Epistemically violent biases in artificial intelligence design: the case of DALL-E 2 and Starry AI

Blessing Mbalaka
The Institute for Pan African Thought and Conversation,
University of Johannesburg, Auckland Park, South Africa

Abstract

Purpose – The paper aims to expand on the works well documented by Joy Boulamwini and Ruha Benjamin by expanding their critique to the African continent. The research aims to assess if algorithmic biases are prevalent in DALL-E 2 and Starry AI. The aim is to help inform better artificial intelligence (AI) systems for future use.

Design/methodology/approach – The paper utilised a desktop study for literature and gathered data from Open AI’s DALL-E 2 text-to-image generator and StarryAI text-to-image generator.

Findings – The DALL-E 2 significantly underperformed when it was tasked with generating images of “An African Family” as opposed to images of a “Family”. The pictures lacked any conceivable detail as compared to the latter of this comparison. The StarryAI significantly outperformed the DALL-E 2 and rendered visible faces. However, the accuracy of the culture portrayed was poor.

Research limitations/implications – Because of the chosen research approach, the research results may lack generalisability. Therefore, researchers are encouraged to test the proposed propositions further. The implications, however, are that more inclusion is warranted to help address the issue of cultural inaccuracies noted in a few of the paper’s experiments.

Practical implications – The paper is useful for advocates who advocate for algorithmic equality and fairness by highlighting evidence of the implications of systemic-induced algorithmic bias.

Social implications – The reduction in offensive racism and more socially appropriate AI can be a better product for commercialisation and general use. If AI is trained on diversity, it can lead to better applications in contemporary society.

Originality/value – The paper’s use of DALL-E 2 and Starry AI is an under-researched area, and future studies on this matter are welcome.

Keywords Epistemic violence, AI bias, Starry AI, DALL-E 2

Paper type Research paper

1. Introduction

Artificial intelligence (AI) has the potential to disrupt the fabric of contemporary society. AI, in essence, is key to transforming the guiding knowledge systems which may be responsible for monitoring human activity. However, the development and deployment of AI systems...
have also raised concerns about the potential for biases. These biases act as a mirror that reflects the innate biases of contemporary society. These biases can occur at various stages of the AI development process, from selecting data to training AI models to designing and implementing AI systems (Bruns, Burgess, Crawford, & Shaw, 2020). Since these AI models are rapidly becoming integral to the very nature of the status quo, it is vital to address these biases to prevent the amplification of existing inequalities.

One aspect that has contributed to the formation of bias is the under-representation of marginalised groups in the data used to train AI models (Buolamwini & Gebru, 2018). Boulamwini and Gebru found facial recognition systems that failed to identify the faces of African American Women. These scholars attested that this poor performance was a result of a lack of diversity within the data sets. The above-noted poor performance suggest that, in some instances, poorly trained AI models can fail to work accurately on racial and cultural minorities. AI models are growing in importance and are assuming roles as judges, human resource screeners and other activities which will shape the human–technology symbiosis (Sourdin, 2018). These concerns raised suggest that AI requires guard rails that work to mitigate these concerns.

Given these potential consequences of embedded AI biases, it is important to ensure that marginalised groups are included in the development and deployment of AI systems. This reasoning is corroborated by Gebru et al. (2021). These scholars advocated for the establishment of an accountability mechanism for AI. This framework was called the datasheet for data sets. The initiative proposed looked to advocate for transparency in AI algorithms. In addition to this imposition, the European Union has since drafted an AI Act to explore this issue of AI biases (Kop, 2021). The world is coming to the realisation that it is important to establish government bodies and informed councils in the area of AI that act to create policies, and one such is the EU AI Acts oversight role (Kop, 2021). This can involve efforts such as the inclusion of diverse data sets in the training of AI models. This paper, therefore, explores the issue in the context of Africa to explore the proliferation and implications of AI biases on African diversity.

AI systems, as per the interconnectedness of the Internet, are a globalised technology that will enter and be required to act within the global diversities. The role of AI systems, such as that of the recent CHAT GPT 3 and 4, has the immense potential to rapidly disrupt the status quo. Whilst image processing systems like StarryAI and DALL-E 2 will rapidly disrupt industries such as the art and content production industry (Oppenlaender, 2022). However, as per the reasoning of the paper thus far, it is important to express how these AI systems could potentially lead to the proliferation of worldly biases, biases adopted from the intended and unintended innate human biases of their creators. The dilemma surfaces when these biases lead to offensive and culturally negligent biases (Ricaurte, 2022).

Unless monitored, AI could mirror what Fanon [1961] (2004) coined as epistemic violence. Fanon expressed how Eurocentric beliefs were legitimised and justified by the colonially oppressed. The negative ramifications of this ideological imposition, according to Fanon, led to the dilution of the black identity and subsequently validated Eurocentric modernity as the ideal trajectory. These same expressions were shared by Boaventura de Sousa Santos who noted the conjecture of how Eurocentric knowledge production could lead to the destruction of knowledge systems (de Dos Santos, 2016). Vázquez (2011) attests to how this reasoning remains relevant in contemporary society. This reasoning is true in the case of AI, especially when unpacking how these AI systems could potentially dilute cultures if they are not sufficiently trained on global diversities.

This paper is thus concerned with exploring how AI systems can lead to the proliferation of worldly biases, which can, unless averted, lead to the dissemination of systems that commit the sin of ideological supremacy and cultural negation from its epistemic violence.

The first section is a brief outline of the emergence of AI. This section intends to briefly highlight the workings and emergence of AI. This section looks to explain how AI works and
expresses the significance of training data. The paper, then, delves into a more critical analysis that explores the notion of epistemic violence, the guiding framework for this study.

This research paper looked to contribute to this imperative academic dialogue to extrapolate the significance of looking into text-to-image generation AI. The paper conducted several experiments which explored the proneness of epistemic violence from two text-to-image generation tools. The results and discussion from these experiments highlight the urgent need to ensure that this particular branch of AI is monitored for algorithmic biases.

Despite this critique of these AI systems, the paper remains adamant that these epistemic dilemmas can be mitigated. The paper, thus, will have a section that proposes potential approaches that remedy the particular area. Such remedies include Gebru et al.’s (2021) algorithmic auditing and equitable culling of algorithms (Jackson, 2018).

This paper generated findings of images from DALL-E 2, which is a text-to-image AI platform developed by Open AI. The other program used in the study was StarryAI, an industry leader in text-to-image generation. The paper wanted to explore how the historical subjugation of Africans and African knowledge has affected AI text-to-image generation tools. The paper found that poor-quality images were rendered when the DALL-E 2 was used to generate images of an African Family; however, slightly better results were found when the family was white (Figure 7). The results that were used to inform the discussion used a statistical analysis of the mean and standard deviation of the images to generate a statistical analysis of the images. The subjective analysis of this image acted to verify or debunk these same findings.

The significantly better image quality from the StarryAI suggests that this specific algorithm has been sufficiently trained and prompted to represent minorities. However, the paper’s look into StarryAI was not infallible. The AI-generated quality image renders of Africans but when it came to representing the results, the paper found that StarryAI significantly underperformed when it was used to generate more culturally specific requests. For example, the study expressed how StarryAI was able to generate images of an African traditional healer, but when a more specific request of “A Zimbabwean traditional healer” was asked, the AI failed to create a more culturally specific render. This inaccuracy suggests that Zimbabwean representation was not well represented in the data set of their guiding algorithm. These results indicate that the marginalisation and representation dilemma should not be universalised to all AI systems because algorithmic infrastructure, which underpins the functionality of each AI system, is unique. These results suggest that the case of algorithmic biases and the subsequent epistemic violence noted in the literature varies in degree.

This poor performance can be used to call for a grander world-scale collaboration to create a participatory approach to AI knowledge production, an approach that acknowledges the need to consult the agglomeration of global diversities. Academia and the social sciences have been delving into context-specific research for centuries. These context-specific approaches to research need to be emphasised in AI research to help mitigate biases.

2. Unpacking epistemic violence

Epistemic violence refers to the ways in which certain knowledge and ways of knowing are privileged while others are marginalised or erased. The Tay chatbot’s lack of parameters and exposure to global prejudices led to the chatbot’s artificial epistemically violent ideological formation. The chatbot Tay also expressed sexist and racist sentiments after it was exposed and influenced by people with either a twisted sense of humour or who resemble those prejudiced beliefs (Davis, 2016; Neff, 2016). This case illustrates the dangers of AI that do not have guard rails or restrictive parameters, and it also indicates how these systems could be harmful to society. The following section looks to unpack how this can lead to epistemic violence.
According to Dotson (2011), epistemic violence can occur in many fields, and this is due to the silencing, subjugation and disregard for divergent thought and belief systems. This is corroborated by Teo (2010), as they also highlight how epistemic violence invalidates divergent thought. The following sections, thus, look to briefly express these arguments.

2.1 The suppression of viewpoints
One way in which epistemic violence can manifest is through the suppression of certain perspectives or viewpoints. This can occur when certain voices are not heard or are actively silenced, either within academia or, more broadly, in society. For example, in the field of psychology, the marginalised have faced a long history of pathologising and invalidating their experiences and knowledge. These groups have included people of colour and the LGBTQ+ community (Teo, 2011; Liu, 2022). This has led to a lack of representation and visibility for these groups within the field and has contributed to the perpetuation of harmful stereotypes and discrimination.

2.2 Academia and the consequence of under-representation
Another way in which epistemic violence can manifest is through the lack of representation of certain groups in research and academia. This can occur when certain groups are under-represented in the research process, either as study subjects or as researchers. For example, research on women's health has historically been lacking, leading to a lack of understanding of women's unique health needs and experiences (World Health Organisation, 1994). Similarly, people of colour and indigenous people are often under-represented in research and academia, leading to a lack of understanding of these groups' unique experiences and challenges.

2.3 The invalidation of the "other" knowledge systems
A third way in which epistemic violence can manifest is through the invalidation or dismissal of certain forms of knowledge or experience. This can occur when certain forms of knowledge or experience are not considered to be "legitimate" or "valid" within the dominant discourse. For example, indigenous knowledge systems and ways of knowing have often been dismissed or marginalised within mainstream academia, leading to a lack of understanding and recognition of the value of these systems (Brunner, 2021; Behl, 2019).

3. Applying the notion of epistemic violence to AI knowledge production
AI data sets are the result of research, and just as within academia, AI research is subject to researcher biases. The biases are the result of researcher biases or shortfalls within the AI algorithmic design (Lee et al., 2019). Researchers have attempted to resolve the biases mathematically (Lee et al., 2019). The issue with this approach is that it negates critical sociological considerations which need reasoning and are difficult to express without text comprehension and sufficient understanding of global world issues and cultures. This approach also negates inert researcher biases. Therefore, it is important to take several steps to address and mitigate the epistemic violence that can occur in AI knowledge production. One approach is to increase representation and visibility for marginalised groups within the AI research and development community and to actively seek out and amplify the voices and perspectives of these groups. Another approach could be to critically examine and challenge the dominant narratives and frameworks within AI research and development and to seek out diversity in the data sets. The issue, however, is that cultural representation and sources from the Global North have been subject to oppression and suppression due to colonial
ideological indoctrination, which Fanon (1952) avidly expresses. This, thus, means that addressing these issues comes with a decolonial agenda.

3.1 Fanonian logic and the dilemma of the colonised mind

Fanon (1952), in the *Black Skin, White Masks*, noted his perspective on how colonial legitimation was centred around the idea of the colonised mind. Fanon argued that the colonised attained a distorted sense of reality in which they were indoctrinated to internalise the values of the colonisers. Subsequently, some of the colonised began to see themselves as inferior and aspire to the coloniser’s way of life. The colonised were forced to reject their own history, culture and identity in order to assimilate into the coloniser’s way of life. This indoctrinated colonial legitimation was epistemically violent to the colonised. Based on this Fanonian rhetoric, the colonised saw a deletion of what they once deemed significant.

Therefore, a reflection on the epistemically violent colonial period can express how historically other forms of knowledge have been subjugated. However, the decolonial movement has led to attempts to exhume these lost knowledge systems. Despite these efforts, some debates do not act as guiding frameworks for AI language design. These knowledge systems need to be decolonised and argued for their relevance. These are ideas that include Oyewumi’s (1997) rejection of Western gender categories and her imposition of a gender-less pre-colonial Yoruba society. These unique perspectives and critical looks into contemporary society could be enhanced if the social sciences and the marginalised are allowed to narrate their cultures. This approach could help preserve the marginalised culture from inevitable time-bound distortions of beliefs.

It is important to ensure that AI encompasses researchers who come from these diverse cultures so that the algorithms can be informed by people who truly understand these cultures. However, as the paper expressed, it is important to be cognizant of the limitations of that approach. Fanon’s logic is important because it emphasises the historical determinants of contemporary thought. If AI systems are to be configured to avoid epistemic violence, it needs to incorporate ethnographic data attained from the consultation of cultural representatives. These representatives could include spiritual mediums and other cultural stakeholders in the AI data collection process.

3.2 The downside of using biased algorithms

According to Benjamin (2019a), AI algorithms and models encompass data sets that lack diversity. Benjamin (2019a) further notes how this neglect can lead to the proliferation of automated racism. As a result of this racism, Benjamin warns that this can lead to the perpetuation of existing inequalities. Fuchs (2018) corroborated this by alluding to how an AI-based judge was, in a particular instance, highly problematic. Fuchs’ reservations are not far-fetched because when one considers his account of how black people are subject to harsher sentences. This data on asymmetrical racial prosecutions was argued by Fuchs (2018) to express how the data could lead to biased training data. This, in essence, means that the AI was trained to express these prejudices. The AI could be trained to act in the same vein as what Hernandez (1990) called racially motivated prosecution biases.

This above-noted case of the racist AI judge is a useful anecdote that can showcase how worldly biases can creep into AI systems. The following section will build on that reasoning and will express how these biases can lead to epistemic violent cultural neglect.

3.3 Unpacking African under-representation in the academia

The lack of Africans in academia could potentially explain the bias-inducing issues of diversity and under-representation. Obeng-Odoom (2020) attempts to explain this under-
representation by alluding to economic reasons. Obeng-Odoom references how Africans divert away from academia in the pursuit of a steady income. This is something that this scholar attests to the historical oppression of Africans who need to work out of desperation and the need to establish wealth. Obeng-Odooms’ statement, upon closer scrutiny, could be unhinged if one interprets it as though academia is the pursuit of poverty. Obeng-Odooms’ argument is limited, but he does touch on a crucial point. Economic reasons are imperative, but they are relevant in the case of access to education. When parents salvage the funds to pay tuition, there is a high likelihood that these students will pursue professions such as accounting and medicine. Maths and computer science are something that only recently have become lucrative.

National Center for Education Statistics (2021) mentioned that only 5% of African Americans study for a bachelor’s degree in computer science, whilst that figure declined to 4% at the master’s degree level. This is a drastically low number when juxtaposed with the 22% in business, 18% in health professions and 16% in history and the social sciences. However, when one explores the same findings and looks at white representation as a control, there is a 2% difference between African Americans and whites. Statistics can be very good liars, and in the case of their study, the 2% of 10, as opposed to the 2% of 30, is clearly asymmetrical. The 2% difference does not consider the ratio of African to white students. The ratio was 1:3 between African and white students.

The population of black pupils, thus, signifies that the representation prevalent in the institution is far less than that of white representation. This is indicative of an obscure representation at universities. This point can be used to explain why systems can be biased. However, the issue of inclusion is far more complex than that because the issue is systemically rooted. The issue with systemic issues is that they can be difficult to resolve quickly and need gradual adaptive policies, which can take time. The drawback of this reality is that the rapidly changing technological climate might not wait for the representation to reach equilibrium. But there are ways which could improve the accuracy of these AI systems, and these methodologies will be discussed further as the paper unfolds.

4. Lack of Africans in STEM and the proliferation of algorithmic biases
Despite the numerous efforts to promote diversity in the field of science, technology, engineering and mathematics (STEM), African people are still under-represented in these fields. According to a report by the National Science Foundation, only 13.4% of African people in the United States earned bachelor’s degrees in STEM fields in 2017 (National Science Foundation, 2017). This lack of representation has serious consequences, as it limits the pool of talent available in these fields and reinforces societal inequalities.

5. Discussing algorithmic biases
Algorithmic biases are systematic patterns of discrimination that occur in machine learning systems as a result of biased data, biased algorithms or a combination of both. These biases can have serious consequences, including discrimination against certain groups of people and the amplification of existing social inequalities. In this section, we will explore the various ways in which algorithmic biases can emerge and consider some of the steps that can be taken to mitigate these biases.

One way in which algorithmic biases can emerge is through the use of biased data. Machine learning algorithms are trained on data sets, and if these data sets are not representative of the population or are otherwise biased, the resulting algorithms may also be biased. For example, if an algorithm is trained on a data set that is predominantly male, it may have a biased view of the world and may make decisions that are unfairly favourable to men.
Similarly, if an algorithm is trained on a data set that is predominantly white, it may have a biased view of the world and may make decisions that are unfairly favourable to white people. These sentiments were expressed by Wellner (2020), who expressed how AI could lead to gendered biases due to the lack of sufficient and inclusive training data sets.

A second way in which algorithmic biases can emerge is through the use of biased algorithms. Even if a machine learning algorithm is trained on a diverse and representative data set, it may still produce biased results if the algorithm is biased. This could be due to various reasons, for example, the algorithm may be designed to optimise for a particular metric, such as accuracy or speed, which can lead to biased results if the metric does not adequately capture the complexity of the problem being solved. Alternatively, the algorithm may be designed to make certain assumptions about the data, such as the assumption that all data points are independent of one another, which can lead to biased results if the assumption is not valid. For example, an AI is trained on images of a family, and only a nuclear family is used to represent the idea of a family; it will likely not include single-mother or father households.

6. Methodology and findings
The paper utilised a mixed methods approach to analyse the research findings. The paper analysed the images quantitatively by analysing the image noise. This was done by calculating the standard deviation and the mean of the images rendered. Noise indicates the dispersion of pixels, and usually, that is indicative of pixilation and less clustered images, which can be of less quality. The juxtaposition between the images of an African man and the white man was analysed through Python. The application was manually coded and built to analyse the region of interest (ROI). The ROI used the coordinates: x, y, width, height = 100,100,200,200. These coordinates were used to calculate the colour and the pixel dispersion between the images. This was achieved by calculating the standard deviation. The higher standard deviation, in comparison to an image(s), is indicative of noise, whilst the lower standard deviation indicates that there is less noise. However, there is a need to explore key variables such as brightness, contrast and saturation too when analysing the image for noise because these may influence the results drastically. This thus means that a visual inspection may be necessary to compensate for the shortcoming of this calculation. The standard deviation noted below encompasses three numbers in parentheses. These numbers (x,y,z) indicate red, green and blue (RGB). The lower standard deviation from each of these respective variables indicates a better quality than that of the higher. The mean below is a set of figures which also explores the RGB. However, this indicates the average colour of the pixels in the image. This data is useful to explore concurrently with the standard deviation because the low mean and high standard deviation value might indicate that an image is noisy or washed out, while an image with high mean and low standard deviation values might be considered oversaturated or lacking in contrast.

The closer the mean and standard deviation, the greater the quality. The rendering of the images above suggests that the image of the white man is better than that of the African man in quality (Table 1). The distinction is not easy to tell. However, the morphed eye of the white man could be contrary to this selection (Figures 1 and 2). This result, thus, emphasises the role of the expert subjective eye because the image quality is not a sufficient measure of accuracy.

The quantitative approach does not factor in aspects such as brightness, saturation and other conditions that can lead to noise. Therefore, a final human-determined verdict is required to assess if these results are of good quality or not. Furthermore, there is another limitation that needs to be considered. In the case of “An African Family 10”, the standard deviation was 80.823235339845, 85.5447135244176, 92.10687084216781 for the RGB elements whilst the mean was 144.34355, 135.804525, 132.966675 (Figure 12). This asymmetrical relationship of a
high mean and a low standard deviation suggests that the image has poor detail and contrast. The eye test confirms this for Figures 3–12. This phenomenon was present across the results of the African Family (Figures 3–12).

### Table 1. DALLE-2 renders by author

<table>
<thead>
<tr>
<th>Image</th>
<th>Standard deviation (R, G, B)</th>
<th>Mean (R, G, B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1 (Author, 2023)</td>
<td>(4.8683065659652645, 4.818732792863517, 4.917488418052812)</td>
<td>(233.5713, 233.409475, 233.307225)</td>
</tr>
<tr>
<td>Figure 2 (Author, 2023)</td>
<td>(69.42574574280277, 62.680061936106185, 63.774447359210996)</td>
<td>(64.4121, 51.0388, 51.691425)</td>
</tr>
</tbody>
</table>

The image lacks detail on the eyes.
The eye renders in the image are not symmetrical.

### 6.1 Experiment 3: images from Starry AI

Starry AI was used to assess its ability to generate quality images. Upon realising that the quality was superb, the study interrogated the program with two culturally specific prompts. The prompt was of “a Zimbabwean traditional healer”. The images rendered lacked cultural specificity and were generic (Figures 13–16) **Table 3.** These results can be noted below. The mean and the standard deviation were also included to help measure the noise and quality. The limitations of which have been noted above.
7. Culturally specific image renders
The results from the image portray a similar quality output to the images rendered by the DALL-E 2, however, the quality of the images from the eye test is more detailed and better in quality. However, the images fail to accurately portray the correct regalia worn by traditional Zimbabwean (Shona) healers (Figures 13–16).

7.1 Xhosa women attire
To further assess the AI’s ability to generate culturally specific images, it generated images of “Xhosa women attire”. The results below rendered were inaccurate. The mean and standard deviation were also analysed to assess the image quality. The results rendered by the DALL-E 2 does not portray accurate and culturally specific renders.

The results rendered are culturally incorrect and depict a generic African attire render (Figures 17–20). The image quality, as per the mean and standard deviation, suggests that the images are washed out and lack sufficient detail. These are problematic findings that indicate a type of hallucination from the AI. Parameters need to be put in place, which adds an accuracy score next to its rendered findings so that algorithmic auditors can monitor, assess and correct such instances.

7.2 Experiment 6: an African Family
The Starry AI was also used to generate images of an African Family, and the AI’s images generally look better than the Open AI variant (Table 2). However, the distortions of the image were less than the ones seen in the DALL-E 2. These images and their mean and standard deviation are illustrated below.

The distortions in white families might be due to a flaw in rendering images of multiple subjects by DALL-E 2. The inaccuracy in these renders in Figures 6-9 in Table 3 suggests that quality might not be the best measure of epistemic violence because similar inaccuracies are prevalent in Figures 10-19 in Table 4. The study has thus found that cultural accuracy within the renders might be the best approach to take in future studies. Another explanation for this could be deliberate sabotage by Open AI to prevent the AI from creating realistic AI images, utilisable for deepfakes.

To conclude, this section of the study looked to measure the proliferation of epistemic violence in text-to-image AI systems. The results were gathered by using Python to analyse the image standard deviation and the mean from the Open AI DALL-E 2 and the Starry AI programs. The results suggest that the images were very similar in the quality of the outputs. However, the study failed to register results which were culturally specific and accurate to the culture in which it was prompted for (Tables 5 and 6). Furthermore, the general sense from the mean and standard deviation indicated that these images were of very poor quality. Generally, the images rendered for the control (the white man) were closer to equilibrium for the mean and standard deviation (Figure 1). This near equilibrium was not achieved for any of the African images. The images of “the family” of the white families were visually better, but did also present distortions. Perhaps the poor performance is a failure of the AI to render multiple subjects. This mutual failure to render images of a white family means that image quality is not a determinant of epistemic violence because generally the images were similarly poor, but the best indicator is the cultural accuracy of the culturally specific prompts. Future studies can build on these lessons. However, the Starry AI’s results suggest that the quantitative analysis does not truly represent the accuracy of aspects such as facial features rendered by the AI. The calculation merely analyses quality but does not consider the visual representation. Starry AI had better results, but this was not expressed in numbers. The only drawback was the above-noted deficiency in cultural specificity.
<table>
<thead>
<tr>
<th>Image</th>
<th>Standard Deviation</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 3</td>
<td>(79.89774718951811, 77.34786447757288, 75.39199299270902)</td>
<td>128.8023, 81.46425, 70.74595</td>
</tr>
<tr>
<td>Figure 4 by author</td>
<td>(87.05271533763646, 75.81664166347784, 76.46798613472234)</td>
<td>(129.176725, 86.499025, 66.79365)</td>
</tr>
<tr>
<td>Figure 4 by author</td>
<td>(83.538585588694, 72.51677721923704, 73.07278193107526)</td>
<td>(139.03585, 108.0891, 84.5358)</td>
</tr>
</tbody>
</table>

The image is of good quality.

The image is of high quality, and the outputs are accurate.

The quality is decent and has sharp detail and colours. The lips are a bit blurry.

Table 2.
An African Family
Starry AI by Author
Image of White Family lacks in detail. Suggests that performance may not be racial.

Another poor performance suggests that the renders are not racial.

The image quality is poor.

Table 3.
Images of white families DALL-E 2 by Author
8. Discussion
This study aimed to investigate the occurrence of algorithmic biases and epistemic violence on African diversity, deconstruct the nature of this phenomenon and propose effective strategies for addressing it. In this section, we will delve into the reasons behind these issues, analyse their underlying structures and outline potential solutions. The discussion begins by highlighting instances of epistemic violence prevalent from the AI programs used in the study. The paper found that the African culture, in the context of the two AI text-to-image generators, has been homogenised into an incorrect depiction of the diversity of the African continent. This was notable in the AI-generated attire of a “traditional healer” and “A Zimbabwean traditional healer”, the outfits rendered are nearly identical (Figures 24-27). This implies that the regalia of traditional healers is synonymous with African regalia. This is not true because the African continent encompasses an agglomeration of diverse regalia, which differs from tribe to tribe. This error was repeated when the AI was prompted to generate images of a Xhosa woman’s attire. The AI hallucinated and generated results that did not accurately depict the Xhosa cultural regalia. These issues of cultural homogenisation and epistemic violence will be, thus, discussed below.

8.1 Homogenisation of African attire
The results from Starry AI and DALL-E 2 emphasises the need for AI regulatory bodies and standardised and inclusive approaches to AI design. This proposed regulatory approach can help to mitigate biases in AI models. The DALL-E 2 AI failed to generate quality images of African people but its alternative, Starry AI, generated images which were significantly better than the DALL-E 2 images (Table 2; Figures 3-5). This indicates that the algorithm used to train Starry AI could be the industry standard. However, upon closer inspection of the results of the regalia, which is meant to depict a Zimbabwean traditional healer, the regalia used by the AI can be seen to be incorrect (Table 6; Figures 24-27). The issue of epistemic violence manifests itself in these results because instead of expressing African diversity, African attire in the Starry AI is homogenised. There are complex nuances that are negated, which risks diluting the unique diversity prevalent within the African continent. It may be necessary for the algorithm to acknowledge when insufficient data is present so that it does not generate images that could inform the uninformed. The AI was prompted to also generate an image of an African Family (Table 2). This request is general and merely looked to explore...
The facial features of the images are poorly rendered for all subjects.

The faces are poorly rendered.

Table 4.
Images of “An African Family” from DALL-E 2 by Author (continued)
The faces are poorly rendered and lack detail.

Table 4. (continued)
Figure 14 by author

The detail on the faces is poorly rendered too.

Figure 15 by author

Body and facial features are present

(continued)

Table 4.

Algorithmic biases in DALLE-E 2 and Starry AI
Figure 16 by author
Poorly rendered facial features and bodies.

Figure 17 by author
The images are poor in detail.

Table 4. (continued)
how the quality is compared with the DALL-E 2 (Table 4). The results, aside from a defective mouth render (Figure 23), were significantly better than the DALL-E 2. The quality rendered suggests that the algorithm used by Starry AI is significantly better trained on African faces. However, it is important to note that African faces are an umbrella of complex nuances which embody African diversity. This can be true for European cultures too. It is thus imperative to ensure that there is the utilisation of diverse data sets that are representative of all humanity.

8.2 Why algorithmic biases and epistemic violence on African diversity occur
Algorithmic biases emerge as a result of the human decision-making process that shapes the development of AI and machine learning systems (Noble, 2018; Crawford, 2021). In particular, the lack of diversity among AI developers and researchers (Holstein, Wortman Vaughan, Daumé, Dudík, & Wallach, 2019), as well as the biased nature of the data sets used for training
<table>
<thead>
<tr>
<th>Image</th>
<th>Standard Deviation</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Figure 20" /></td>
<td>(70.36818509263405, 55.49149892119232, 53.90641956507716)</td>
<td>(94.132075, 57.51625, 53.859525)</td>
</tr>
<tr>
<td><img src="image2.png" alt="Figure 21" /></td>
<td>(74.64295375930017, 72.98319322763524, 75.0702356003688)</td>
<td>(163.012975, 162.4386, 163.06095)</td>
</tr>
</tbody>
</table>

**Table 5.**
Xhosa women attire generated by DALL-E 2 by Author

(continued)
these systems (Buolamwini & Gebru, 2018), have contributed to perpetuating stereotypes and perpetuating epistemic violence on African diversity. Additionally, the global domination of Eurocentric world views in the technology sector (Mavhunga, 2019) may lead to the under-representation or misrepresentation of African cultures, histories and epistemologies. However, it is important not to universalise these cases but to instead use a case-by-case approach that unpacks instances of bias-induced epistemic violence.

8.3 Deconstructing the nature of the phenomenon of AI biases

Understanding the nature of algorithmic biases and epistemic violence requires an examination of the socio-technical systems that produce and maintain these issues (Benjamin, 2019b; Eubanks, 2018). This involves a critical analysis of the role that power structures, colonial legacies and socio-economic factors play in shaping the technology landscape (Mbembe, 2017; Ndlovu-Gatsheni, 2018). It is crucial to recognise that technology is not neutral; it is embedded in the cultural, historical and political contexts that inform its development and deployment (Winner, 1980). A key aspect of this deconstruction is to investigate the relationship between the developers of AI systems and the people who are affected by them (Costanza-Choic, 2020). By acknowledging the unequal power dynamics at play, we can identify points of intervention to address biases and promote more equitable outcomes. Furthermore, analysing the ways in which algorithmic systems may reproduce...
The quality of the images is good. Regalia is hallucinated.

Table 6.
Zimbabwean traditional healer by Author

<table>
<thead>
<tr>
<th>Image</th>
<th>Standard Deviation</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 24 by author</td>
<td>(76.44784491860574, 78.6310486610074, 84.8848051635159)</td>
<td>(152.657525, 112.173875, 99.020325)</td>
</tr>
<tr>
<td>Zimbabwean Traditional Healer 2</td>
<td>(123.699625, 86.14125, 79.19455)</td>
<td>67.6134453293803 7, 74.4264666023503 6, 76.125154489118 8</td>
</tr>
</tbody>
</table>

(continued)
and amplify existing social inequalities is essential for understanding the broader implications of these technologies on African diversity (O’Neil, 2016).

8.4 Approaches to resolving the AI biases
Numerous approaches have been mentioned in the literature that looks to resolve the dilemma of bias in AI systems, an issue that this paper argues leads to epistemic violence seen in Tables 5 and 6. One such approach pertains to ensuring that researchers in AI systems are diverse (West, Whittaker, & Crawford, 2019). The other approaches interrogate the data sets

<table>
<thead>
<tr>
<th>Algorithmic biases in DALLE-E 2 and Starry AI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 6.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Figure 26 by author</th>
<th>Figure 27 by author</th>
</tr>
</thead>
<tbody>
<tr>
<td>The image quality is good, but the regalia is incorrect.</td>
<td>Quality image, but the regalia is incorrect.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Zimbabwean Traditional Healer 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(131.8583,99.485,83.1716)</td>
</tr>
<tr>
<td>(71.0541,3358372,91,72.002748154</td>
</tr>
<tr>
<td>9816,72.85097175084302)</td>
</tr>
</tbody>
</table>

| 119.287725,95.904425,67.67705 |
| (77.692749700691,77.14625738845082,70.82428118426358) |
to assess if the data set data is equitably representative of global diversities (Hajian, Bonchi, & Castillo, 2016). However, representative data sets are difficult to implement if certain epistemologies or ways of knowing are vilified by Eurocentric reasoning. Tuck and Yang (2012) corroborate this by stating that marginalised knowledge needs to be decolonised and included in the AI data set.

Based on the above-mentioned approaches, it can be thus delineated that algorithmic biases and epistemic violence on African diversity is a complex phenomenon that demands critical examination and targeted interventions. By adopting a multi-pronged approach that addresses the root causes of these issues and promotes equitable technology development, we can work towards a more inclusive and just AI landscape that respects and values African and global diversity. Especially, in the case of DALL-E 2, the poor-quality image renders were the most notable flaw, but the illustration of African diversity was also under-represented in the renders (Tables 2, 4, 5 and 6). The DALL-E 2 images looked like pixilated stock images that failed to generate basic human features in Table 4 (Figure 10-19). These two programs differ, but the hypothesis of how AI biases can lead to epistemic violence, in the context of this study, is arguably true because of the incorrect cultural representation (Tables 5 and 6).

There are numerous measures which can be used to prevent algorithmic biases. The following matrix in Table 7 looks to highlight some of the literary accounted approaches.

9. Conclusion
Diversity in AI research is crucial because of its irrefutable benefits, which include a wider range of perspectives and ideas to the table, which can lead to more innovative and creative solutions to problems. This is especially important in the field of AI, where new ideas and approaches are constantly needed to advance the discipline. The paper also noted that diversity could help to prevent bias in AI systems. The argument was that if the people developing and training AI systems are all from the same background, they may bring their own biases and assumptions to the process, which can result in biased AI systems. Therefore, by having a diverse team, it is more likely that these biases will be recognised and addressed.

Diversity in AI design and training processes is imperative because it contributes to the proliferation of AI systems which are just and fair. The lack of diversity can lead to the creation of AI, which is not suited to other unique contexts. However, as argued in the paper, the lack of underappreciation of African value systems, as per Fanon's reasoning, may be neglected due to a colonially induced legitimation by very Africans. This problematises the romanticism of representation as the ideal remedy. The issue of representation has a few limitations, such as the negation of the voices of those who may live in regressive societies and the overemphasis on higher-status leaders. The logic of this critique is centred on the notion that all societies have a hierarchy and that hierarchy can lead to under-representation of the marginalised views of those structurally subjugated by those very hierarchies.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data curation</td>
<td>Baker (2019) mentioned that this would be the identification of inaccuracies in the data set. It is a process that can be done manually by identifying instances of bias (Kleinberg, Ludwig, Mullainathan, &amp; Sunstein, 2018)</td>
</tr>
<tr>
<td>Data Augmentation/forecasting</td>
<td>Generating synthetic outputs based on the inputs (Zhang, Lemoine, &amp; Mitchell, 2018; Zhang, Li, &amp; Zhang, 2018). These outputs are rendered from evidence-informed AI predictions which simulate outputs based on the the training data</td>
</tr>
<tr>
<td>Cultured curation</td>
<td>Data is selected by culture experts or people of a particular culture</td>
</tr>
</tbody>
</table>

Table 7. AI bias mitigation matrix
Furthermore, the paper discussed how indigenous African values could be repackaged and re-legitimised through the practice of cultural commodification. This approach, of course, has limitations in the form of cultural appropriation and dilution of cultural beliefs, so it is important to ensure that cultural commodification remains true to its roots and is not diluted cheaply into losing its original symbolic and metaphysical meanings. The generic results rendered by DALL-E 2 and StarryAI illustrate this dilution of meaning, and it is imperative to ensure that there are guard rails put in place to circumvent this dilemma.

The profit incentive is a great motivator for many in this market-centred environment, but it needs to be done through the correct methodologies. These methodologies could benefit significantly from adopting a participatory approach. The apparent issue of cost may, of course, deter funding, but maybe social media could help to close that gap.

If AI is made knowledgeable of the varying belief systems, it can become better utilisable in various geopolitical contexts. This essentially would mean that AI applications and products would be marketable to other groups within society. If this is achieved, then the business sector could potentially endorse further research in diversity in the AI space because the knowledge will begin to have a sense of marketable value. The downside of this approach is that these market actors may potentially mistranslate or unintentionally or intentionally offend diverse groups, all in the pursuit of profit.

Aside from the profit incentive, it is essential to remain cognizant of the need to create applications which are culturally offensive. So it is perhaps best to depoliticise AI into not having any human-like guiding beliefs but to, instead, work proficiently within the plethora of global diversities and knowledge systems. However, one could interpret such an approach as being censorship, but AI should not be programmed to have an artificial guiding belief system or political position. It needs to be framed and instilled with guard rails surrounding topics that can be offensive to some.

In the case of DALL-E 2 and StarryAI, the failure to sufficiently represent the historically subjugated in its image render is indicative of this lack of guard rails. The subsequent culturally offensive renders show how a lack of training in subjugated knowledge can lead to pseudo-representation and overgeneralised conceptions of African cultures.

To refrain from being a doomsday naysayer, the paper identified technological solutions, such as adversarial debiasing, counterfactual fairness and disparate impact remover. These technologies need to be made an industry standard to help mitigate the issue of AI algorithmic biases.

The AI systems are in their infancy and still require development, but that development should be cognizant of marginalised belief systems and ways of knowing to ensure that the systems work within the plethora of global diversities.

9.1 Areas for future research

(1) Gathering sufficient training data can be problematic, but the use of computational social sciences could be used to gather and render culturally specific data from social media platforms such as Twitter. This can be done through Python and its respective API (Application Programming Interface).

(2) Another proposition could be that the AI systems could become personalised and trained by their users. Even within the generalised and universalised conceptions of diversity, there exists splintered and subjugated thought which may not be exhumed by the training algorithm.

The results generated from the paper’s use of DALL-E 2 indicate that the systemic factors of under-representation of the marginalised might be depicted in the results of such AI systems. However, maybe future studies can explore if AI forecasting can be used to try and
compensate for the insufficient diverse data sets. If solutions are not advocated for and implemented, then the AI-driven future may be riddled with inaccuracies, inaccuracies which could be hazardous for the general public.

References


Further reading


About the author
Blessing Mbalaka is a junior researcher and emerging scholar at the University of Johannesburg. He attained his honours degree in 2021 and has since been actively involved in activism and journalism. He is an avid scholar and a firm lover of the Sicilian defence chess opening. Blessing Mbalaka can be contacted at: bjmbalaka@gmail.com