Data Ethics In-Class Activity (May 23)

Today we will discuss two of the Ethics subtopics, Data Privacy and Validity. Refer to the corresponding articles listed in [the previous activity document](https://docs.google.com/document/d/13J_-nHDxBU_2zlobJNP78J2SnImEXvBPPqFicYffYHw/edit#). There is no need to submit any written report about the topics or about the discussion. Be sure to stay in your breakout room even if you finish your group discussion early, as there is a class discussion at the end of the session.

## Data Privacy

Your job is to discuss the Data Privacy topic first within your breakout group (for approximately 30 minutes) and then join the main room for a full class discussion of the topic.

Within your group:

1. Who read any of the articles about Data Privacy?

Rucha Kulkarni

1. Which article(s) did you read?

How one country blocks the world on data privacy

1. Each person who read any of the articles should take a few minutes to give an overview of the article, specifically:
   1. Who wrote the article, what is their role or point of view?

Written by Nicholas Vinocur, the author tells how ireland misused the loop hole in the data privacy

* 1. What are the main points of the article?

How how much important it is for the privacy of the data

* 1. What are the strong points of the article?

We must not follow the loop holes regarding the privacy of a data

* 1. What are the weak points (if any)?

No

* 1. What did you learn from it or take away from it?

Privacy of our own data is very important and it must be protected

1. All others in the group then should discuss and ask questions about the article.
2. Be ready to discuss the following questions with the full class
   1. What is the GDPR?

It is a organisation called as General Data Protection Regulation

* 1. GDPR is a European effort, how does it relate to the USA?

This GDPR applies to the companies outside EU

* 1. How might a Data Engineer be involved in GDPR compliance?

Here data engineers are regulated to use or store or process any data related to individuals.

* 1. Discuss the following questions:
     1. Popups everywhere. It’s annoying and the average internet user has no idea how to control/configure data privacy consent, so they just agree to everything.

Yes, they do that. Atleast there must be a small line of sentence which should be clearly visible, so that the average people understands what are they understanding, instead of very big page of consent

* + 1. Companies are scared, so they are spending bajillions protecting themselves. Bajillions that could be spent on things that actually benefit customers.

They are always in need of profit, so they protect themselves

* + 1. The whole thing is toothless. Only Ireland can bring an actual judgment, and they are in the pocket of big tech. So there have not been many significant cases or judgements so far.

Ireland is being played by a big tech companies

* + 1. It requires private data to be transparent and easily accessible by the users, and that makes it easier for hackers to obtain private data by impersonating users.

No, never do that!!!!

## Validity

Discuss the Validity topic first within your breakout group (for approximately 30 minutes) and then join the main room for a full class discussion.

Within your group:

1. Who read articles about Validity?

Tarun Kumar Yerra

1. Which article(s) did you read?

6 Problems That Make Data Unreliable and How to Fix Them

1. For each article read by at least one person in the group:
   1. Who wrote the article, what is their role or point of view?

Transform solution

* 1. What are the main points of the article?

There are multiple ways data becomes unreliable, hence the data must not be outdated

* 1. What are the strong points of the article?

The data becomes unreliable only because of human errors

* 1. What are the weak points (if any)?

N/A

* 1. What did you learn from it or take away from it?

We must always ensure that the data is correct and unreliable

1. Discuss the following questions:
   1. The articles list many problems with data validity. Which of these problems could be helped by a Data Engineering approach?

Try to automate data entry and integrate different software programs in order to streamline data syncing. This way, all your data is updated simultaneously and you can also avoid human errors

* 1. What specifically could/should a Data Engineer do to address the challenges listed in these articles?

Data should regularly cleanse and ensure that it is all stored on encrypted cloud-based storage locations

## Submit

Create a copy of this document (or create a new document if you prefer), and use it to answer the following question.

For each of the four major areas of Data Ethics, mention a situation that you have experienced that involved the corresponding area of Data Ethics. Say whether or not (in your opinion) the issue was handled satisfactorily. Finally, state how you might improve the handling of Data Ethics in similar situations in the future.

Use the in-class assignment submission form to submit your response(s).