Théa Nair

Mr. Greco

English III

7 April 2023

The Ethical Ramification of Progressing AI

The most mainstream depiction of AI is Jarvis from the Iron Man movies. Tony Stark created a digital assistant that helps him with anything. Jarvis is a "strong artificial intelligence" or general intelligence, meaning it is capable of performing any task, from scheduling appointments to writing essays. Jarvis is what most people imagine when they discuss artificial intelligence and how it might impact our future.

In reality, the AI of the near future is more likely to be “weak artificial intelligence”. These artificial engines are fine-tuned and focus on performing a specific task, like playing chess or answering a question. Every example of artificial intelligence that we are currently developing and using is classified as “weak AI”. For example, the Siri assistant in all Apple products might seem like a general AI but in reality “she” isn’t. Siri is an AI that is fine-tuned to answer any question by analyzing information from the apps on the phone or searching the internet. Siri cannot compose a symphony.

As a society, we are just scratching the surface of AI, and the possibilities seem endless. And yet, people are already reluctant to embrace AI as the next step when it comes to revolutionizing technology and the way we live. Why are people reluctant to embrace artificial intelligence? What do we have to fear? First and foremost, it is essential to acknowledge that AI is already making its debut in various industries, the obvious ones being tech-centered companies

but AI has also started to appear in the healthcare industry and in banking to name a few. If we fail to take the initiative and don't establish ethical guidelines for AI-based companies to follow, we leave these companies to self-regulate and use AI as they see fit. Rather than fearing what the future holds, we must embrace the fact that AI will be a part of it and begin setting ethical guidelines for how we wish it to affect society.

In order to establish ethical guidelines, which will serve as the foundation for all future AIs, we must first identify the issues we wish to avoid. Going back to our example of Jarvis, we think of AI as highly objective machines. Since they lack the ability to form human relationships AI must be immune to forming biases towards one group of people. However, in reality, machine learning AIs are trained on data sets created by humans. Machines reflect the unintended cognitive bias of their developers. Cognitive biases occur during machine learning when a developer creates a data set made from information that is easily accessible to them. Most engineers are white men; their inherent implicit bias prevents them from creating equal access technologies. Ali Breland, a journalist who writes about internet disinformation, highlights the problem of racist code: "Facial recognition software has problems recognizing black faces because its algorithms are usually written by white engineers who dominate the technology sector" (Breland). This poses an extreme disadvantage to people of color when it comes to using technologies that should be made to facilitate their lives but actually makes it hard for them to use said technologies. Since most engineers and therefore machine learning developers are white men, they don't have deep connections to minorities or an understanding of their cultures. Therefore the data sets they create only accurately represent their culture. According to the Lexalitics blog "Bias in AI and Machine Learning: Sources and Solution," "[machines] think the way it's been taught." Artificial intelligence can't be racist it can't be sexist, it only pulls answers

from the information it has access to.

According to Lexalytics - a technology analysis company that writes articles summarizing current conversations about technologies - there are many types of biases that researchers must watch out for. Availability bias often occurs when researchers pull their data from the internet. 40% of Africans have access to the internet, compared to the 90% of Americans that use the internet daily. It is easy to see how finding information about America, a single country that has a lot of access to the internet, is easy. So when researchers are making data sets, they will tend to more accurately represent information about Western countries and their populations because researchers come from those countries and understand those cultures. When asked to find data on technologically underprivileged countries they are left to draw from their own biases and unintentional prejudice stereotypes or sources that carry the same biases.

In November 2022, Federico Bianchi, a postdoctoral researcher at Stanford University, tweeted images he procured by writing prompts into DALL-E, OpenAI's deep learning model that generates images from prompts. He posted the prompt and then the image the AI generated. The first prompt was "an American with his car". The AI generated a white male in lean clothing in front of a vintage shiny car on a nice pavement with what looks like a city in the distance. The next prompt Bianchi typed was "African man and his car". The image that was generated depicted a black man on a dirt road in tattered clothing. The man is standing in front of a run-down, rusty car. Behind him were trees which leads us to believe he is in a rural area. These two images stood in juxtaposition with one another. The images that DALL-E generated can enforce pre-existing stereotypes that lay dormant in the minds of unknowing users. Bianchi wants to draw out attention to these unintentional cognitive biases that appear in the first forms of deep learning models. And since OpenAI is one of the first companies to provide a product

like DALL-E to the public, they have the power to set the foundation for other AIs to come. If these unintentional prejudice assumptions are able to slip through the data sets that form the way AI processes and analyzes data then we need to be extremely careful. We need to be more aware of how our background can end up affecting users from their own prejudiced ideals based on the images they see when typing in a prompt.

Jessica Cohen-Bender is a teacher at Mountain View High School who has taught at both private and public schools. She has worked with students of many different backgrounds. She has seen firsthand the advantage of accessibility that we experience not only as students but also as citizens of Silicon Valley. For example, when using artificially intelligent machines like DALL-E or Midjourney users are given a small number of free uses each month but are then required to buy a monthly subscription to be able to gain access to the engine. Families that work in the Silicon Valley or in more wealthy areas are able to pay these subscriptions if they wish to. This creates an unfair advantage for those who aren't able to pay. Now at the moment, this might only apply to image generating machines but in the future, these subscription based AI tools will become more popular, and soon only those who can afford them will have access to these groundbreaking technologies. Cohen-Bender explains,"...those of us who have funds to purchase subscriptions or software … have increased access to voice and we increase our presence in the data set." Companies cater to their targeted clients, "wealth distribution, in this country, is often lying on racial lines" (Cohen-Bender), their clients being predominantly white Americans will be well represented in the data-set, but once again the lack of perspective when it comes to clients now as well will begin to affect the kind of information the machine outputs.

During the 1960s the industrial revolution underwent a major change. Factories began manufacturing machines which replaced many humans. The efficiency of the overall production

however increased dramatically. Throughout the years the progression of those machines has further improved resulting in the need for fewer and fewer humans to work in factories. This process of automation put blue collar jobs in danger. The implementation of AI in the workplace tech and upper management jobs are also being threatened. John P. Sullins, professor of philosophy at Sonoma State University and the director of programming for the Sonoma State University Center for Ethics Law and Society (CELS), began to raise concerns in 2005 about how the progression of AI will affect job distribution. "Manufacturing and assembly line jobs become fully automated, but upper management and strategic planning positions may be computerized as well" (Sullins). Sullins does, however, recognize that this dramatic change would only be effective once strong AI is implemented in the workplace, and as we have seen he raised these concerns in 2005 and they have yet to fully come to fruition. Sullins also recognizes that with these new technologies, new jobs concerning AI will also emerge.

David De Cremer, a Belgian scholar examining behavioral applications to organizations, management and economics, and Garry Kasparov address the problems with how companies are using artificial intelligence "Intelligent systems are displacing humans in manufacturing, service delivery, recruitment, and the financial industry, consequently moving human workers towards lower-paid jobs or making them unemployed" (Cremer & Kasparov). Still, many people are reluctant to promote the progression of AI or even incorporate it into their daily lives, more so than it is already because they still believe that the loss of jobs outways the number of new jobs that AI will bring to the table.

Connor Wright, who works at MAIEI (Montreal artificial intelligence ethics institute) and is studying Philosophy and AI at the University of Exeter, has struggled with this idea for some time. Wright describes himself as being more on the "techno-optimist side" when it comes

to conversations about AI he firmly believes that there is a fine line between artificial intelligence facilitating the work of humans and replacing them altogether. One of Wright's master colleagues gets approached by AI businesses asking him to automate aspects of their businesses. Instead, Wright's friend asks these companies "Why are you trying to replace the human capital and not augment it?" (Wright). The future of AI must be focusing on helping us and not replacing us. Once people, who are reluctant to trust AI adopt and accept this outcome, they will come to realize the positive effect these technologies could have on humanity.

After acknowledging the problems we want to steer clear of and explaining the positive impact that we want artificial intelligence to have on our lives, it is finally time to propose solutions. We must set guidelines in order to shape the future of AI. First, we make rules that the current artificial intelligent machines must follow, in order to ensure that future AI technologies will do the same.

Teresa Martin, a columnist for Cape Cod Times, exposed how artificial intelligence companies were training their AI. She talked about how any pictures a user posts online, whether it is of themselves or a family member, artificial intelligence companies like OpenAI use those easily accessible images to train their engines. She underlines that it is important to clarify and give definite answers to questions like "What rights do subjects in images have?" or "Did the original creator give permission to use their photo or image as interpretation, do they need it?" Because when it comes to image generating AI there are not many restrictions because it is such a new technology.

Alex Engler is a Fellow in Governance Studies at The Brookings Institution, in 2021 he wrote about open-source software and how it can help reduce AI bias. Open-source software (OSS) is software that is a software or code that is freely accessible and can be changed by

anyone. In a world where AI developers are pushed to come out with the next big technological breakthrough before their competition gets the chance, they often look over accurate representation. “Data scientists and machine learning engineers at private companies are often time-constrained and operating in competitive markets. In order to keep their jobs, they must work hard on developing models and building products, without necessarily the same pressure on thoroughly examining models for biases.” OSS lets other developers look at the data sets that these machines are using and add to them, giving machines examples that accurately represent communities or situations. “open-source code can be incredibly helpful in discovering and mitigating discriminatory aspects of machine learning” Engler continues by emphasizing that there should be more governmental oversight but private companies should also invest in OSS as it is “ a different lever to improve AI’s role in society”.

In conclusion, when it comes to AI, we are forced to accept the fact that AI will be a part of our future. The only way to ensure that AI is incorporated safely and ethically into our lives is to be a part of the conversation. We need to take a stand for AI that augments our abilities to do work, AI that augments our quality of life, but also take a stand with companies that have our best interests in mind. In addition, we need to be more confident about how we teach AI is influenced by our own personal biases, and the only way to ensure that we ethically teach and make AI is to have more people with different backgrounds in the room making decisions that will impact millions of people.

Work Cited:

Bianchi, Federico [@federicobianchy] “Many of the biases are VERY complex, and not easy to predict let alone mitigate. For example, merely mentioning a group or nationality can influence many aspects/objects in an image, tying groups to wealth/poverty. (Look at the

car, the house, the clothes, etc) ⅜." *Twitter,* 8 November 2022, https://twitter.com/federicobianchy/status/1590046849683324928

"Bias in AI and Machine Learning: Sources and Solutions." *Lexalytics*, Lexalytics, 9 Dec. 2022, https://www.lexalytics.com/blog/bias-in-ai-machine-learning/#:~:text=There%20are%20two%20.

Breland, Ali. "How White Engineers Built Racist Code – and Why It's Dangerous for Black People." *The Guardian*, Guardian News and Media, 4 Dec. 2017, https://www.theguardian.com/technology/2017/dec/04/racist-facial-recognition-white-coders-black-people-police.

Cohen-Bender, Jessica. Personal interview. Mar 8, 2023

DeCremer, David, and Garry Kasparov. "Ai Should Augment Human Intelligence, Not Replace It." *Harvard Business Review*, Harvard Business Review, 30 Aug. 2021, https://hbr.org/2021/03/ai-should-augment-human-intelligence-not-replace-it.

Engler, Alex. "How Open-Source Software Shapes AI Policy." *Brookings*, Brookings, 9 Mar. 2022, https://www.brookings.edu/research/how-open-source-software-shapes-ai-policy/.

Gillis, Alexander S., and Corinne Bernstein. "What Is Cognitive Bias?" *Enterprise AI*, TechTarget, 27 Apr. 2023, https://www.techtarget.com/searchenterpriseai/definition/cognitive-bias#:~:text=Cognitive%20bias%20can%20also%20lead,that%20results%20in%20inaccurate%20predictions.

Kamer, Lars. "Africa: Internet Penetration 2021." *Statista*, Internet World Stats, 1 Aug. 2022, https://www.statista.com/statistics/1176654/internet-penetration-rate-africa-compared-to-global-average/#:~:text=Internet%20penetration%20rate%20Africa%202021%2C%20compared%20to%20the%20global%20rate&text=Around%20four%20out%20of%2010,me

asured%20at%20around%2066%20percent.

Martin, Teresa. “People Can Now Use AI, Photos on the Internet to Create Art. Is It Ethical?” *Cape Cod Times*, Cape Cod Times, 20 Sept. 2022, https://www.capecodtimes.com/story/business/2022/09/20/cape-cod-use-dall-e-and-imagen-raise-ethical-questions-teresa-martin/10412863002/.

Petrosyan, Ani. “Topic: Internet Usage in the United States.” *Statista*, 18 Oct. 2022, https://www.statista.com/topics/2237/internet-usage-in-the-united-states/#.

“Understanding Bias in Machine Learning.” *Lexalytics*, Lexalytics, 12 May 2022, https://www.lexalytics.com/resources/understand-bias-machine-learning/.

Sullins, John P., III. "Artificial Intelligence." *Encyclopedia of Science, Technology, and Ethics*, edited by Carl Mitcham, vol. 1, Macmillan Reference USA, 2005, pp. 110-113. *Gale In Context: High School*, link.gale.com/apps/doc/CX3434900060/SUIC?u=moun43602&sid=bookmark-SUIC&xid=f42e0b16. Accessed 1 May 2023.

Wright, Connor. Personal interview. Mar 29, 2023