"In the base argument of the ethical dilemma on data ethics described in the essay, give your own arguments on which side you support. It is essential to use the concepts and methodology related to the essay. How will you extend your reasoning to support the need for an objective framework you mentioned earlier?"

Abstract:

In the previous essay, I briefly described the Ethical problem I am concentrating on. The issue is that when we are using data collected through unethical means in our analysis, we are trying to use as much data as possible to get closer to reality. Instead, suppose we only use the data collected by ethical means (with user consent, lying within terms and conditions, performing harmless experiments etc.). In that case, we are constraining ourselves or restricting ourselves to understanding the truth. Both these actions (using and not using the data collected by unethical means) have disadvantages and advantages; Therefore, it is difficult to pick a side in this case. Nevertheless, it is essential to choose one in the long run. In this paper, I'll try to use ethical concepts like overall harm, virtue ethics vs consequentialism etc., to provide some logical reasoning for picking one from the two. However, my base reasoning is dependent on virtues of standard scientific methodology, like replicability, reproducibility etc. Since this is the last evaluation, I am giving a spoiler here; I am going to argue for leaving the data collected by unethical means out of data analysis. Apart from this, as described in the essay, I will also extend this reasoning to argue for a need for an objective framework to be followed by the researchers, analysts, companies, tech giants etc., to regulate the methods being used by them for data collection. Make a note of the following assumptions and constraints before diving further into the paper.

Terminology and Assumptions:

In this paper, whenever I say 'Unethical data', I am referencing the data collected by unethical means. Whenever I say 'ethical data', I am referencing the data collected by ethical standards. Also, moving forward, Please try not to distinguish between 'unethical data collection practices' and 'using data collected from unethical means in analysis'. In the scope of this paper, I am assuming that if someone collected unethical data, that is for trying to commodify it for their good and for helping or pushing humanity to get closer to reality.

Before Diving into the reasoning part, let us first define what exactly is unethical or ethical data: Unethical data or data collected through unethical means represents data that causes harm to one or more subjects involved with it. Based on our previous discussions, one might assume the fact that unethical data only deals with privacy because violating privacy might harm (due to reasons like identity theft, targeted advertising, feeling of insecurity, liable to scams and other reasons mentioned by batchmates in their presentations) an individual. But that is not one hundred per cent true. Data analysis is not restricted to the field of big data, which focuses on digital data. There are other means for collecting unethical data, such as

Experiments on animals and sometimes even on humans, which causes overall harm to the subjects involved.

I hope you got a brief idea about what unethical data is. Now it is not hard to extend this idea to Ethical data; the process should not cause any overall harm to the subjects of the data. There were many questions on what kind of data is Unethical or ethical in our previous essay presentation sessions. To give a correct Impression, I included the paper's version of unethical and ethical data, which are essential for understanding the paper better. Now let us know to familiarise ourselves with the virtues of science.

Ethical Reproducibility: An virtue of Science?:

We have discussed in our previous discussions that Humans have virtues such as Wiseness, Kindness, Virtue of being patient, Virtue of being courageous etc. If we look at the definition of virtue, it is not very difficult to claim that anything can be defined with the help of some virtues associated with it. For example, we define a person as brave means we expect that person to show bravery under all circumstances. It can be seen that both are interrelated; that is, the definition can often be considered as a basis for defining virtue. We can extend this reasoning to identify the virtues of science. Even the methodology of science has some desired features such as empiricist accuracy, reproducibility, consistency etc. All these qualities are desired characteristics of scientific methods. So by definition, if there is any scientific methodology out there without any of these virtues mentioned is not considered as an ethical scientific methodology.

Now, what is this term Reproducibility? It is nothing but the ability of any researcher or scientist in academia to exactly mimic the results or outcome in following a particular sequence of steps stated by the scientific methodology. For a better understanding, consider this example; You have 100 different cooks in an empire, and the emperor asked these 100 chefs to prepare biryani using his mother's recipe. If all the cooks cooked biryani with the same taste as the emperor's mother, that recipe would be scientifically Reproducible else not.

But this explanation does not complete this sections job; It is to explain "ethical" Reproducibility. This term ideally comes from the field of biomedical research; Ethical Reproducibility demands reporting all the research methods used in a study transparently with critical engagement from the researcher itself.[2] How and why ethical reproducibility is required and beneficial for science is out of scope for this paper, but you can look more into this topic here. Nevertheless, I will try to explain them briefly in the next few

Sentences. The need for Ethical Reproducubility comes from the need for generalising a theory to prove its credibility for the greater good of all the subjects of the science. But within this paper's scope, it is essential to understand that ethical reproducibility is something that acts as a virtue of science, and any ethical study must have this virtue for the study to claim itself moral.

Scientific Analogy & A new View of Dilemma:

There are two primary arguments, one supporting the use of unethical data and one rejecting the use of ethical data. I am not going to explain the arguments once again, but in this section, I will elaborate more on the moral dilemma I described in the essay presentation.

The statement supporting the use of unethical data is gaining support from the getting closer to the reality. If we look at this supporting argument, 'discovering the truth of reality is nothing dissimilar from what the scientific method does. Data analysis that can be generalised as Scientific methodology is something that helps us to understand society better. Therefore we can apply the virtues of science described above to critically analyse the ethical aspects of data analysis. Let us redefine the problem so that one can truly appreciate the role of the virtues of science in resolving the conflict described in the essay.

The problem with the argument supporting unethical data collection practices is that it is based upon the utilitarian theory to maximise the overall utility of all the agents involved in the system. Although some people might get Negatively affected by using unethical data practices, many people would benefit from getting closer to reality, often, almost all of the population.

The consequentialist approach is action dependent which suggests choosing the action that offers the greater good. Since most of the population gets benefited from using unethical data, it is only advisable to do so right; then how are we going to break the tie? The virtue of science saves the day.

Usage of Unethical data: Lack of Ethical Reproducibility

In this section, we will see how practical ethics helps us to break the conflict by proving that including unethical data in data analysis is immoral. It is evident that using or not using unethical data in our data analysis are two mutually exclusive possibilities, and both of

them belongs to different scientific methodologies. Now let us try to point out the critical difference between these two methodologies (not the obvious ones!).

Consider the case where you are a data scientist or undergrad researcher who is using unethical data analysis for a study. If your scientific methodology is virtuous, your study should be ethically reproducible. First, the people you have stolen data unethically will not sit and stay quiet if you make it public that you are using data collected from a community or a region. So it is improbable you would be transparent enough in describing your research methods for the study. Apart from this, there can be some other prestige-issue reasons like you yourself want to protect your identity as a respectable academician who is virtuous and therefore do not intend to let the public know that you are not what you are pretending to be.

Ultimately due to all these kinds of reasons, you would not remain transparent enough compared to that of an ethical researcher who indeed contributes to the scientific world by making their work utterly transparent so that it is replicable and reproducible. Given the above reasoning, one might try to contradict the statement by arguing, what if the researcher is not prone to these prestigious reasons? What if he publicly accepts the fact that he did use unethical data in his research?. Still, the researcher revealing he used unethical data does not make the scientific methodology transparent or reproducible. Even if he admitted that there is unethical data involved in this study, how can other researchers and his colleagues out there can arrive at the results without the particular data set used by the owner in the first place? Nowadays, it is not surprising to see how Machine learning and Artificial Intelligence models and algorithms are data-driven. There can be significant variations in results with minor variations in the data set being used for most of the algorithms. Therefore taking all these claims in to picture, I argue that whenever a person uses unethical data in his research or data analysis, his / her scientific methodology lacks the virtue of Reproducibility.

Lacking Virtue & Being Unethical: A Conclusion Argument

In this section, I will conclude the critical section in our paper that it is immoral to use data collected from unethical means with a simple extension of the results obtained in the previous area. Now we know that using unethical methods for data collection results in a lack of virtue Reproducibility let us understand the impacts of lacking virtue in a person's scientific methodology. As I mentioned earlier in our discussions, it is essential for any researcher to respect the virtues of

Science and should make sure he is not violating any of them while doing his research.

Any scientific methodology that lacks any of science's desired virtues does not guarantee that its outcome will help us achieve what science helps us achieve; getting closer to the truth or understanding the man-nature relationships clearly. If you look at the second argument supporting unethical means of data collection, the consequentialist stance is gaining its support from the results being more appropriate for understanding the truth.

But once we reason with this virtue of science stance, the supporting argument no longer holds. Because lacking virtues of science in any scientific methodology not only guide us on a wrong path, it may further deviate humanity from the course which is the shortest to understanding reality. To conclude, a simplified-standard form of this argument looks something like this

[**P1**] Scientific Methodology has some virtues which define science and are essential to get the correct outcome.

[**P2**] Reproducibility is a virtue of science.

[**P3**] Unethical Data collection lacks Reproducibility.

[**C1**] Unethical Data collection Lacks one virtue of science. [**P2 & P3**]

[**P4**] Whenever a scientific methodology lacks any virtue of science, it is highly probable that that methodology will not help us get closer to reality.

[**C2**] unethical data collection probably condemns us to reach closer to reality. [**C1 & P4**]

[**P5**] Consequentialist approach in the argument only holds when using unethical data helps us get closer to reality.

[**C3**] Consequentialist approach in the argument supporting unethical data collection practices does not hold. [**P5 and C2**]

And just like that, we broke the conflict. 🙂 👉

Role of Virtue ethics in Modern-day data collection practices

If we look at the above standard form of argument, one can quickly identify the drawbacks of the discussion. The first thing is the inclusion of probability into the game. Whenever we say something is probably true, that does not mean it will always be true. Who knows, there might be a possibility where even though there is some intuition of using unethical data in research, it might result in an outcome that helps us closer to reality. There are many experiments in the past, such as the famous Milgram's experiment, that is indeed helpful for the field of psychology even though the methods followed for proceeding with the study are unethical. One thing is clear; there are a lot of inconsistencies involved in this debate around data collection. The primary reason for this is the distinctions in unethical practices involved in these studies. It is not practically possible to evaluate each and every study for unethicality from a consequentialist stance. (due to the reasons such as the number of the agents in the system as described in the essay).

So instead of critically evaluating each and every study associated with data practices for ethical loopholes, we can do better. Taking a virtue ethics stance, we can define the virtues for research associated with data. These virtues act as desirable features for the study. It will also be easy for the pioneers in the field to check the credibility of their research by referencing these virtues throughout the study. After all, they all are human beings, and they can make mistakes accidentally. But if we make this set of virtues into an objective framework that must be valid throughout the research process.
This objective framework not only helps the scientists involved in the study to be ethical, but it also helps the scientific community as this framework keeps unethical practices in check, which in turn helps the scientists and researchers to maintain the virtues of science. There is a new emerging field called critical data studies (CDS) that is trying to introduce the same; several philosophers and computer science professors around the globe are coming together to develop advancements in this field for building this set of virtues of data related studies. That involves entries such as the data collection practices, transparency, replaceability, reproducibility, scepticism, consent etc.

To conclude, in these last few papers and presentations, I talked about the ethical problem we are trying to tackle. The problem exists because of the different policies being suggested by other moral theories. However, I solved the conflict using a practical ethics

Standpoint talking about the virtues of science and trying to argue for the lack of virtue of reproducibility. Extending this reasoning, I further argued that any scientific methodology without these virtues results in a deviation from path to reality and therefore eliminating the second argument supporting ethical data. Apart from this, the last part of this paper concentrates on the need for an objective ethical framework and the benefits that it offers. And finally, conclude the essay with further readings related to critical data studies.

References:

[1]
"Reproducibility and Replicability in Research | National Academies." https://www.nationalacademies.org/news/2019/09/reproducibility-and-replicability-in-research (accessed Nov. 29, 2021).

[2]
Anderson, J., Eijkholt, M. & Illes, J. Ethical reproducibility: towards transparent reporting in biomedical research. *Nat Methods* 10, 843–845 (2013). https://doi.org/10.1038/nmeth.2564

[3]
"Ethical_Considerations_in_Data_Collection.pdf." Accessed: Nov. 29, 2021. [Online]. Available: http://www.njhealthmatters.org/content/sites/njhc/Ethical_Considerations_in_Data_Collection.pdf

[4]
"Ethical Dilemma," Corporate Finance Institute. https://corporatefinanceinstitute.com/resources/knowledge/other/ethical-dilemma/ (accessed Nov. 29, 2021).

[5]
L. Floridi and M. Taddeo, "What is data ethics?," *Phil. Trans. R. Soc.*

A., vol. 374, no. 2083, p. 20160360, Dec. 2016, DOI: 10.1098/rsta.2016.0360.

[6]

A. Bilal, S. Wingreen, and R. Sharma, "Virtue Ethics as a Solution to the Privacy Paradox and Trust in Emerging Technologies," Apr. 2020. DOI: 10.1145/3388176.3388196.

[7]

S. Leonelli, "Locating ethics in data science: responsibility and accountability in global and distributed knowledge production systems," *Phil. Trans. R. Soc. A.*, vol. 374, no. 2083, p. 20160122, Dec. 2016, DOI: 10.1098/rsta.2016.0122.

[8]

"Big Data- Can Virtue Ethics Play a Role unimelb.pdf." Accessed: Oct. 19th, 2021. [Online].Available:https://minerva-access.unimelb.edu.au/bitstream/handle/11343/227494/BigData-%20Can%20Virtue%20Ethics%20Play%20a%20Role%20unimelb.pdf?sequence=1&isAllowed=y

[9]

"Statistical Methods in Artificial Intelligence lecture transcript"

[10]

"Oxford Uehiro Prize in Practical Ethics: Can Science Ethically Make Use Of Data Which Was Gathered By Unethical Means? | Practical Ethics." http://blog.practicalethics.ox.ac.uk/2020/03/oxford-uehiro-prize-in-practical-ethics-can-science-ethically-make-use-of-data-which-was-gathered-by-unethical-means/ (accessed Nov. 29, 2021).