Privacy Concerns and the Regulatory Future of Big Data

William Barker, Scott Breitbach, Peyton Mosbaugh

Introduction

As technology has progressed in the last 50 years, computers have shrunk, processing power has increased, and computers have integrated themselves into nearly every aspect of our lives, from smart homes and cars, to social networks, online shopping, and even our Internet searches. This increase in computer technology has led to an explosion in data due to the ease of collection and the ever-decreasing cost of processing and storing large data sets.

These data are used primarily by corporations, governments, and researchers and for a wide variety of purposes. Ideally, the information generated from this data will benefit both those doing the data collection and analysis and the individuals or groups from whom the data is being collected [13].

Generally speaking, data used for research is good for everyone because the end goal is typically the pursuit of knowledge, though there can still be unintended biases, both in the data and in the algorithms we build, which we need to constantly be on the lookout for. Oftentimes we might be mistrustful of corporations [6] and governments collecting our data because corporations are often concerned primarily with their bottom line and governments might want to control their populace, however they can be good too. Targeted marketing from corporations can help a customer find a useful product that they didn't know exists and governments often use traffic data to plan roads in order to relieve congestion.

With all this data being collected, security and privacy are major concerns that arise [10]. Problems can arise both intentionally (e.g. through those using the data for unethical purposes or through third-party data hacking), and unintentionally (e.g. when inherent data biases aren't accounted for or when sufficient security measures aren't taken).

In order to protect against these privacy and security risks [19], we need to utilize a multi-pronged approach. We will need regulations to maintain checks on the industry as well as standards for the industry to follow. For those working with the data, training around ethics [13], biases, and security as well as building security into the software architecture itself [1].

Many have already started down this path, but as with any rapidly developing technology, some are lagging behind. Here we will look at the current state of privacy and security guidelines and regulations around big data as well as where it looks like the future of data privacy regulation is headed [14].

Why is this Data Science

As data creation and collection continue to rapidly expand, data scientists are struggling with figuring out where the "fine line" between using relevant data for their companies and maintaining the privacy of the consumers. Highly regulated and private environments require a lot of labor in order to write the custom controls in the code needed to avoid the risk of violating privacy regulations. However, with data increasing exponentially, the policies and regulations on how it can and should be used and collected are constantly changing to try to keep up with the fast-paced world of data science. If people feel as though they lack necessary privacy with their data collection and use, then the field of data science will be shown in a negative light to the general public instead of highlighting how useful it is and will be in all aspects of our world.

Source: https://www.fedsmith.com/2017/03/24/epa-investigating-employee-covert-activity/

Deliverables

Currently there are several approaches being pursued within the big data framework to protect privacy, such as incorporating security and privacy controls directly within the code during processing [26] and utilizing encryption in order to ensure security of the data [27] or even user-facing software approaches such as showing the usage of the data directly and requiring authentication [28]. Trust models can also be used to measure security strength of cloud services [29]. Data brokers are those that collect personal data from numerous sources and then sell that data to business worldwide. Often times this happens without consumer knowledge and can be harmful and discriminatory. While US policy makers have been urged to regulate these brokers, they remain unregulated, whereas the European Union has introduced broad data privacy guarantees for consumers and a data protection regime that limits data brokers' activities if consumers' consent has not been obtained for specific purposes. [18]

We plan on delivering a white paper, highlighting these concerns where we will discuss:

1. Building security and privacy protection into the data architecture
2. New regulations and standards being implemented in the EU
3. How these regulations lead to changes in how data is collected
4. Possible future regulations.

Acknowledgements

Thank you to Professor Shankar Parajulee and our DSC 500 classmates for their support

Literature

1. Alfredo Cuzzocrea. 2014. Privacy and Security of Big Data: Current Challenges and Future Research Perspectives. In Proceedings of the First International Workshop on Privacy and Secuirty of Big Data (PSBD '14). Association for Computing Machinery, New York, NY, USA, 45– 47. DOI:https://doi.org/10.1145/2663715.2669614
2. Cho, Do-Eun, er al. "Double Privacy Layer Architecture for Big Data Framework." International Journal of Software Engineering and It's Applications, vol. 10, no. 2, 2016, pp. 271-278., doi:10.14257/ijseia.2016.10.2.22.
3. Covert, Quentin, et al. "Towards a Triad for Data Privacy." Proceedings of the 53rd Hawaii International Conference on System Sciences, 7 Jan. 2020, doi:10.24251/hicss.2020.535
4. Kambatla, K., Kollias, G., Kumar, V.,& Grama,A.(2014). Trends in big data analytics. Journal of Parallel and Distributed Computing, 74(7), 2561-2573.
5. Lei Xu, Chunxiao Jiang, Jian Wang, Jian Yuan and Yong Ren, "Information Security in Big Data: Privacy and Data Mining," in IEEE Access, vol. 2, pp. 1149-1176, 2014. https://ieeexplore.ieee.org/abstract/document/6919256
6. Lui, Yang, and Connor Greene. "The Dark Side of Big Data: Personal Privacy, Data Security, and Price Discrimination." Digital Transformation in Business and Society, May 2019, pp145-153., doi:10.1007/978-3-030-08277-2_9.
7. Sharma, Anil, and Gurwinder Singh. "A Review of Big Data Challenges and Preserving Privacy in Big Data." Advances in Data and Information Sciences, vol. 94, 3 Jan. 2020, pp. 57-65., doi:10.1007/978-981-15-0694-9_7.
8. Shui Yu, "Big Privacy: Challenges and Opportunities of Privacy Study in the Age of Big Data," in IEEE Access, vol. 4, pp. 2751-2763, 2016. https://ieeexplore.ieee.org/abstract/document/7485855
9. Strang, Kenneth David, and Zhaohao Sun. "Big Data Paradigm: What Is the Status if Privacy and Security?" Annals it Data Science, vol. 4, no. 1,2017, pp. 1-17., doi:10.1007/s40745-016-0096-6.
10. Tene, Omer, and Jules Polonetsky. "Privacy in the Age of Big Data: A Time for Big Decisions." Vol. 63, 2 Feb. 2012.
11. Jacob Metcalf and Kate Crawford. 2016. Where are human subjects in Big Data research? The emerging ethics divide - Jacob Metcalf, Kate Crawford, 2016. (June 2016). Retrieved May 1, 2020 from https://journals.sagepub.com/doi/full/10.1177/2053951716650211
12. Mandy Chessell. Ethics for big data and analytics. 2014. PDF. (2014).
13. Scott Beardsley, Luis Eriquez, ferry Grijpink, Sergio Sandoval, Steven Spittaels, and Malin Strandell-Jansson. 2014. Building Trust: The Role of Regulation in Unlocking the Value of Big Data. *The Global Information Technology Report* (2014), 73–80.
14. Daniel Gozman, Wendy Currie, and Jonathan Seddon. 2015. The Role of Big Data in Governance: A Regulatory and Legal Perspective of Analytics in Global Financial Services. *SSRN Electronic Journal* (2015), 1–54. DOI:http://dx.doi.org/10.2139/ssrn.2752561
15. Sharyn Ohalloran, Sameer Maskey, Geraldine Mcallister, David K. Park, and Kaiping Chen. 2016. Data Science and Political Economy: Application to Financial Regulatory Structure. *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2, 7 (2016), 87–109. DOI:http://dx.doi.org/10.7758/rsf.2016.2.7.06
16. D. Mendelson and D. Mendelson. 2017. Legal protections for personal health information in the age of Big Data – a proposal for regulatory framework. *Ethics, Medicine and Public Health* 3, 1 (2017), 37–55. DOI:http://dx.doi.org/10.1016/j.jemep.2017.02.005
17. Chih-Liang Yeh. 2018. Pursuing consumer empowerment in the age of big data: A comprehensive regulatory framework for data brokers. *Telecommunications Policy* 42, 4 (2018), 282–292. DOI:http://dx.doi.org/10.1016/j.telpol.2017.12.001
18. Jensen, M. Challenges of Privacy Protection in Big Data Analytics. Proc. of BigData Congress, 2013.
19. Betgé-Brezetz, S., Kamga, G.-B., Dupont, M.-P., and Guesmi, A. End-To-End Privacy Policy Enforcement in Cloud Infrastructure. Proc. of CLOUDNET, 2013.
20. Machanavajjhala, A., and Reiter, J.P. Big Privacy: Protecting Confidentiality in Big Data. ACM Crossroads 19(1), 2012.
21. ang, M., Yoon, M., Chang, J.-W. A Privacy-Aware Query Authentication Index for Database Outsourcing. Proc. of BigComp, 2014.
22. Agrawal, R., and Srikant, R. Privacy-Preserving Data Mining. Proc. of SIGMOD, 2000.
23. Ishibuchi, H., Yamane, M., and Nojima, Y. Learning from Multiple Data Sets with Different Missing Attributes and Privacy Policies: Parallel Distributed Fuzzy Genetics-based Machine Learning Approach. Proc. of BigData Conference, 2013.
24. Agrawal, D., El Abbadi, A., and Wang, S. Secure and Privacy-Preserving Database Services in the Cloud. Proc. of ICDE, 2013.
25. Arasu, A., Eguro, K., Kaushik, R., and Ramamurthy, R. Querying Encrypted Data. Proc. of SIGMOD, 2014
26. Wu, C., and Guo, Y. Enhanced User Data Privacy with Pay-By-Data Model. Proc. of BigData Conference, 2013.
27. Rizvi, S., Ryoo, J., Liu, Y., Zazworsky, D., and Cappeta, A. A Centralized Trust Model Approach for Cloud Computing. Proc. of WOCC, 2014

Conclusion

A majority of Americans describe themselves as confused, concerned, and lacking control when describing their relationship with their data privacy. As people are constantly tracked through their smart devices, some feel that the data collection poses more risks than benefits. Most of their concerns are centered around their lack of knowledge and understanding of how their data is being used by major companies. The future of regulations can either help people feel more secure due to clear guidelines or the insecurities can increase, leading to ill feelings towards data collection and data science.