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?
    a. Cesar Ochoa Navarro
    b. **Mazin Ashfaq**
2. Which article(s) did you read?
    a. **I read "Rethinking Patient Data Privacy In The Era Of Digital Health."**
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?
        i. **Lisa Bari Daniel P. O'Neill**
    b. What are the main points of the article?
        i. **HIPAA is outdated and needs to be updated to adapt to the new sources and handlers of data**
    c. What are the strong points of the article?
        i. **Demonstrating the problem with HIPAA and how it is applied today to the progressing world of data handling**
    d. What are the weak points (if any)?
        i.
    e. What did you learn from it or take away from it?
        i. **We need to seriously consider the security and privacy of our health data.**
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?
        i. **"The GDPR is an EU data privacy law that went into effect May 25, 2018. It is designed to give individuals more control over how their data are collected, used, and protected online. It also binds organizations to strict new rules about**

> **using and securing the personal data they collect from people, including the mandatory use of technical safeguards like encryption and higher legal thresholds to justify data collection."** - Source

    b. GDPR is a European effort, how does it relate to the USA?
        i. **Any data of a european is under this law regardless of where they live. Any company handling their data is affected, aka google facebook** amazon all the major companies.
    c. How might a Data Engineer be involved in GDPR compliance?
        i. **Data engineers will have to consider compliance when processing data and transforming.**
    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?
    a. **Mazin Ashfaq**
7. Which article(s) did you read?
    a. **"6 Problems That Make Data Unreliable and How to Fix Them"**
8. 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?
        i.
    b. What are the main points of the article?
        i. **Goes through 6 issues when it comes to validity of data**

      c. What are the strong points of the article?
          i. **Concise and to the point**
      d. What are the weak points (if any)?
          i. **Lacking evidence and support for the claims. Also formatting is really bad.**
      e. What did you learn from it or take away from it?

9. 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?
      b. What specifically could/should a Data Engineer do to address the challenges listed in these articles?

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).