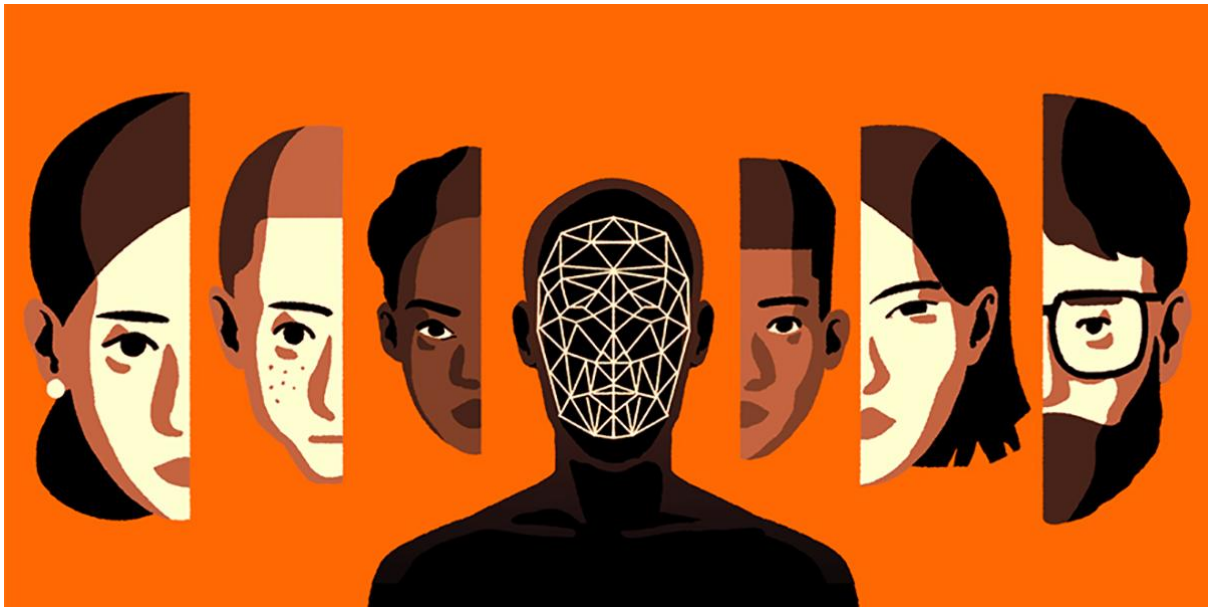
Artificial Intelligence - A Threat or A Blessing?

Artificial intelligence is more than just a system based on computer science. It is a revolutionary technological advancement that replicates neurons in the human brain using complex system's called neural networks. However, with great change comes many pros and cons, many of which revolve around ethics and social responsibility.

What is AI bias?

"Machine learning bias, also sometimes called algorithm bias or AI bias, is a phenomenon that occurs when an algorithm produces results that are systemically prejudiced due to erroneous assumptions in the machine learning process." - Margaret Rouse, writer for TechTarget's IT encyclopedia.

Bias in AI refers to situations where machine-learning systems discriminate against particular groups of people regarding their race, gender, biological sex, nationality, or age. AI bias can appear in job selection and hiring, judicial systems, healthcare systems, facial recognition technologies and many other areas.



How can machine learning systems develop bias?

A general misconception revolving around machine learning systems is that they *learn by themselves.* A supervised machine learning system is given training data in the form of statistics, images, voice recordings etc. and is taught to make decisions based on this data. Bias can creep into this training data through biased human decision that include historical or social inequalities, sexual orientation, race, and gender inequalities. The bias lies not in the machine learning algorithm, *but in the way it is taught.*

On the contrary, an unsupervised machine learning system is not given any training data. It is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It enters a problem blind, and uses logical operators to guide it. An unsupervised machine learning system can also develop a bias through data that it analyzes. By spotting trends in society, using its logical operators, an unsupervised machine learning system's decision making capabilities can be altered to favor or oppose a certain group(s) of people.

It is important to recognize that the machine learning systems are *not* at fault here. AI models behave based on how we tell them to. They learn from the data we feed

them, and often times non-neutral AI models reflected the bias of those delivering the data. Bias in artificial intelligence is all of our responsibility.

Examples of AI bias

COMPAS

One of the most popular cases of AI Bias is the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) algorithm used in US court systems to determine whether a defendant would become a recidivist - a term used to describe criminals who reoffend.

Research conducted by computer science enthusiasts Jeff Larson, Surya Mattu, Lauren Kirchner and Julia Angwin shows that black defendants were often predicted to be at a higher risk of recidivism than they actually were and that white defendants were often predicted to be less risky than they were. Furthermore, their analysis showed that when controlling all factors, such as prior crimes, age, and gender, the AI model was 45% more likely to assign a higher risk of recidivism to a black defendant.

Facebook Ads

Facebook's rules before 2019 allowed its advertisers to intentionally target adverts based on race, gender, ethnicity, and religion. For example, the Facebook algorithm would recommend jobs as a secretary or as a helper in a nursing home to female adverts. To male adverts, particularly those from minority backgrounds, the algorithm would recommend jobs as a taxi driver, or as a janitor.

As a result, Facebook settled five lawsuits, and promised to change the way it manages the advertisements for housing, employment, and credit that run on its platform. Advertisers would no longer be allowed to target adverts based on gender, race, or religion. However, an article by Molly Callahan hints that "the algorithm Facebook uses to deliver advertisements can still skew toward specific demographic groups—despite the changes the company made."

Nevertheless, after research done by a team of computer scientists, Northeastern professor Alan Mislove says, "It’s very hard right now, in the sense that the protected identities [against which discrimination can occur] really permeate our society. It’s going to be much harder and much more subtle than simply removing certain features at the outset."

Amazon's Employment Tool

In 2015, Amazon's software engineers discovered a big problem in their new online recruiting tool. The algorithm was prone to rejecting a woman's job application. According to sources that told Reuters, "The glitch stemmed from the fact that Amazon's computer models were trained by observing patterns and resumes of job

candidates over a 10 year period, largely from men." This taught the algorithm that the male candidates were preferable. Reuters' correspondent Jeffrey Dastin explains the Amazon algorithm's tendency to disregard any job application with the term "women's" in it "because the company has hired so many male engineers or software developers."

Amazon never solely relied on these online recruiting tools and disbanded the unit that created it by the start of 2017. It now uses a lighter version for administrative chores.

What can we learn from this?

It is important to recognize that these AI biases reflect the mindset of our society. Equality should be a basic attribute of society, and an integrated feature of machine learning systems. According to Terrence S, an MBA student, the data that's used in a machine learning system has to be bias-free and the engineers that are creating these algorithms need to make sure they're not leaking any of their own biases, in order to create non-biased algorithms.