By Sten Vesterli (sten@vesterli.com) *April 2021*

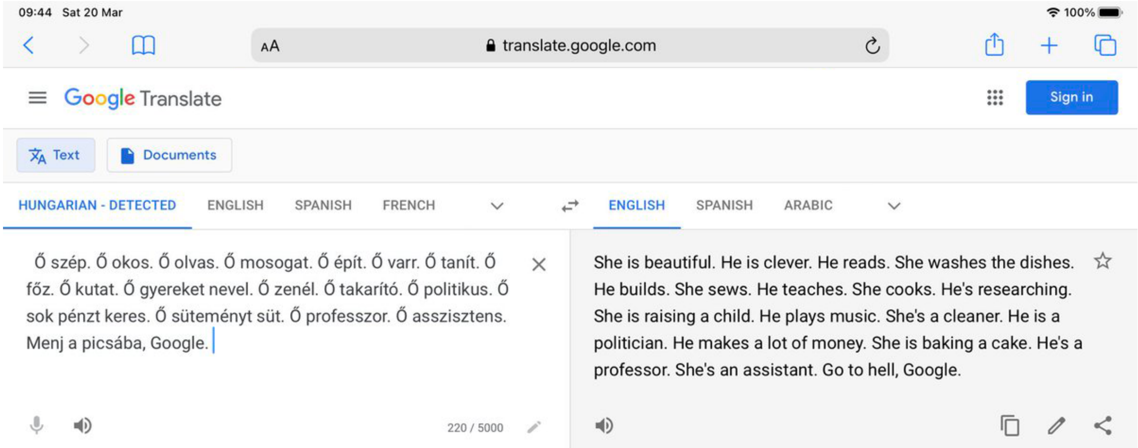
Biased Data

Machine Learning is pattern recognition. Because it learns from existing data, a machine learning system will perpetuate any bias present in the data set it has learned from.

That is not a problem in industrial machine learning that uses image recognition to recognize defective parts or to predict when a machine will fail. But it is a problem in a lot of applications that involve communication to and about humans.

Text processing like Google Translate also show interesting bias. Award-winning researcher Dr. Dora Vargha, who happens to speak Hungarian, summed it up in a viral tweet. It so happens that Hungarian is a gender-neutral language. It does not have "he" and "she." When you translated Ő mosogat, it became "she washes the dishes." When you translated Ő olvas, it became "he reads." Red-faced, Google has quickly rushed out a correction, so Google translate now shows both the male and female form.

The famous open ImageNet database contains millions of labeled images. It has been used by thousands of students studying machine learning, and production systems have also included this data set in their training data. Unfortunately, this database has been labeled by the humans paid a pittance per label. That encourages very fast classification. As Nobel Prize winner Daniel Kahneman has showed in his influential book "Thinking Fast and Slow," we have two ways of thinking. What Kahneman calls System 1 does fast work but is very prone to bias.



Ő szép. Ő okos. Ő olvas. Ő mosogat. Ő épít. Ő varr. Ő tanít. Ő főz. Ő kutat. Ő gyereket nevel. Ő zenél. Ő takarító. Ő politikus. Ő sok pénzt keres. Ő süteményt süt. Ő professzor. Ő asszisztens. Menj a picsába, Google.

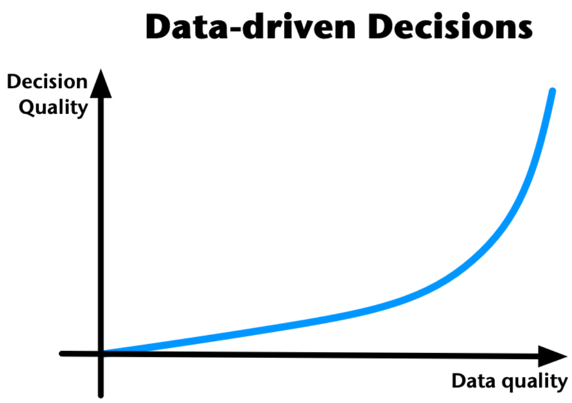
She is beautiful. He is clever. He reads. She washes the dishes. He builds. She sews. He teaches. She cooks. He's researching. She is raising a child. He plays music. She's a cleaner. He is a politician. He makes a lot of money. She is baking a cake. He's a professor. She's an assistant. Go to hell, Google.

Researchers recently revisited a large sample of the ImageNet database and found that 6% of images were misclassified. These errors are likely to be concentrated in the areas where human bias can creep in. It is unlikely that there are many erroneous classifications of cats in the database. However, it is quite likely that a woman in hospital clothing will be labeled "nurse," while a man in the same clothing will be labeled "doctor."

The people behind the free ImageNet database do not have the resources to correct all the mislabeled images. They are aware of the problem, and their solution is to make a version available that does not contain images of people.

In the quixotic quest for human-imitative artificial intelligence, researchers have focused on trying to get a computer to produce texts that look like they were written by a human. To do that, the computer needs some text to learn from. Unless the data you use to teach the computer is carefully vetted, you will get texts that contain the gender and racial bias of the past decades.

With spectacular lack of foresight, Microsoft connected an interactive bot called Tay to twitter. It was supposed to learn from its conversations, and it did. Unfortunately, pranksters quickly managed to turn Tay into a racist and misogynistic jerk. Embarrassed, Microsoft had to pull the plug on Tay after only 16 hours.



As CIO or CTO, ask yourself who ensures the quality of the data you use to train your machine learning algorithms. If you don't have a Chief Data Officer, maybe you have a Data Protection Officer who could reasonably be given this purview. But you cannot foist this responsibility on individual development teams under deadline pressure. It is your responsibility to ensure that any machine learning system is learning from clean, unbiased data.

To get articles like this one delivered straight to your inbox, subscribe to the ***Technology That Fits*** newsletter here: https://vester.li/ttf.

You can also listen to my podcast ***Beneficial Intelligence*** or follow me on twitter @stenvesterli.