**Ethical Storyboarding Worksheet**

**Describe a fictional person who was positively affected by a model trained with these data**

Samantha keeps up with the news and reads articles regularly. She is skeptical of how significant and potentially dangerous the coronavirus is. After reviewing the data displaying the most popular words in 2020, Samantha realizes the coronavirus needs to be taken seriously and starts taking precautions to keep herself safe. She now makes a point to stay informed of current events globally and is more aware of the threat.

**Describe a fictional person who was negatively affected by a model trained with these data**

Christopher is a cautious person who constantly analyzes news headlines. Upon viewing the dataset showing covid as the most popular word of 2020, he rushes to the nearest store. He begins buying as much water and toilet paper as possible for fear that they soon may be unavailable. These excessive purchases performed in mass by rattled citizens ultimately cause businesses to run out of stock quickly.

**Describe at least two sources of bias the particular model in your story could have**

There is a built-in assumption that words like “COVID” and “pandemic” will frequently be used in our model. This brings about the risk of confirmation bias in our model, as there is a chance that this assumption may change how one analyzes data within the model. This is a form of bias that may skew how one views the data currently in the model. In addition to the risk of confirmation bias, we’re only using one source (CNN). What if the data analyzed in the business pages of CNN drastically differed from the data from articles of other websites?

**In a paragraph, describe ways to mitigate bias in our model**

Using data from multiple sources reduces bias and helps establish the credibility of the model. Taking note of the time and how data is collected may improve our ability to identify bias in our model and avoid it. While the pandemic has massively impacted the lives of everyone in the world in the past year and a half, we must try to approach the data with an objective approach.