

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?
Zhengmao Zhang
2. Which article(s) did you read?
E-scooters becoming forum for mobility data privacy decisions and legislation
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?
Kaye, Kate,
 - b. What are the main points of the article?
Privacy concerns brought about by emerging modes of transportation such as ride-sharing, scooter, etc. are creating challenges for US legislatures at all levels
 - c. What are the strong points of the article?
New laws could preempt city control
 - d. What are the weak points (if any)?
Nope
 - e. What did you learn from it or take away from it?
Multi-state governments are not very clear about the division of rights
4. All others in the group then should discuss and ask questions about the article.
 - a. Is there anyway to protect our privacy?
5. Be ready to discuss the following questions with the full class
 - a. What is the GDPR?
General Data Protection Regulation (EU)
 - b. GDPR is a European effort, how does it relate to the USA?
In December 2019, Politico reported that Ireland and Luxembourg – two smaller EU countries that have had a reputation as a [tax havens](#)

and (especially in the case of Ireland) as a base for European subsidiaries of U.S. **big tech** companies – were facing significant backlogs in their investigations of major foreign companies under GDPR, with Ireland citing the complexity of the regulation as a factor. Critics interviewed by *Politico* also argued that enforcement was also being hampered by varying interpretations between member states, the prioritization of guidance over enforcement by some authorities, and a lack of cooperation between member states.

Reference:

https://www.wikiwand.com/en/General_Data_Protection_Regulation#/Impact

- c. How might a Data Engineer be involved in GDPR compliance?
When data engineers want to access sensitive data from Europe, they may also need to request higher permissions from users, which incurs many other costs for engineers.
- 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.
True, So we should come up with a standard to show the permissions we need to users who don't know this knowledge.
 - ii. Companies are scared, so they are spending bajillions protecting themselves. Bajillions that could be spent on things that actually benefit customers.
Because user data determines their distribution of information, this directly affects their core advertising revenue.
 - 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.
As a core tax haven for tech companies outside Europe, their policies are too sweet. lol
 - 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.
tbh, this is also to log in to the account. It depends on the level of encryption and security of these companies.

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. Yixuan
7. Which article(s) did you read?
 - a. How Trustworthy Is Big Data?
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. Timo Elliott - Big data is not trustful
 - b. What are the main points of the article?
 - i. Big data is not trustworthy because of data completeness, accuracy, credibility and ethic.
 - c. What are the strong points of the article?
 - i. It has a lot of opinion on why big data is not trustworthy
 - d. What are the weak points (if any)?
 - i. Didn't have a lot of detailed information so it's hard to understand
 - e. What did you learn from it or take away from it?
 - i. Big data is not trustful
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?
 - i. Data validity - Data validation can help improve data validity
 - ii. Data completeness - Data validation can also help with completing the missing data
 - b. What specifically could/should a Data Engineer do to address the challenges listed in these articles?
 - i. Make sure we do data validation before we load the data

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).