

3.4 Case Analysis

Introduction

In Jason Tan's article "The Fine Line of LinkedIn Data Scraping: Legality, Consequences, and Best Practices", legal and ethical problems of data scraping are discussed. Data scraping is the process of gathering massive amounts of data from websites. In the article, Tan particularly looks over the issue that data scraping violates LinkedIn's terms of service. He informs readers that LinkedIn has certain policies in place to prevent scraping in order to protect their user's privacy and maintain their trust. In doing such practices, LinkedIn can ban your account for not following their conditions. In the case that an HR department of a mid-sized private company wants to use data that is scraped from LinkedIn in order to develop training materials for new Hiring Managers, I believe that it would be wrong of the company to breach LinkedIn's policies. In this Case Analysis I will argue that Kantian Deontology will show us that this company should not use scraped data because going against the agreements set by LinkedIn would also mean violating ethical principles of respect and consent, as suggested by Immanuel Kant.

Discussion of first article

One of the concepts explained in Micheal Zimmer's article "'But the data is already public': On the ethics of research in Facebook" is about having integrity and maintaining ethical practices. Zimmer argues that even though people may post their information on a public platform, it does not necessarily mean that it is okay for everyone to take advantage of it. In the article, we analyze the "Tastes, Ties, and Time" (T3) Facebook research study where researchers gathered a huge amount of data taken from Harvard students' Facebook accounts to track their

online behavior and interests. The problem presented in this situation is that the T3 researchers failed to get any informed consent from those who fell victim to the study. Zimmer states that “As made apparent to the position of some of the T3 research team that their data collection methods were unproblematic since the “information was already on Facebook”, future researchers must gain a better understanding of the contextual nature of privacy in these spheres, recognizing that just because personal information is made available in some fashion on a social network, does not mean it is fair game for capture and release to all” (Zimmer, pg. 12). While individuals know that they are posting publicly on a public platform might be a valid argument from people on the T3 team, they fail to fulfill their duty to conduct ethical research by ignoring the fact that users who post on Facebook would expect their information to stay on Facebook and Facebook only.

Regarding our case analysis of the HR department, using scraped data to create training materials without explicit consent would not be ethical considering LinkedIn’s purpose is solely to be a professional social networking platform. Tan states that “LinkedIn can yield valuable insights for business and marketers” (Tan). Although it is not illegal to scrape data on websites, LinkedIn has strict rules that disapprove of it and going against those rules would not uphold the ethical standard proposed by Kantian Deontology. Additionally, another issue lies behind the intention of data scraping methods. Similar to how data from Facebook users were used in a different way than they were intended by the T3 team, the HR department wants to use scraped data for corporate training and not with the intention of networking and job searching. Here, the concept of remaining rooted in ethical practices that can be found in Zimmer’s article can be applied to the case analysis.

From a Kantian Deontology perspective, the HR department's actions aren't ethically correct. By scraping and using the data for training material without any form of consent, they are disregarding LinkedIn's original intentions and breaching the company's terms of service. If a situation like this were to happen it would make users lose their trust in LinkedIn, knowing that their data has been taken advantage of. Kantian Deontology emphasizes points about being respectful to others, suggesting that it is our duty to comply with rules and being respectful of them. If the HR department in this case were to have respect for LinkedIn and its users, they wouldn't violate their rules and take advantage of user data. The right thing to do for HR would have been to ask permission from LinkedIn or the users themselves.

Discussion of second article

In the chapter "Civilian Casualties: Justice in the Age of Big Data" of Cathy O'Neil's book, *Weapons of Math Destruction*, we explore ideas regarding the ethical issues presented by algorithms based on predictive programs such as PredPol. One of the main concepts discussed in this excerpt is the idea of feedback loops. A feedback loop can occur when the results of the system, such as predictive policing (PredPol), are pumped back into it, repeating certain patterns and trends overtime. "PredPol doesn't focus on the individual. Instead, it targets geography. The key inputs are the type and location of each crime and when it occurred" (O'Neil, ch. 5). This can create skewed data results and can show racial biases. Overtime, the same communities get unfairly targeted over and over again. Of course, most crimes can be found in poverty stricken areas for their nuisance and petty crimes. This makes it an easy place for the police to target. According to O'Neil, "These nuisance crimes are endemic to many impoverished neighborhoods. In some places police call them antisocial behavior, or ASB. Unfortunately, including them in the model threatens to skew the analysis. Once the nuisance data flows into a

predictive model, more police are drawn into those neighborhoods, where they are more likely to arrest more people” (O’Niel, ch. 5). Low-level crimes, such as underage drinking, that continue to populate predictive models will send more cops back to the same neighborhood and create the effect of uneven policing. O’Niel then continues to state that this problem creates a detrimental feedback loop that almost never fixes itself. “The policing itself spawns new data, which justifies more policing. And our prisons fill up with hundreds of thousands of people found guilty of victimless crimes. Most of them come from impoverished neighborhoods, and most are black or Hispanic. So even if a model is color blind, the result of it is anything but. In our largely segregated cities, geography is a highly effective proxy for race” (O’Niel, ch. 5). The purpose of the predictive police software was to deter serious crimes. So, why are such low level crimes taken record of at all? For police departments, they are under the assumption that nuisance crimes create an atmosphere of lawlessness in the community.

The concept of feedback loops and uneven data bias can be applied to our case analysis regarding the HR department’s plan to scrape data from LinkedIn for training materials. Similar to how PredPol creates repeating patterns of inequality, scraping LinkedIn data using certain algorithms could create a feedback loop during the hiring process. For example, scraped data could target a specific group of people with prominent backgrounds in the long run if only those with outstanding track records are hired. This creates a problem for those who don’t have as much experience or underprivileged background because, of course, the algorithm excludes them if they aren’t as “interesting”. From a Kantian Deontological perspective, The HR department’s actions would, again, not be ethically reasonable to go against LinkedIn’s terms and conditions. In order to approach this situation with Kant’s views in mind, you would have to emphasize

fairness in decision making. The hiring managers should look into every candidate's portfolios with fairness in mind to ensure that they don't exclude certain groups and bring about inequality.

Conclusion

In conclusion, applying Kantian deontology and key concepts from both articles to our case analysis allows us to understand why their actions would be ethically wrong. The principles of Kantian deontology highlights respect for everyone, which the HR department disregards when defying LinkedIn's terms of service. Also, if they had respect for all their applicants, they would go through all of their resumes without having to use scraped data to choose their employees. This can make the hiring process unfair because it can create data bias, producing systemic disadvantages for those with underprivileged backgrounds. Some people may argue that, naturally, those with outstanding resumes should be hired so data scraping can't be that bad. However, this overlooks the unfairness in the process. People who come from impoverished environments don't get the same chances to build such impressive resumes even though they may have equal, or even greater, potential. Focusing only on the top resumes automatically excludes those at the bottom. With the views of Kantian deontology, the HR department should give everyone a shot at showing their capabilities and what they have to offer for the company.

References:

O'Neil, Cathay. (2016) *Weapons of Math Destruction*, "Justice in the Age of Big Data"

Tan, Jason. (2024) "The Fine Line of LinkedIn Data Scraping: Legality, Consequences, and Best Practices"

Zimmer, Micheal. (2016) "'But the data is already public': On the ethics of research in Facebook"