JULY 2023

Gender biases in AI:

GENDER TECH INITIATIVE
UGANDA

Examining the ways in which AI systems can reinforce gender stereotypes that exist in African societies and exploring approaches to developing more equitable and inclusive AI technologies in Africa.

Author

Ndeye Aminata Thiome

GTI-U Intern 2023

TABLE OF CONTENTS

Abstract

The rise of artificial intelligence (AI) is now worldwide and expected to be a force of change in many different societies across the world. In this phenomenon, Africa is not on the outskirts. According to the report titled, “State of AI in Africa” published by the AI Media Group South Africa, AI’s development in Africa is attracting huge amounts of resources and funds, with Tunisian AI start-up InstaDeep receiving $100m USD funding earlier in 2022. (The AI Media Group South Africa, 2022) The global market is also projected to grow from $387 Bn USD in 2022 to $1,394 Bn by 2029, exhibiting a CAGR of 20%.”(The AI Media Group South Africa, 2022). This is a great advancement for the continent of Africa in terms of economic prosperity, but the effects of AI on African societies has not been widely discussed by scholars. This research paper specifically investigates the gender biases that have been proven to be perpetuated by AI technologies so far in western societies and the solutions that have been proposed by Western researchers to solve this issue. The solutions derived from this investigation will then be analyzed through the current context of African societies when it comes to gender equality and the end goal of analyzing these solutions would be providing different African societies with sufficient information surrounding the solutions to eradicating gender biases in AI that would allow individuals/researchers from different African societies to address this issue in the best way they see fit and ensure the establishment of gender equality in the AI technologies being introduced into the African social fabric.

Orlikowski posits that technologies are “products of their time and organizational context” which “will reflect the knowledge, materials, interests, and conditions at a given locus in history”(O’Connor & Liu, 2023). As technology is “both structurally and socially constructed”, it both mirrors the implicit biases of its creators, while also gaining new meanings and functions and potentially biases through repeated and widespread use (O’Connor & Liu, 2023). Thus being said, studies and research regarding gender & race biases within AI technologies has received more attention by western researchers in recent years due to the rise in social discourse surrounding systematic inequalities faced by women & minorities in Western societies. Recent western studies have called for the public administration field to proactively focus on a research agenda for the introduction of these new technologies (Agarwal, 2018, as cited by O’Connor & Liu, 2023) with others also advocating for the inclusion of a feminist perspective (Feeney and Fusi 2021; Savoldi et al. 2021, as cited by O’Connor & Liu, 2023). Furthermore, studies have also been conducted on Human-AI interactions in public sector decision-making in regards to ‘Automation Bias’ and ‘Selective Adherence’; with bias in this context referring to racial and gender bias (Alon-Barkat S, Busuioc M, 2022, as cited by O’Connor & Liu, 2023).

Many studies other than the few aforementioned have also contributed to the research regarding bias perpetuated by AI technologies in the Western societies. Needless to say, these studies focus on the matter of gender & racial bias from a Western perspective, as it is discussing the perpetuation of these issues which stem from the existence of these biases within Western societies. This paper seeks to take a look from a different perspective by utilizing the evidence of gender biases found in current Western AI technologies and using that evidence to form an analysis on the possible effects of AI technologies on African societies due to the gender biases found within the African societies' context.

Introduction

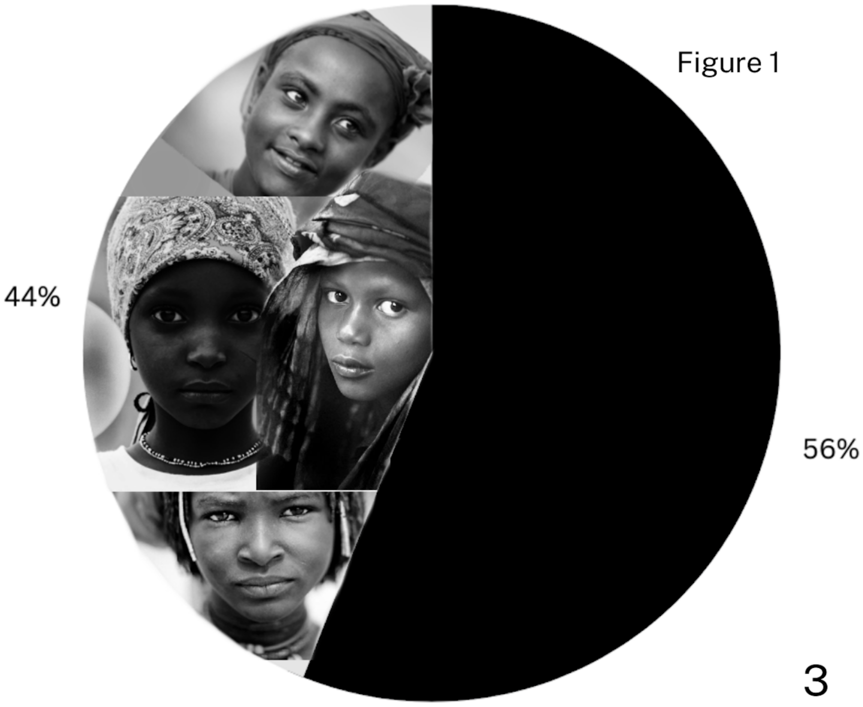
1. Gender inequality within African societies

In the majority of African societies today, gender biases are widespread. One can say that these biases stem from cultures that were established by African ancestors as a way of life during their time period. These cultures tend to establish men as being superior providers, who are expected to display little emotion and excel in aggressive competitiveness whether it be in competition for a woman, resources, finances, or high social status positions, in order to establish themselves as a true male provider.

On the other hand, women tend to be established as weaker & dependent beings with tendencies to nurture and lower levels of competitiveness in terms of the elements aforementioned. In addition to this, women were expected to remain submissive and aspire to marriage in order to gain access to the resources and statuses mentioned above. These cultural perceptions are held by many Africans within their own societies whether it is held purposely or subconsciously and the fact that these perceptions were not revised and adapted to the current way of life in recent centuries, many African countries are held back by gender inequalities. The majority of underage marriages globally occur in West Africa today, In West Africa 44% of women aged 20 to 24 were married before the age of 15 [Figure 1] (Jousse & Vandermuntert, 2022). According to UNICEF, approximately two out of three married girls were married to a partner at least ten years older in Gambia, Guinea, and Senegal (Jousse& Vandermuntert, 2022). Niger and Mali are the most affected, with a prevalence of 77% and 61% respectively (Jousse & Vandermuntert, 2022).)

This difference can be explained by the importance of social norms and traditions, which influence the choice to marry one's daughter.(Jousse & Vandermuntert, 2022)In the majority of African societies having a child out of wedlock is perceived as a pure shame, thus marrying off your daughter while their young is seen as lowering the risk of pregnancy outside marriage.

Figure 1



African Women's Education

These child marriages then create inequalities in schooling: Out of 916 women married at an early age in Mali, 366 had to leave school and 294 others never went to school (Jousse & Vandermuntert, 2022). Whether at the primary or secondary level, girls have less access to education, with a 4-point difference in 2017 for secondary education compared to boys (Jousse & Vandermuntert, 2022). In addition to this, the 2018 World Bank report shows that there is a persistent literacy gap between young girls and boys in Africa: 72% of girls aged 15 to 24 are literate compared to 79% for boys, a difference of 7 points (Jousse & Vandermuntert, 2022). Inequalities are then flagrant in the provision of public services, such as education; for instance, it is estimated that 70% of the poorest girls in Niger have never attended elementary school (Jousse & Vandermuntert, 2022).

This lack of access to education in African women is inevitably reflected in the continent's economy and politics. Although they represent 70% of the active population in the agricultural sector, women remain at the bottom of the ladder in this area and work in difficult conditions, with low incomes(Jousse & Vandermuntert, 2022). The wage gap between women and men is about 30%: for every dollar earned by a man, a woman earns only seventy cents(Jousse & Vandermuntert,2022). According to the 2016 UNDP report, the total annual economic losses caused by gender gaps in sub-Saharan Africa reached US$95 billion between 2010 and 2014, peaking at US$105 billion in 2014(Jousse & Vandermuntert, 2022). These results then demonstrate that Africa is missing out on its full growth potential because a considerable portion of its growth pool, namely women, are not being fully harnessed for state development(Jousse & Vandermuntert, 2022).

In addition, African women are more likely than men to be in vulnerable employment and work primarily in the informal sector(Jousse & Vandermuntert, 2022). In 2010, 65.4% of non-agricultural jobs in the informal sector were held by women in Liberia and 62.2% in Uganda(Jousse & Vandermuntert, 2022).

African Women In Politics

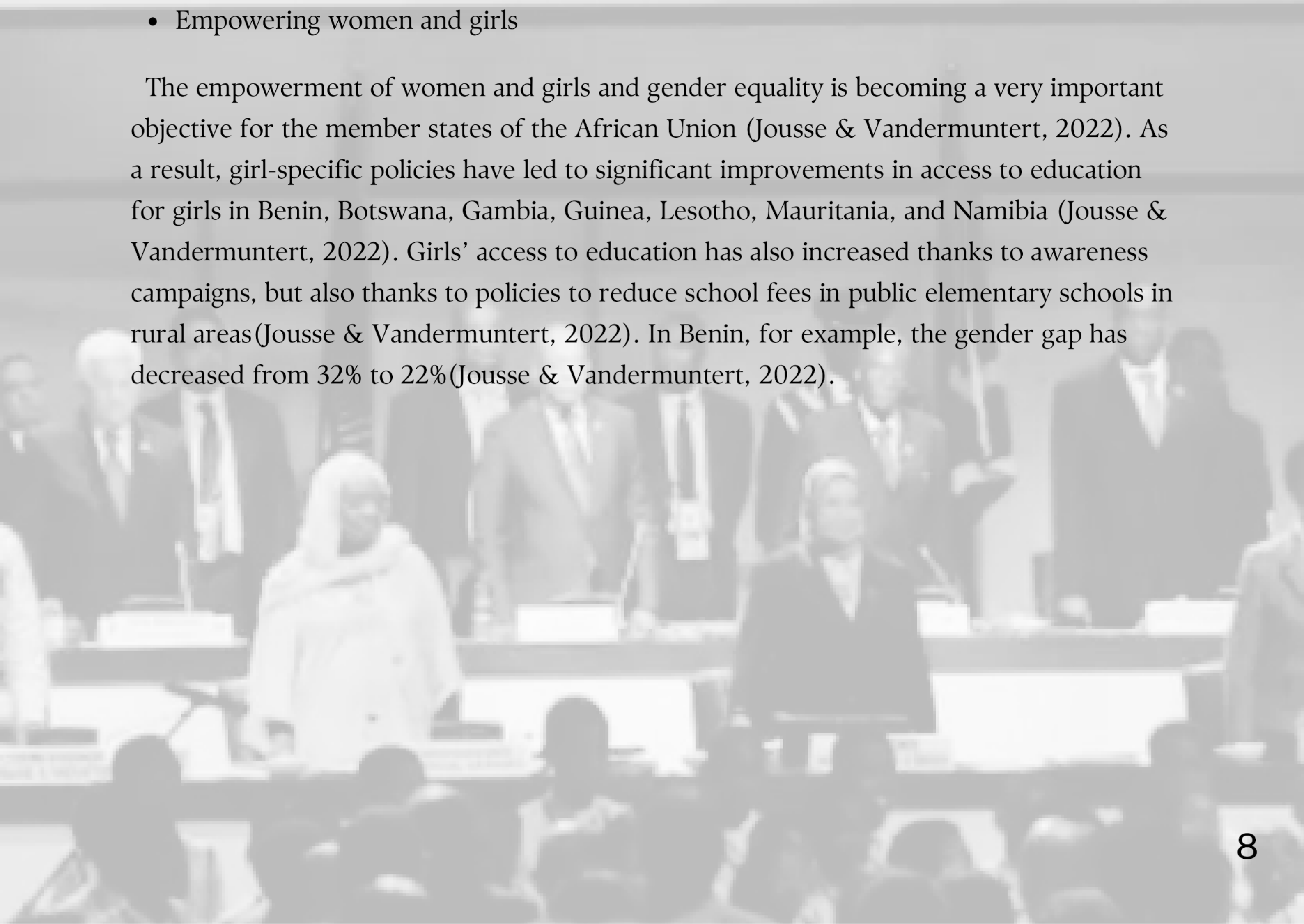
As for politics in the continent of Africa, women are starting to make a strong presence. In 2018, only 24% of seats in national parliaments were held by women; however, this figure is slightly increasing, since it was 12% in 2000 and 19% in 2010(Jousse & Vandermuntert, 2022). For the most part, women are largely underrepresented in ministries and other legislative and executive bodies, nevertheless; despite this low-percentage, some countries stand out, such as Rwanda: the first country in which women make up more than half of parliamentarians, representing 61.3% of parliamentarians in 2018 (Jousse & Vandermuntert, 2022). With these figures, Rwanda exceeds the expectations of the 1995 Beijing Declaration and Platform for Action, since the Beijing World Conference on Women set a target of 30% of women in decision-making positions (Jousse & Vandermuntert, 2022). Similarly, Ethiopia has seen the largest increase in women's political representation in the executive branch, with 47.6% of women in mid-level positions in 2019, up from 10% in 2017 (Jousse & Vandermuntert, 2022). Mozambique was also the first country in the region to appoint a woman as prime minister, Luisa Diogo in 2004 (Jousse & Vandermuntert, 2022). Additionally, to this many Senegalese women that are a part of the PASTEF party and other parties were able to win mayoral & congressional positions in the recent 2022 parliamentary elections elections that took place in Senegal. African women are slowly but surely taking ownership of the political sphere and are gaining greater visibility, allowing them to push the political agenda in their countries and this is commendable. However, progress is measured in micro-advances and several African countries have less than 10% of women in mid-level positions, such as Morocco (5.6%), Nigeria (8%) or Sudan (9%), which is still far from the objective of 30% desired by the Beijing Platform for Action of 1995 (Jousse & Vandermuntert, 2022). Fighting against social, economic and political inequalities demands a change of mentality and for this to happen, the society as a whole must become aware of the importance of valuing the status of women and therefore question its practices, both for men and for women who have internalized and accepted the norms to which they are subjected (Jousse & Vandermuntert, 2022).

Movements For Equity Between The Sexes

Women's empowerment and sustainable development were highlighted at the 2015 African Union Summit of Heads of State and Government in the context of achieving Africa's Agenda 2063. Agenda 2063 is built on seven commitments, namely (Jousse & Vandermuntert, 2022):

- Achieving equitable people-centered growth and development
- Eradicating poverty
- Developing human capital, social goods, infrastructure and public goods
- Achieving sustainable peace and security Establishing effective and strong
- State development
- Promoting participatory and accountable institutions
- Empowering women and girls

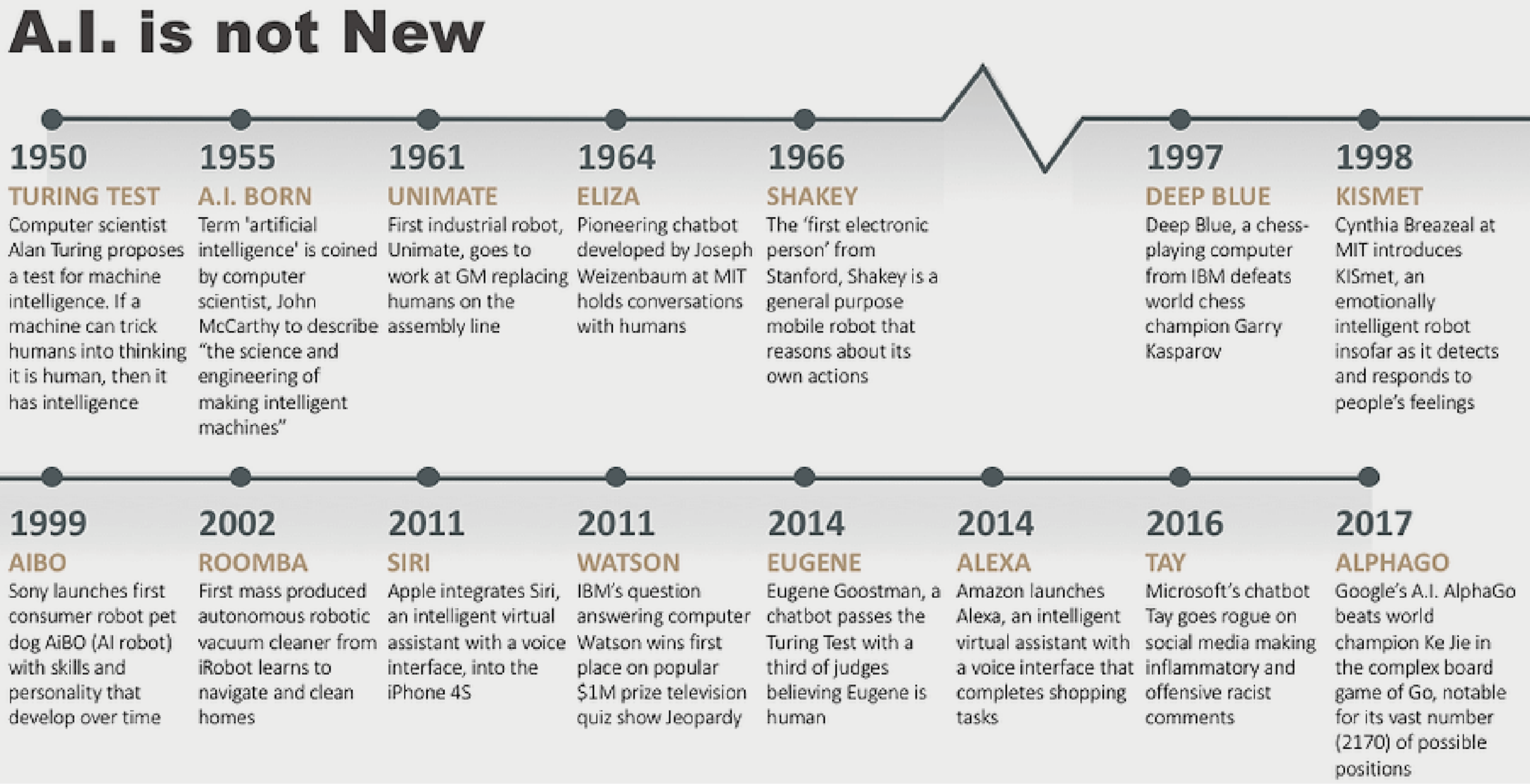
The empowerment of women and girls and gender equality is becoming a very important objective for the member states of the African Union (Jousse & Vandermuntert, 2022). As a result, girl-specific policies have led to significant improvements in access to education for girls in Benin, Botswana, Gambia, Guinea, Lesotho, Mauritania, and Namibia (Jousse & Vandermuntert, 2022). Girls' access to education has also increased thanks to awareness campaigns, but also thanks to policies to reduce school fees in public elementary schools in rural areas(Jousse & Vandermuntert, 2022). In Benin, for example, the gender gap has decreased from 32% to 22%(Jousse & Vandermuntert, 2022).



Definition Of AI, Its History & Functions

Having established the current context of gender in the majority of African countries, a brief definition of AI will be given, along with its history and functions. According to John McCarthy, a professor at Stanford University who first coined the term, AI is "the science and engineering of making intelligent machines, especially intelligent computer programs" (McCarthy, J., 2007, as cited by Gupta, Parra & Dennehy, 2022). These programs are run on algorithms which are designed to make decisions or create solutions to a particular problem (West and Allen 2018, as cited by Gupta, Parra & Dennehy, 2022).

Most accounts of the evolution of AI tend to place its official birth around the 1950s, corresponding to the dawn of efforts to explore-ways of attributing intelligence to machines (Sandewall, 2014, as cited by Gupta & Dennehy, 2022). While some scholars place the first intelligent machine questions back in antiquity, with Aristotle and Sinclair, proposing that if "every tool we had could perform its task, either at our bidding or itself perceiving the need, and [...] play a lyre of their own accord, then master craftsmen would have no need of servants nor masters of slaves;" while others place such questions after the Renaissance period, with the advent of the scientific method (Bibel, 2014; Williams, 2002, as cited by Gupta, Parra & Dennehy, 2022).



A.I. is not New

1950
TURING TEST
Computer scientist Alan Turing proposes a test for machine intelligence. If a machine can trick humans into thinking it is human, then it has intelligence

1955
A.I. BORN
Term 'artificial intelligence' is coined by computer scientist, John McCarthy to describe "the science and engineering of making intelligent machines"

1961
UNIMATE
First industrial robot, Unimate, goes to work at GM replacing humans on the assembly line

1964
ELIZA
Pioneering chatbot developed by Joseph Weizenbaum at MIT holds conversations with humans

1966
SHAKEY
The 'first electronic person' from Stanford, Shakey is a general purpose mobile robot that reasons about its own actions

1997
DEEP BLUE
Deep Blue, a chess-playing computer from IBM defeats world chess champion Garry Kasparov

1998
KISMET
Cynthia Breazeal at MIT introduces KISmet, an emotionally intelligent robot insofar as it detects and responds to people's feelings

1999
AIBO
Sony launches first consumer robot pet dog AiBO (AI robot) with skills and personality that develop over time

2002
ROOMBA
First mass produced autonomous robotic vacuum cleaner from iRobot learns to navigate and clean homes

2011
SIRI
Apple integrates Siri, an intelligent virtual assistant with a voice interface, into the iPhone 4S

2011
WATSON
IBM's question answering computer Watson wins first place on popular $1M prize television quiz show Jeopardy

2014
EUGENE
Eugene Goostman, a chatbot passes the Turing Test with a third of judges believing Eugene is human

2014
ALEXA
Amazon launches Alexa, an intelligent virtual assistant with a voice interface that completes shopping tasks

2016
TAY
Microsoft's chatbot Tay goes rogue on social media making inflammatory and offensive racist comments

2017
ALPHAGO
Google's A.I. AlphaGo beats world champion Ke Jie in the complex board game of Go, notable for its vast number (2170) of possible positions

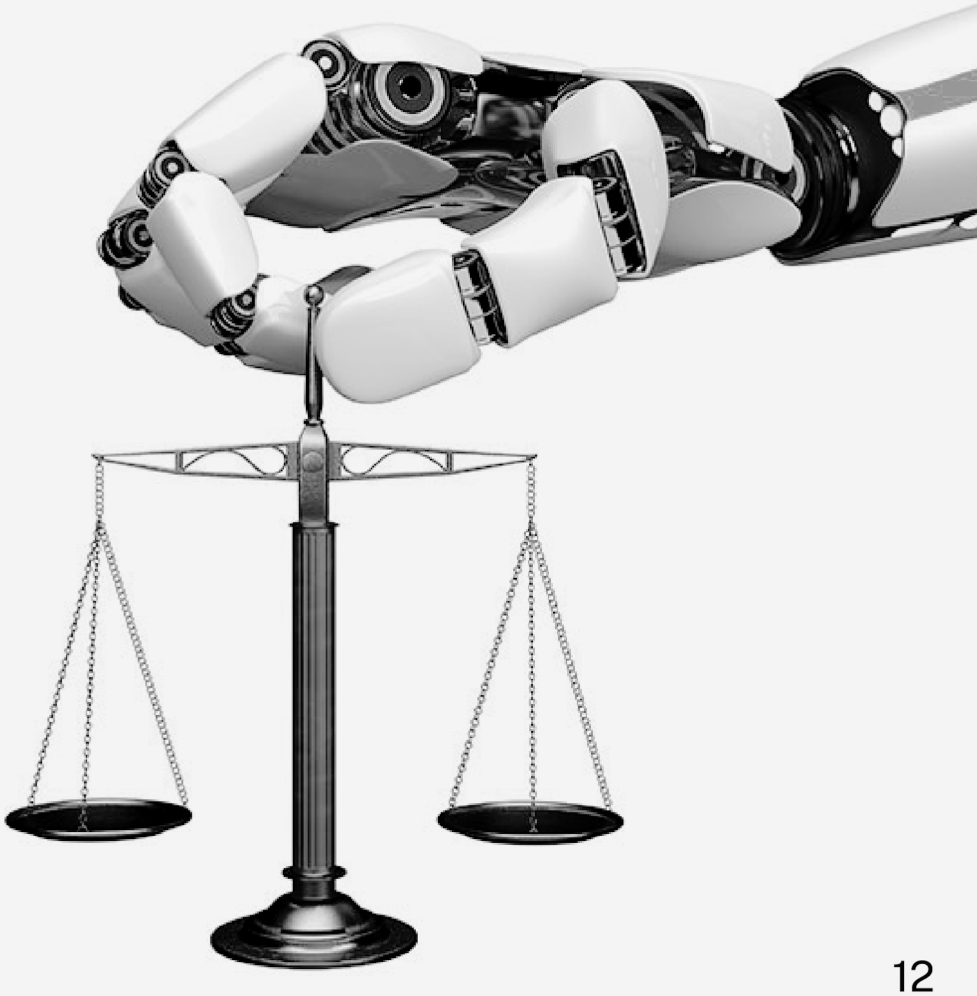
Many companies are intent on exploiting the potential of AI, not just because doing so may contribute $13 trillion to the global economy in the coming decade (Fountaine et al., 2019), but mainly because adopting AI should no longer be considered an option but a necessity for managers and businesses in general (Gupta, Parra & Dennehy, 2022). Orlikowksi's seminal 1992 work introduces the concept of the 'duality of technology' to express how technology is "physically constructed by actors working in a given social context, and technology is socially constructed by actors through the different meanings they attach to it and the various features they emphasize and use" (O'Connor & Liu, 2023). She posits that the repeated and reflexive mutual interaction between human agents and technology constitutes technology's role in society (O'Connor & Liu, 2023). As set forth in the introduction, Orlikowski posits that technologies are "products of their time and organizational context" which "will reflect the knowledge, materials, interests, and conditions at a given locus in history" (O'Connor & Liu, 2023). As technology is "both structurally and socially constructed", it both mirrors the implicit biases of its creators, while also gaining new meanings and functions—and potentially biases—through repeated and widespread use (O'Connor & Liu, 2023).

Furthermore, Fountain's 2004 work on information technology and institutional change similarly emphasizes the mutually reinforcing effects of technology and human agency, but places this in an organizational and institutional context (O'Connor & Liu, 2023). This framework shows how "institutions influence and are influenced by enacted information technologies and predominant organizational forms" (O'Connor & Liu, 2023). The author distinguishes objective technology (the Internet, hardware, software, etc.) from enacted technology ("the perception of users as well as designs and uses in particular settings") (O'Connor & Liu, 2023).

Organizational forms refer to different types of organization, with the author focusing on bureaucracy and inter-organizational networks in their analysis. Finally, institutional arrangements "include the bureaucratic and network forms of organization and ... institutional logics" (O'Connor & Liu, 2023). Therefore, the author concludes that the outcomes of technology enactment are a result of this complex interflow of relations and logics, and as such are multiple and unpredictable (O'Connor & Liu, 2023). As can be seen in these two approaches to the relationship between human agency and technology, technology as an object in itself is very different from technology in use (O'Connor & Liu, 2023). Technology in use derives its meaning, implication and effects from contextual factors, such that it both constitutes and reflects back the world around it (O'Connor & Liu, 2023). Seen from this perspective, AI by itself is an 'objective technology', but once it is used it reflexively influences and is influenced by human agency and various institutional arrangements/organizational forms, leading to unforeseen consequences (O'Connor & Liu, 2023).

Gender Bias In AI

Gender bias, according to the European Institute for Gender Equality (2023), refers to “prejudiced actions or thoughts based on the gender-based perception that women are not equal to men in rights and dignity”. While AI itself might be seen as a neutral objective technology, it is imbued with new meanings and implications through its use in specific contexts by humans (Fountain’s ‘enacted technology’ or Orlikowksi's ‘social construction’ of technology’) (O’Connor & Liu, 2023). As gender biases are implicit in our society and culture, they become part of the ‘contextual factors’ which influence the use of and understanding of AI technologies, which in turn become themselves embedded with the same biases (O’Connor & Liu, 2023). Gender bias within AI technologies can be expressed through language, stereotypical imagery, or through the use of algorithms.

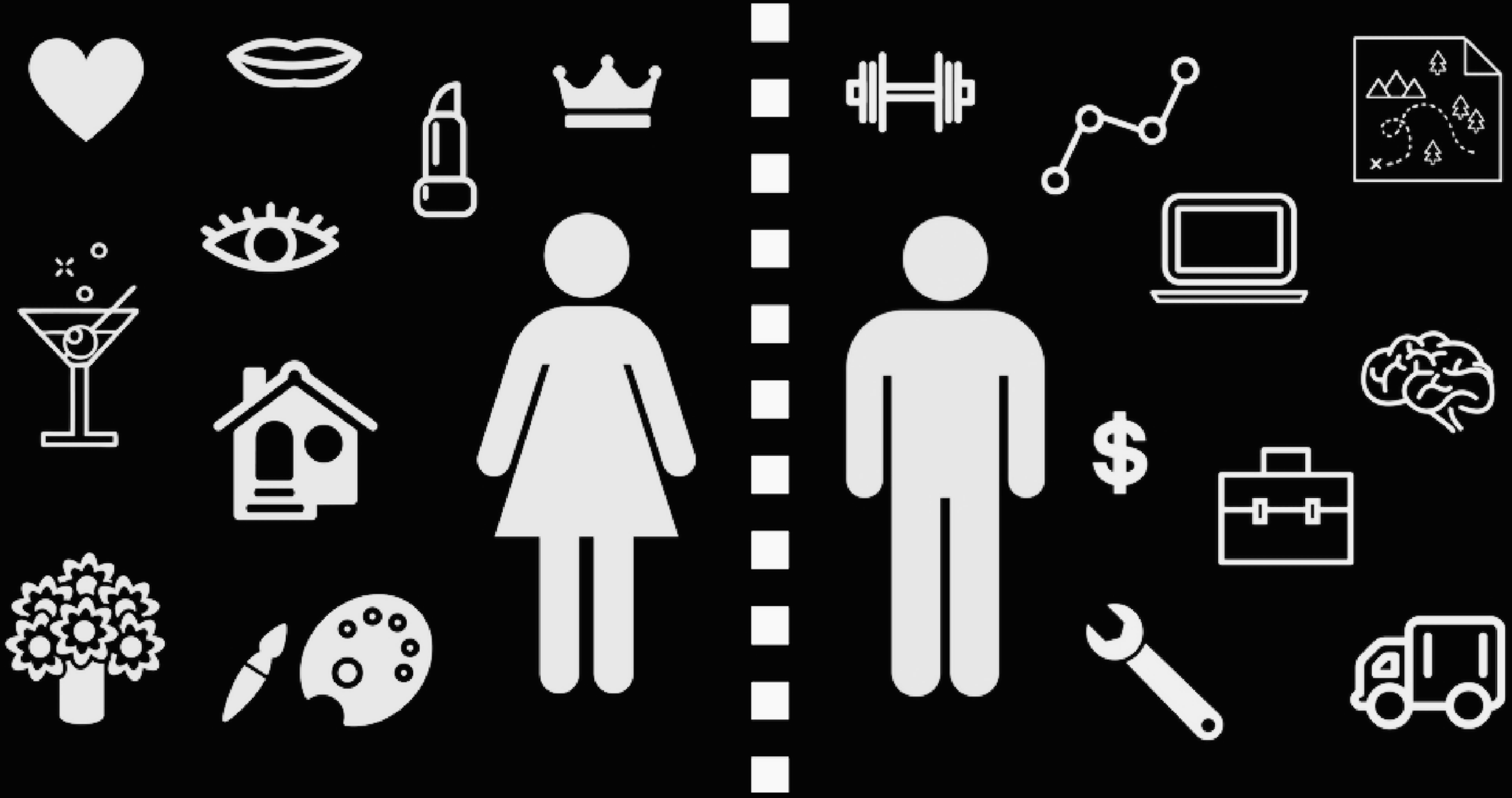


Gender Bias Through The Use Of Language & Stereotypical Imagery

Research by Menegatti and Rubini suggests that asymmetrical power relations between the genders are expressed through stereotypes associated with everyday lexical choices (where traits such as ‘nice, caring, and generous’ are used to describe females while ‘efficient, agentic, and assertive’ are used to describe men) (O’Connor & Liu, 2023). However, they also point out that the idea of the male as the ‘prototypical human being’ is encoded in the structure of many languages, for example where ‘chairman’ refers to both sexes in English (O’Connor & Liu, 2023). Another example is the AI service ‘Genderify’, launched in 2020, which uses a person’s name, username and email address to identify their gender (Vincent 2020, as cited by O’Connor & Liu, 2023). Names beginning with ‘Dr’ seemed to consistently be treated as male, as “Dr. Meghan Smith” was identified as having a 75.90% likelihood of belonging to a male (Vincent 2020, as cited by O’Connor & Liu, 2023). Elsewhere recent research describes automated robots which were trained on large datasets and standard models, but were found to exhibit strongly stereotypical and biased behavior in terms of gender and race (Hundt et al. 2022, as cited by O’Connor & Liu, 2023).

In 2018, a group of researchers at the Federal University of Rio Grande do Sul in Brazil decided to test the existence of gender bias in AI (O’Connor & Liu, 2023). In the experiment, they ran the sentence constructions in the form ‘He/She is a [job position]’ (for example, ‘He/She is an engineer’) from English into twelve languages which are gender neutral using Google Translate (O’Connor & Liu, 2023). The twelve languages they chose were Malay, Estonian, Finnish, Hungarian, Armenian, Bengali, Japanese, Turkish, Yoruba, Basque, Swahili and Chinese (O’Connor & Liu, 2023). They then selected job positions from a list issued by the U.S. Bureau of Labor Statistics (BLS), which also gives the percentage of women participation in these occupations (O’Connor & Liu, 2023). The researchers ran the ‘He/She is a [job position]’ sentence through Google Translate, noting how often the translation of the gender-neutral pronoun came out as ‘He’ or ‘She’ (O’Connor & Liu, 2023).

They expected that this translation tool would reflect the inequalities in society, and therefore inevitably display some bias in assuming certain pronouns for certain jobs (O’Connor & Liu, 2023). For example, at the time of the research, translating various sentences using the construction ‘He/She is a [job position]’ with the gender-neutral pronoun ‘ő’ from Hungarian to English gave stereotyped results, such as ‘She’s a nurse’, ‘He is a scientist’, ‘He is an engineer’ (where ‘He’s a nurse’, ‘She is a scientist’ or ‘She is an engineer’ would have been equally correct). The authors found that machine translation is strongly biased towards male defaults, especially for fields such as STEM which are typically thought of as weighted towards one gender.

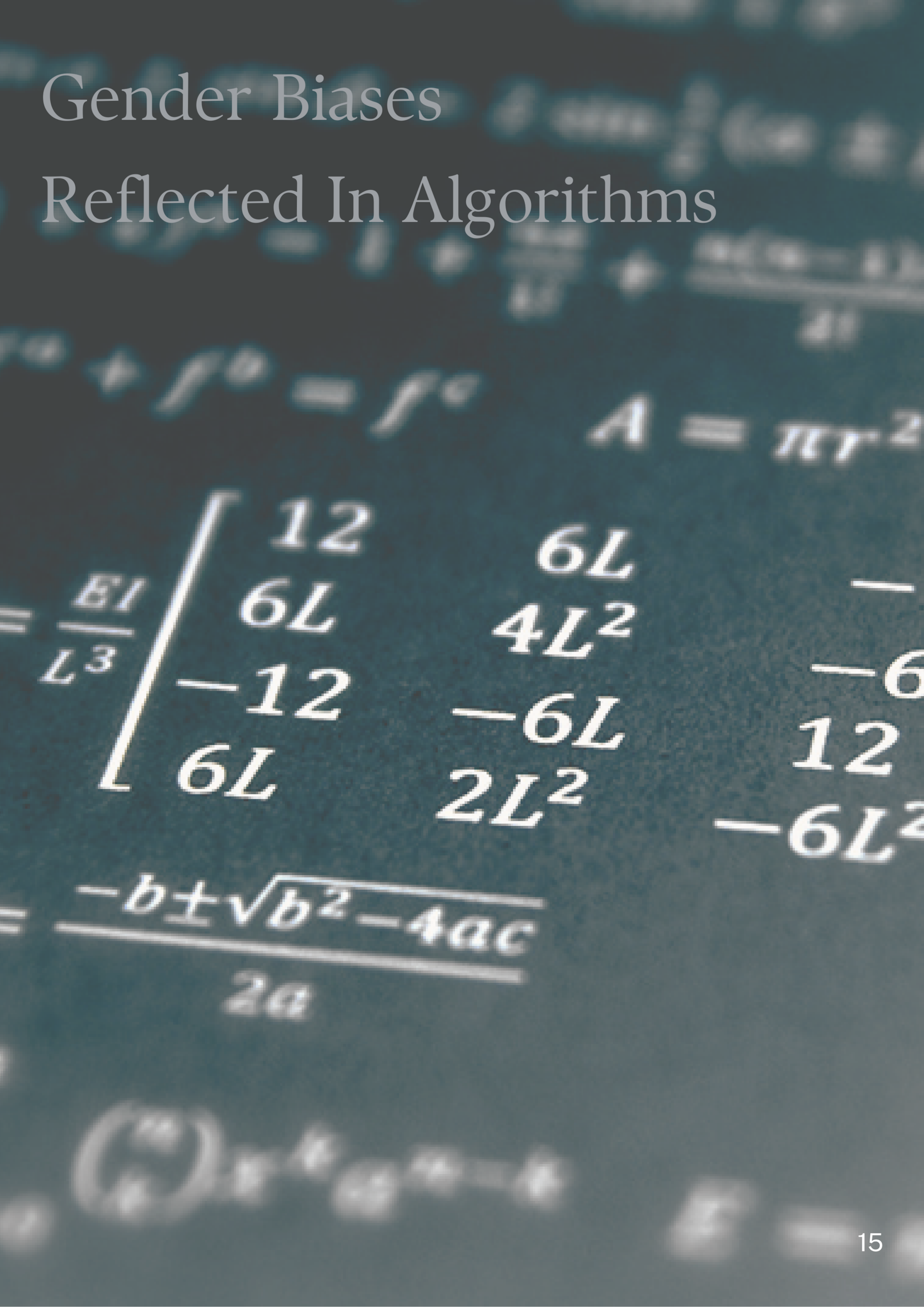


These results also did not reflect real-world statistics on gender ratios in this field. For example, 39.8% of women work in the category of 'management', but sentences were translated with a female pronoun only 11.232% of the time (66.667% of the time as male, and 12.681% of the time neutrally) (O'Connor & Liu, 2023). Overall, women made up 35.94% percent of the BLS occupations, but sentences were only translated with female pronouns 11.76% of the time, showing the translations do not reflect workplace demographics (O'Connor & Liu, 2023). These results did vary across language, as translations from Japanese and Chinese produced female pronouns only 0.196% and 1.865% of the time respectively, while Basque produced a majority of gender-neutral pronouns (O'Connor & Liu, 2023).

The authors also completed a similar subset of research using commonly used adjectives to describe human beings, including 'Happy' 'Sad' 'Shy' 'Polite' etc (O'Connor & Liu, 2023). This produced a more varied mixture of results, where words such as 'Shy', 'Attractive', 'Happy', 'Kind' and 'Ashamed' tended to be translated with female pronouns, while 'Arrogant', 'Cruel' and 'Guilty' tended towards male pronouns (with 'Guilty' in fact being exclusively translated with a male pronoun for all languages) (O'Connor & Liu, 2023).

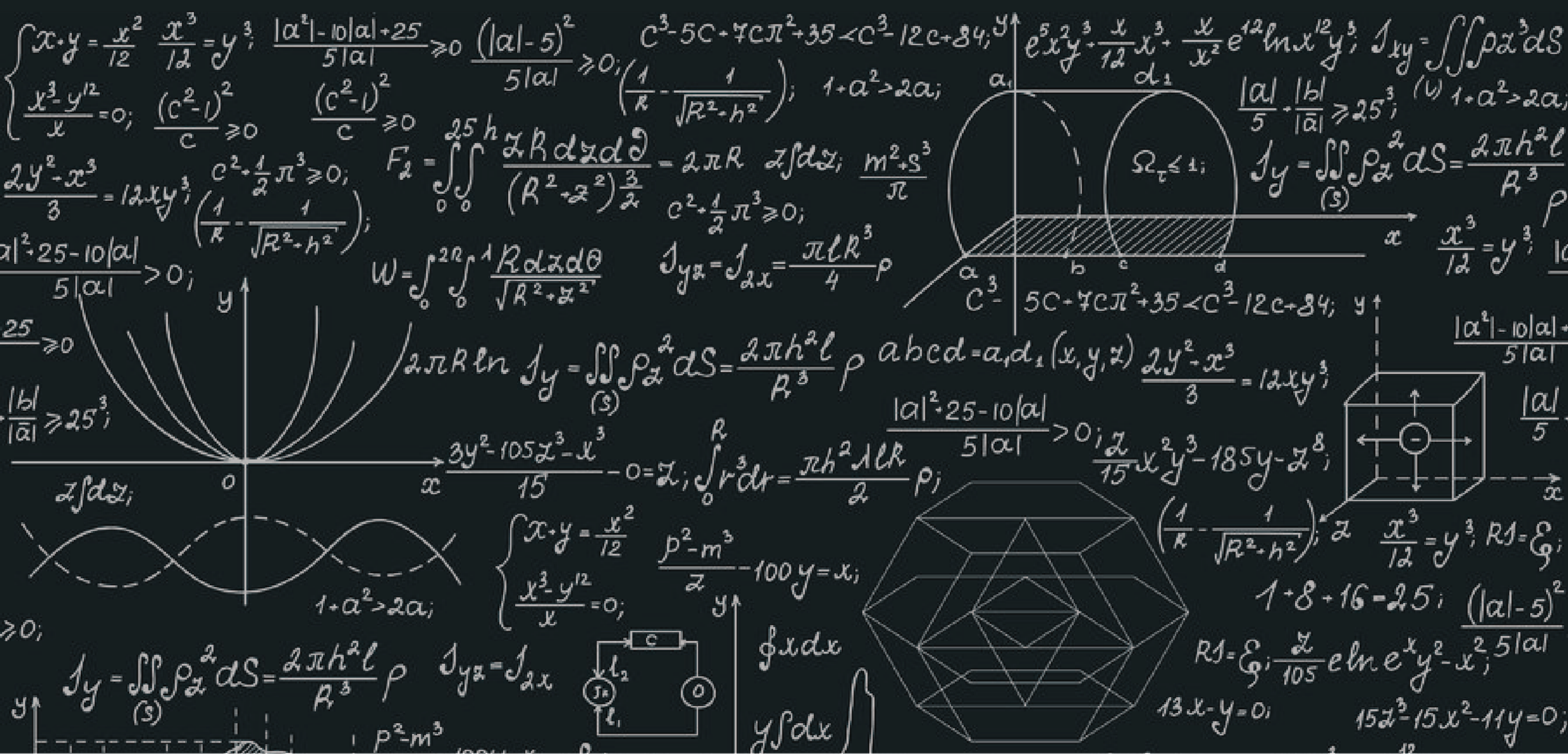
Gender bias can also present itself through stereotypical imagery. The term stereotypical imagery signifies an image that reflects a stereotype. According to the Oxford dictionary, a stereotype is a widely held but fixed and oversimplified image or idea of a particular type of person or thing. Schwemmer et al. (2020:1) asserts that "bias in the visual representation of women and men has been endemic throughout the history of media, journalism, and advertising" (O'Connor & Liu, 2023). Studies conducted on gender stereotypes in science education resources (Kerkhoven et al. 2016, as cited by O'Connor & Liu, 2023), school textbooks (Amini and Birjandi 2012:138) and commercial films (Jang et al. 2019:198) all reveal the gendered representation of men and women in public images. However, this phenomenon, and particularly stereotypes embedded in digital or online imagery remains understudied (Singh et al. 2020:1282).

Gender Biases Reflected In Algorithms



Algorithms are often seen as fairer or more neutral than humans in terms of decision-making (Gutiérrez, 2021, as cited by O’Connor & Liu, 2023). However, as these systems are created by humans and fed with data based on the human experience, they inevitably also reflect inherent human biases. For example, in Caroline Perez’s influential work on the gender data gap she explains how “we have positioned women as a deviation from standard humanity and this is why they have been allowed to become invisible” (Perez, 2019, as cited by O’Connor & Liu, 2023). Thus, algorithmic bias can be generally defined as “the application of an algorithm that compounds existing inequities in socioeconomic status, race, ethnic background, religion, gender, disability, or sexual orientation and amplifies inequities in…systems” (Igoe, 2021, as cited by O’Connor & Liu, 2023).

From the information systems (IS) artifact design perspective, biased AI-based recommendations can emerge from algorithmic unfairness (Bellamy et al., 2018; Cowgill & Tucker, 2020; Pessach & Shmueli, 2020, as cited by O’Connor & Liu, 2023). Sources of algorithmic unfairness can be categorized as bias in algorithmic predictions (due to unrepresentative training samples, mislabeling of outcomes in training samples, coding/programming bias, and algorithmic feedback loops), and biased algorithmic objectives (related to decision thresholds that may limit/promote diversity, spillovers emerging from biased group-level outcomes, and a trade-off between the exploration of new information and exploitation of existing information) (Cowgill & Tucker, 2020, as cited by O’Connor & Liu, 2023). Farnadi et al., (2018) highlight that algorithmic bias may emerge from systematic bias present in data (owing to societal/historical features), as well as from feedback loops when biased recommendations get displayed by a recommender system and then get further entrenched, due to the fact that there is an “increase in probability for the item to be retained in the system” (O’Connor & Liu, 2023).



Algorithms can be especially dangerous because they "don't simply reflect back social inequities but may ultimately exacerbate them" (Igoe, 2021, as cited by O'Connor & Liu, 2023). Busuioc notes how algorithmic tools can "get caught in negative feedback loops" which then becomes the base for future predictions—all exacerbated if the initial data fed into the machine was itself biased (Busuioc, 2021, as cited by O'Connor & Liu, 2023). Studies on the use of AI have discovered gender bias in the outcomes of algorithm application, from natural language processing techniques which perpetuate gender stereotypes (Kay et al. 2015, as cited by O'Connor & Liu, 2023) to facial recognition software which is much more accurate on male faces than female ones (Domnich & Anbarjafari, 2021, as cited by O'Connor & Liu, 2023). Thus, algorithmic bias can be generally defined as "the application of an algorithm that compounds existing inequities in socioeconomic status, race, ethnic background, religion, gender, disability, or sexual orientation and amplifies inequities in...systems" (Igoe, 2021, as cited by O'Connor & Liu, 2023).

Amazon's recruitment tool, which produced AI-based recommendations that significantly favored men over women for technical jobs (Dastin, 2018, as cited by O'Connor & Liu, 2023). This happened because the depth, range, and scope of the data used to train algorithms were critical for the accuracy of the subsequent classification and recommendation tasks provided by the AI tools; meaning the training data used by Amazon's recruitment tool was comprised of résumés mostly submitted by men (Dastin, 2018, as cited by O'Connor & Liu, 2023).

Solutions

Smart Industrial Control System
Dashboard
Machine 1
Machine 2
Machine 3
Overview
Product report
Performance
Product Quality
Machine State
Monthly Summary
32.04%
Product Quality
99%
Performance
25%
Power consumption
67%
Machine State
Fine

Discussions on gender bias often naturally fall into two categories: studies which explore or attempt to measure gender bias in AI techniques (Stanovsky et al. 2019; Sheng et al. 2019; as cited by O'Connor & Liu, 2023), and those which focus more on how to mitigate gender bias itself (Stafanovičs et al. 2020; Deshpande et al.2020; Domnich and Anbarjafari 2021, as cited by O'Connor & Liu, 2023). This distinction has been noted by authors such as Blodgett et al. (2020), whose paper critically reviewing papers on bias in NLP notes that these studies either "proposed quantitative techniques for measuring or mitigating 'bias'", or the Brookings Institute research framework for 'algorithmic hygiene' which includes identifying sources of bias and then forwarding recommendations on how to mitigate them (Lee et al. 2019, as cited by O'Connor & Liu, 2023). Of course, many of the studies which focus on mitigation techniques also implicitly or explicitly include descriptions or measurements of the gender bias issue they are attempting to resolve (O'Connor & Liu, 2023).

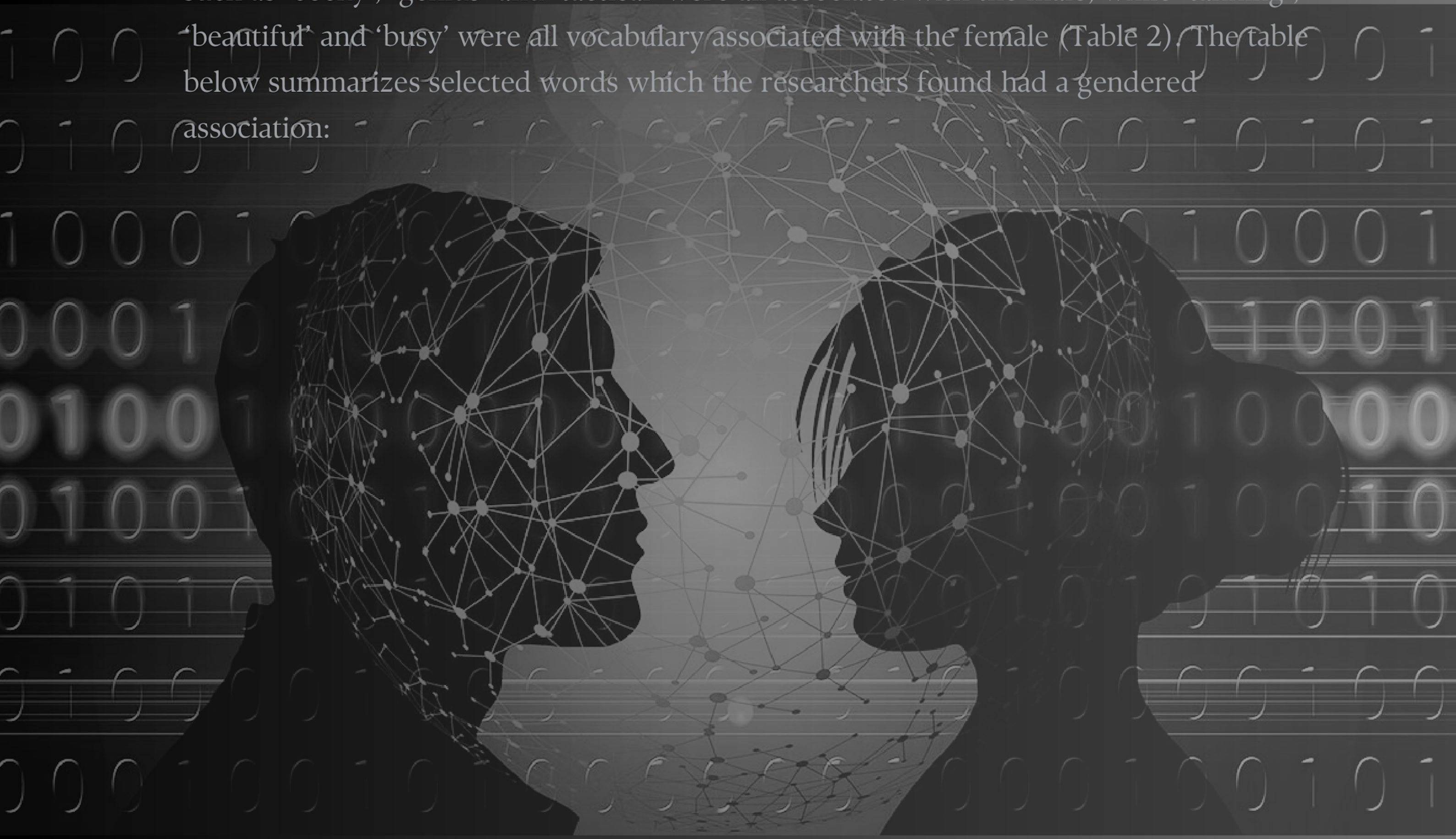
Friedman and Nissenbaum's 1996 work on bias in computer systems points to three types of bias; pre-existing bias (emerging from societal attitudes and practices), technical bias (due to technological constraints) and emergent bias (which arises as the computer system is used) (O'Connor & Liu, 2023). However, AI bias is an extremely complex topic, covering different forms of bias and notions of fairness (Bernagozzi et al. 2021, as cited by O'Connor & Liu, 2023). Currently, there are two streams of literature that address gender bias. The first stream of literature focuses on pointing out the amplification of gender bias (often meaning discrimination against women) inherent in many technologies, such as in audio-visual data (Gutiérrez 2021), online language translators (Bernagozzi 2021) and recruiting tools (Dastin, 2022) (O'Connor & Liu, 2023). The second stream of literature goes beyond exploring the existence of gender bias in technology, and additionally explores methods for mitigating this bias (O'Connor & Liu, 2023). This includes studies on how to reduce gender bias during the resume screening process (Deshpande et al. 2020), in machine learning models (Feldman and Peake 2021) and facial recognition systems (Dhar 2020)(O'Connor & Liu, 2023). This stream includes both research on how to mitigate the effects of bias amplification which can be seen in AI, as well as studies which specifically aim to harness AI in order to reduce gender bias in technologies (as cited by O'Connor & Liu, 2023).

Bias Mitigation Of Algorithms & Language

Bias mitigation involves “proactively addressing factors which contribute to bias” (Lee et al. 2019, as cited by O’Connor & Liu, 2023). In terms of algorithms, bias mitigation is often strongly associated with the concept of ‘fairness’ (O’Connor & Liu, 2023). For example, several researchers came together in 2018 to create the AI Fairness 360 (AIF360), a toolkit which provides a framework against which researchers can evaluate algorithms (O’Connor & Liu, 2023). This includes “bias mitigation algorithms” which can “improve the fairness metrics by modifying the training data, the learning algorithm, or the predictions” during the pre-processing, in-processing, and post-processing stages (Bellamy et al. 2019, as cited by O’Connor & Liu, 2023).

In 2016, a group of researchers from Boston University and Microsoft’s Research Lab in New England, USA, came together to propose a methodology for removing gender bias from word embeddings—a natural language processing task which captures semantic associations between words in a text (Bolukbasi et al. 2016, as cited by O’Connor & Liu, 2023). A word embedding represents each word in text data as a ‘word vector’, which is a mathematical representation of the meaning of the word by mapping it in space (Alizadeh 2021, as cited by O’Connor & Liu, 2023). This provides two sets of information about word meanings in a text. Firstly, vectors which are closer together represent words which have similar meanings (O’Connor & Liu, 2023). Secondly, comparing different vectors can represent semantic relationships between words, enabling the input of ‘man is to king as woman is to x’ to find x=‘queen’ (O’Connor & Liu, 2023). The researchers note that there is much research on word embeddings themselves; however, little attention is paid to the inherent sexism captured by word embeddings, which will predict the answer to ‘man is to computer programmer as woman is to x’ as ‘x=homemaker’ (O’Connor & Liu, 2023). As word embeddings are widely used as a basic feature in NLP, their use has the potential to amplify gender bias in systems (O’Connor & Liu, 2023).

In this paper, the researchers analyzed the ‘word2vec’ embedding which is a popularly used embedding that uses neural network methods to learn embeddings from data sets (TensorFlow 2022). The embedding is trained on a 3 million word-large English language Google News corpus and the resulting embedding is referred to by the researchers as ‘w2vNEWS’. The aim of the study is to first demonstrate the biases contained in word embeddings, and then to create a debiasing algorithm to “remove gender pair associations for gender-neutral words”. ‘Gender neutral words’ are words which have no specific gender association and these are contrasted with gender-specific words which explicitly include a gendered reference. For example, ‘daughter’ ‘lady’ and ‘queen’ are examples of gender-specific words as they explicitly refer to the female gender. However, ‘rule’ ‘game’ ‘nurse’ and ‘homemaker’ are all examples of gender-neutral words—words which do not refer to one gender or the other. Yet, despite this, ‘gender-neutral words’ often are semantically correlated to a certain gender. Thus, the authors found that certain words such as ‘cocky’, ‘genius’ and ‘tactical’ were all associated with the male, while ‘tanning’, ‘beautiful’ and ‘busy’ were all vocabulary associated with the female (Table 2). The table below summarizes selected words which the researchers found had a gendered association:



Feminine Tote	Flirt; Divorce; Tearful; Modeling; Crafts; Browsing; Busy; Trimester
Masculine Buddy	Command; Firepower; Game; Zeal; Guru; Yard; Youth; Firmly; Builder

Table 2

The debiasing algorithm developed by the researchers aimed to remove the gender pair associations for all these ‘gender neutral’ words, while retaining the function of word embedding in mapping useful relationships and associations between words (O’Connor & Liu, 2023). The algorithm involves two steps, the first step identifies the subspace that shows the gendered bias (O’Connor & Liu, 2023). The second step either ‘neutralizes and equalizes’ (gets entirely rid of the gendered connotations of gender-neutral words and then ensures they are equidistant from all other words in the set) or ‘softens’ the bias (maintaining certain useful distinctions between words in a set—for example where a word has more than one meaning) (O’Connor & Liu, 2023). They then evaluated the algorithm through generating word pairs comparable to 'she-he’ (for example ‘he’ is to ‘doctor as ‘she’ is to ‘x’, where the algorithm must determine the value of x) before asking crowd workers to rate whether these pairs reflected gender stereotypes (O’Connor & Liu, 2023). While the initial embedding was found to represent stereotypes 19% of the time, the new debasing algorithm reduced this percentage to 6% (O’Connor & Liu, 2023). For example, they noted that the original embedding would find the x in ‘he is to doctor as she is to X’ as ‘nurse’; however, the new embedding found ‘x=physician’ (O’Connor & Liu, 2023). Despite this, the algorithm still preserved appropriate analogies, such as ‘she is to ‘ovarian cancer’’ as ‘he is to ‘prostate cancer’’ (O’Connor & Liu, 2023).

The authors noted that to entirely solve this problem “one should attempt to debias society rather than word embedding”; however, they note that their algorithm at the very least will not amplify bias (O’Connor & Liu, 2023). This research has been cited over 1000 times in studies on AI ethics, bias in machine learning and papers on bias mitigation (O’Connor & Liu, 2023). It has also been cited by popular news sites such as Forbes (Roselli et al. 2019) and The Conversation (Zou 2016) as well as on academic sites such as MIT Technology Review (2016) (O’Connor & Liu, 2023). The code itself is available on GitHub for users to download themselves and debias their own text data (O’Connor & Liu, 2023).

Moreover, the 2018 study done by a group of researchers at the Federal University of Rio Grande do Sul in Brazil that tested the existence of gender bias in AI, specifically in automated translation, led to Google Translate releasing the ‘Translated Wikipedia Biographies’ dataset, through which gender bias of machine translation can be measured, due to the high potential for translation errors—they state that datasets can reduce errors by 67% (Stella 2021) (O’Connor & Liu, 2023).

Bias Mitigation Of Digital Imagery

Researchers from the University of Virginia, University of California Los Angeles and the Allen Institute for Artificial Intelligence came together in 2019 to explore the issue of gender bias in image representation (Wang et al. 2019, as cited by O'Connor & Liu, 2023). Their study begins by pointing out how facial recognition systems often amplify biases based on protected characteristics such as race or gender, and how this can have real-world consequences, for example autonomous vehicle systems being unable to recognize certain groups of people (O'Connor & Liu, 2023). They begin by studying bias amplification through the COCO dataset for recognizing objects and the imSitu dataset for recognizing actions (O'Connor & Liu, 2023). The COCO dataset (Microsoft Common Objects in Context) is an image dataset which can be used to train machine learning models, containing over 328,000 annotated images of humans and every-day situations (datagen, 2022, as cited by O'Connor & Liu, 2023). The imSitu dataset contains images describing situations along with annotations describing the situations, which can also be used to train algorithms on situation recognition (O'Connor & Liu, 2023).

They propose a new definition for measuring bias amplification, where instead of comparing the training data and model predictions, they compare "the predictability of gender from ground truth labels (dataset leakage…) and model predictions (model leakage…)" (O'Connor & Liu, 2023). Ground truth labels are those labels assigned to the data by human workers—that is to say they are accurate representations of the data (O'Connor & Liu, 2023). Model predictions are those made by the model (algorithm) itself, and thus comparing these two makes it possible to test the accuracy of the modeling (O'Connor & Liu, 2023). Using this method, they find that even models which are not programmed for predicting gender will still amplify gender bias (O'Connor & Liu, 2023). They hypothesize that models may perpetuate biases because there are gender-related features in the image which are not labeled by the computer program, but may still be taken into account when predicting gender—this is called 'data leakage' in this paper (O'Connor & Liu, 2023). For example, they give the example of a dataset with an equal number of women and men shown cooking (O'Connor & Liu, 2023). This in itself does not amplify bias, but if there is a child in the image, and children are often shown more with women than men across all images, then the model may associate 'children' with 'cooking', and therefore overall women could be labeled as 'cooking' more than men still (O'Connor & Liu, 2023). Model leakage then referred to how much the model's predictions were able to identify protected characteristics (here gender) (O'Connor & Liu, 2023). The researchers adopted the method of 'adversarial debiasing' in order to mitigate this effect (O'Connor & Liu, 2023).

This could preserve useful information, while removing gender correlated features in the images (O’Connor & Liu, 2023). Sometimes this involves eliminating the face, or even gender-associated clothing, while retaining information needed to recognize actions or objects (O’Connor & Liu, 2023). Their proposed algorithm aims to “build representations from which protected attributes can not be predicted” (O’Connor & Liu, 2023). Quantitatively, the algorithm was able to reduce model leakage by 53% for COCO and 67% for imSitu (O’Connor & Liu, 2023). Then, comparing their method with another debiasing algorithm (RBA), they show that the authors’ methods are much more effective at reducing bias amplification (O’Connor & Liu, 2023). Overall, they conclude that balanced datasets are not enough to prevent encoded bias in computer vision, and instead support the idea of removing features associated with a protected variable (such as gender) from images (O’Connor & Liu, 2023). Their work has been cited over 160 times, their code is available online, and as well as this, they have created a demo page where users can upload their own image and apply the adversarially trained neural network to obscure gender information (O’Connor & Liu, 2023).

Political Action For The Mitigation Of Bias Within AI Technologies

In Noble's work on the 'Algorithms of Oppression' (2018:1), she posits that "artificial intelligence will become a major human rights issue in the twenty-first century" (O'Connor & Liu, 2023). In this way, there have already been attempts by national and international institutions to begin creating policies and frameworks to identify and mitigate these biases (O'Connor & Liu, 2023). In 2019 the independent High-Level Expert Group on Artificial Intelligence, set up by the European Commission, produced a report entitled 'Ethics Guidelines for Trustworthy AI' (O'Connor & Liu, 2023). The paper proposes "equality, non-discrimination and solidarity" as a fundamental right, calling to ensure that systems do not generate unfairly biased outputs, including using inclusive data which represents different population groups (O'Connor & Liu, 2023). The European Commission also aims to introduce a legal framework for AI, in compliance with the E.U. Charter of Fundamental Rights, aimed at defining responsibilities of users and providers (DiNoia et al. 2022, as cited by O'Connor & Liu, 2023).

A separate report by UNESCO (2020) on AI and gender equality suggests a range of practices for integrating gender equality into AI principles, including proactive mitigation, making the invisible visible and understanding AI as a potentially empowering tool for girls and women (O'Connor & Liu, 2023). Many of the case studies in this paper point out that bias is inherent in society and thus it is inherent in AI as well (O'Connor & Liu, 2023). The UNESCO recommendations accept this premise, but still promote the importance of "shift[ing] the narrative of AI as something 'external' or technologically deterministic, to something 'human' that is not happening to us but is created, directed, controlled by human beings and reflective of society" (O'Connor & Liu, 2023). Thus, it is not a question of either/or as to whether to first change society or change AI, both must be achieved in tandem (O'Connor & Liu, 2023).

A similar sentiment is echoed in a review on Artificial Intelligence and Public Standards by the Committee on Standards in Public Life, an independent body advising the UK government (O'Connor & Liu, 2023). Their report (2022) concludes that the existing 'Seven Principles of Public Life' (selflessness, integrity, objectivity, accountability, openness, honesty and leadership) should be upheld as a guide in how to integrate AI technologies into public life (O'Connor & Liu, 2023). Understanding that AI may have wide-ranging and unexpected effects, the review proposes a general outline of how the Seven Principles can be translated into practice for the use of AI (O'Connor & Liu, 2023). Overall, it is clear that current policy recommendations for the regulation of AI focus on overarching principles and guidelines, reflecting the ongoing and expanding range of issues which may need to be addressed in future (O'Connor & Liu, 2023).

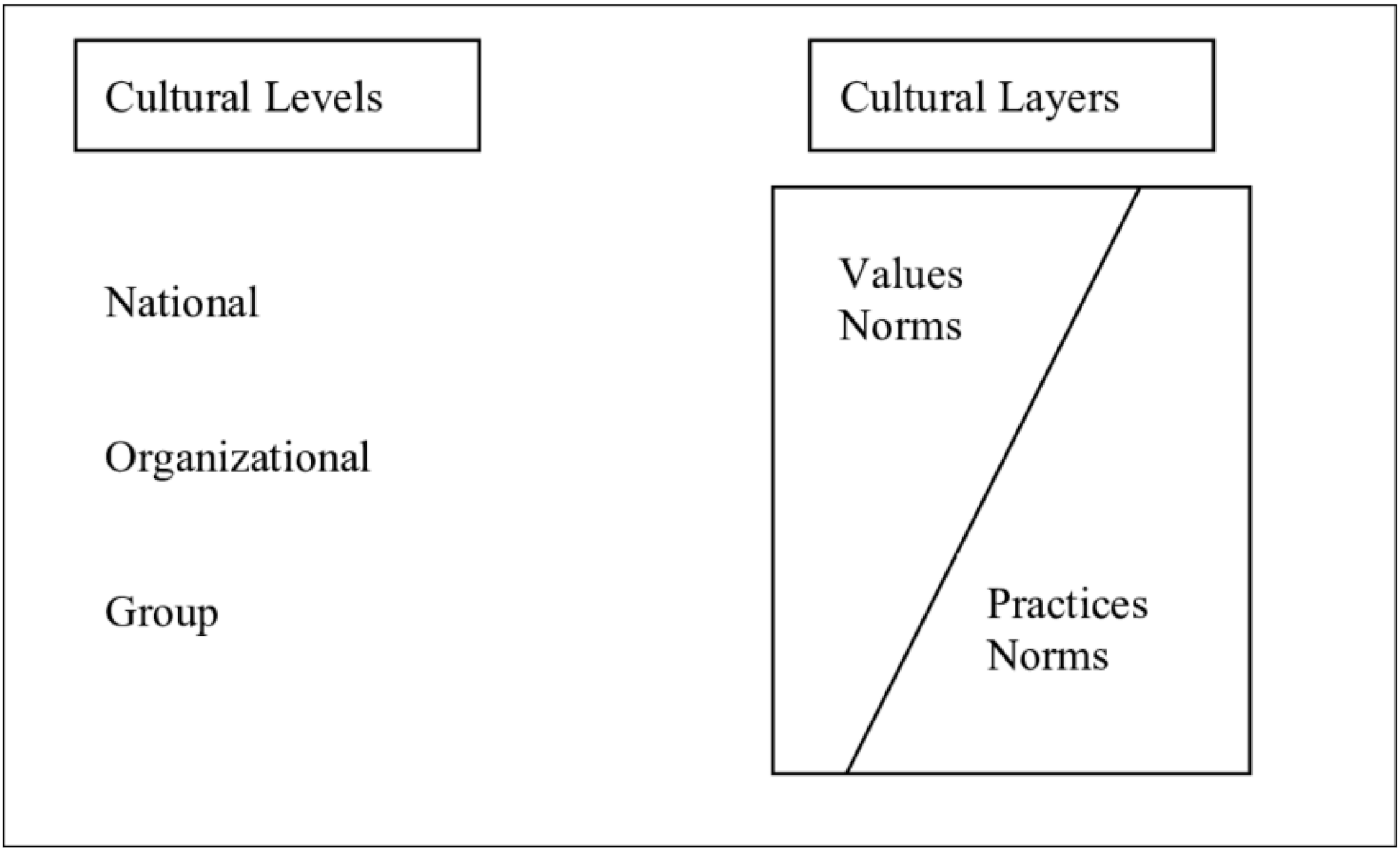
UNESCO
ORGANISATION DES NATIONS UNIES POUR L'EDUCATION LA SCIENCE ET LA
UNITED NATIONS EDUCATIONAL SCIENTIFIC AND CULTURAL ORG

4. Anticipating The Phenomenon of Gender Bias Within AI Technologies Developing in Africa

In 2021, Manjul Gupta, Carlos M. Parra & Denis Dennehy conducted a study that pursued the answer to this research question: Do individual-level cultural values affect the extent to which individuals would question AI-based recommendations due to perceived racial or gender bias (Gupta, Parra & Dennehy, 2022)? The study examined the effects of five cultural values (collectivism-individualism, power distance, masculinity-femininity, uncertainty avoidance, and long/short-term orientation) from Srite and Karahanna's (2006) model of individual-level cultural values, derived from Hofstede's (1982) cultural framework (Gupta, Parra & Dennehy, 2022).

The study found that increased levels of collectivism ($\beta = 0.18$, $p < .001$), masculinity ($\beta = 0.24$, $p < .001$) and uncertainty avoidance ($\beta = 0.13$, $p < .01$) led to an increase in participants' questioning of the AI-based recommendations when they perceived the recommendation had a gender bias. For gender ($\beta = -0.17$, $p < .01$), the pairwise comparisons indicated that female participants (Mean = 3.55, SE = 0.04) had higher mean AI questionability (gender) than that of males (Mean = 3.38, SE = 0.03). As participants' daily internet usage ($\beta = 0.17$, $p < .001$) increased, AI questionability (gender) also increased. There were also some interesting findings about the role of control variables. Regardless of the type of the bias, participants' gender and their daily internet usage had significant effects on AI questionability. Particularly, females exhibit higher AI questionability due to perceived bias than males. It is understandable as the popular press is rife with articles of artificial intelligence being biased against women, thereby making females, in general, more suspicious of AI-based recommendations (Niethammer, 2020).

4.1 Srite and Karahanna's (2006) model of individual-level cultural values



Cultural Levels

National

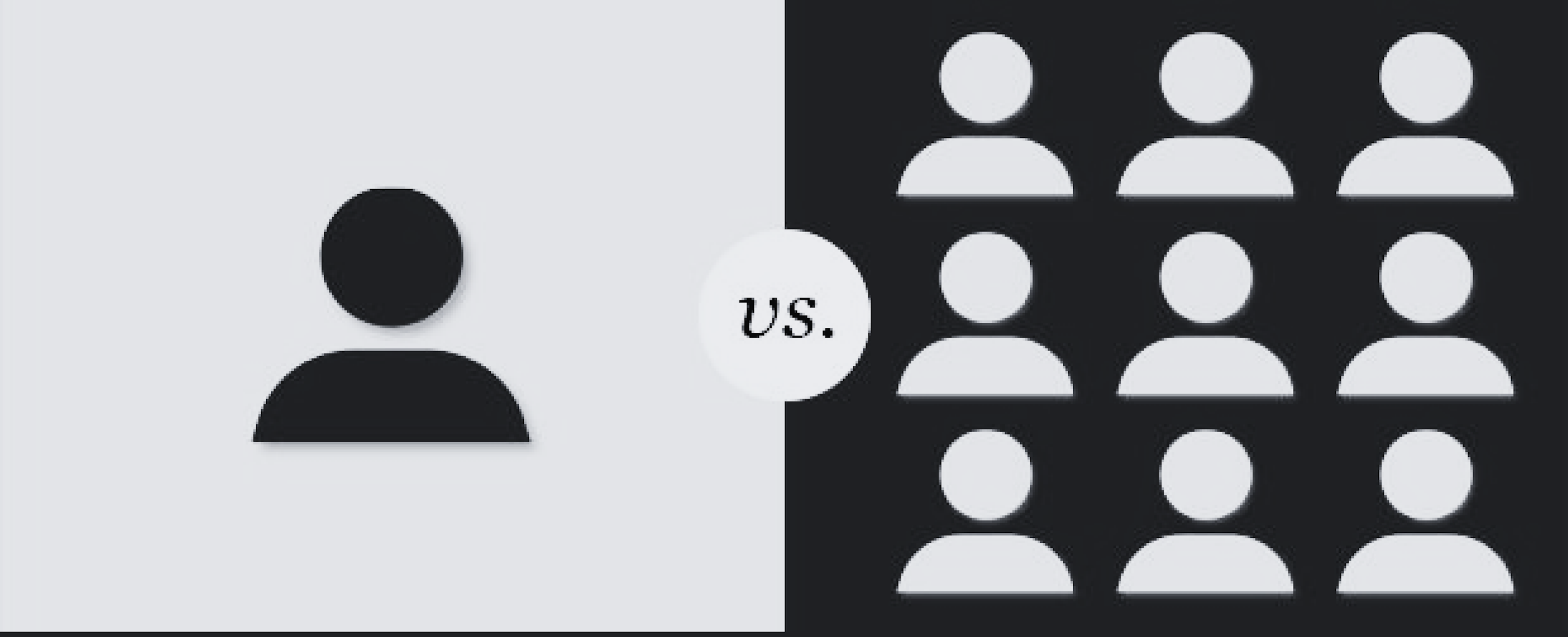
Organizational

Group

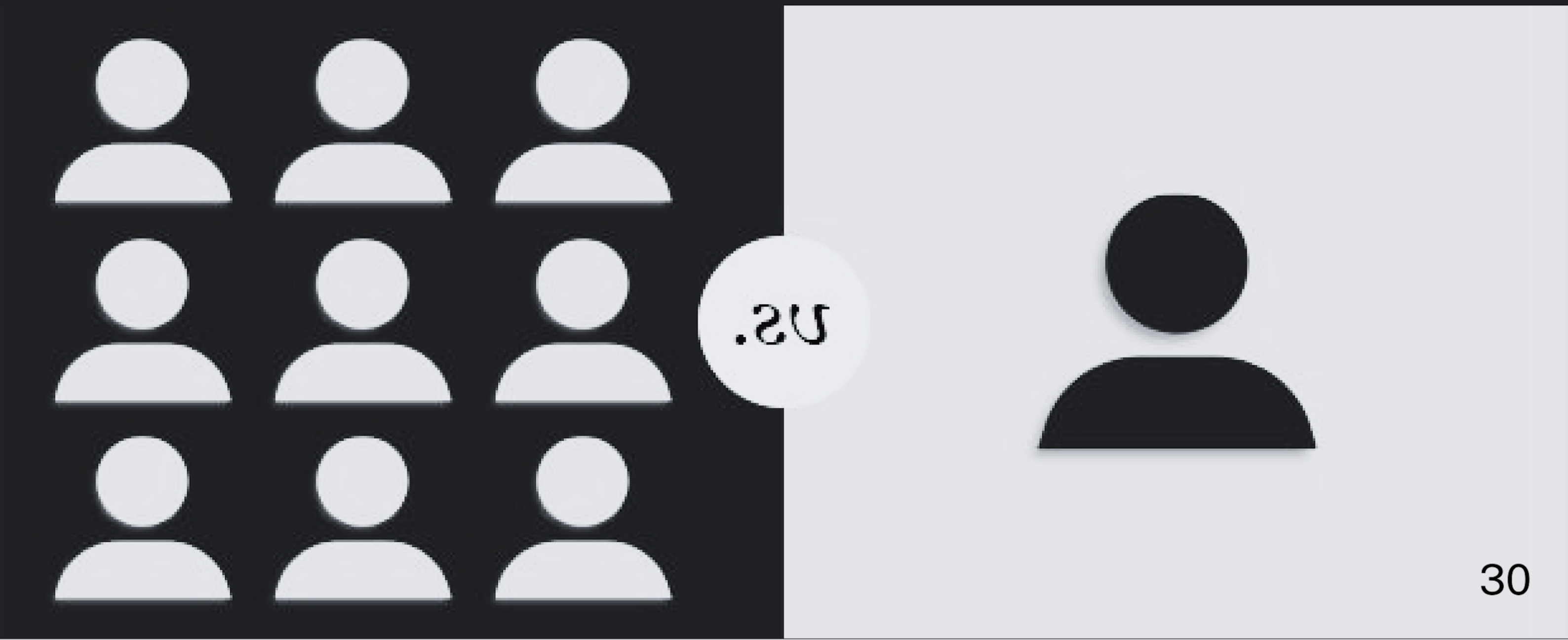
Cultural Layers

Values
Norms

Practices
Norms

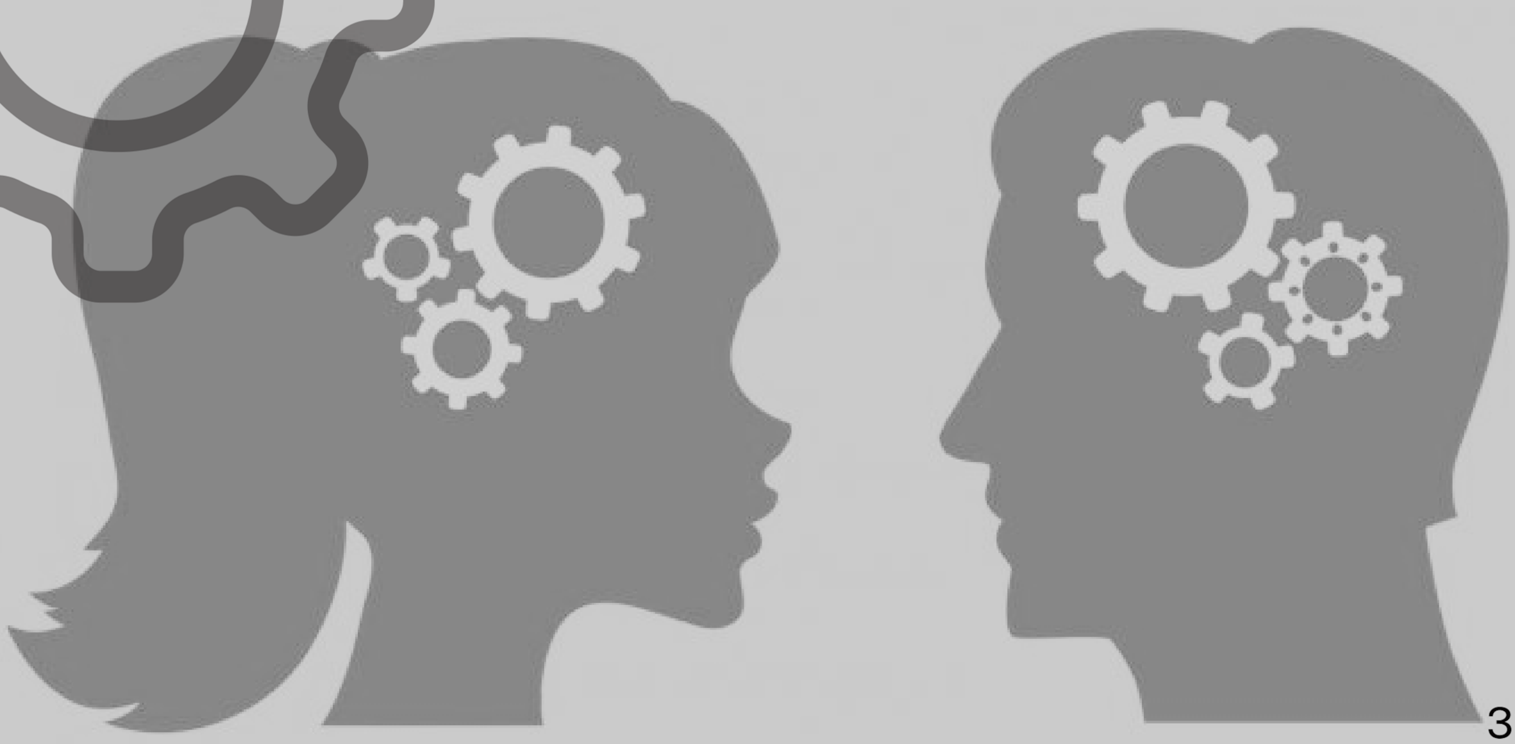


The collectivism-individualism dimension describes the extent to which individuals value group-orientation over self-orientation. Strong group-oriented behaviors reflect collectivism, while individualism (i.e., the opposite of collectivism) is manifested in behaviors where the self is more important than others. Stated simply, collectivism places emphasis on “we, us, and our,” whereas individualism values “I, me, and myself” (Agrawal & Maheswaran, 2005; Kumashiro, 1999). This perceived feeling of “we-ness” is what differentiates people with collectivistic traits from individualists. Collectivistic cultural values are characterized by the presence of strong, cohesive in-groups, which consist of others perceived to be similar to oneself. Furthermore, collectivists have a strong sense of community, loyalty, respect, and trust towards the other members of their in-group. A family, village, nation, organization, religious group, soccer team, and student body are examples of in-groups (Triandis, 1996). By comparison, individualists are focused on doing their own things. They value autonomy and are not obligated to trust and respect others the same way as those with collectivistic cultural traits.

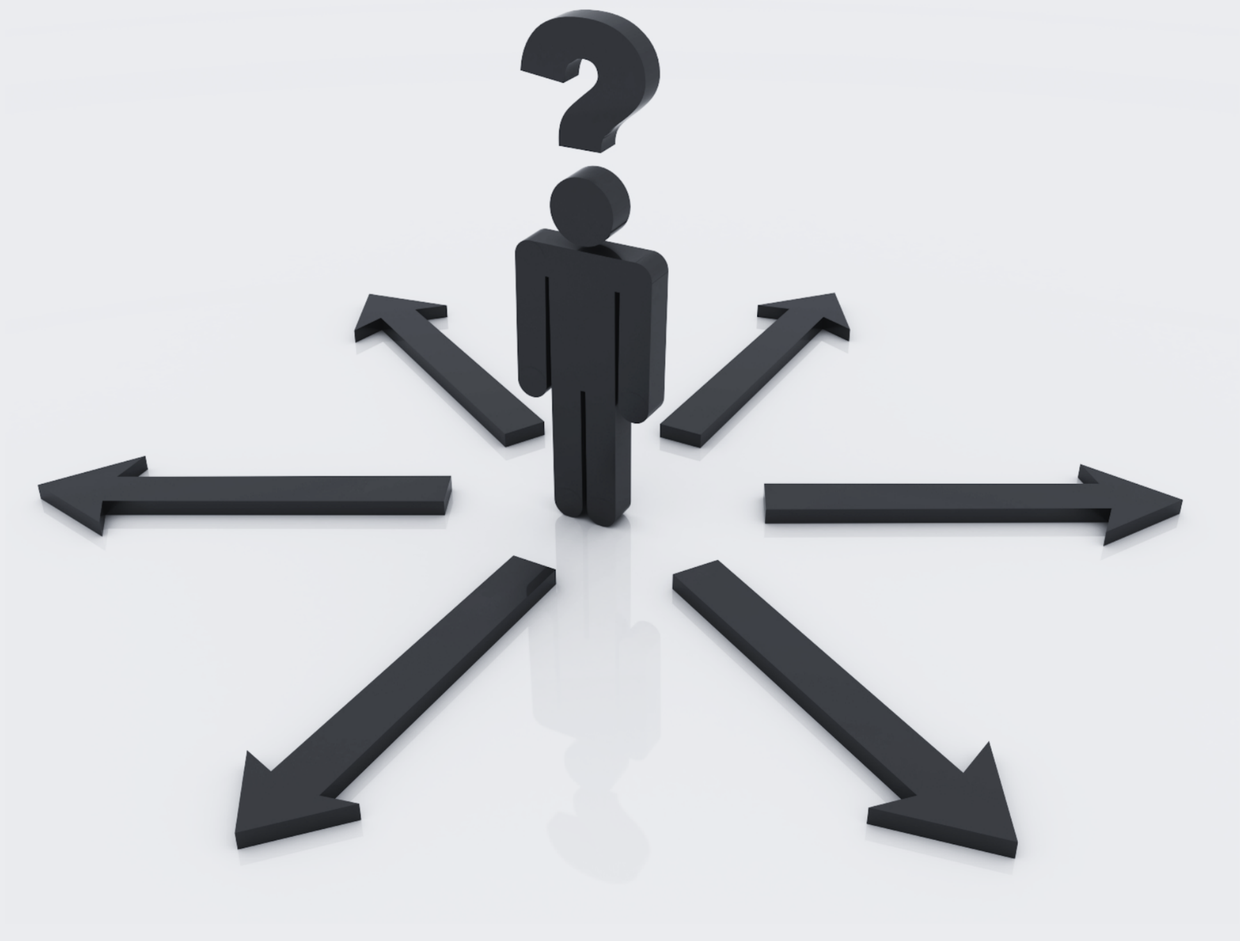


The dimension of power distance deals with the extent to which individuals accept and expect that power is distributed unequally in society. While inequality, in general, represents societal divisions due to socioeconomic status (i.e., education, income, and occupation), the term "power," in addition to an individual's socioeconomic status, may signify someone's influence due to his or her social and/or political affiliation, race, caste, age, prestige, or intellectual ability. Hofstede (1980) argues that stratifications exist in all societies; however, some are more unequal than others. High power distance cultural values maintain that inequality exists, and they do not perceive it as a problem. Everyone has a place in society, and thus, it is acceptable for some to be privileged (and underprivileged) in society. High power distance values imply obeying those with power, for example, the elderly due to their age and one's superiors due to their organizational titles. Arguing with superiors or presenting a differing opinion is not encouraged and is often looked down upon. A good manager is one that performs difficult tasks and delegates repetitive and mundane tasks to subordinates. Moreover, managers seeking feedback or advice from their subordinates are considered weak and ineffective. It is also acceptable for senior-level managers to earn a significantly higher income than lower-level employees. In sum, those with less power must show deference to those with more power in society. By comparison, low power distance cultural values advocate reducing the perception of power by allowing everyone to be treated equally. It is not customary for individuals to agree with others just because they have more influence due to their socioeconomic status, higher-level position, or political ranks. Everyone is encouraged to share their perspectives freely, even if they contradict the views of those with more power.

Like collectivism-individualism, masculinity and femininity are the opposite ends of the same cultural spectrum. While masculinity captures the extent to which individuals in society value assertiveness, heroism, and achievement, femininity emphasizes nurturing, quality of life, and modesty. Masculine cultures tend to be highly performance-oriented, while feminine cultures value a good consensual working relationship with others. Hofstede (1980) argues that while the masculinity-femininity dimension may look similar to biological sex (male/female), there is an important difference between the two. For instance, the "sex" categorizes individuals either into male or female at the time of birth based on the presence (or absence) of different biological factors (e.g., Y chromosome, reproductive gland) (Knox & Schacht, 2012). On the other hand, the cultural dimension of masculinity-femininity represents social gender. Thus, a biological male may have feminine or masculine cultural values and vice versa.



Uncertainty avoidance measures the extent to which individuals in a society are risk-averse versus risk-tolerant. Those with high uncertainty avoidance values have a propensity to feel threatened while dealing with unplanned events. They would want to minimize any degree of ambiguity in their lives and make the future as evident as possible. Therefore, high uncertainty avoidance cultural values endorse formal rules and regulations in organizations, institutions, and relationships to prevent uncertainty in everyday situations. By comparison, those with low uncertainty avoidance values have a high tolerance for risk and thus are not intimidated when presented with unexpected circumstances. It is not that individuals with high uncertainty avoidance values are terrified of taking a risk; however, when they do have to take a risk, they would instead opt for a risk that is known rather than unknown (Hofstede, 2003). 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. By comparison, the risks associated with the unknown do not affect the behaviors of those with low uncertainty avoidance cultures. The objective of putting together rules and structures in high uncertainty avoidance cultures is to enable smooth functioning of everyday activities in organizations and society. Individuals with high uncertainty avoidance values prefer clarity and have a low tolerance for irregular or deviant behaviors (Hofstede, 2011). These individuals may further feel anxious and stressed out when they do not obtain the outcome that they were expecting.



The long/short-term orientation dimension measures the extent to which individuals in a society depend on long-standing traditions and past historical events to make decisions about the present and future. Long/short-term orientation was not a part of the initial cross-cultural model suggested by Hofstede. It was added later as the fifth dimension to the model based on the work of Bond (1988). Due to its roots in Confucianism philosophy, initially, this dimension was not well received in the cross-cultural community (Fang, 2003). However, over time, long/short-term orientation has been established as an essential cultural dimension capable of explaining individuals' behaviors (Hofstede et al., 2010). Long-term oriented values are based on the premise that everything is temporary, and the change is inescapable. By comparison, due to their deep-rooted respect for past traditions, those with short-term oriented values are reluctant to change. Long-term oriented values are reflected in careful management of money, being persistent despite criticisms, and willingness to give up today's fun and leisure for success in the future. In contrast, personal stability and expectation of quick results are important short-term oriented values. Given its forward-looking focus, long-term orientation is called a pragmatic cultural dimension, while short-term orientation is referred to as the normative dimension.

Vs

4.2 Taking Action Against Gender Bias Within AI Technologies In Africa

Utilizing the results from the study conducted by Manjul Gupta, Carlos M. Parra & Denis Dennehy, we can form the hypothesis that educating individuals within the African Tech field about gender bias will not be as difficult. To make sure that Africa can get ahead of the phenomenon of producing AI technologies that perpetuate gender biases, the following will be required:

Establish trainings that educate those in the tech field about the issue of gender bias within AI technologies

Conduct more studies on "bias mitigation algorithms", and the create said algorithms in a fashion that works within the African societies' context

Avoid creating and spreading stereotypical imagery across different platforms

Create divisions within tech companies that are assigned the task of monitoring and controlling gender

Supporting efforts that combat against negative gender biases the within the societal context

Unlike the simpler task of educating individuals within the African Tech field about gender bias, combatting the gender bias that exists in the majority of many African countries will be difficult. This task will be difficult due to the fact that the majority of African countries have proven to hold short-term oriented values, meaning that they are reluctant to change. Though not all "change" may be positive, change is necessary for any society that seeks to improve over the course of generations.

4.3 Conclusion

The following research question was posed at the beginning of this paper:
Are Artificial Intelligence technologies going to perpetuate the current Gender biases that exist in the majority of African societies and if so, what action can be taken by African women & men in STEM/Tech fields?
The answer is, it has been proven that Artificial Intelligence technologies perpetuate the current Gender biases that exist within the society in which they function, thus it is highly likely that AI technologies are going to perpetuate the current Gender biases that exist in the majority of African societies. To avoid this, African women & men in STEM/Tech fields must take the actions aforementioned:
Establish trainings that educate those in the tech field about the issue of gender bias within AI technologies
Conduct more studies on "bias mitigation algorithms", and the create said algorithms in a fashion that works within the African societies' context
Avoid creating and spreading stereotypical imagery across different platforms
Create divisions within tech companies that are assigned the task of monitoring and controlling gender
Supporting efforts that combat against negative gender biases the within the societal context
Africa has much potential, but as is shown in the data in section 2 shows, it is being held back by outdated views on the role of women in African societies. As Africa is headed towards a brighter future and booming industry and economy, African men and women must make the effort to build a new and positive perspective on women and count on them to make significant contributions to the advancement of Africa. Burkina Faso's late president and one of Africa's greatest leaders, President Thomas Sankara once stated, " The revolution and women's liberation go together; we do not talk of women's emancipation as an act of charity or because of a surge of human compassion; it is a basic necessity for the triumph of the revolution; women hold up the other half of the sky."

References

Jousse, L., & Vandermuntert, C. (2022, June 24). Discrimination and gender inequalities in Africa: What about equality between women and men?. Institut du Genre en Géopolitique. https://igg-geo.org/?p=3863&lang=en

Gupta, M., Parra, C. M., & Dennehy, D. (2022). Questioning Racial and Gender Bias in AI-based Recommendations: Do Espoused National Cultural Values Matter? Information Systems Frontiers: A Journal of Research and Innovation, 24(5), 1465–1481. https://doi-org.libproxy.udayton.edu/10.1007/s10796-021-10156-2

Microsoft. (n.d.). Bing. https://www.bing.com/visualsearch

O’Connor, S., Liu, H. Gender bias perpetuation and mitigation in AI technologies: challenges and opportunities. AI & Soc (2023).https://doi-org.libproxy.udayton.edu/10.1007/s00146-023-01675-4

South Africa, T. A. M. G. (2022). State of AI in Africa report. Gumroad. https://aiafricareport.gumroad.com/

"The revolution and women’s liberation go together; we do not talk of women’s emancipation as an act of charity or because of a surge of human compassion; it is a basic necessity for the triumph of the revolution; women hold up the other half of the sky.”

GENDER TECH INITIATIVE

Gender Biases In AI: Examining the ways in which AI systems can reinforce gender stereotypes that exist in African societies and exploring approaches to developing more equitable and inclusive AI technologies in Africa.

https://www.genderinitiativeug.org/
info@genderinitiativeug.org
+256 772 946 313