



## AI Bias in Data Training

Sarjas Gauhar Singh

[mypublishedpaper@gmail.com](mailto:mypublishedpaper@gmail.com)

Heritage Xperiential Learning School, Haryana

### ABSTRACT

*This research paper talks about AI bias in data training and how it creates age, gender and cultural discrimination. This paper also talks about how spreading awareness about AI bias can help mitigate the issue. It examines how biased training data distorts decision-making in various fields like hiring, healthcare and law enforcement. This paper shows us the need for transparency, accountability and awareness in AI systems and how mitigating data bias is essential for creating an AI system that is fair, responsible, and that can be held accountable in case of any biased decisions and output.*

**Keywords:** *AI Bias, Biased Training Data, Algorithmic Discrimination, Awareness and Bias Mitigation, Transparency and Accountability.*

### INTRODUCTION

Artificial Intelligence (AI) algorithms are widely used by businesses and organisations to make decisions that impact individuals and the society as a whole. These decisions can impact and influence everyone, everywhere, and anytime. (Ntoutsis et al. 2) But these decisions made by AI systems can lead to rejection of jobs or biases in many systems. There are many reasons for the biases in AI systems, which include biases in data training, algorithm bias and societal bias.

AI Bias is a problem that is faced when Artificial Intelligence systems make decisions. These decisions can be very impactful to individuals, societies and organisations. They can influence everyone, everywhere and anytime. But these decisions can lead to a rejection of a job for someone, just because the AI system that is being used is biased. There are many causes for AI Bias, which include biases in data training, problems with the algorithm or biases created by the society that have been picked up by the AI system. For example, in 2018, Amazon created an AI system to review and accept or reject resumes. (Hakimi et al. 3) This AI system was trained by past data of job applicants that had been rejected, which mostly had the data of men. Because of this, the AI system rejected the resume wherever it found the word "women" in it.

We should care about AI bias because it can lead to many problems and conflicts in society. As mentioned before, decisions made by AI can influence everyone, everywhere and anytime. Biases made by AI may lead to the enhancement of existing biases in society or may even create new biases. As we saw in the case study of Amazon, the AI system rejected any woman that has applied for the job because of the AI being trained on past data. This can lead to a person being highly qualified for the job being rejected for the job, and Amazon losing out on a person who is highly qualified for the job. This says that AI bias can lead to an unfair decision for both parties.

AI bias may also lead to a wrong diagnosis of dangerous diseases, which can have a high impact on a person's life. In Uganda, an app was made to diagnose skin diseases because there were fewer doctors compared to the number of patients in the country. (Louis Kamulegeya et al. 755) This app was based on AI to have faster and more easy diagnosis of skin diseases in Uganda. This is a great example of when an app is created for the marginalized and is based on non-biased artificial intelligence. It is important for us to care about this topic because it can be on the matter of life and death, as shown with healthcare.

"One concrete illustration of the dangers of bias in AI can be found in a case study conducted by researchers at MIT and Microsoft. They discovered that facial recognition systems from major technology companies exhibited significant racial and gender biases.8 The algorithms were found to be less accurate when identifying people with darker skin tones and women compared to those with lighter skin tones and men. This reveals a deeply troubling reality: if these biased systems are used for law enforcement purposes, innocent individuals from marginalized communities may be wrongfully targeted based on their appearance." (Mensah 2). "Moreover, Buolamwini and Gebru's exploration of "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification" (Buolamwini and Gebru, 2018) provides empirical evidence of racial and gender disparities in commercial gender classification systems. By systematically evaluating the performance of these systems across diverse demographic groups, the study highlights the urgency of addressing intersectional biases to ensure equitable outcomes." (Hakimi et al. 3) Gender classification systems have biased data, and by seeing the performances of these systems, it highlights the urgency of addressing biases in AI training datasets.

One key finding of the review by Hakimi "is the identification of different types of bias in AI algorithms, including but not limited to gender bias, racial bias, and socioeconomic bias." (Hakimi et al. 5). "The European Commission's "Ethics guidelines for trustworthy AI" (EU-HLEG-AI, 2019) outlines fundamental principles for the development and deployment of AI technologies. Emphasizing transparency, accountability, and societal benefit, these guidelines provide a roadmap for practitioners and policymakers to navigate the ethical dimensions of AI." (Hakimi et al. 3-4) Transparency, accountability and social benefit need to be emphasized.

Transparent datasets will allow anyone to check and confirm if the datasets have any corruption or biases included. The companies that have made the AI algorithm should be held accountable if any biases in datasets are found and they should find a solution to that problem or even be punished.

“The review underscores the importance of education and awareness initiatives concerning bias in AI. Educators, industry professionals, and policymakers can collaborate to develop programs that raise awareness about the ethical implications of AI technologies. This includes educating developers on best practices for creating unbiased algorithms and fostering a culture of responsible AI use.” (Hakimi et al. 8) Making the society, especially the younger part of it, aware about biases in AI and its datasets can help finding solutions to not having bias datasets. Educators are important for teaching the next generation of AI developers about AI bias awareness and how they can be mitigated. This can be done by including ethical AI principles and biases in future curriculums. If the training datasets are non-biased, the AI can be more reliable and trustworthy, it can have more non-biased and better responses, and it won't affect people who are getting denied jobs because the training datasets have biases against them.

“In another case, Google's Ads tool for targeted advertising was found to serve significantly fewer ads for high paid jobs to women than to men (Datta, Tschantz, & Datta, 2015) (gender-bias).” (Ntoutsis et al. 2) To address this issue, we need to feed unbiased and more recent data. This is because the data taken for this could have been taken from older job applications which mostly took men as compared to nowadays where men and women get equal job opportunities and get paid equally to men. Companies also should audit and check their datasets every few months to check if any corrupt or biased data is being shown or this can be done by an outside organization to ensure transparency.

“Bias in language has attracted a lot of recent interest with many studies exposing a large number of offensive associations related to gender and race on publicly available word embeddings (Bolukbasi, Chang, Zou, Saligrama, & Kalai, 2016) as well as how these associations have evolved over time (Kutuzov, Øvrelid, Szymanski, & Velldal, 2018).” (Ntoutsis et al. 5) The datasets of the AI have offensive language related to gender and race. This has led to the AI being offensive towards a particular race or gender. The problem arising with this is that the AI is promoting offense towards a particular race and gender which can impact and influence the real world.

“For instance, the COMPAS system for predicting the risk of re-offending was found to predict higher risk values for black defendants (and lower for white ones) than their actual risk (Angwin, Larson, Mattu, & Kirchner, 2016) (racial-bias).” (Ntoutsis et al. 2) This can lead to innocent black defendants having a false risk value for re-offending and could lead to criminal justice sanction while guilty white defendants get a lower risk of re-offending. Due to this, black defendants will have a higher chance of getting jailed or being given extra jail time. This will also lead the people to lose trust in the government as because of the AI driven systems the decisions are being made biased.

“As most commercial AI systems depend on data collected from a multitude of public and private sources (such as Twitter, open-source datasets), societal inequities arising from prejudiced beliefs, actions, and laws may be reflected in these systems.” (Chu et al. 2) Using open source datasets like X or private blogs for using data can lead to biased datasets as personal opinions, biases and views can make the data more biased and corrupt. What people write on social media or personal blogs is from their own perspectives, from their own context, educational background and how they are influenced by opinions and discussions around them. The responses of AI can be the result of personal opinions on topics, which others might not agree with, resulting in the accountability of AI dropping. “The dangers of bias in AI systems cannot be overstated. As artificial intelligence becomes increasingly integrated into society, it is crucial to recognize the potential harm that biased AI can inflict on individuals and communities. Bias in AI refers to the unfair or unjust treatment of certain groups based on race, gender, or other protected characteristics. This bias can manifest itself in various ways, such as discriminatory hiring practices, biased loan approvals, or even deadly decisions made by autonomous vehicles.” (Mensah 2)

### **CULTURAL DISCRIMINATION**

Cultural discrimination is when the AI system is trained with datasets that make the decision made by the AI favourable or biased against certain cultures or religions

“It seems troublesome that a society's implicit biases (in perhaps too many realms of daily life) may be exacerbated through the use of AI-based recommendations... A reasonable question to ask is whether some individuals would be more likely to question an AI-based recommendation if they happened to perceive it as biased (specifically in terms of race or gender).” (Gupta et al. 1466) This says that the society's biases can be made worse or can also arise due to biases in AI. These AI biases can influence people's opinions and push existing biases further. The question being asked is will an individual question biased responses by the AI, especially in terms of racial and gender bias. “We aim at elucidating the effects of individuals' cultural values on AI questionability due to perceived bias, which we understand and operationalize here as the extent to which individuals are likely to question racially or gender biased AI-based recommendations.” (Gupta et al. 1466). “AI-based recommendations may discriminate against some members of society more than others... One such concern that we examine in this study is the extent to which individuals, owing to their individual-level cultural values, would be likely to question AI-based recommendations when perceived as racially or gender biased.” (Gupta et al. 1467)

“Understood this way, we agree with Fonseka (2017) that academics are in arrears of holding AI accountable. In particular, as Ågerfalk (2020, p. 5) suggests, “there are good reasons to worry about misuses of AI,” given its potential to perpetuate society's inequalities and injustices through implicit biases due to race, gender, and sexual orientation (Manyika et al., 2019).” (Gupta et al. 1466). “When individuals with high uncertainty avoidance cultures come across a biased AI-based recommendation, they will likely question it. This is because of the inherent unforeseen risks associated with believing in the AI-based recommendation that seems discriminatory.” (Gupta et al. 1470) People that have high uncertainty avoidance cultures are people that feel threatened by uncertainty and the unknown. They will question AI biased recommendations and they will lose the trust in the AI, because it was biased.

### **AGE DISCRIMINATION**

Age Discrimination talks about how AI systems can be trained to view a certain age group higher than the other and how it can detect people to be their incorrect age by stereotypical facial features.

“The notion of digital ageism is used to refer to the extension of ageism into the realm of the design, development, deployment, and evaluation of technology, and how AI and related digital and socio-technical structures may produce, sustain or amplify systemic processes of ageism (Billette et al., 2012; Chu et al., 2022a, 2022b, 2022c, 2022d; Nyrup et al., 2023).” AI systems can sustain or amplify ageism by embedding biases from the data they are trained with. For example, facial recognition systems have problems with recognising older faces compared to younger ones. Even digital social media systems are designed keeping in mind the youth, which makes it less accessible to the older society. These patterns create biases which the AI systems catch onto and amplify digital ageism.

“Factors that can be considered to contribute to digital ageism involve excluding older adults from the development or design processes (Ashley, 2017; Neary and Chen, 2017; Sourdin and Cornes, 2018), replicating uneven power dynamics between older and younger people, and result in algorithms and/or products that are not optimised for them.” (Chu et al. 2) Such processes may further negatively impact older adults’ desire to use digital technologies or services, thereby generating less training data to better understand the needs of this demographic and further perpetuate digital ageism.

“Previous research has identified several well-known biases and inaccuracies in our ability to estimate age from facial appearance. For example, there is a significant decrease in accuracy when estimating age from the faces of middle-aged adults (40–60) compared to young adults (20–40) and of older adults (ages 60–80) compared to middle-aged adults. This decrease in accuracy could be accounted for by the fact that genetic and environmental factors have a larger impact on the appearance of the face as people grow older, and that there is considerable variance in these effects on the apparent age of the face.” (Ganel et al. 1) As humans are less accurate with older faces, the AI will also have higher error rates because the training data is labeled by humans. This could make AI wrongly classify older people’s ages, affecting how they are treated in applications. “Recently, there has been a growing interest in automated age estimation using artificial intelligence (AI) technology<sup>16</sup>. The current platforms use machine-learning algorithms based on training with a large set of photos to achieve the most accurate performance in age estimations. The current interest in age estimation by AI is part of an overall attempt to extract various visual features automatically from faces, features that include identity, expression, and gender as well as other information that can be gleaned from the face.” (Ganel et al. 2) Datasets the AI is trained on may have underrepresented certain groups, like AI trained mainly on younger faces can overestimate ages of younger people.

“The specific interest in age estimation is also boosted by recent commercial incorporation of automatic age estimation technology for different uses, including age verification in retail outlets that is now being implemented in different countries.” (Ganel et al. 2)

## **GENDER DISCRIMINATION**

Gender discrimination in AI bias takes about how an AI system can favour one gender over another based on the training dataset it's trained on, having a historical influence as well.

“Studies have shown that AI-generated content often aligns male pronouns with professions such as “engineer,” “scientist,” or “CEO,” while associating female pronouns with roles like ‘nurse,’ ‘teacher,’ or ‘homemaker’ (Blodgett et al., 2020. This biased representation not only reflects historical disparities but also influences societal perceptions, potentially discouraging gender diversity in various fields.” (Khan 32). “Some sentiment analysis models have been found to rate statements associated with female names more negatively than those associated with male names. This issue extends to AI-driven hiring tools, where algorithms trained on biased data have demonstrated a tendency to favor male candidates over female candidates in recruitment processes (Mehrabi et al., 2021). In extreme cases, biased AI systems have contributed to discriminatory decisions in critical sectors such as banking, healthcare, and law enforcement.” (Khan 32) AI systems can unknowingly record and amplify gender bias from biased data, and when it consistently shows this gender bias, it gets normalized for society- people lose faith, especially amongst marginalized groups.

“Many voice assistants, such as Apple’s Siri and Amazon’s Alexa, have historically been designed with female-sounding voices and programmed to respond in submissive or apologetic manners. This design choice reinforces gender stereotypes related to service and obedience, raising ethical concerns about the role of AI in perpetuating gendered societal norms (Crawford, 2021).” (Khan 32) The choice of keeping the voice assistant’s being designed with female voices on purpose, reflects the social and cultural bias of women being more submissive and apologetic.

“One of the foundational studies in this area was conducted by Bolukbasi et al. (2016), who demonstrated that word embeddings in NLP models, such as Word2Vec, encode gender biases. Their work illustrated how AI systems learn associations like “man is to computer programmer as woman is to homemaker,” highlighting the deep-seated biases present in training data. This discovery led to further investigations into the sources of such biases, with scholars pointing out that large-scale datasets used for AI training predominantly reflect historical and cultural gender norms.” (Khan 33) The statement “Man is to computer programmer as woman is to homemaker” is incorrect and it says that the data suggests that only men are computer programmers and only women can be homemakers. Bolukbasi states that debiasing of training datasets is crucial. To remove the biases of the training datasets we should improve and enhance the quality of data and ensure its not favoring or opposing a particular gender or community.

“A key aspect of gender bias in AI is its manifestation in occupational stereotypes. Mehrabi et al. (2021) highlighted how language models reinforce gendered job associations, where words like “leader,” “doctor,” and “engineer” are more commonly linked with men, while “nurse,” “teacher,” and “assistant” are associated with women. Such biases have implications beyond linguistic representation, affecting automated hiring systems, recommendation algorithms, and AI-generated content. For instance, Amazon’s AI-based hiring tool, which was trained on historical hiring data, was found to systematically disadvantage female candidates by downgrading resumes that contained words such as “women’s” or were associated with female dominated fields (Crawford, 2021).” (Khan 33-34). “Biased AI can influence hiring practices, academic admissions, and loan approvals, disproportionately affecting women and non-binary individuals. Moreover, biased language models can shape public discourse by subtly influencing how information is presented, potentially skewing narratives in ways that favor dominant societal groups.” (Khan 32) Biased AI models can influence public opinions as users read and use the information given by AI on a day to day basis. These public opinions can heighten already existing biases in society and even create newer stereotypes against a certain community.

“Efforts to mitigate gender bias in AI have led to the development of several bias detection and debiasing techniques. Researchers have proposed methods such as adversarial training, fairness-aware algorithms, and balanced dataset curation to reduce biases in NLP models. For example, bias mitigation frameworks like IBM’s AI Fairness 360 and Google’s Perspective API aim to identify and correct biased outputs in AI systems (Mehrabi et al., 2021).” (Khan 33). “Another promising approach involves interdisciplinary collaboration between computer scientists, linguists, ethicists, and policymakers. By incorporating diverse perspectives in AI development, researchers can design models that are more inclusive and representative of different gender identities. Additionally, increased transparency in AI design, such as open source bias auditing tools, can enable greater accountability and fairness in AI-driven decisionmaking.” (Khan 33)

### **AWARENESS ABOUT AI BIAS**

Transparency and accountability are important in using and developing AI responsibly. Being transparent allows people to understand how an AI system works and how it reaches its decisions. Without this openness, users cannot identify potential errors or biases in the system. Accountability ensures that developers and organizations are held responsible when their AI technologies cause harm. (Mensah 2) When AI systems are open, users can understand how decisions are made, check for mistakes, and identify unfair or biased output which helps prevent discrimination and misuse. Accountability ensures when AI makes harmful or unethical decisions, someone is held responsible and is able to fix issues or handle consequences.

“In 2020, California passed legislation requiring companies to provide detailed documentation about their facial recognition technology’s performance across different demographic groups before law enforcement agencies could use it.<sup>11</sup> This transparency measure aims to hold companies accountable for any biases present in their facial recognition software.” (Mensah 2)

“Transparency and accountability are vital when it comes to the ethical considerations of AI systems, especially in terms of bias mitigation. As AI becomes increasingly integrated into our society, it is crucial that we ensure its fair and responsible implementation.<sup>14</sup> The potential for bias in AI systems has raised concerns as they have been shown to reflect the biases present in the data they are trained on. One concrete illustration of this issue is the case of Amazon’s recruitment tool, which was developed using machine learning algorithms to review resumes and identify top candidates for job positions.<sup>15</sup> However, it was later discovered that the system had a gender bias as it consistently downgraded resumes from female applicants. This example highlights the need for transparency in AI systems so that biases can be identified and addressed.” (Mensah 4) Transparency in AI training data is important because it helps people to understand how AI is making the decisions and it helps build trust and accountability. If non-transparent can lead to wrongful arrests or discrimination, but when the data is open, experts can look for biases in the data and improve to mitigate the biases.

“In addition to concrete illustrations, relevant authorities have recognized the significance of transparency and accountability in AI systems. The European Union’s General Data Protection Regulation (GDPR)<sup>17</sup> has established guidelines on transparency by requiring organizations to provide individuals with clear information about how their data will be processed using automated decision-making systems like AI algorithms. This regulation aims to ensure that individuals understand how decisions affecting them are made, reducing opacity surrounding AI processes.” (Mensah 4) Important global organizations and laws emphasize transparency and accountability in AI because it often makes decisions directly affecting people’s lives. Laws aim to build public trust, protect individuals’ rights and privacy, prevent discrimination or unethical use of data, make AI systems explainable so users and regulators can hold organizations responsible for the outcomes.

“Furthermore, case laws have also emphasized transparency and accountability within AI systems. In a landmark ruling by New York City’s Commission on Human Rights<sup>18</sup> against an employment agency utilizing an algorithmic hiring tool, it was determined that if an employer uses an algorithmic tool during hiring processes, they must disclose this information to applicants upon request along with an explanation of how the tool works.” (Mensah 4) Laws have emphasized transparency and accountability in AI systems making decisions. For example, New York City’s Commission on Human Rights rules that if an employer uses an algorithmic tool during the hiring processes.

“The ethical considerations of AI systems, specifically bias mitigation, transparency, and accountability, are of utmost importance to ensure their fair and responsible implementation in society.” (Mensah 5). “Various techniques are being employed to reduce bias in AI systems.<sup>22</sup> These include using diverse training data that represents a wide range of demographics, making algorithmic adjustments to reduce discriminatory outcomes, and implementing transparency measures to identify and rectify biases effectively.<sup>23</sup>” (Mensah 5-6). “One approach involves diversifying the training data used during the system’s development phase. By including a wide range of demographic groups and perspectives within the dataset, it becomes possible to ensure that the resulting AI system does not favor any particular group unfairly.” (Mensah 6)

“For example, a research project led by Google used this technique to improve the accuracy of their speech recognition system for underrepresented dialects. <sup>35</sup> By including more diverse voices during model training, they were able to minimize bias against certain accents or dialects.” (Mensah 7). “By providing clear explanations of how these algorithms function, developers can ensure that users understand and have confidence in the decisions made by AI systems. For instance, in the healthcare industry, where AI is increasingly employed to assist in diagnosing diseases and recommending treatments, transparency becomes paramount.” (Mensah 10)

### **CONCLUSION**

Artificial intelligence had the power to shape the future society, helping in decisions for healthcare, employment, education etc. These decisions are based on the datasets the AI is trained on. However, as shown in this case study, AI is not neutral. If the datasets that the AI is trained on are corrupt, which favors or goes against a certain individual or a group of individuals, it can affect the targeted group of people. Bias in AI systems can cause discrimination based on culture, gender, age or societal, which originate from the data the AI is given. These biases can reflect on or create new inequalities and prejudices present in society that create harm for certain individuals and groups as a whole.

Due to these biases, addressing AI bias is not optional, but essential. AI bias threatens fairness, equality and trust in technology. Because of AI bias, societies can become more divided and individuals could be denied opportunities they deserve. The examples of Amazon’s hiring algorithm, biased facial recognition systems, age-misclassifying models, and gender-skewed language tools tell us that biased decisions by AI, driven by biased data, lead to biased decisions that can affect lives in many ways.

To build an ethical and responsible AI system, we need to prioritize transparency, accountability and awareness. Developers must ensure that non-biased, diverse and representative datasets are used in building these AI driven tools. Companies should be held accountable for making biased AI systems affecting others and they should offer transparency to the datasets, keeping a check on them.

By increasing awareness about AI bias and educating future developers, policy makers and everyday users, we can make a non-biased AI system that serves fairly for everyone. By acknowledging the dangers of AI bias and eliminating them, we can build more inclusive and reliable. AI has huge potential to benefit humanity, only if we make it fair, equal and trustable.

## REFERENCES

- [1] Chu, Charlene H., et al. "Age-related bias and artificial intelligence: a scoping review." *Humanities & Social Sciences Communication*, vol. 10, no. 510, 2023, p. 17. *Google Scholar*, <https://www.nature.com/articles/s41599-023-01999-y.pdf>. Accessed 29 September 2025.
- [2] Ganel, Tzvi, et al. "Biases in human perception of facial age are present and more exaggerated in current AI technology." *Scientific Reports*, vol. 12, no. 22519, 2022, p. 10. *Google Scholar*, <https://www.nature.com/articles/s41598-022-27009-w.pdf>. Accessed 6 October 2025.
- [3] Gupta, Manjul, et al. "Questioning Racial and Gender Bias in AI-based Recommendations: Do Espoused National Cultural Values Matter?" 2021, p. 17. <https://link.springer.com/content/pdf/10.1007/s10796-021-10156-2.pdf>. Accessed 18 September 2025.
- [4] Hakimi, Musawer, et al. "A Comprehensive Review of Bias in AI Algorithms." *Nusantara Hasana Journal*, vol. 3, no. 8, 2024, p. 11. *Google Scholar*, file:///Users/sarjassingh/Downloads/1.-abdul-wajid-fazil-1-11%20(1).pdf. Accessed 25 July 2025.
- [5] Khan, Varda. "Artificial Intelligence and Gender Bias: Analyzing Algorithmic Discrimination in Language Models." *Journal of Gender, Power and Social Transformation*, vol. 1, no. 2, 2024, p. 10. *Google Scholar*, <https://www.researchcorridor.org/index.php/jgpst/article/view/329>. Accessed 31 October 2025.
- [6] Louis Kamulegeya, et al. "Using artificial intelligence on dermatology conditions in Uganda: a case for diversity in training data sets for machine learning." 2023, p. 11. Accessed 26 May 2025.
- [7] Mensah, George Benneh. "Artificial Intelligence and Ethics: A Comprehensive Reviews of Bias Mitigation, Transparency, and Accountability in AI Systems." *Africa Journal For Regulatory Affairs (AJFRA)*, vol. 10, no. November, 2023, p. 26. *Google Scholar*. Accessed 7 November 2025.
- [8] Ntoutsis, Dr. Erini, et al. "Bias in data-driven artificial intelligence systems—An introductory survey." 2019, p. 14. Accessed 25 June 2025.