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. 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?
   I - Cesar Ochoa Navarro, did
2. Which article(s) did you read?
   I read the articles “How one country blocks the world on data privacy and “Is Pokémon Go racist? How the app may be redlining communities of color”

3. Each person who read any of the articles should take a few minutes to give an overview of the article, specifically:
   a. Who wrote the article, what is their role or point of view?
   b. What are the main points of the article?
   c. What are the strong points of the article?
   d. What are the weak points (if any)?
   e. What did you learn from it or take away from it?
4. All others in the group then should discuss and ask questions about the article.
5. Be ready to discuss the following questions with the full class
   a. What is the GDPR?
      GDPR is Eurpoes new data protection law that essentially regulates how data is collected, processed and who owns the data.
   b. GDPR is a European effort, how does it relate to the USA?
      The GDPR massively affects the USA since its scope is defined that “if you process personal data of EU citizens or residents, or you offer goods or services to such people, then the GDPR applies to you even if you’re not in the EU.”
      For example Facebook and Instagram are in Europe and provide a service so they must comply with the GDPR.
   c. How might a Data Engineer be involved in GDPR compliance?
      A data engineer would have to follow GDPR standard on how data is processed and data protection. Specifically any action performed on data, whether

automated or manual and they also regulate when you are allowed to process data

d. Discuss the following questions:
    i. 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.

    ii. Companies are scared, so they are spending bajillions protecting themselves. Bajillions that could be spent on things that actually benefit customers.
    iii. 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.
    iv. 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.

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:

6. Who read articles about Validity?
   Mazin Ashfaq
7. Which article(s) did you read?
8. "6 Problems That Make Data Unreliable and How to Fix Them"
9. For each article read by at least one person in the group:
    a. Who wrote the article, what is their role or point of view?
    b. What are the main points of the article?
    c. What are the strong points of the article?
    d. What are the weak points (if any)?
    e. What did you learn from it or take away from it?
10. Discuss the following questions:
    a. The articles list many problems with data validity. Which of these problems could be helped by a Data Engineering approach?
       Data should be free from bias which includes human bias.Reliability and validity of the sources as well. Can be fixed by further validating sources.
    b. What specifically could/should a Data Engineer do to address the challenges listed in these articles?

They should hire data security analyst to audit the data. Hire others that are non biased for the data. Make sure sources are peer reviewed. Have lots of assertions in place to further validate the code.

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).