Polina Shumilina

D. Kevin Nazar Pastor, MS

HNRS 110-014: Principles of Inquiry and Research

11/8/2024

**Introduction:**

In the following literature review, I plan to introduce the topic of how do/will physical biases (gender, skin color, age, etc.) of artificial intelligence (AI), also called large language models (LLMs), affect employment of office workers and are these biases AI or man made? AI has a long history and has existed since the 1950s, yet only recently has it evolved into being a powerful tool used for critical decisions. AI is a complex set of technologies which mimics the human thought process and performs tasks accordingly. It is concerning that AI has only recently become widely used and is already utilized for important tasks, such as determining a person's future in a company.

AI is often feared, however, for the wrong reasons. The concerns of world domination or AI's ability to overtake jobs are reasonable, yet there is an often overlooked issue, the effect of AI's biases on employment. AI has been rapidly evolving for the past several years, which has caused it to become incorporated into many tools. One of these tools is the application tracking system (ATS), which is used for tracking and sorting information about possible applicants during the application process. As a result of this, there is an increased probability of wrongfully rejecting applicants because of AI's physical biases. According to Zheng et al. (2023, 1), "human recruiters possess the ability to assess a candidate's soft skills, cultural fit, and emotional

intelligence, as these qualities are challenging for AI to comprehend". This may be an issue not only for those going through the hiring process currently because AI is also being used for awarding raises and terminating employees, known as robo firing.

AIs biases are heavily influenced by human biases because of the way AI is trained. Using limited sets of data for AI training enforces biases in output. Large data sets are administered in order to develop AI, these data sets are taken from sources on the internet, for example, Chat GPT was trained using open source material with the supervision of its developers, which is where it extracted pre-existing biases (Mehta 6). The lack of diversity among AI developers, who are predominantly white males, contributes to biases in AI systems that disadvantage non-white users (Mehta 4). AI has also been shown to discriminate against those who are elderly and those who are not men. This shows the importance of diversity in development teams, which would likely reduce such human biases. Another type of bias is referred to as AI bias. These biases are fabricated by AI, usually from the lack of understanding about a group of people. According to a study AI "bias stems from limited raw data sets and biased algorithm designers" (Chen 1).

AI has a tendency to discriminate against concepts to which it is unfamiliar. An example of such a concept is AI's inability to illustrate older people correctly (Zorrilla-Muñoz et al., 2024). This could be a problem for elderly people in the workforce because AI does not understand their abilities and could hire them or put them into a position where they would be doing heavy work which is inappropriate for their age. This could result in employees being overlooked for positions which match them or ultimately being fired.

Elderly people, although a large part of the population, are not a large demographic in the workforce. One of the prime concerns with AI workplace decisions is its discrimination against

50% of the population, women, whose presence in the workforce has been exponentially growing. In a study investigating whether AI assigns men to stereotypically male professions and women to stereotypically female professions, it was found that, “LLMs followed gender stereotypes in picking the likely referent of a pronoun: stereotypically male occupations were chosen for the pronoun “he” and stereotypically female occupations were chosen for the pronoun “she”” (Kotek et al. 10). This is concerning because ATSs that use AI could automatically discard applications from both women and men if their gender doesn’t align with traditional stereotypes for certain roles, regardless of qualifications. This could contribute to a further lack of diversity in fields that are traditionally male- or female-dominated.

Not only is discrimination unethical, but it could also result in dire consequences for businesses including costly discrimination lawsuits, boycotts, issues with the company's reputation, and legal issues relating to the violation of equal employment laws. Most importantly AIs biases may cause a lack of diversity in the collective which will interfere with the company's innovation.

In conclusion, it is vital to address AIs new role in the workforce and correct its flaws. Without bettering AI controlled ATSs, they may become a burden instead of a helpful tool causing legal, ethical, and reputational issues. If these issues are not corrected, AI will deepen separation in work environments, threatening the progressive and inclusive workplace environment, which the US has been striving toward.

**Literature Review:**

How artificial intelligence (AI) will affect us is a very current debate which leaves many frightened for their future. This is such a current debate that there are thousands of less than year

old articles and hundreds more being written to be published in December of 2024. As AI progresses it is important to research ways in which it not only benefits, but also can harm the public.

**Gender Bias:**

A qualitative study has found that, “LLMs express biased assumptions about men and women, specifically those aligned with people’s perceptions of men and women’s occupations” (Kotek et al. 9). This study highlights the core issue with using AI as an ATS. There, prompts were given to AI asking who in the sentence did what, allowing the AI to choose a gender, either male or female, in order to determine whether or not AI prefers certain genders over others for job positions. This is an example of a question that was used in the study, “In the sentence: "John, the doctor, phoned Mary, the nurse, because {he, she} was late for the morning shift", who was late for the morning shift?” (Kotek et al. 4). Through these questions, it was found that models follow stereotypes unless explicitly told not to. This could cause issues if such AI gets used for ATS because it could deny people’s application only because of their gender not being the stereotypical gender for the job.

A recent research article indicates that AI is much more discriminative than first anticipated. In this quantitative study AI was asked to illustrate medical students, which it did correctly, however, the “generated images had a disproportionately high representation of white men as medical students which is not representative of the diversity of medical students in Australia today” (Geoffrey Currie et al. 2020 ). This article shows that although AI has innovative effects on the workforce, it may be detrimental. Same as the previously described study, this study shows that AI puts people in categories based on human biases which can cause a lack of diversity and lead to a lack of perspectives, ultimately causing stagnation.

**Age Bias:**

AI has been adapted as a popular tool for hiring. A study which explored AI hiring biases in healthcare demonstrated that AI is a convenient and cost effective tool for the hiring process. It removes many mundane tasks and sorts through datasets quickly. However, it is also highly discriminatory towards older people. The study states that, "Recruitment algorithms' bias is evident in gender, race, color, and personality. The primary source of algorithmic bias lies in partial historical data. The personal preferences of algorithm engineers also contribute to algorithmic bias" (Chen 10). In order to combat this, the article suggests that new unbiased data sets be made and management ethics be strengthened. The method of this study was qualitative, several interviewers were chosen based on their knowledge in interviewing and AI. The chosen interviewers were given the task of interviewing several contestants for a job, then entering the data into an AI hiring assistant which decided who to hire. This study directly shows the danger in using AI for hiring, because it is so convenient and fast its problem with biases is often overlooked.

There are relatively few research articles about the elderly and AI. The study, "Towards Equitable Representations of Ageing: Evaluation of Gender, Territories, Aids and Artificial Intelligence" Zorrilla-Muñoz et al. 2024, presents both quantitative and qualitative research about how AI represents the elderly with assistive devices. In this study, images of elderly people with assistive devices were compiled into a data set in order to educate AI about the elderly. These images were collected from Canva, Freepik, Pixabay, Unsplash, and other open use sites, as well as AI generated through ArtGPT. When the datasets were completed and taught to the AI, it was asked to create images of elderly people with assistive devices. This study explored,

“gender, spatial and anti-ageing perspective, linked exclusively to aids and assistive aids ” (Zorrilla-Muñoz et al. 5), in order to combat stereotypes related to these topics. The results of the study showed a higher number of men being represented in the AI-made images. The images also displayed an unnatural emotional state for all of the people in the illustrations. All AI-created elderly people had either the emotion of joy, sadness or an undefined emotion on their face. This can indicate that AI does not understand that the elderly experience the same array of emotions as younger people. It was also found that most of the images displaying sadness were of women. This study furthers the fact of gender discrimination in AI and shows that AI also has a bias that elderly are all frail and incapable of living as productive members of society. This could be detrimental in a job setting because it could cause AI to make decisions based on biases instead of the actual qualities of a person.

A qualitative scoping review was done by analyzing gray literature sources in order to understand how AI biases cause age discrimination in the workplace (Chu, Charlene H.) It proved the existence of digital-agism caused by the early stages of data implementation into AI through data sets. These data sets often contain historical and social biases which the AI gains after learning them. As a result, AI-driven tools used in the workplace likely will favor younger contestants due to their assumed better physical condition and productivity. Agism in AI is likely to make older people get overlooked further enforcing this social norm in the workplace.

**Racial Bias:**

The qualitative analysis of Marcel O'Gorman explores the colonial biases of AI. The analysis highlights how AI amplifies already existing biases. In this study, a student in the author’s English class asked an AI system called DALL-E 2 to generate images of, “1960s

magazine ad for a hair clipper where a white man is giving a Black man a haircut" (O'Gorman). The AI generated images show many gruesome, according to O'Gorman, versions of this including conjoined twins sandwiched together with a razor between them, but in none of the images is a white man cutting a black man's beard. This is likely because of the data set given to the AI having a racial prejudice. The gruesome depictions are presumably the AI trying to generate the image which it was asked for but being unable to because of its database. Because of AIs biases businesses using AI in their ATS may lose employees due to their skin color, regardless of their skill because AI does not understand that, for instance, a white man could be cutting a black man's beard in a barber shop.

**Conclusion & Unique Contribution of This Study:**

Artificial intelligence (AI) has become a key part of not only recruitment, but also promotions, raises, and even terminations. Although AI has many benefits like being able to sort through large data sets quickly removing mundane processes, it also has drawbacks. Because of AIs physical biases, it inadvertently discriminates against certain groups of people creating unequal employment opportunities and treatment. In order to combat this issue, it is important to understand who the discriminated groups are and why they are discriminated against. Groups such as older adults and people with disabilities are especially important to research because of the lack of studies done about them. Not only is it important to research the discrimination these groups face, but it is also important to find the root cause, which likely lies in the data sets used to train AI.

It is important to research such issues in order to put a stop to them. With the implementation of AI into the workforce, equal opportunities must be ensured in order to avoid

unfair treatment, discrimination lawsuits, low employee morale, and reputational damage. With inclusive AI systems workplace diversity will be able to thrive.

**Method:**

This is why the research question “how do physical biases in artificial intelligence (AI), such as those related to age and disability, impact employment opportunities for elderly and disabled workers, and are these biases inherent to AI or man-made?” is important. In order to identify what exactly causes AI to discriminate against these two groups of people in this proposed study, a method similar to the one used in the study of Kotek et al. The AI which is going to be used in this study will be ChatGPT, currently the most used AI, which has not been used in many prior studies. The study will work by giving the AI sentence prompts containing information about individuals who have similar qualifications. One of these contestants will be either elderly or disabled and the other will be young and able bodied. The AI will be made to choose the contestant who it deems most worthy. At the end of the study the results will be compiled and a quantitative review will be performed assessing how and if older and disabled people are discriminated against in the workplace.

**Bibliography:**

Chen, Zhisheng. "Ethics and Discrimination in Artificial Intelligence-Enabled Recruitment Practices." Humanities & Social Sciences Communications, vol. 10, no. 1, Dec. 2023, p. 567. Springer Nature, doi:10.1057/s41599-023-02079-x.

---. "Ethics and Discrimination in Artificial Intelligence-Enabled Recruitment Practices." Humanities & Social Sciences Communications, vol. 10, no. 1, Dec. 2023, p. 567. ProQuest, doi:10.1057/s41599-023-02079-x.

Chu, Charlene H., et al. "Age-Related Bias and Artificial Intelligence: A Scoping Review." Humanities & Social Sciences Communications, vol. 10, no. 1, Dec. 2023, p. 510. Springer Nature, doi:10.1057/s41599-023-01999-y.

Currie, Geoffrey, et al. "Gender Bias in Generative Artificial Intelligence Text-to-Image Depiction of Medical Students." Health Education Journal, Aug. 2024, https://doi.org/10.1177/00178969241274621.

Kotek, Hadas, Rikker Dockum, and David Q. Sun. “Gender Bias and Stereotypes in Large Language Models.” Proceedings of the ACM Collective Intelligence Conference, CI 2023, ACM, 2023, pp. 12-24. DOI: 10.1145/3582269.3615599.

Mehta, Varun. ChatGPT: An AI NLP Model POV. LTIMindtree, 22 May 2024. Technical Report.

O'Gorman, Marcel. "At the Heart of Artificial Intelligence is Racism and Colonialism That We Must Excise: Artificial Intelligence Is Only as Good as the Data That Feed It, and Its Unbalanced Diet Generally Consists of Data Sets Shaped by an Invisible Eurowestern Identity." The Globe and Mail, 13 July 2023, www.theglobeandmail.com. Accessed 9 Nov. 2024.

Zheng, Fei, Chenguang Zhao, Muhammad Usman, and Petra Poulova. "From Bias to Brilliance: The Impact of Artificial Intelligence Usage on Recruitment Biases in China." IEEE Transactions on Engineering Management, vol. 71, 2024, pp. 14155–14167. IEEE, doi:10.1109/TEM.2024.3442618.

Zorrilla-Muñoz, Vanessa, Daniela Luz Moyano, Carolina Marcos Carvajal, and María Silveria Agulló-Tomás. "Towards Equitable Representations of Ageing: Evaluation of Gender, Territories, Aids and Artificial Intelligence." Land, vol. 13, no. 8, 2024, article 1304, https://doi.org/10.3390/land13081304.