

Social Impacts Inquiry Assignment

Name: Aryan Raval

1) Choose and try to answer the "clarifying question" from one of the classmates in your group.

I choose Nicolas Ardila's clarifying question in the article titled "Fairness and Abstraction in Sociotechnical Systems" which was presented in 2021 at the conference on Fairness, Accountability, and Transparency (FAccT).

Nico's clarifying question: Are there more traps or pitfalls that researchers could fall into when creating ML models?

My response:

Yes, there are more potential traps or pitfalls that researchers could fall into when creating machine learning models. As Nico stated in the summary, the authors of the paper identify several specific traps related to abstraction and modular design, such as the framing trap and the portability trap, but many other potential sources of bias or unfairness could arise in different contexts or for different types of models.

Some other potential traps or pitfalls that researchers could face include the data bias trap which is if the training data used to develop the machine learning model is biased or unrepresentative, the model may perpetuate or even amplify those biases. The other one is the feedback loop trap which is if the outputs of the machine learning model are used as input to other systems or processes, those systems may reinforce the biases or inaccuracies of the original model, creating a feedback loop that amplifies the negative effects. One more example is the interpretability trap which is if the machine learning model is not transparent or interpretable, it may be difficult to understand how its decisions are being made, making it harder to identify and correct sources of bias or unfairness.

These are a few examples of the many potential traps or pitfalls that researchers could face when creating machine-learning models. I feel that to address these challenges, it is important to take a comprehensive and context-specific approach to model development, incorporating diverse perspectives and considering the potential impacts of the model on all stakeholders.

2) Investigate and report on your findings with respect to the "substantive critique" of a different one of the classmates in your group.

I choose Muhammad Ahsan Tahir's substantive critique in the article titled "Can AI be racist? Color-evasiveness in the application of machine learning to science assessments" by Tina Cheuk, published in the journal Science Education in 2021.

Ahsan's critique: Machine learning requires a lot of data. So an important question not addressed is how we protect minorities against bias when we do not have much data on minorities as they are minorities.

My response:

I feel that the critique, that machine learning requires a lot of data, and the challenge of protecting minorities against bias when there is limited data on them, is an important issue that has been raised in discussions around AI and equity.

In the article, the author acknowledges the potential for bias in AI, particularly concerning race. However, the article does not specifically address the issue of limited data on minorities and the challenge this poses for protecting against bias.

The concern is valid, as machine learning algorithms can only make accurate predictions based on the data they are trained on. If there is limited data on minorities, the algorithm may not accurately represent their experiences or characteristics, leading to biased results.

I feel that to address this challenge, researchers and developers must work to ensure that data used in AI systems are representative of the entire population, including minorities. This can be done through intentional data collection efforts that are inclusive and diverse, as well as through strategies such as data augmentation and transfer learning.

Moreover, it is also important to acknowledge the limitations of AI and the need for human oversight in decision-making. AI should not be viewed as a replacement for human judgment, but rather as a tool that can augment and support decision-making processes.