The Limits of Machine Learning

Ma. Mercedes T. Rodrigo

In 2014, on my birthday, I opened my browser to find this. It was a Happy Birthday greeting from Google. My kneejerk reaction was: OK, Google, this is creepy. How did you know it was my birthday? In retrospect, my question was naive. My birthday is just about the most elementary thing Google knows about me--I supplied this information when I created my account, That shouldn't have been a surprise.

Still, the experience made me wonder: How much does Google know about me? About any of us? How worried should we be?

Let's tackle these questions one by one. First: How much does Google know about us?

A lot.

Google tracks our emails, photos and videos we save, all the documents on our GDrives. What we watch on YouTube. And this is not a secret. They are totally transparent about this--if you want to see for yourself check the privacy page of your account and it will tell you in black and white about all the data they collect. Based on this data, Google surmises things about us. If you want to know what it thinks about you, go to your ad settings page. Put all that data from hundreds of millions of uses together and google, facebook and similar companies can predict What we will most likely buy, where we will most likely go, what we will most likely do, what we will like to watch, and how we can be influenced or nudged.

It does this through a process called machine learning. Machine learning is a subfield of artificial intelligence that takes massive quantities of data and uses sophisticated algorithms to gain insight about you and me.

In my own field of artificial intelligence in education, we use machine learning to find students who are frustrated, students who are struggling, students who have disengaged with the learning materials.

Once we find them we use another set of models to try to bring them back on task. to help them make use of the resources available to them, to scaffold and guide their learning better.

At its best, machine learning should help us notice things we might otherwise miss, find relationships that we hadn’t seen, and discover creative solutions to complex problems.

Machine learning can be good.

How worried, then, should we be?

We should be aware that machine learning has limits. For today, let me discuss six of these limitations.

Limit #1: Data is inherently biased.

Much of the data fed into machine learning algorithms comes from WEIRD countries. I'm not trying to be derogatory here. WEIRD is an acronym that stands for Western, Educated, Industrialized, Rich, Democratic. People like you and me from developing countries are not always represented.

Limit #2. Machine-learned models tend to be opaque.

They are black boxes, but in some cases, you could peer into them and understand their logic. As machine learning becomes more sophisticated, the models become more difficult to interpret and explain.

In the United Kingdom, high school graduates take what's called the the A-levels. These are high-stakes exams whose scores universities use to determine who they will accept. Because of COVID-19, however, the 2020 A-levels were cancelled. Instead, the UK government used an algorithm to determine student test scores. The algorithm considered factors like assessments from teachers, students' rankings relative to their peers, and the prior performance of the school. As a result, 40% of students--about 280,000--received grades that were lower than what their teachers prescribed. Protests erupted and the government was forced to make a u-turn. the students who were downgraded received the grades that their teachers gave instead.

This brings us to limit #3. Outcomes can be discriminatory. Remember I said that data is biased? If you are starting with a data set that is already biased, the machine learned model becomes a mathematical representation of that bias.

In the documentary film Coded Bias released in 2020, a computer science researcher from MIT named Joy Buolamwini talks about how facial recognition software doesn't recognize black faces. Why? Because the data used to train these models was predominantly composed of images of caucasian people. Joy is a person of color. For facial recognition software to recognize that she has a face, she had to wear a white mask.

Even subtle biases in language are codified. Have you ever used Google translate? Try translating

"siya ay doktor" to English. This translates to "he is a doctor". Now try translating "siya ay nars". This translates to "she is a nurse." Doctors are men. Nurses are women. The models capture biases.

Limit #4. Applications of machine learning can be manipulative and controlling.

In 2019 the Wall Street Journal released a video on how China uses AI for education. Children in a primary school were asked to wear EEG headbands with colored lights in the forehead. Red meant that the student was concentrating. Blue meant the student was distracted. Data was sent in real time to both the teacher and the parents.

On the plus side, teachers said that students concentrated more overall and studied harder. However, some children reported that they were punished by their parents for not concentrating enough.

After the news story broke, there was a public uproar and the school discontinued the headbands' use.

Limit #5. The effects of these models can cascade and scale.

Remember my story about the A-level exams in the UK? 40% of students were directly affected--about 280,000 students in all. Universities were also affected. Many had already sent out admissions offers to applicants, based on the old scores.

Protests erupted and the government was forced to make a u-turn. they changed the scores of the students who were previously downgraded. These students were given grades that their teachers gave instead. And universities were suddenly inundated with previously-rejected applicants who were now qualified for admission.

In 2014, there was a paper published about massive Facebook experiment. Facebook took two groups of their users, Group 1 and Group 2 and they subtly manipulated these groups' newsfeeds. They reduced the number of positive posts seen by Group 1 and reduced the number of negative posts seen by Group 2. The result was that Group 1, those who had seen fewer positive posts, produced fewer positive posts and more negative posts. The opposite was true for Group 2. Those who saw fewer negative posts produced fewer negative posts and more positive posts. The bottom line: manipulate the newsfeed and you manipulate feelings. They did this to 700,000 Facebook users.

This brings me to perhaps the most critical of Limits: 6. Applications of machine-learned models

reflect the values of those who own or control the models.

In 2016, a company called Cambridge Analytical made headlines because of its use of Facebook data to influence the US presidential campaign. They used what they learned about user profiles to send targeted political messaging and advertising to influence the vote. There were allegations as well that Cambridge Analytica used similar tactics to our own Presidential elections here in the Philippines. Now, exactly how influential CA was is the subject of debate, but what I'd like to highlight with this example is that people are the ones controlling how machine learning models are used.

Models are not neutral. How they are used, how they are applied, these decisions are controlled by people with agendas.

So how worried should we be?

Ask yourself:

Who are the people behind the models?

What ends do they serve?

Do I trust them?

That's how worried you should be.

The stories that I told you--the A-levels in the UK, the headbands in China, Google translate, facial recognition software, even Cambridge Analytica--they have a common here: Us. Humanity. Machine learned models have blind spots and they do not have the self-awareness to realize what these blind spots are. So it's up to people, it's up to us to point out these limitations and to demand truth, transparency, fairness, ethics of these algorithms and from the people who control them.

Cathy O'Neil, the author of the book Weapons of Math Destruction, puts it perfectly:
"[Machine-learned models] codify the past. They do not invent the future. Doing that requires moral imagination, and that's something only humans can provide. We have to explicitly embed better values into our algorithms, creating … models that follow our ethical lead."