stored indefinitely for future uses when new algorithms are produced that can use the data for prediction, and that time the users cannot be sought for their consent (Saqr, 2017).

The collection of data on individual online habits and predilections for financial gain through the sale of this information or for customized advertisement while publicly promoting it as beneficial for service provision to users is a main ethical dilemma that any IT professional is likely to encounter (Jurkiewicz, 2018). Transparency should guide IT technicians to avoid this ethical dilemma when working with data as required of computing professionals under the ACM professional responsibility that they "foster public awareness and understanding of computing, related technologies, and their consequences." (ACM Code of Ethics and Professional Conduct, 1998). Consequently, the users should be informed of the intentions of the collected data and the

processes involved in data collection, storage, and analysis. The users should also be informed of the involvement of third parties in their data processing and that their data would only be sold to third parties after seeking their consent.

Personally, I recognize the vast range and amount of data collected by modern technologies as a double-edged sword, in that, while it is useful in improving quality of life, proprietary uses may pose an ethical dilemma concerning consent by the user. Currently, big data employs a concept of obtaining consent by default, where users agree to terms and conditions without reading. Such terms openly defy the codes of ethics and professional conduct stipulated under ACM and IEEE. In my professional life, I would advocate for active acquisition of consent through lobbying for IT companies to comply with regulations on privacy policy language, data storage, opt-in and opt-out language, and data sharing. This will ensure the responsible use of data that consent was obtained and limits how the data may be used, ensuring actions when the data is abused. By creating a data-sharing loath policy, a company will ensure limited access to user data by third parties.

References

ACM Code of Ethics and Professional Conduct. (1992). *Communications of the ACM*, *35*(5), 94–99. https://doi.org/10.1145/129875.129885

Hand, D. J. (2018). Aspects of data ethics in a changing world: Where are we now? *Big Data*, *6*(3), 176–190. https://doi.org/10.1089/big.2018.0083

Jurkiewicz, C. L. (2018). Big Data, big concerns: Ethics in the Digital age. *Public Integrity*, *20*(sup1). https://doi.org/10.1080/10999922.2018.1448218

Pugh, E. W. (2009). Creating the IEEE code of ethics. *2009 IEEE Conference on the History of Technical Societies*. https://doi.org/10.1109/hts.2009.5337855

Saqr, M. (2017). *Big Data and the emerging ethical challenges*. Researchgate. Retrieved from https://www.researchgate.net/publication/320741378_Big_data_and_the_emerging_ethical_challenges.