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?
2. Which article(s) did you read?
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?
    b. GDPR is a European effort, how does it relate to the USA?
    c. How might a Data Engineer be involved in GDPR compliance?
    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?
7. Which article(s) did you read?
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?
    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?
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.

Data Privacy: When Siri was on, I used to always get ads that relate to something I've had a conversation about. It was weird that Siri was always listening. I was not super happy about it but I had to use my phone so I turned Siri off for good. At least that was an option.

Validity: I've read amazon product reviews that gave 5 stars but were really generic, not specific and probably fake (They weren't verified buyers). I found it hard to trust these reviews but at least Amazon will tell you if they are verified buyers or not.

Fairness: My first name starts with a C and my last name starts with an A and those letters appear near the top of an alphabetically ordered name list. Thus, I find that quite often because my name is near the top I get called first or have to start first (not in this class though) and I found this to be unfair.

Ownership: When gathering images to use for an image classification neural network, my team made sure the images were free stock images. Most of the websites with image datasets allowed anyone to freely use the images for non-commercial purposes.

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