Adam Johnson

Professor: Dr. Shamin Hunt

PHIL 355E

01 October 2023

Case Analysis 2.4: User Data

In Danny Palmer's article "What is the GDPR", I learned about the General Data Protection Regulation that was enacted in 2018. The scope of the GDPR applies to any organization operating inside the EU or any organization that offers services to consumers or businesses in the EU. The GDPR focuses on transparency of user data and how it's used, requiring organizations and those who collect data to follow conditions set out by the GDPR or face fines. Under the GDPR those who are collecting data are required to protect that data from misuse and exploitation. The GDPR addresses privacy concerns about personally identifiable information differently than has been done before. The GDPR expands on PII by identifying specific technology based identifying information such as an IP address as personal identifying information. In this case analysis I will argue that the deontology viewpoint shows us that the United States should follow Europe's lead because user data is personal and attached to an individual. This data should be protected and used ethically, not just for profit but also with respect and consideration for the people who the data is collected from.

In Zimmers analysis of the T3 researchers and their actions he identified many issues with how they conducted the data apprehension, analysis and handling of the data. Zimmer identified several privacy violations including, errors in personal information, unauthorized secondary use, amount of personal information collected and improper access to personal information. One of the core themes with Zimmer's analysis was that the researchers, collegiate officials and even Harvard's institutional review board didn't understand what personal identifying information actually was and how certain information could be considered PII. The researchers thought they had wiped the data clean of all PII by just removing names and identification numbers from their dataset before releasing it. They believed that all the data they collected was open to the general public and anyone could access so it shouldn't be considered private, they had permission from the college to collect the data as well as from facebook. The researchers did not seek outside security evaluations or help before releasing the data and tried to avoid responsibility by saying they were just sociologists.

At the time of Zimmer's analysis, the GDPR was not a regulation but the EU still had the Data Protection Act 1998 which maintained a similar definition of PII. In Zimmer's article he address the EU's perspective on PII which stated in the Data protection directive " 'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity;" (European Union, Article 2). Even when Zimmer uses the older version of this definition of PII he finds that the information released by the T3 researchers were in violation of PII standards. The researchers were specifically logging and releasing cultural and social data that could be easily used to identify groups and people within them. Once the public was able to identify the school based on its released programs and majors list it wouldn’t be too difficult to start identifying students who were a part of the social and cultural groups from the data that was being collected and released. This is where the GDPR regulations would have helped the students involved in the study with legal recourse and rights or to help the researchers identify the errors in the ways they collected and used the data for their research. The GDPR would have required the researchers to adhere to the guidelines and practices or face legal and monetary consequences. If the United States had something similar to the GDPR or most notably the ability to specifically define PII in a way that it protects people from exploitation it could have helped researchers, Harvard and Facebook understand what is and it’s allowed. It’s important that the US has a clear definition of PII as an understanding of what is expected since organizations and businesses aren't necessarily known for their ethical and moral approaches to things. Zimmer’s analysis shows that in terms of a deontologist viewpoint the researchers did not respect and give dignity to the students it tracked and monitored for four years. The researchers didn’t notify the students before or after that their data was being collected and analyzed, while also releasing the data to the public that contained many aspects of PII that could be related back to the students showing a blatant disregard for their personal privacy. Due to the nature of the research the students had no way of knowing what was happening and then no recourse when it came to stopping or preventing the data collection from happening in the first place. Which showed the researchers lack of respect and denial of dignity to the students it collected data from.

Elizabeth Buchanan reviewed the results of a twitter research on ISIS/ISIL by Matthew Curran Benigni. The article analyzed how the researchers collected the data and identified people associated with the terrorist organizations by using an Iterative Vertex Clustering and Classification (IVCC) model. This type of model uses artificial intelligence, algorithms and input to create a social network of twitter users that were active within the ISIS/ISIL communities online. It was able to analyze how users retweeted or reposted messages used hashtags and interacted with groups on twitter that would put them in a category for use in the research. Buchanan makes a case for the legitimacy of the data by stating that “The context and foci of the research are well-defined. The methods are technically valid and reliable.” (Buchanan, Pg.2). She believes in the researchers and their work the ability for the algorithms to produce accurate results and for the data to be reliable and verifiable. But then she makes a clarification on the ethics of the research stating “The ethics of the methods, however, are less clear, and part of this is the novelty of this form of research” (Buchanan, Pg. 2). A major concern for Buchanan were the ethics of data collection and mining and how modern algorithms can be used to target specific groups, political parties, anyone for anything without a context. Buchanan’s review touches on the bigger topic of how the data is collected and how it will affect the future. The collection of this data from twitter is a part of an algorithm data collection that involves massive amounts of information and data and it was stated in the article that “researchers would “conclude that seeking informed consent from all 119,156 participants is “impracticable” (Buchanan, Pg. 2). There are many concerns for this type of data collection and how the data can be used. How does a user know if their data is being used for advertising purposes or for the government or an AI to collect data on them and put them in an identifiable group for use you are unaware of? Buchanan ends her review with several good questions that have no real answers. She wants to know how data scientists consider their users. If the users are considered subjects or participants and if they are, what rights do they have? Another big question was “Who has the authority and power to decide how and in what contexts the IVCC and related mining methods are used, and to what end, in what context? (Buchanan, Pg. 3). I think a deontologist viewpoint would say that the algorithmic methods of data collection are immoral because they use data to create a depiction of someone through interactions on the internet, creating labels and associating people with groups and causes they may have nothing to do with. This can lead to a miss characterization of the individual using them and their data as a means, while the individual may or may not have any idea that it is even happening, giving them no recourse or input into how they are viewed and associated based on the data that was mined through AI methods. Buchanan’s article adds to the concerns of how data is currently being mined and used but also how it will be used in the future and things like the GDPR are just the initial steps in approaching current technology and its methods. So not only does the US need to adapt something similar to the GDPR but it also needs to advance on it to help include future methods of data collection like IVCC that completely change how data is collected and used.

Zimmer showed that privacy within research is a complicated thing to achieve. Even when researchers set out with good intentions and think they are doing the right thing, they may underestimate the impact specific data has. Buchanan showed that modern technology is posing a privacy concern and there isn't much being done to control or regulate it. In each case something like the GDPR helps researchers and consumers control their data and to understand what data is actually important. One of the most important aspects of something like the GDPR is the clarification or lack thereof concerning personally identifiable information. The GDPR's foundation for personal data is stated as "any information, relating to, and identified or identifiable, natural person" (GDPR). This definition can be subject to interpretation but sets the consumer up with the ability to understand their rights to privacy concerning data and helps the individual. The ability to identify what is personal data and what isn't so that personal data can be protected is something that the US should be doing just as the EU has. The US needs something on the level of the GDPR just as a starting point in the right direction for defending personal data and regulating the use of personal data by organizations from harming the individual, intentional or not. This can also help provide users with the dignity and respect they deserve, by not allowing organizations to use them for profit without any recourse due to regulations and individual rights.

References

European Union Data Protection Directive 95/46/EC. 1995 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31995L0046

Zimmer, Michael. “But the Data is already public”: on the ethics of research in facebook. June 2010. Springer Science and Business Media B.V. 2010.

Buchanan, Elizabeth. Considering the ethics of big data research: A case of Twitter and ISIS/ISIL. December 2017.

GDPR

https://gdpr.eu/eu-gdpr-personal-data/