Technoracism: The Inherent Racism Within AI and How It Affects People of Color

by

Caya Carter

A Thesis submitted to

The Director of the Honors Program and the Honors Council of
Elizabeth City State University
in partial satisfaction of the requirements for the
Honors Program

27 April 2023

Thesis directed by

Dr. Karen Strother-Jordan, Ph.D.

Department of English and Digital Media

And

Director of University Honors Program

Abstract

As the usage of artificial intelligence systems is on the rise, so are racist inaccuracies. The algorithms and technology of these systems leave out or discriminate against racial minorities due to a lack of consideration for their physical or cultural circumstances, which leads to the technology working ineffectively. This thesis looks at various instances of racist bias within multiple artificial intelligence systems. It summarizes a sample of the vast information on this massive fault within technology over the past five years and seeks to explain the social effects of this issue and how it perpetuates racial, sexist, and even ageist bias. However, this thesis does specifically focus on racial bias. This paper seeks to ask the question of how do algorithms and inputs of search engines perpetuate discriminatory social constructs, and to what extent. The study conducted within this research paper looks at Google Trends and the related topics and queries in the United States and how they continue to perpetuate racism and prejudice.

Acknowledgments

A huge thank you to my advisor, Dr. Karen Strother-Jordan, and the director of the ECSU University Honors Program, Dr. Dolapo Adedeji for their patience and guidance through this process. I would also like to thank my parents, friends, and family for their emotional and mental support throughout the duration of my research.

Table of Contents

Introduction	4
Review of the Literature	5
Methodology	17
Findings	17
Analysis, Recommendations & Conclusions	21
Works Cited	24

Introduction

As mankind advances day-to-day, so does the technology we use. It becomes integrated into everything we use such as faucets, doors, and watches to software like TikTok filters, search engines, and facial recognition software. However, these artificial intelligence (or AI) systems are not perfect. In order for most of these systems to operate they must be based on the testing, algorithms, and planning of a team of people. While many companies strive to include all demographics in their designs, there is a level of bias that can be incorporated into these AI processes. Due to undersight, lack of diversity, and algorithmic bias, many of the items and programs we use every day have racism programmed into them. While some may just see this as an unfortunate outcome that will become fixed over time, others are starting to see the issues that persist today. This is not just an issue of nonfunctional water faucets or wonky search results. This is a real issue permeating and negatively affecting minorities' outlook in financial, health, and hiring fields. Ruha Benjamin aptly refers to this new set of discrimination laws in *Race After Technology* as the "New Jim Code"(x).

Algorithmic bias is the systematic and unfair exclusion or discrimination of certain racial, gender, age, sexual orientation, ethnic, and/or religious groups through an algorithmic pattern or computer program. In this case, we will be focusing on the fields of race and gender.

Looking over the recent research on this topic, two books have been instrumental in providing a look at the current state of racism, algorithmic bias, and perpetual stereotypes in AI and computer programs: Ruha Benjamin's *Race After Technology* and Safiya Umoja Noble's *Algorithms of Oppression: How Search Engines Reinforce Racism*. Both books lean more towards including information about the prejudices around African Americans but also present other observations and research regarding minorities both in the United States and out.

As artificial intelligence has been incorporated into more fields, various studies have been done regarding issues of bad medical discernment through judgments of AI. This has been specifically prevalent with the recent developments of artificial intelligence interference in COVID-19 (Chase 2020, Leslie et.al 2021, Shachar et. al 2020). This is also true in instances where AI has been used to increase surveillance in the capacity of criminal justice (Zajko 2021, Noriega 2020). This paper will seek to summarize the various fields in that AI's algorithmic bias puts various minorities of both races and sex at a major disadvantage.

What is Algorithmic Bias?

To truly understand the drastic nature of this issue, algorithmic bias must first be explained. Algorithmic bias, or The New Jim Code as Ruha Benjamin describes it, is the patterned prejudice of an artificial intelligence system due to the skewed data that it was fed while being developed or trained. This issue is particularly harmful to minorities as it preserves and perpetuates the same social discrimination that AI claims to solve. The biggest problem is that this issue can start a never-ending cycle of bias and discrimination in fields such as healthcare (see Figure 1.) These faulty AI are a direct reflection of the society its developers claim to change and improve.

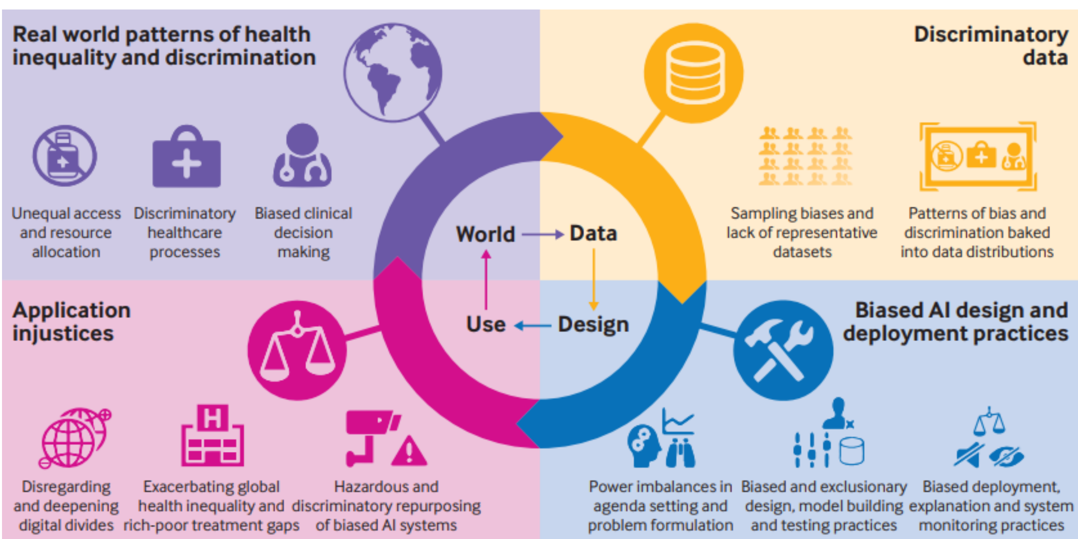

Fig.1 - Leslie, David, et al. " Cascading Effects of Health Inequality and Discrimination Manifest in the Design and Use of Artificial Intelligence (AI) Systems." *BMJ*, 2021, https://www.bmj.com/content/bmj/372/bmj.n304.full.pdf. Accessed 2023.

This happens when they are either fed biased data, designed with the inherently discriminatory beliefs and practices of those it was developed by, or when these design teams are not diverse enough to eliminate the oversight that including minorities may have eliminated. This can have drastic real-life consequences as this can further disadvantage religious, racial, and gender identity groups that are already hurting. This is especially true for machine learning systems like chatbots and suggestive search engines that can be changed based on the frequency of who is interacting with the technology. For example, back in 2016, Microsoft released a Twitter AI named Tay that learned from users that interacted with it. In less than 24 hours, the AI was tweeting out majorly offensive tweets (Fig.2) that users had coaxed its algorithm into repeating.

Fig.2 - Mellor, Gerard. “User @gerardmellor Documents the Complete Change in Response of Microsoft's Twitter AI, @TayAndYou, after 24 Hours.” *Twitter*, 24 Mar. 2016, https://twitter.com/geraldmellor/status/712880710328139776/photo/1. Accessed 2023.

Review of the Literature

Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification (2018)

This journal article evaluates AI's algorithmic bias within facial recognition systems. *Gender Shades* is a study that examined the accuracy and disparities in commercial gender classification systems, and how this bias affected darker-skinned individuals and women. Facial recognition has been used recently to decide who is hired, granted loans, released early from prison, and other major decisions that usually are given to human morals and decisions.

Buolamwini and Gebru state that the data sets that many facial recognition systems use for training allow for prejudice to continue to be perpetuated (1). This error can be attributed to minorities being underrepresented in these data sets and other various errors. Lack of representation in these data sets makes these facial detection programs unable to read the faces of people of color accurately, thus resulting in major deficits that European faces don't receive. More and more programs continue to be produced such as Faception, a program that claims to report an individual's IQ or predisposition to commit terrorism all from just a picture of your face. As programs like this continue to be used to make life-changing judgments and decisions like this, minorities are constantly left with a massive disadvantage. With data sets like these purposely placing negative connotations to non-European physical characteristics, racism, and even featurism continue to be maintained.

In this paper, a study was conducted to determine how various programs or classifiers read and sort the faces of darker males and females and lighter males and females (see Fig.3).

The results ended with an observation that these classifiers performed better on male and lighter faces than their counterparts (8). It concluded with stress on fairer algorithms needing to be developed before something with such a large percentage of error (8.1% – 20.6% on females and 11.8% – 19.2% on darker-skinned subjects (11)) could be used in such important institutions as criminal justice, healthcare, and financial fields.

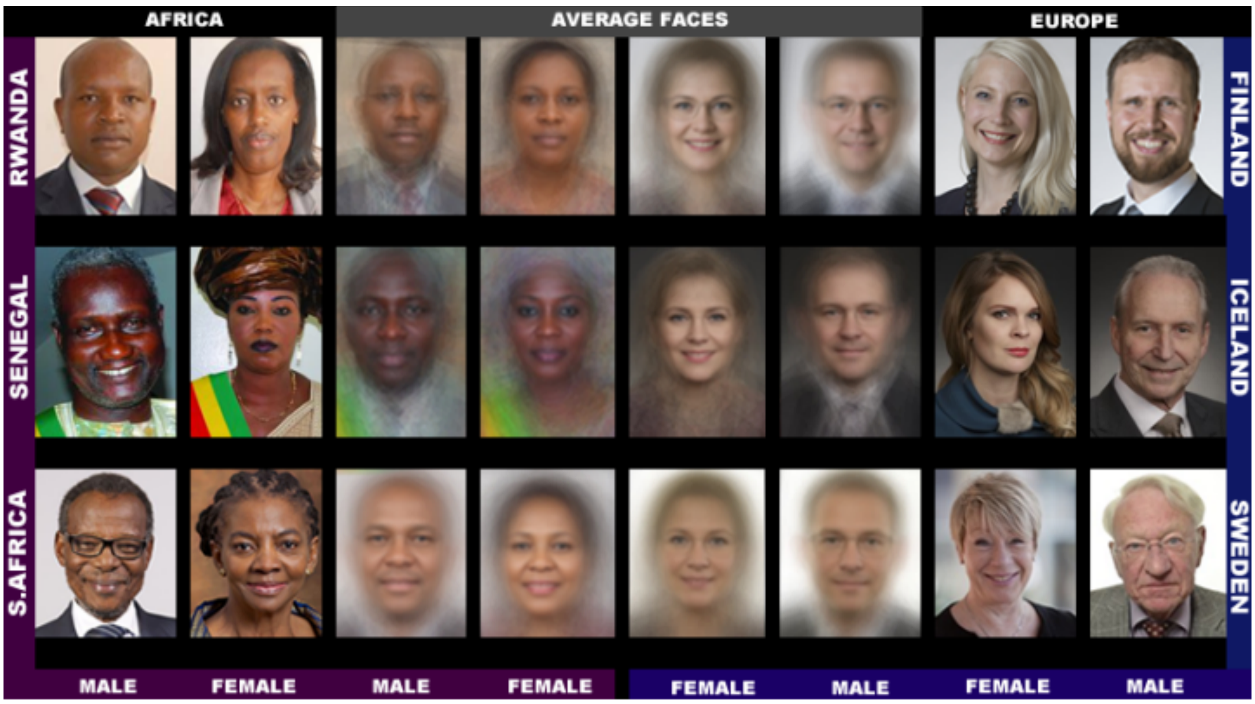

Fig.3 - Buolamwini, Joy, and Timnit Gebru. "Example images and average faces from the new Pilot Parliaments Benchmark (PPB). As the examples show, the images are constrained with relatively little variation in pose. The subjects are composed of male and female parliamentarians from 6 countries. On average, Senegalese subjects are the darkest skinned while those from Finland and Iceland are the lightest skinned." 2018, https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

The Ugly Truth about Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity (2018)

Howard and Borenstein explore the possibility of bias and discrimination in the development and use of robots and artificial intelligence that have been released to the general public. They begin by recognizing the excitement and zeal around the benefits of these technologies to increase efficiency and improve life as we know it. That being said, it is argued that one must also acknowledge the possibility of negative consequences, specifically concerning the amplification of society's biases and preconceptions (1523).

It is noted that these creations can only be as unbiased as their creators, and therefore we must hold these creators accountable (1531). If those who develop these systems hold prejudiced views, those same inclinations will be reflected in what they created. For example, if a predominantly white and male team structures these AI systems, they may not work as accurately for women and people of color. These kinds of problems are shown in facial recognition software that fails to accurately identify people of color or may even racistly confuse them with animals (1525). This is also an issue with public adaptable chatbots that change according to the people who interact with them, thus eventually kicking out racist and sexist responses from those that feed such sentiments into it.

To prevent these complications from being retained in robots and AI, they argue that we all must be aware of our own biases and continue to address them. This also includes making more diverse development teams and data sets. It can never be fully eliminated but continuing to correct these incorrect assumptions will progressively lower the amount of error within these programs.

Algorithms of Oppression: How Search Engines Reinforce Racism (2018)

This book by Safiya Umoja Noble is an extensive look at how search engines perpetuate and bolster systemic biases, stereotypes, and discrimination. She debates that search engines are not entirely unbiased and neutral, but are actually influenced by corporate, political, and societal interests, oftentimes at the cost of remaining discriminatory (5). She intentionally focuses on how marginalized communities (like POC and women) are often misrepresented or altogether excluded from these search results.

Noble argues how these search algorithms can be used to further stereotype and judge minority populations. There are examples provided where “searching on ‘black girls’ surfaced “Black Booty on the Beach” and “Sugary Black Pussy” to the first page of Google results, out of the trillions of web-indexed The Future pages of Information that Google Culture Search crawls” (Fig.4) (64). These results perpetuate hypersexualization over results that would be more positive or have a higher amount of actual correlation.

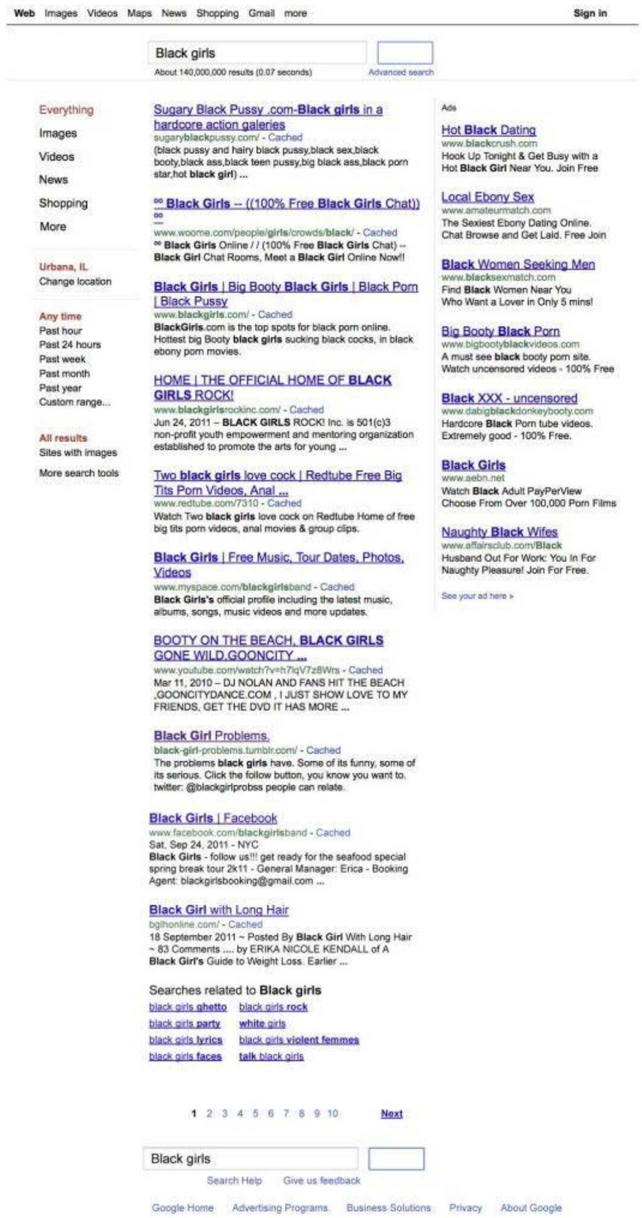

Fig.4 - Noble, Safiya U. "First Page of Search Results on Keywords 'Black Girls.'" *ProQuest Ebook Central*, 18 Sept. 2011, https://ebookcentral.proquest.com/lib/ecsu-ebooks/reader.action?docID=4834260. Accessed 2023.

This book also explains the impact of commercial interests on search algorithms. Noble asserts that search engines are made to favor and promote content that makes advertisers money rather than correlating with the search the user has put in (35-36). This turns into a massive problem that can foster a narrow and stereotypical view of the world and misinforms the public, all for the sake of corporate greed.

Noble demands greater honesty and accountability in the programming and implementation of search engines. Noble is also a supporter of the diversification of the workforce behind these engines as a means to address this issue. She stresses the importance of educating the public on the impact of these search engines and urges readers to verify, question, and evaluate the roots of these biases in order to improve these societal consequences.

Racism, responsibility, and autonomy in HCI: Testing perceptions of an AI agent (2019)

This research paper concentrates on the impact of race in human-computer interactions (HCI) and the resulting perceptions of how people feel about AI agents. Joo-Wha Hong argues that the use of AI agents in HCI can enforce racist biases and discrimination. Hong urges that it is important to confront these biases as a means of ensuring fairness and equity.

This study tested participants' perceptions of an AI agent that was made to predict crime. The agent included blatantly racist predictions. This study resulted in the AI agents receiving a similar amount of blame to human agents, despite a lack of autonomy. The people in the study expected the same amount of fairness from an AI agent as they do from a human one as well (83). Overall, this study asserts that the public perception and trust of AI is high, and just as scary considering it makes clearly racist judgments. Hong suggests that in order to properly correct this issue, the prejudice in AI must continue to be addressed.

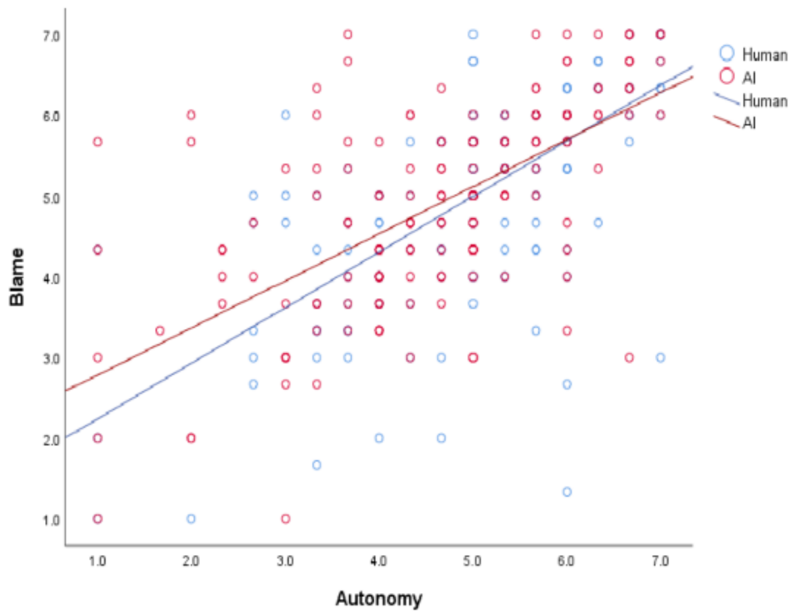

Fig.5 - Hong, Joo-Wha, and Dmitri Williams. "The regression analysis of the relationship between the autonomy of the crime predictor and the level of blame."*Elsevier.* 2019, https://reader.elsevier.com/reader/sd/pii/S0747563219302389?token=racis&originRegion=us-east-1&originCreation=20230427183223.

Feminist AI: Can We Expect Our AI Systems to Become Feminist? (2019)

In this paper, Wellner and Rothman look at the sexist bias within AI systems and where it comes from and explore steps that can be taken to improve this dilemma. When translating texts from Turkish, where a neutral gendering was used, to English, various occupations like doctor and soldier were automatically assigned to males whereas other jobs like teacher and nurse were automatically classified to females (192). Just like racially-biased algorithms perpetuate racially-biased ideals, gender-biased systems do the same exact thing.

In order to get out of this cycle, they suggest four solutions. The first is to completely remove gender from datasets. This allows for the algorithm to focus specifically on the actions and input of the user without adding sexist implications. The second calls for transparent algorithms, which would be AI programs that clearly state which factors led to certain decisions, this would then allow for humans to detect where the bias is (200). The next solution is to build systems that actively fight against the bias of the data sets that they are fed. This may include certain features being taken out of programs so as to completely inhibit the machine from making a deliberately sexist choice or suggestion. The last solution is human involvement. This allows for the AI to do its job, however, it is ultimately aided by human interference so as not to completely place vital decisions in the hands of automated programs. Overall, there are plenty of ways to make AI feminist, it is just a matter of how and when.

Race After Technology: Abolitionist Tools for the New Jim Code (2019)

In this book, Ruha Benjamin seeks to inform the reader of the recent development of discrimination within the internet and AI systems and algorithms and how to go about fixing and fighting these plights. It explores how technology can sustain the presence of bigotry within the digital world and the real one. The author brings to light a new form of digital oppression she names "The New Jim Code" (5). Benjamin brings to light various predicaments in which minorities are left on the sidelines or affected negatively by the discernment of AI systems.

This book critiques different fields of discrimination such as ethnic naming affecting hireability (5), the overwhelming bias towards White people to win AI-judged beauty contests (see Fig.6) (50), and the blatant exclusion of certain cultural norms from data sets (79.) She also mentions the digital divide, which is the lack of access to technology and other digital means, which further instill the divide between societal, racial and economic classes.

In conclusion, *Race After Technology* is a hard-hitting and information-packed read that discusses how technology affects the way we view and analyze race. It points out faults while also offering solutions.

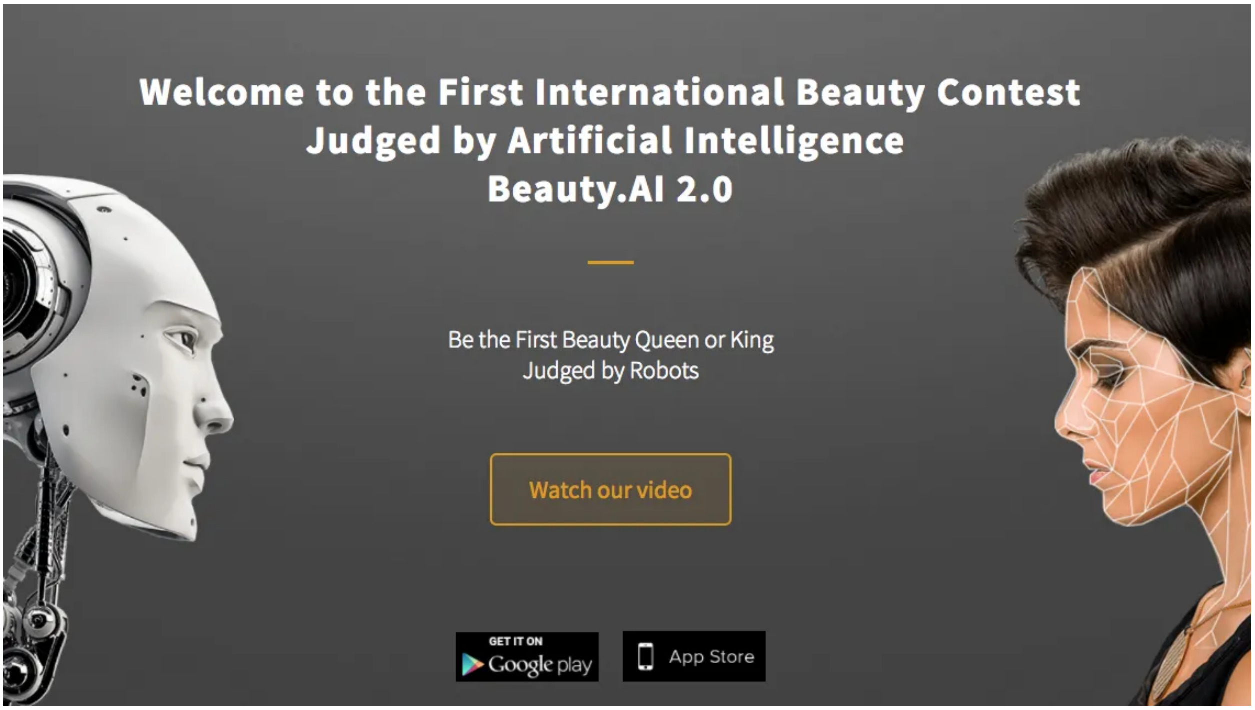

Fig.6 - beauty.ai. "Beauty.AI 2.0"

Artificial Intelligence and Race: A Systematic Review (2020)

Written by Channarong Intahchomphoo and Odd Erik Gundersen, this paper takes a comprehensive look at the research on where race and artificial intelligence overlap. They examine a wide range of 36 studies in the field and observe four main interactions that come up between the topic of race and how AI interplays with it. The four main interactions are as follows: "(i). AI causes unequal opportunities for people from certain racial groups (see Fig.7), (ii). AI helps to detect racial discrimination, (iii). AI is applied to study the health conditions of specific racial population groups, and (iv). AI is used to study demographics and facial images of people from different racial backgrounds" (74).

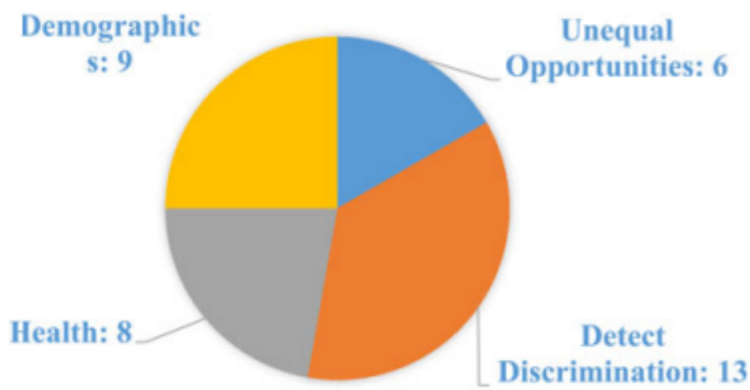

Fig.7 - Intahchomphoo, Channarong, and Odd Erik Gundersen. "Relationships between AI and race and the number of papers."*ProQuest.* Jun. 2020, https://www.proquest.com/docview/2442778504/fulltextPDF/914085F4C79D4F05PQ/1?accountid=10717. Accessed 2023.

In general, this paper highlights the importance of acknowledging the intersection of race and AI. It is a helpful tool in summarizing the current research on this affair and emphasizes the need for deeper thought and consideration in the evolution and usage of AI systems.

Does "AI" stand for augmenting inequality in the era of covid-19 healthcare? (2021)

David Leslie's team summarizes the effects of AI in the field of healthcare, specifically concerning COVID-19. It speaks on the risks and consequences of over half of the states in the US not submitting the COVID-19 data of minority citizens. The data set is incomplete and therefore leads to faulty reports and algorithms. This continues to perpetuate the cycle and reinforce algorithmic bias in healthcare. The chart below explains the risks and remedies of this issue.

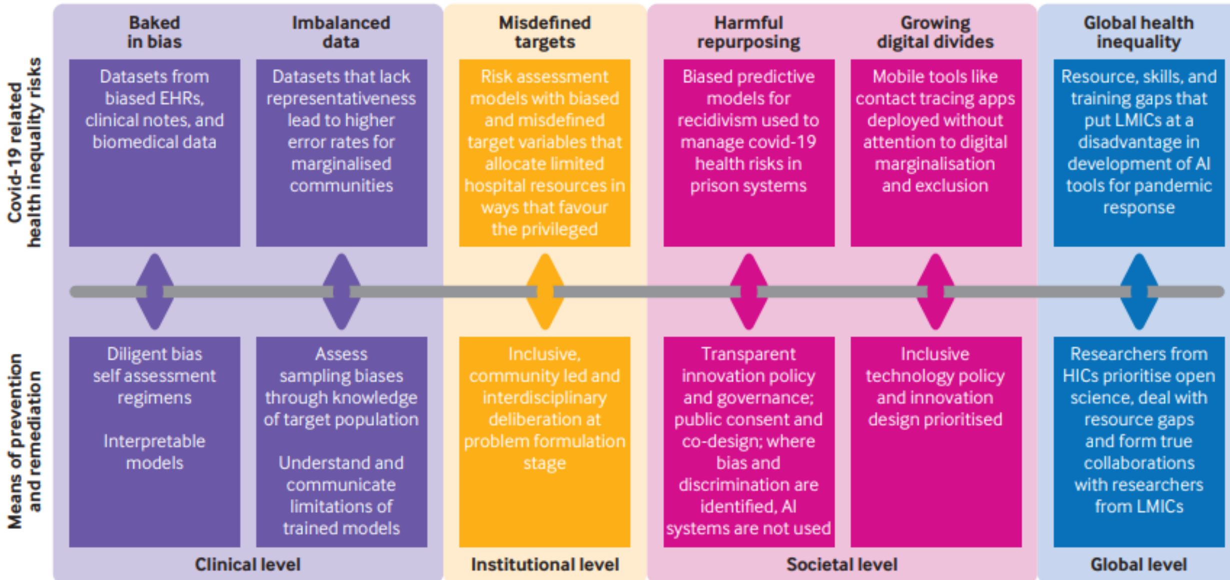

Fig.8 - Leslie, David, et al. “Risks of, and remedies for, developing and deploying artificial intelligence (AI) systems safely. EHRs=electronic health records; HICs=high income countries; LMICs=low and middle-income countries.” *BMJ*, 2021, https://www.bmj.com/content/bmj/372/bmj.n304.full.pdf. Accessed 2023.

Methodology

Using Google Trends, I will input descriptor keywords (i.e. search topics like Black woman, Black man, ugly man, why are Asian women so, etc,) to identify the related queries and topics. The search parameters of the results will overlook the frequency of search and correlation over the past five years in the United States. I will then discern the amount of negative related queries in comparison to their keyword counterparts (i.e. why are White women so, White men, etc). I will put this information into graphs to show my findings.

Findings

As a result of searching through Google Trends data, I have indeed found a negative pattern when it comes to people of color and related search queries. These related search queries show exactly what people search for most in relation to a specific topic. This is important because Google Search's suggestions tend to cater towards the highest trending correlating searches. Users that search for the terms that I entered in have also searched for or been suggested the related queries. The frequency at which they search for or are suggested and then search for these topics varies from a range of positive neutral and negative results.

When entering in the search term "why are black people so" some of the highest related queries finish the phrase off with loud (100%), racist (86%), mean (79%), tall (58%), and stupid (44%.) In the top ten results, ⅘ of the results are of negative connotations and the other ⅕ relates specifically to the stereotype of Black people's athletic prowess.

Table 1 - "why are black people so"

Related query	Relative correlation score
loud	100
racist	86
mean	79
tall	58
stupid	44
fast	43
angry	42
rude	38
good at running	28
hated	28

When typing in "why are white people so" into Google Trends, many of the same results pop up, but with different correlations. There are three queries that are less related specifically to finishing the prompt. When attempting to type in the prompt, "why are asian people so," Google Trends reports that there is not enough data to show the related queries, therefore suggesting this is not a highly searched question.

Table 2 - "why are white people so"

Related query	Relative correlation score
racist	100
mean	84

angry	49
evil	41
tall	34
south africa	27
why are black people so mean	26
rude	18
weird	17
why are black people so tall	16

Upon searching "ugly people" the following results show up in the table below. Terms relating to both White people and Black people show up in the top 25, but only Black people show up in the top 10 and they show up twice, both within the top five. In contrast, when searching for "beautiful people," Trends reveals no particular race within the top ten.

Table 3 - "ugly people"

Related query	**Relative correlation score**
pictures	100
ugly people pictures	99
ugly pictures	96
ugly black people	83
black people	77
fat people	69
ugly fat people	68
fat	68

funny people	41
funny	40

When looking at the gender bias side of things, I had to switch up my terms, by comparing the related queries for the search terms, “why are black women so,” “why are black men so,” “why are white women so,” “why are white men so,” “why are asian women so,” “why are asian men so.” When it comes to the related queries for both the White women and the Black men questions, the results do not correlate directly to the search and instead seem to reflect homework questions. For White men, in the top ten their top result finishes off the phrase with angry (100), the rest of the related queries are also homework questions. For Asian women, there are two results; beautiful (100) and attractive (95). For the query with Asian men, that search has not been conducted as frequently to have enough data to show. However, with Black women, there are eight related queries with six of them directly relating to finishing off the phrase.

Table 4 - “why are black women so”

Related query	**Relative correlation score**
fat	100
mean	96
rude	52
ugly	48
loud	40
why are so many black women single	36

Discussion

Analysis

In summary, this data pool has a predisposition to associating negative terms more so with terms concerning Black people and women, with the results especially harming Black women. Within the United States in the past five years, search results have been constantly affected by these negative associations. The associations will continue to permeate the search engine's suggestions due to this algorithmic bias. These negative connotations have roots in racism and misogyny as stereotypes about Black women have sat in American society for decades. Users' searches constantly associate Black people with being loud and ignorant and Black women with being overweight, unattractive, and undateable through these related queries.

When it comes to searches and related queries concerning Asian people, the data sets simply aren't there. However, for Asian women, the results were minimal but appeared positive. However, this could be due to the fetishization of Asian women as well.

I was quite surprised to see a large amount of negatively related queries for White people as I did not expect that. I wish more data could have been provided regarding the Asian demographic. For some keywords, such as "why are black men so," and "why are white women so" the related queries appear actually to be quite unrelated.

In order to be less biased, Google Trends must be more transparent. Raw data would be preferred to the keywords' relativity to each other and releasing the results of smaller data sets. We cannot combat issues within search data without being able to look at it properly.

Conclusion

Overall, the potential bias within AI can be incredibly harmful. Although developers may not mean to make AI systems operate this way, it must be solved. AI systems can only be as unbiased as the people that made them and the data that it is trained on. Biased data sets result in biased decisions which filter right back into the AI's algorithm. Altogether, it creates a vicious cycle that without human interaction could have grave consequences in the real world. Before we rely more on artificial intelligence's decisions we must work to make these programs as unbiased as possible. Solutions can be various things such as diversifying development teams and data sets, reporting inappropriate suggestions in search engines and constantly updating and adapting systems to serve people correctly and without prejudice.

In conclusion, preprogrammed racism within AI has become a major concern for people of color as it reinforces biases, prejudices, and even fetishizations that harm minorities today. The lack of ethnic and gender diversity in data sets, algorithms, and even software development teams has resulted in AI systems that amplify the issues that minorities face which leads to negative outcomes in a whole array of fields such as healthcare, criminal justice, hiring and so much more. Additionally, there needs to be greater transparency and accountability put on the companies we trust to make these algorithms. Ultimately, in order for all to enjoy and reap the benefits of AI, it must be equal, accessible, and beneficial for everyone.

Works Cited

Benjamin, Ruha. *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity, 2019.

Buolamwini, Joy, and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." *Proceedings of Machine Learning Research*, vol. 81, 2018, pp. 77–91. *Gender Shades*, http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf.

Dancy, Christopher L. "Using a Cognitive Architecture to Consider Antiblackness in Design and Development of AI Systems." *Proceedings of the 18th International Conference on Cognitive Modelling*, 2022, pp. 65–72.

Hong, Joo-Wha, and Dmitri Williams. "Racism, Responsibility, and Autonomy in HCI: Testing Perceptions of an AI Agent." *Computers in Human Behavior*, vol. 100, 2019, pp. 79–84. https://doi.org/10.1016/j.chb.2019.06.012.

Howard, Ayanna, and Jason Borenstein. "The Ugly Truth about Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity." *Science and Engineering Ethics*, vol. 24, no. 5, 2018, pp. 1521–1536. https://doi.org/10.1007/s11948-017-9975-2.

Intahchomphoo, Channarong, and Odd Erik Gundersen. "Artificial Intelligence and Race: A Systematic Review." *Legal Information Management*, vol. 20, no. 2, 2020, pp. 74–84. https://doi.org/10.1017/s1472669620000183.

Leslie, David, et al. "Does 'AI' Stand for Augmenting Inequality in the Era of Covid-19 Healthcare?" *BMJ*, vol. 2021. https://doi.org/10.1136/bmj.n304.

Mellor, Gerard. “Twitter User @gerardmellor Documents the Complete Change in Response of Microsoft's Twitter AI, @TayAndYou, after 24 Hours.” *Twitter*, 24 Mar. 2016, https://twitter.com/geraldmellor/status/712880710328139776/photo/1. Accessed 2023.

Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press, 2018. *ProQuest Ebook Central*.

Wellner, Galit, and Tiran Rothman. “Feminist AI: Can We Expect Our AI Systems to Become Feminist?” *Philosophy & Technology*, vol. 33, no. 2, 2019, pp. 191–205. https://doi.org/10.1007/s13347-019-00352-z.