Laura Opsahl-Ong

Part of what initially drew me to data science was the idea that everything in our world is driven by patterns. Down to the electrical signals that cause us to make decisions, everything is part of multiple very complex systems. And if we had all of the data in the world to be able to map out those patterns and systems, we could theoretically predict everything. While we are very far away from that reality, as we slowly make improvements in computing power, predictive power, and our ability to collect data, we get slowly closer to it. One on hand, this progress can be incredibly helpful to us by essentially creating new knowledge, but on the other hand, the knowledge created can be used by bad faith actors and is often made using intrusive methods.

So many problems are created by ignorance, or lack of knowledge, either on a personal or institutional level. The fact is that there is just a lot of information that doesn't exist in the world. The beautiful thing about data science is it can create new knowledge. If you give a country's population dataset to a data scientist, they can tell you information like the extent to which gender is a predictor for education level, information which might not have existed before. That is part of what attracted me to data science. Growing up in the age of the internet, it feels like all of the knowledge in the world is just one google away. Nothing excited or frustrated me more than the realization that there are some things that we just don't know. The innovation of data science has been essential in increasing our knowledge as a species. For instance, we can know where crops are failing around the world using spatial data techniques to analyze satellite images. Having this knowledge allows us to then know where to best allocate resources, potentially helping many struggling farmers.

But what happens when this knowledge falls into the wrong hands? Through my work at geoLab, the spatial data research lab here on campus, I have done crop analysis for an NGO that was hoping to know how effective their work with farmers was. However, my entire team and I had to sign NDAs and work on secure computers because of the concern that either the data on the farms where the NGO was working or the results of our analysis could somehow fall into Boko Haram's hand via a possible hacking. While that possibility felt unlikely to me, it could have been deadly to the NGO's workers or the farmers that they work with. This illustrates the dangers of data science in the wrong hands. Knowledge is power, and that power can be used for good or bad. Just look at China. A few years ago they started implementing social credit scores, where your credit score would be based on your "social worth". Companies use data on your purchases as well as ratings of your social interactions to determine the extent to which they believe you are contributing to society. Your social credit score will then drastically impact what benefits you can receive in the world, like the speed at which you would be seen at a

hospital. China has also rolled out millions of government controlled security cameras. Using extremely advanced facial recognition technology, they are able to punish J-walkers or traffic offenders, but this comes at the expense of recording and tracking all of their citizen's everyday movements. Both of these actions have been decried by the rest of the world as huge oversteps and violations of privacy. The sad truth is that although data science can do a lot of good in the world, considering how difficult collecting and storing large amounts of data is, it is a tool that is strongest in the hands of the powerful. For instance, there's no way that my laptop and I can do as much as Google and their vast network of programmers and supercomputers. So the extent to which data science is used for good is primarily up to large corporations and countries.

It is important to remember that data science is not an inherently positive or negative thing. It is a tool, much like fire or electricity. And like fire and electricity, while we can use it positively, we must also attempt to curtail its negative uses. Already we have seen Europe and even China pass data privacy laws, which will hopefully help with this. I also think that the good intentions of most data scientists will help put us on the right track. While big countries and corporations might have the most resources, there are thousands of well-intentioned researchers out there who are advancing the field and using their skills to improve the world. Getting to learn and work with so many of those people here at William and Mary these past four years has given me faith in the benefits of data science.