O'Leary, Denyse. "Has AI Been Racist?," *Mind Matters*, October 3, 2019

MIND MATTERS NEWS

Has AI Been Racist?

AI is, left to itself, inherently unthinking, which can result in insensitivity and bias

by Denyse O'Leary

A current exhibition in New York, courtesy of the Cooper, Hewitt, Smithsonian Design Museum, "Face Values: Exploring Artificial Intelligence," explores the ways AI projects can exhibit unconscious bias, for example:

> "R. Luke DuBois's installation, Expression Portrait, for example, invites a museumgoer to sit in front of a computer and display an emotion, such as anger or joy, on his or her face. A camera records the visitor's expression and employs software tools to judge the sitter's age, sex, gender and emotional state. (No identifying data is collected and the images are not shared.) We learn that such systems often make mistakes when interpreting facial data."
>
> "Emotion is culturally coded," says DuBois. "To say that open eyes and raised corners of the mouth imply happiness is a gross oversimplification."

His concern is that these biases can find their way into surveillance systems. And what about evaluation systems and studies? Well…

Last week the creators of ImageNet, the 10-year-old database used for facial recognition training of A.I. machine learning technologies, announced the removal of more than 600,000 photos from its system. The company admitted it pulled millions of photos in its database from the Internet, and then hired 50,000 low-paid workers to attach labels to the images. These labels included offensive, bizarre words like enchantress, rapist, slut, Negroid and criminal. After being exposed, the company issued a statement: "As AI technology advances from research lab curiosities into people's daily lives, ensuring that AI systems produce appropriate and fair results has become an important scientific question."

That's more than half of the face photos in the database, according to another story on the biases encoded into programs like ImageNet, which was launched by Stanford University in 2009. The problem is that the bias is not exhibited by the user of the program; it is encoded into it. Unless the coders are chosen with care, that won't be easy to solve.

But, beyond that, are we sometimes overthinking this? Researchers at the University of Canterbury in New Zealand believe, based on their open-access study that we attribute race to robots:

Most robots currently being sold or developed are either stylized with white material or have a metallic appearance. In this research we used the shooter bias paradigm and several questionnaires to investigate if people automatically identify robots as being racialized, such that we might say that some robots are “White” while others are “Asian”, or “Black”. To do so, we conducted an extended replication of the classic social psychological shooter bias paradigm using robot stimuli to explore whether effects known from human human intergroup experiments would generalize to robots that were racialized as Black and White. Reaction-time based measures revealed that participants demonstrated ‘shooter-bias’ toward both Black people and robot racialized as Black. Participants were also willing to attribute a race to the robots depending on their racialization and demonstrated a high degree of inter-subject agreement when it came to these attributions.
– BARTNECK, C., YOGEESWARAN K, SER QM, WOODWARD G, SPARROW R, WANG S, EYSSEL F, “ROBOTS AND RACISM” AT UC RESEARCH REPOSITORY (OPEN ACCESS)

Robots are said to be “racialized as white,” which raises a question: Were they racialized otherwise, might a different sort of offense be taken? “Robot,” applied to humans, is not usually a term of approval. In any event, shades of plastic should be an easy problem to address compared to, say, the difficulty of finding unbiased sources of judgment about human characteristics.

And then sometimes the whole thing just goes full Twitter on us all. Like Microsoft’s chatbot Tay, reminiscent of American singer-songwriter Taylor Swift. A chatbot frames responses based on previous threads of conversations. So it can keep the chit going as long as a user wishes to chat.

Under a different name, Tay had worked out well enough among lonely young people in China, who often lack siblings and alas, if they are men, marriage prospects. But within 16 hours of its U.S. introduction, the Microsoft team had to shut Tay down; its ability to be nasty in a freer environment—Twitter-level nasty—exceeded human toleration.

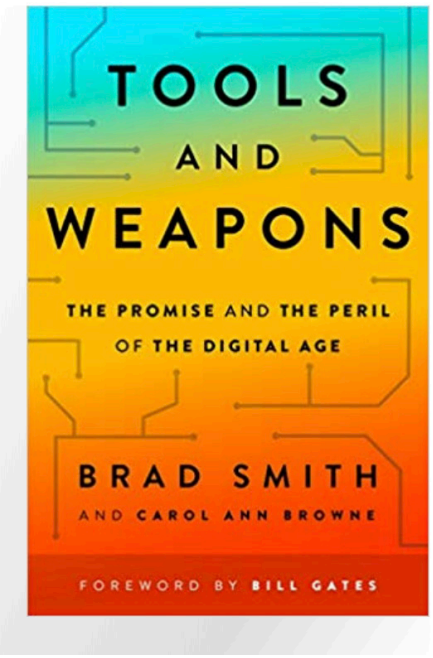
Then, mid-uproar, Taylor Swift threatened a lawsuit:

As [Microsoft president Brad] Smith recalls in his book:

“I was on vacation when I made the mistake of looking at my phone during dinner. An email had just arrived from a Beverly Hills lawyer who introduced himself by telling me, “We represent Taylor Swift, on whose behalf this is directed to you.”… He went on to state that “the name ‘Tay,’ as I’m sure you must know, is closely associated with our client.”… The lawyer went on to argue that the use of the name Tay created a false and misleading association between the popular singer and our chatbot, and that it violated federal and state laws.”

Smith adds that Microsoft’s trademark lawyers disagreed, but Microsoft was not interested in fighting Swift, so the company immediately began discussing a new name. He took the incident as a noteworthy example of “differing cultural practices” in the U.S. and China.

– JENNINGS BROWN, “TAYLOR SWIFT THREATENED LEGAL ACTION AGAINST MICROSOFT OVER RACIST AND GENOCIDAL CHATBOT TAY” AT GIZMODO



Having learned its lesson. Microsoft then released Zo instead. Zo, we are told, is the ultimate in correctness. But not many people have been inclined to pay attention, which makes it difficult for the virtuebot to lead us all by example.

It seems that we face two separate problems: One is an AI problem: Just any available data swatched into systems may embody prejudices that only becomes evident in use. Garbage In; Garbage Out is not obsolete. One thinks of the Amazon “sexist pigs” uproar and the Google “gorilla” fiasco, neither of which resulted from the direct intentions of the systems’ providers at the time.

But second, sometimes people forget that all of these electronic entities are mere products of the human imagination. None of them is a person. Not one.

This is not a new problem. It happens all the time with the creations of art and literature. For example, Superman may have been racist, as some now say, but he wasn’t a human being. He originated in the imagination of a man who had worked briefly as a child for the Toronto Star (and so Clark Kent was a mid-twentieth-century newsman of the sort that man remembered from his childhood…) To be angry about Superman is to be angry with cultural history, not with a person.

If we think change is needed, it must still start with the people, not the cultural artifacts or the programs.