Intro

I

Zimmer's article about Tastes, Ties, and Times details the project's failure to protect students' privacy as well as collect their data in an ethical way. By showing us T3's failures, we can learn what not to do when collecting and using people's data. O'neil's chapter about predictive policing talks about how data can be used and how it can have adverse effects if it is used incorrectly. The adverse effects come in the form of disproportionately targeting specific minority demographics and impoverished areas. Using the concepts in these two articles, as well as Kantian Deontology, I will argue that Kantian Deontology shows us that this company can potentially scrape data in an ethical manner by avoiding the common ways in which privacy violations can occur, providing complete transparency and a choice to opt in to those who will have their data collected, and being careful about what data is being used to train their hiring managers as to not cause harm outside of their company.

a.)

In Zimmer's article, he talks about the Tastes, Ties, and Times project and how their collection of data from college FaceBook students violated their privacy. Although the T3 project put in a good faith effort to keep the student's information private, their actions were not sufficient to maintain student privacy. This is because T3 deemed themselves capable enough to ethically collect and use the student's data without actually consulting privacy experts throughout their research. It is very difficult to anonymize data, and Zimmer points this out in his article. Data is difficult to keep anonymous because data can be very unique to people. The college that was studied sometimes had only one student from a certain place or ethnicity. Zimmer also

points out that social networks and cultural tastes are like a digital fingerprint, and usually no two people will have the same networks and cultural tastes.

In his paper, Zimmer lists four categories in which privacy violations can occur. The first category was the amount of personal information collected. This violation occurs when too much personally identifiable information is collected. The second category was improper access to personal information. This violation occurs when data is given to someone who is not authorized to receive it. The third violation is unauthorized secondary use, and occurs when data is used in a way that it was not originally intended to be used. The fourth category was errors in personal information, and occurs when people are not allowed to correct errors on information about themselves.

b.)

Using Zimmer's concepts of privacy violation and the nature of personally identifiable information, I will analyze the case of a company scraping data from LinkedIn to create training materials for new hiring managers. This case can violate a person's privacy as much as the T3 project did. But if the company does decide to scrape data from LinkedIn, they have the opportunity to do so more carefully than T3 did. First, they would need to get permission from LinkedIn to start scraping, but data scraping is against LinkedIn's terms of service so it would probably not happen. Second, assuming they get permission to scrape data, they would need to consult experts in privacy to see what kind of data would not compromise the privacy of LinkedIn users.Third, they would need to get explicit permission from the people that they are collecting data from and inform them of what they intend to do with the data, and who they are going to share the data with. They would also need to let the people correct errors in their

collected information. If the company does not do all of these four steps, then their data collection policies could potentially violate a person's privacy.

c.)

From a Kantian Deontological perspective, the company would be ethical in their scraping of data if they follow the steps mentioned above. This is because the company is fulfilling their duty to respect others. One of the central tenets of Kantian Deontology is the ability for people to reason, and to do so people need to be properly informed before they make decisions. This means that lying or omitting information is unethical because it takes away a bit of a person's ability to reason by not giving them entirely correct information. If the company provides complete transparency and gives the opportunity for people to consent, then they will have scraped their data in an ethical fashion. Another category to determine if an action is ethical or not in Kantian Deontology is the categorical imperative, which asks the question, "How would the world be if everyone did what I am going to do"? If every company decided to collect data in such a transparent manner, nobody's right to privacy would be violated and progress could be made in an ethical manner.

II

a.)

The chapter "Civilian Casualties in the Age of Big Data" covers policing based on different policing models. O'neil talks about how underfunded police departments would use technology to more effectively police certain areas. The end result was that more crime was stopped. But in the process of stopping crime by using predictive software like PredPol, large issues were ignored because the specific numbers that PredPol cared about were going up. The issue was that even though more crime was stopped, it was only because they were stopping

crime in certain areas. The more crime that was stopped in those areas meant that it was more likely that crime was going to occur in that area. This leads to a cycle where crime is stopped, more crime data is added to that area, more cops are sent to the area because more crimes are caught there, and finally more crime is caught because of the increased police presence. Another issue with the predictive models was that all crime recorded was treated the same no matter what crime was committed. This means that the data treated violent crime as equal to victimless crimes. The data that was being recorded served as a good indicator for the economic class and ethnicities of the area even though the data was not based on individual characteristics like ethnicity. This led to a skewing of arrests where minorities ended up being arrested more frequently. The predictive software and police take a very utilitarian view in that they only care about stopping more crime. By stopping more crime, they are making the world a safer place. But by over policing certain areas they end up hurting a community because they are disproportionately putting millions of minorities in jail .

b.)

The things that happened in the O'Neil paper can happen when scrapping large demographics anywhere. If LinkedIn were scraped and fed into a model, we would see many commonalities happen when the model develops its training. Depending on the criteria that the company has its model use to develop training, the model may see some qualities as more desirable than others like how the models in the O'neil paper saw stopping more crime as being the end goal with no regard for the other consequences that happen when stopping the crime. If the company is not careful with what information they feed to the model, they could find themselves developing training materials that further their goal of training new hiring managers. The training may result in practices that are good in the short term, but bad over a longer time

period. For example, while they may train good hiring managers that make the company more money, they could also be contributing to things like a bad job market where it is increasingly difficult to get hired. And since the immediate results they see pad their numbers, they will keep on doing what makes the numbers good while ignoring the effects that are not tracked by the numbers. The model could similarly judge people by the demographics that they live in instead of their individual human aspects. This could lead to certain demographics being ignored in the hiring process, or perfectly capable candidates being passed over because they do not meet the metrics defined by the model.

c.)

From a Kantian Deontological perspective, the company scraping the data would have to be very careful about the data that they are collecting and how they are using the data to be ethical. They would need to find the right data analysts and market professionals to make sure that their actions are not affecting the world outside of themselves. Using Kantian Deontology, the company would be ethical if they made sure that the data that they collected did not unintentionally discriminate against certain demographics because to fulfill Kant's categorical imperative, they need to act as if everyone else were to do what they do. If everyone else took into consideration how the job market would be if they also used scraped data from LinkedIn to conduct their training, then the world would have a healthy job market with more equal opportunities. But if every company scraped data from LinkedIn and implemented it into their training in a careless way, it could lead to certain demographics being treated unfairly during the hiring process.

Conclusion

In conclusion, a company can ethically scrape data from LinkedIn provided they consult professionals in privacy, provide complete transparency to the users of what data they are taking, how they are using the data, and be very careful in how they use the data they collect to not isolate any specific demographics. I do however believe that these expectations are unrealistic in this case because LinkedIn does not allow data scraping according to their terms of service. It could also be argued that even though the people accepting the agreement have as much knowledge as possible about what happens to their data, they are not really professionals in the field of privacy and data. This could mean that these people are not properly equipped to make decisions about what happens to their data in this case because they do not know the harm that may arise from giving this company their data. Another flaw is that it is impossible to quantify some aspects of people by only using numbers like a model would, so the model used to develop training materials may get some things wrong. By treating a person as a number instead of a person, it goes against the aspect of Kantian deontology where people need to be treated as an end to themselves.