Occurs over time	Can be positive or negative	Leads to change	Can be planned or spontaneous

Student-led Research Results

We decided to interview Bobby Azarian who is a cognitive neuroscientist and is a science writer who has had works published in The Atlantic,The New York Times, BBC, Scientific American, Slate, The Daily Beast, and the Huffington Post.

| Question:	Answer:
As artificial intelligence develops over time, is there a capability that they will inherit qualities like racism or discrimination based on race?	
What changes should be made to ensure that A.I. can adapt over time to be humane,	

objective, and morally or ethically sound?	
Does A.I. inherit qualities through its environment or its internal programming?	"To address your first question about whether AI will inherit our racial biases and prejudices, the answer is, they will by default if we don't address this issue directly. Essentially, if you recognize that advanced AI often learns from "reading" tons and tons of human generated content, like that which is on the Internet, you will see that it is inevitable. Since the Internet contains documents and articles written by humans, who have inherent biases, they tend to inherit these biases, and this shows when they are making judgments or associations. I wrote an article just weeks ago about this exact issue, called, "[Study finds AI systems exhibit human-like prejudices](#)". Unfortunately our biases are embedded in our language and cultural history. Special care must be taken by the programmers of these systems to ensure that these biases are corrected for

	when the AI system is making decisions."
Is the capability of A.I. consistent across cultures around the world or do cultural or racial factors influence how an A.I. thinks and acts?	"Until we have Strong AI, which should have morals and ethics that emerge naturally, we will have to be very meticulous in the way we write programs to ensure ethical and moral behavior. We won't be really creating moral machines, just machines that follow our instructions which ideally have built in rules that result in the appropriate moral and ethical decisions. For example, if a machine is used as a weapon on the battlefield, it will not be an intentional agent, it will blindly follow it's programming, so the responsibility of it's moral behaviors fall on the humans writing its software. A robot soldier should be programmed to distinguish between a civilian and a threat, for example, and to treat them appropriately. But these decisions are a result of its programming — it has no true intent. To ensure that moral and ethical behaving AI software is written,

	there should be non-partisan and diverse ethics committees in place that oversee these issues. Complex moral dilemmas that will affect us will inevitably arise. It is not as easy as making machines that will know how to act humanely or morally in every new situation, unfortunately. Since they aren't truly aware or self-aware, they are just making decisions based on predetermined algorithms, most new cases will have to be planned for in advance, when the program is being written. Until we know how to make AI where true intention and subjectivity emerges, the ethics and morals must be programmed as computational rules. "
	"You ask if AI inherits properties through the environment or its internal programming, which is an astute question. It is a combination of the software's code and the information it is processing from its

|	environment. Input is converted to output via specific algorithms, so the computer learns from the environment but its "learning program" is created in advance. What's neat about modern AI is that the neural networks they operate according to are modeled after the brain, and therefore have "plasticity". AI learns as humans learn in that associations are reinforced by what it learns from input from the external environment. In this way, it can 'learn' our negative stereotypes. But it does not understand or experience anything subjectively."

Changes over time - A.I has changed from the past and it will continue to improve in the future.
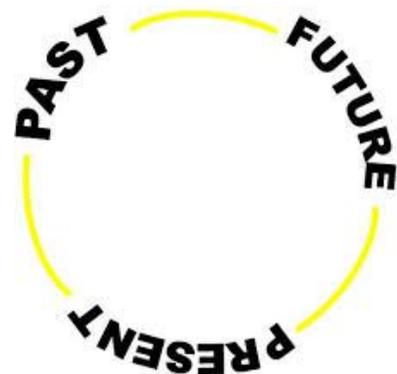
- A study conducted by Aylin Caliskan found that, "African-American names in a program were less associated with the word 'pleasant' than white names."(Resnick, 2017) The program was taught to skim the internet and learn the definitions of words through process of elimination and context clues. It would learn what the word "water" was the more it was used next to the words "bottle" or "ocean." There is so much stuff on the internet talking about blacks being treated badly. Over time, the AI starts making connections about African Americans being unpleasant and makes the AI form racial biases through the internet.
- According to Resnick (2017), "Like a child, a computer builds its vocabulary through how often terms appear together." AI is like a child that learns through its environment and through its programming. The AI starts creating racial stereotypes and over time shows signs of implicit bias.

- According to an interview my group conducted, Bobby Azarian said, " Unfortunately our biases are embedded in our language and cultural history." Over time people have written thousands of articles which contain human bias. Machine programs that are programmed to learn through analyzing articles on the internet learn about these racial biases and start to make correlations. This makes them develop biases and stereotypes about certain cultures.

## Ethics- A.I can bring both good things and bad things to occur
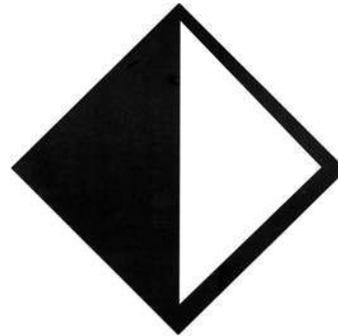
- In an article by Brian Resnick(2017) he says that, " Already AI is making its way into the health care system, helping doctors find the right course of treatment for their patients." This can be both good because AI has been

known to make more accurate decisions than humans. It can also be negative if AI continues being sexist or biased. According to Resnick(2017), health data also contains historical bias like how it's been known that more men get surgeries than women. This can make the difference between AI choosing to give women less surgery even when needed. This icon represents how A.I can be good and can also be bad in different situations.

- According to Crawford, " Machines are increasingly being given the same kind of tasks; to make certain predictions about segments of the populations, often based on visual algorithms."(Lynn, 2017) This can be positive because if they get the AI to predict correctly then it can improve living conditions and reduce crime rate around the areas that need it the most. It can also be negative though if AI is forming conclusions based on informations that contains biases and stereotypes. This ruins the technology and makes it ineffective.

- According to Bobby Azarian after an interview we conducted, he says, " It is a combination of the software's code and the information it is processing from its environment." This can be both positive and negative because if AI is programmed correctly and is learning well through its environment then it can really help and fulfill its true purpose but if the AI is programmed with flaws and is learning through its environment to be racist then the technology

becomes useless and is not helpful..

Contributions- AI can really change the way of life for many people
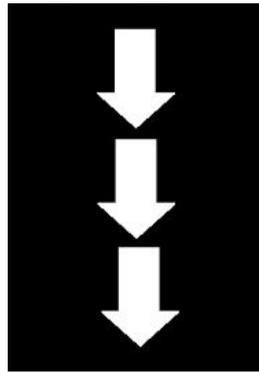
- In an article by Brian Resnick(2017) he says, " Increasingly, Caliskan says, job recruiters are relying on machine learning programs to take a first pass at resumes. And if left unchecked, the programs can learn and act upon gender stereotypes in their decision-making." This can lead to change because if AI is making decisions off of gender stereotypes this can lead to a huge increase in men engineers or a huge increase in women that are nurses. Even if a woman is very smart the AI might pass up on her just because she is not a man. If informational charts show that more men are engineers or that more women

are nurses, it might discourage both men and women to apply for jobs that they are more likely not to get.

- In a speech given by Kate Crawford she says, " Because, even great tools can be misapplied and can be used to produce the wrong conclusions, and that can be disastrous, if used by those who want to centralize their power and erase their accountability."(Lynn, 2017) Giving the wrong conclusions can give people the wrong idea of something that might not be true. According to Lynn,(2017), "Crawford said that a study done on Chicago's predictive policing efforts showed that the technology was 'completely ineffective at predicting future crime'" The way the technology predicts future crime and is ineffective at it just creates false statements that can change how people live. People in area might now be heavily policed and harassed more often just because of a wrong conclusion given by the huge amounts of historical crime data that was used by the AI to come up with these conclusions.
- According to an article by Hannah Devlin(2016) she states, " Concerns have been growing about AI's so-called "white guy problem" and now scientists have devised a way to test whether an algorithm is introducing gender or racial biases into decision-making." This can change AI and make it much safer and reliable to use once the problem is fixed. This can lead to a lot of change and make AI that chooses jobs a lot more reliable to use. They still have to find a

way to actually make AI not racist. It only detects bias.

Language of the Discipline- AI is the theory and development of machines that can do human tasks without the help of humans

- According to Crawford she states that, " Human trained data contains all of our biases and stereotypes and AI and machine learning can be used in ways we don't even realize." This shows that AI can be spontaneous because their is still a lot that is happening to AI that we don't know about. Sometimes without knowing AI is creating racial stereotypes and biases.
- In an article written by Samara Lynn(2017) she says, " She ended the

discussion by stating the need for a new resistance movement that actively monitors and brings awareness of the ways in which AI can harm society especially in the hands of dictators or those who would use the technology to manipulate others." If someone uses AI to create fake racial stereotypes and convinces the country to believe them then this could cause problems. AI being racist can also be planned and have devastating effects on the mindsets of some people.

- One of the first international beauty contests to be judged by AI happened and the results were very shocking. According to Sam Levin(2016), the beauty contest did not like women of dark skinned beauty and didn't consider them to be beautiful. Nobody planned for this and it just happened and shocked a lot of people, including the judges. The group that designed, Beauty.AI, did not plan the algorithm to treat light skin as one of the things looked for in women to determine their beauty. Most of the time AI is racist or has some racial bias, it is due to an accident and wasn't done purposefully.

REFERENCES PAGE

Lynn, S. (2017, March 16). Artificial Intelligence as a Weapon for Hate and Racism. Retrieved from
http://www.blackenterprise.com/technology/artificial-intelligence-weapon-hate-racism/

Resnick, B. (n.d.). How artificial intelligence learns to be racist. Retrieved April 17, 2017, from
https://www.vox.com/science-and-health/2017/4/17/15322378/how-artificial-intelligence-learns-how-to-be-racist

Levin, S. (2016, September 8). A beauty contest was judged by AI and the robots didn't like dark skin. Retrieved from
https://www.theguardian.com/technology/2016/sep/08/artificial-intelligence-beauty-contest-doesnt-like-black-people

Devlin, H. (2016, December 19) Discrimination by algorithm: scientists devise test to detect AI bias. Retrieved from
https://www.theguardian.com/technology/2016/dec/19/discrimination-by-algorithm-scientis

Driving Question:	How might the adaptation of artificial intelligence continue patterns of humans faults such as racial bias and discrimination?
Names of student researchers: Daniel Gonzalez, Matthew Sarabia, and Jordan Nava	
Grade level: 8	Mentor teacher: Ms. Park

Special notes about handling and or placement or presentation space needed:

ENDING QUESTION

Ending question – Where do I fit in?
What can I contribute as a future computer engineer to ensure that AI doesn't continue human faults of bias and discrimination?