

Needs Grading

In NLP there has been research on automatically estimating the demographics (e.g. gender, age) of people based on their language use, for example in social media.

Give an example of a potential 1) beneficial use and 2) harmful use.

Selected 1) Research to explain how diseases evolves

Answer: 2) Models make decision-making general so individuals might fall under undeserved bias

Correct

Answer: There are many possible answers, but an example could be:

1) Beneficial use: To help social science research. For example, research may be interested in studying behavior in online platforms, but often researchers need to take demographics into account. Another example is analyzing public opinion, where knowing the demographics can help you get better estimates for the whole population.

2) Harmful use: Using this information for political influence (e.g. targeted messages). As another note: an additional risk is that misidentification (e.g. gender) can be quite harmful for individuals.

Response [None Given]

Feedback:

Needs Grading

When someone says that a ML system is like Clever Hans, what do they mean with this?

Selected We might reach a certain correct answer not necessarily using right signals. In seeking more answers, the

Answer: invalid signals might fail to give an (correct) answer as they do not possess the capacity for it. They were simply not meant for the task at hand.

Correct

Answer: That the system hasn't actually learned the intended task even though performance numbers may be high. Instead, it has learned to use some shortcut/spurious signal.

Response [None Given]

Feedback:

Needs Grading

Are any aspects of the lecture material unclear, do you have follow-up questions about this, or are there any topics you would like to learn more about?

I will try to address this by adding more background material to blackboard.

Leave this blank if you do not have any questions.

Selected Answer: Thanks!

Correct Answer: [None]

Response Feedback: [None Given]

Wednesday, November 4, 2020 10:34:14 PM CET