

The HR department of a mid-sized private company wants to use data scraped from LinkedIn to develop training materials for new Hiring Managers. Is this okay? If not, why, and if so, under what conditions?

Data scraping refers to the process of extracting enormous amounts of data from websites. Undoubtedly, it can be very beneficial for many businesses and organizations as it provides valuable insights from large volumes of data, saving a lot of time and effort. So, the HR department of a mid-sized company wanting to use the process of data scraping from LinkedIn in order to develop their training materials may hold good intentions, however, it is still not okay to just perform data mining on millions of users. Although it is not illegal, it is still so morally wrong, regardless of the valuable outcomes. Furthermore, allowing data scraping holds great potential for cyber threats and misuse of users data. So, in the case of the HR department, while the intent of data scraping from LinkedIn is not bad, it would still raise serious ethical concerns. Therefore, in this case analysis, I will argue that Kantian deontology shows that scraping of users data is unethical; companies like LinkedIn should never permit data mining under any circumstances or for any reason.

LinkedIn is a platform used by millions of people worldwide for various purposes including professional networking, career development, and job hopping. Its databases contain users' valuable personal data such as name, date of birth, contact information, professional history, education information etc. At the time of profile creation, users provide this information with the expectation of data protection and for valuable purposes like career development, not to be exploited by some private company. Even if the company's intent for data scraping is to develop training materials for their new Hiring Managers, from a Kantian perspective, this action is still so morally wrong because it would treat the LinkedIn users as a means to an end, rather than end to themselves. In this case, the moral duty of the company is to treat every user with dignity and respect, regardless of any outcome or consequences. Furthermore, the act of permitting data scraping without the users' consents can not be morally justified as it violates the individuals' right to autonomy, completely disregarding users' decision making capabilities about their own information. Also, data extraction creates opportunities for malicious actors to exploit users' information, leaving them exposed to potential harm. Endangering users' data for the sake of making training material is just fundamentally wrong. Zimmer concluded in his article "*But the data is already public: On the ethics of research in Facebook*" that, even with good intentions in protecting user data, because of lack of competencies and limitations, the researchers of "Taste, Ties, and Time" (T3) project still failed to protect users information which was extracted from Facebook accounts of college students. Making the matter even worse, they extracted students' information without any prior consent; they just exploited students' autonomy

and treated them as mere tools for gathering information in their project. Some may argue that data is already available publicly, normalizing the act of data mining for its intended purposes. Even though users' information is widely available on LinkedIn publicly, it doesn't mean that any private company could exploit it in the pursuit of their own personal gain. In the article, Zimmer also emphasizes that public accessibility to users' data doesn't automatically grant ethical permission for any kind of unrestricted use. Indeed, the extraction of a big chunk of user data from LinkedIn for the purpose of developing training materials is simply an act of exploiting their vulnerable users. As a Kantian, it is morally wrong to treat users as the means to an end; users are humans and they must never be treated merely as instruments for the benefit of others. Moreover, according to Kant's principle of universality, "act so that the maxim of your action can be willed as a universal law", moral action should be to the one that could be universally applied without any contradiction. So, in this case, data extraction for developing training materials violates the law of universality. Additionally, the worst case scenario occurs when data extraction is carried out without users' explicit consent; unfortunately, in most cases, data mining takes place without any direct interaction with the users. Now, from a Kantian perspective, if everyone mines data for whatever reason, then, it would not hold the universality principle. Therefore, I continue to argue that data extraction may be not illegal, but fundamentally wrong and unethical.

As I mentioned above, in most cases, data gets extracted without explicit consent from the individuals to whom it belongs. For example, according to Zimmer, researchers of the "Taste, Ties, and Time" (T3) project only contacted the Facebook company and the targeted university. None of the college students had any idea about this project or any data extraction process. In order to serve a group of researchers' own interests, extracting a big amount of students' data without their consent was just unethical. Proponents might allow data mining by arguing that it holds a great potential to offer benefits to society, specifically when there is limited time and resources. However, it is not always the case. Sometimes, it holds more harm than benefits. O'Neil showed a very good example to prove this claim. He talks about PredPol, a software used for predicting crime, that seemed like reducing the crime rate in Reading, Pennsylvania. With limited officers, this software turned out to be very beneficial for the police department. A similar software called CompStat was used in New York. Both software uses huge amounts of data from the history of crimes and predicts when and where crime is going to happen in future. Based on that, police officers were patrolling more in those geographical areas during specific times, arresting more criminals. It might seem reasonable to consider data extraction as a valuable tool which would benefit society. However, the problem emerges when not all crimes are recorded and when patrolling areas zero in on marginalized communities. Moreover, they were focussing on nuisance crimes rather than violent crimes. Crimes which are not as serious as homicide, burglary etc. O'Neil called it "nuisance" crimes. They are mostly petty crimes such as

panhandling, selling or consuming small amounts of drugs. Unlike violent crimes, these types of crimes often do not get recorded if a police officer wasn't present to witness the crime. As a result, even with the right intentions, the model reaches inaccurate or misleading conclusions. Moreover, it empowers police officers to target marginalized and impoverished communities, arresting more and more poor people for petty crimes. So, the question arises, how does it benefit society? A society where poor people are criminalized for petty crimes while violent crimes are left unnoticed and overlooked. Based on Kantian perspective, we want a society where every human being is respected with dignity and autonomy; we want a society where some people's achievements don't come at the expense of others. Data mining may offer some benefits, but we should think about whether it is beneficial for everyone in society; and whether the benefits come at the expense of others. Because, data is not just random assigned numbers, characters or strings; they represent actual people's information. In some cases, data includes more than personal identifying information. For instance, data in the "Taste, Ties, and Time" project include student's demographic, cultural, and relational information. So, it is not worth taking risks. If a data breach happens, it is the real people who would suffer and bear the consequences. Then, how can the risks to others be justified for the benefit of some people? This action is immoral as there is no equality, no dignity, no honor whatsoever. Therefore, data extraction on LinkedIn by a company to develop training materials is unethical.

In conclusion, data extraction may provide some benefits; however, extracting users data from LinkedIn or any other social media platform without explicit consent is immoral and unethical for many reasons. From Kant's perspective, it is wrong because users would be treated as mere means rather than as ends to themselves. So, it is not right to extract users' data for a private company just to develop training material for their new hiring managers; it is morally wrong to exploit users to serve its own interests. Moreover, data mining is unethical as it can not be applied universally and it doesn't serve the good-will of everyone. Even if the intentions for data mining is not bad, it still has potential to harm users. Morality should be based on goodwill for everyone, regardless of consequences or outcomes. Lastly, I want to conclude this case analysis with Kant's words, "an ideal world would be where dignity, honor and justice becomes the reality we all share".

Note: This is my own work and I have proofread it.

Signature: *Reema Chakma*

Date: 04/08/2025