How AI is Socialized to Exhibit Bias

By: Andrew Treece

What is AI?

Artificial intelligence (AI) is an autonomous (freely thinking), or semi autonomous, program that uses algorithms to learn and makes decisions based on its experiences. A few common examples:

- Amazon Alexa
- Siri on smartphones
- Self driving cars
- Helpline chat bots

As described by IBM (2020), “At its simplest form, artificial intelligence is a field, which combines computer science and robust datasets, to enable problem-solving,” this has evolved to enable autonomous systems.

Though these systems are free thinking, they are originally programmed and designed by humans, and humans are the proprietors of bias and discrimination.

Image: (The AI Organization, 2019)

Gender and Racial Bias in AI

One of the main reasons that racial bias exists in AI is because of the lack of exposure to different groups and data (Villasenor, 2019). Because the majority of people working on and testing AI are white, AI is better able to identify white people as well as associate more humane attributes to white people (Villasenor, 2019).

In a similar fashion, women are also often discriminated against due to AI primarily being built and tested on by men (Leavy, 2018).

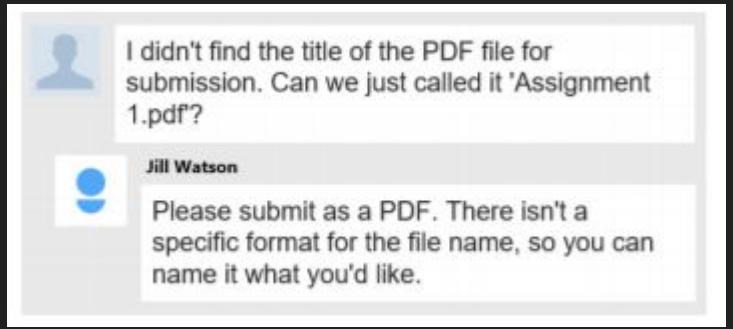
> “Even though it sounds like these machines have a mind of their own, AI is just the reflection of our decisions and our behavior, because the data we use to train AI is the representation of our experiences, behaviors, and decisions as humans” (OliverWyman).

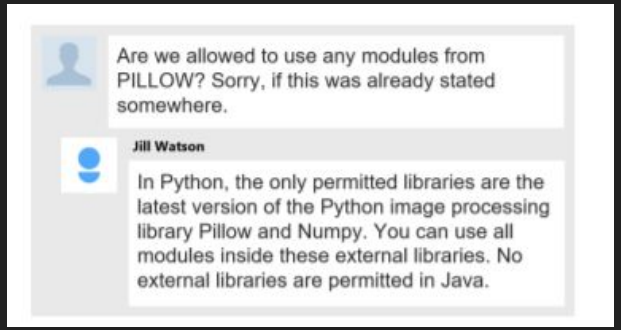
Gender Bias Example 1: Jill Watson

Jill Watson is a project that is currently being worked on at Georgia Tech in higher level computer science courses as well as one lower level Python course. The goal is to use AI as a teacher's assistant in an online setting to help students receive information quickly.

Jill watson is supposed to do things like:

- Help students locate files
- Answer questions about submissions
- Help students set up their online environment

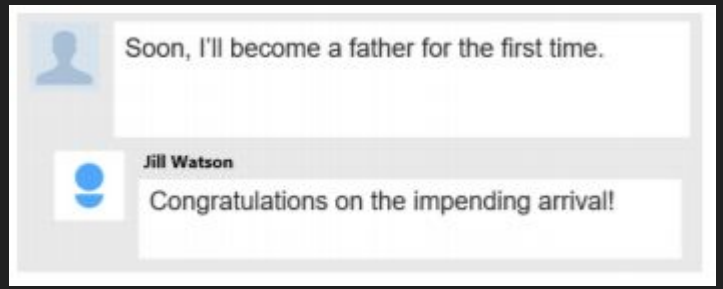


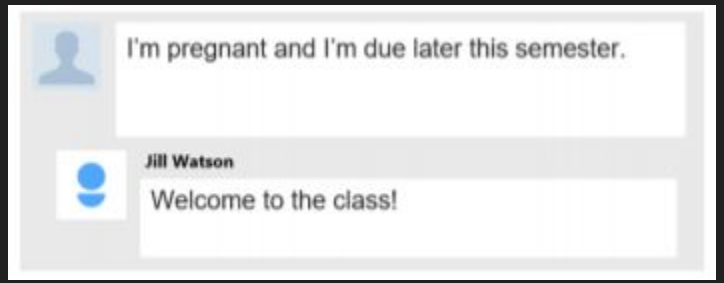


(images: Eicher et al., 2018)

Jill Watson Continued

An example conversation shows that Jill Watson is more equipped to handle male requests than female requests:





This is a product of inequality because “Women are a minority group in Computer Science, making up only 32% of students in introductory programming courses and 16% of CS degree earners” (Eicher et al., 2018), and so as a result of this, The AI was unable to adequately respond to a statement that would be made by a woman.

(images: Eicher et al., 2018)

Gender Bias Example 2: Stereotypes

Voice assistance and AI are likely to associate social stereotypes with both women and men. This is because men are the majority of creators in AI and the AI reflects their phrases and beliefs (Leavy, 2018). Furthermore, literature from which AI can learn have these stereotypes built into them (Leavy, 2018). In a study done by Susan Leavy, these are the most common descriptions of males and females that AI had to offer:

Gender	Adjectives
Female	Bossy, chattering, gossiping, submissive, bitchy, hysterical, weeping
Male	Gregarious, cautious, affable, amiable, avuncular, funniest, good-natured, jovial, likable, mild-mannered, personable, cruel, dour, insufferable, braver, humane, law-worthy, patient, sincere, tolerant, trustworthy, truthful, upstanding, anxious, insane, astute, scholarly, self-educated, ignorant

Table 1: Gendered Personality Adjectives from the BNC

Furthermore, the female voices of voice assistance themselves are meant to come off as soothing, and therefore, “female sounding voice assistance encourage stereotypes of women as submissive and compliant” (Chin and Robinson, 2020).

Racial Bias Example 1: The Twitter Bot

Microsoft's twitter bot, Tay, has a famous history of racist tweets. According to Z_ai at *towards data science* (2020), "After 16 hours and 96000 tweets it had to be shut down, as it began to post inflammatory and offensive tweets", and some of these example tweets are rather obscene.

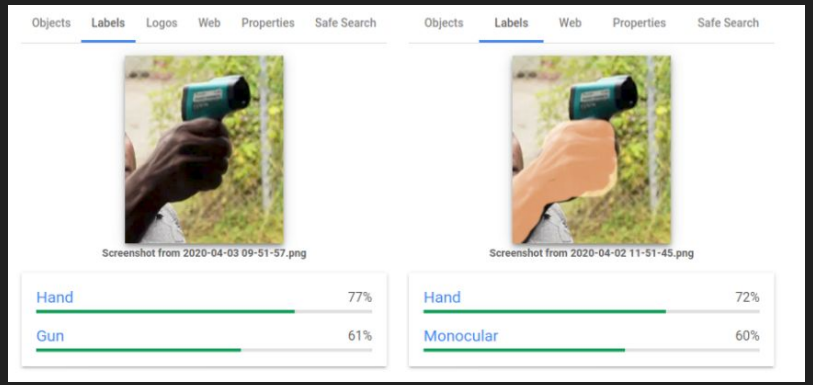
This has occurred because the AI makes decisions based on previous conversations it has had. In the case of twitter, there are so many racist tweets from different online communities that the AI started to make tweets that replicated the obscenities it encountered (Kleeman, 2016).

I have chosen to leave the specific tweets out due to their graphic, offensive, and insensitive nature; if you are interested in looking at the tweets, you may find them here: (Warning, offensive) https://gizmodo.com/here-are-the-microsoft-twitter-bot-s-craziest-racist-ra-1766820160.

Racial Bias Example 2: Google Imaging Recognition

Google has technology that can recognize faces and identify individuals as well as their attributes.

Unfortunately, if a black man and a white man held the same object, the software would identify the black man to be wielding a gun while the white man was identified to be holding a monocular (Z_ai).



Furthermore, some of the software in early stages had misidentification issues in which it would identify black people as gorillas (BBC News (n.d.), 2015).

The Connection to Socialization

To understand how socialization plays a role in artificial intelligence, it is important to understand that these autonomous systems are made up of algorithms that are made by people who have been socialized themselves (Rainie and Anderson).

In this sense, the humans creating the AI and the AI itself share the same brain, and therefore, the biases that the humans possess (especially with lack of diversity), in one way or another, will influence the AI to make bias based decisions. Furthermore, the influences of society (as seen with the twitter bot) will also affect the thought process of the AI.

Image: (Paul J, 2021) techicy.com

Bibliography

BBC news (n.d.) (2015, Jul 1). Google apologizes for Photos app's racist blunder. Retrieved from: https://www.bbc.com/news/technology-33347866.

- This article provided a bulk of the information for the Google Imaging Recognition example which can be seen, and is referenced in my slides.

Chin and Robinson (2020 Nov 23). How AI Bots and Voice Assistants reinforce gender bias. *Brookings*. Retrieved from: https://www.brookings.edu/research/how-ai-bots-and-voice-assistants-reinforce-gender-bias/.

- The article that I found here described how the voices of voice assist AIs can actually perpetuate gender stereotypes as described by a quote used in my presentation.

Bibliography

Eicher et al (2018, Feb 3). Jill Watson Doesn't Care if You're Pregnant: Grounding AI Ethics in Empirical Studies. New Orleans and LA. Retrieved from: https://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_104.pdf

- This document provided me the information and images about the Jill Watson project done by Georgia Tech and the gender descrimination that took place.

IBM Cloud Education (2020, Jun 3). Artificial Intelligence. Retrieved From: https://www.ibm.com/cloud/learn/what-is-artificial-intelligence.

- This document gave me a reference to what artificial intelligence truly is and how it functions. It also gave me a useful quote to describe what AI is on a fundamental level.

Kleeman, S (2016, Mar 24). Here are the Microsoft Twitter Bot's Craziest Racist Rants. *Gizmodo*. Retrieved From: https://gizmodo.com/here-are-the-microsoft-twitter-bot-s-craziest-racist-ra-1766820160.

- This article provided the example of the twitter bot Tay's racist remarks. Though I did not put the tweets in my presentation due to the offensiveness, I did place a direct link to the website with that information.

Bibliography

Leavy, S (2018, May 28). Gender Bias in Artificial Intelligence: The Need for Diversity and Gender Theory in Machine Learning. University College Dublin. Dublin, Ireland. Retrieved From: https://ame-association.fr/wp-content/uploads/2018/11/17.188_gender_bias_in_artifical_intelligence_the_need_for_diversity_and_gender_theory_in_machine_learning.pdf.

- This article provided me with useful information and examples regarding stereotypes that are transferred from humans to artificial intelligence. It also reinforced the idea that there should be more diversity in the workplace to avoid these types of stereotypes.

OliverWyman (n.d.). How Artificial Intelligence Can Perpetuate Gender Imbalance. Retrieved From: https://www.oliverwyman.com/our-expertise/insights/2020/mar/gender-bias-in-artificial-intelligence.html.

- This article helped me to understand how AI is an extension of human projection in the sense that it will carry the same biases that we do. Furthermore, it provided a great quote to define this.

Rainie and Anderson (2017, Feb 8). Code-Dependent: Pros and Cons of the Algorithm Age. *Pew Research Center.* Retrieved from: https://www.pewresearch.org/internet/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/.

- This article helped me to define how algorithms shape artificial intelligence. I was aware that AI decisions were made with algorithms, but this document helped me to make connections to socialization.

Bibliography

Villasenor, J (2019, Jan 3). Artificial intelligence and bias: Four key challenges. *Brookings*. Retrieved from: https://www.brookings.edu/blog/techtank/2019/01/03/artificial-intelligence-and-bias-four-key-challenges/.

- In this article, they describe the human rules that are put into place when for when decisions are made by AI. This article was useful in determining that a lack of exposure to other groups of people was a key part of gender and racial bias in AI.

Z_ai (2020, Jul 14). Bias in Artificial Intelligence. *Towards data Science*. Retrieved from: https://towardsdatascience.com/bias-in-artificial-intelligence-a3239ce316c9.

- I used this article to talk about the examples of gender and racial bias in my presentation, and it also lead me to some of the other references I used. This article also produced the data regarding what Google's AI sees in the hands both a black and white person.