Artificially Generated Minorities (AGMs): The Veneer of Algorithmic Bias Correction

Ibifuro Robert Jaja[1[0000-0003-1101-8591]]

Centre for Philosophy and Artificial Intelligence Research (PAIR), Friedrich-Alexander-Universität Erlangen-Nürnberg, Nürnberger Str. 74, 91052, Erlangen, Germany
Department of Philosophy, Rivers State University, Rivers State, Nigeria

Ibifuro.jaja@fau.de

Abstract. Algorithms often reinforce societal biases and stereotypes. This is especially concerning for minorities, who are disproportionately impacted by it, thereby threatening their further marginalization. Data fundamentalists frame this issue of algorithmic bias as stemming from data bias, indicated by the underrepresentation of some groups (minorities) in the datasets. Consequently, measures adopted to address algorithmic bias have been data-focused. A relatively recent data-focused measure adopted to address this issue is the deployment of what I term artificially generated minorities (AGMs)—synthetic data used to increase the representation of underrepresented groups (minorities) in algorithms' training datasets. Data fundamentalists make two central claims about AGMs, which I term the representation claim, which holds that AGMs are representative of minorities, and the normative intervention claim, which holds that the deployment of AGMs addresses algorithmic bias. In this paper, I argue that AGMs do not meet these claims, particularly in the context of algorithmic recruitment. First, I demonstrate that AGMs do not capture the experience of historic and systemic oppression, which defines minority status. Hence, I contend that they do not meaningfully represent minorities. Second, I demonstrate that while AGMs facilitate the realization of the futuristic component of an adequate normative intervention, they undermine the reparative component. Thus, I contend that AGMs do not adequately address algorithmic bias. Finally, I briefly highlight that the failure of AGMs to meet these claims indicates that a data-focused framing of algorithmic bias is overly simplistic and does not account for all the complexities involved in the issue of algorithmic bias and its correction, particularly in the context of algorithmic recruitment.

Keywords: Algorithmic Bias, Artificially Generated Minorities (AGMs), Data Fundamentalism, Representation Claim, Normative Intervention Claim.

1 Introduction

Algorithms are known to replicate and reinforce existing societal biases and stereotypes. This is especially concerning for minorities, who disproportionately bear the brunt of algorithmic bias [1]. To exemplify, the now-disbanded Amazon hiring tool was found to be biased against women in its recommendation of candidates for technical roles. Additionally, an investigation into the correctional offender management profiling for alternative sanctions (COMPAS) revealed that the software assigns higher risk scores to Black offenders compared to white offenders with similar profiles [2]. Moreso, facial recognition systems tend to perform poorly on darker-skinned people. As these systems become mainstream, there is the risk of further marginalization of minorities.

Data fundamentalists frame this issue of algorithmic bias as stemming from data bias, indicated by the underrepresentation of some groups (minorities) in datasets [3, 4, 5].[1] Following this popular framing, most measures proposed and implemented to resolve the issue of algorithmic bias have focused on data, in an attempt to circumvent its unrepresentativeness [60]. Buolamwini and Gebru's [5] study showed that facial recognition systems perform poorly on individuals with darker skin tones due to their underrepresentation in the datasets used for training the systems. In their study, they oversampled individuals with darker skin tones in the datasets and reported that this led to improved model performance. The results of this study are used to support data fundamentalism [26].[2]

A relatively recent data-focused approach adopted to address algorithmic bias is the deployment of what I term artificially generated minorities (AGMs). By AGMs, I mean synthetic data generated through computational processes deployed to increase the representation of underrepresented (minority) groups in algorithms' training datasets [6,7,8,9]. Data fundamentalists claim that the deployment of AGMs improves the representation of minority groups in datasets and, consequently, addresses the issue of algorithmic bias that disproportionately affects minorities [10,11].

This paper aims to show that AGMs do not meet this claim, particularly in the context of algorithmic recruitment. To this end, I analyze the claim, which I split into two parts: the representation claim, which holds that AGMs are representative of minorities, and the normative intervention claim, which builds on the representation claim and, thus, holds that AGMs address algorithmic bias. Concerning the representation claim, I begin by demonstrating that AGMs embody a superficial understanding of what con-

---

1 The term data fundamentalist was coined by Crawford, Miltner & Gray [41] to describe scholars who are of the view that sufficient big data is the solution to algorithmic bias.

2 As a reviewer helpfully pointed out, Buolamwini and Gebru emphasize additional remedies for algorithmic bias, such as mechanisms for consent and structural redress. However, as I note, the findings of their study have often been taken up as evidence of the importance of balanced datasets in addressing algorithmic bias.

stitutes minority status, mainly reducing minority status to physical traits often associated with minorities. In contrast, I show that it is the experience of historic and systemic oppression that constitutes minority status. Thus, I assert that AGMs are not meaningfully representative of minorities. Regarding the normative intervention claim, I begin by stressing that an adequate normative intervention must include a futuristic component that focuses on preventing algorithms from replicating biases against minority groups and a reparative component that focuses on enabling algorithms to redress historical and systemic biases and injustices that minorities endure. I then show that while AGMs enable the futuristic component, they undermine the reparative component. This, as I explain, results in the insidious reinforcement of algorithmic bias and the attendant inequalities that disproportionately affect minorities. As I point out, AGMs create the illusion of equal footing as they fail to capture the experience of historic and systemic oppression that shapes minorities and sets them apart from dominant groups. Consequently, algorithms trained with AGMs will misinterpret disparities exhibited by minorities as personal shortcomings rather than as outcomes of systemic oppression needing redress. I, therefore, assert that AGMs do not adequately address algorithmic bias.

Furthermore, I briefly highlight that the failure of AGMs to meet these claims indicates that a data-focused framing of algorithmic bias is overly simplistic and does not account for all the complexities involved in algorithmic bias. I contend that there is a need for a more robust framing of algorithmic bias that accounts for all the complexities involved in the issue. However, I do not make any specific suggestions for the sake of scope.

This paper will be divided as follows: In §2, I explore what AGMs are in greater detail. In §3, I analyze the representation claim. In §4, I analyze the normative intervention claim. In §5, I make further comments about the implications of my argument. In §6, I conclude.

2 Artificially Generated Minorities (AGMs)

2.1 What are AGMs?

Data fundamentalists view unrepresentative data as the cause of bias and stereotypes in AI systems [12,13,14,15]. In datasets drawn from real-world contexts, privileged social groups are overrepresented, while minority groups, such as women and racial or ethnic minorities, are often underrepresented [16,1,17]. According to data fundamentalists, when this skewed dataset is used to train algorithms, the systems replicate patterns in the dataset, overfitting to dominant groups, thus reproducing bias and stereotypes reflected in the data [14,18]. For example, the bias that the now-disbanded Amazon hiring tool exhibited towards women was diagnosed as stemming from the unrepresentativeness of data, specifically the underrepresentation of women in the training datasets used to train the system.[3]

AGMs are deployed to circumvent this unrepresentativeness of natural datasets. They are artificially generated datasets or synthetic datasets created using methods such as generative models, agent-based modeling, neural networks, and the synthetic minority oversampling Technique (SMOTE)[4] and deployed to improve the representation of minorities underrepresented in datasets to make datasets more representative and balanced [19, 20]

AGMs are deployed to make datasets representative in two ways. First, through data augmentation. AGMs are used to augment natural datasets where datapoints of minorities are scarce [21,22]. The aim here is to rebalance training datasets without necessarily discarding existing natural data. For instance, in recruitment, artificially generated resumes have been used to improve the representation of gender and racial minorities in natural datasets. Second, AGMs are used to replace natural data. This involves the entire substitution of natural datasets with artificially generated data. While not yet a popular practice, this is seemingly a futuristic move in AI development. Sam Altman, OpenAI's CEO, has predicted that "soon all data will be synthetic," suggesting that artificially generated data will eventually become the primary source for training future models.[5] Besides enhancing the representativeness of datasets, this shift toward the replacement of natural datasets with artificially generated data is motivated by concerns over the privacy risks associated with the collection and processing of natural datasets [23, 24].

2.2 AGMs: The claims

There are two key claims made about AGMs, which I refer to as: (1) the representation claim and (2) the normative intervention claim. I explain them in subsequent paragraphs.

The representation claim holds that AGMs are representative of the real-world minority groups. This claim underpins the practice of deploying AGMs in datasets to increase the representation of minorities. For instance, in recruitment, where minorities, particularly women, are often underrepresented in technical and leadership roles, artificially generated profiles of women (AGMs) are deployed to improve the representation of women in the datasets. Implicit in this practice is the assumption that AGMs are representative of minorities. More explicitly, AI developers and deployers of AGMs have committed to the representation claim.[6] To exemplify, Kjell Carlsson, Vice President and analyst at Gartner, specializing in AI, data science, and advanced analytics, voiced this claim as he says:

[3] See [55]
[4] See [42, 22, 43, 44] for further discussions.
[5] See [56]
[6] See [45, 46, 47]

> You can create synthetic, additional versions with variations of underrepresented groups in your data. In my clinical trial, I don't have enough people who are a certain ethnicity, age, or gender. Increasing representation with sufficient variety rebalances the dataset. I can create synthetic versions of these individuals with additional variations around them.[7]

The assertions by AGM deployers, as well as the practice of deploying AGMs to improve the representation of minorities in datasets, purport the idea that AGMs are representative of minorities.

The normative intervention claim holds that AGMs address algorithmic bias. This claim builds on the representation claim. It thus holds that by making datasets more representative or balanced, AGMs address the issue of algorithmic bias [21,20]. In other words, the deployment of AGMs is viewed as a normative intervention in the issue of algorithmic bias. Several AI researchers and developers/deployers commit to the normative intervention claim. For instance, Lee, Jaiswal, and Byrne [11] declare that the use of AGMs mitigates bias in hiring models. In their study, they deployed AGMs—artificially generated data to represent gender minorities. After retraining the model with the datasets supplemented with AGMs, they reported hiring rates of 61.54% for males and 58.82% for females, with an adverse impact ratio of 95.59%, indicating no statistically significant gender bias. Based on this result, they concluded that artificially generated data could be effectively deployed across real-world contexts to address bias. Upwage, a recruitment firm, drew a similar conclusion after deploying AGMs—artificially generated data to tackle gender bias in its hiring algorithm. First, they generated artificial profiles of women to supplement the datasets of roles where women were underrepresented. These datasets, augmented with AGMs, were then used to retrain the algorithm. Following this measure, the company boasted of having resolved the issue of algorithmic bias in hiring.[8]

In this section, I have explored what AGMs are, their purposes, and the various claims made about them. In the subsequent sections, I will turn to analyze these claims.

3 Are AGMs Representative of Minorities?

As discussed in the previous section, there is an implicit and explicit claim that AGMs are representative of minorities. Thus, the question posed as the title of this section is a probe into the representation claim. This question, however, is a layered one, and as such, I will unpack it in steps.

---

[7] See [57]

[8] See [58]

3.1 Defining Minority Status: Between What AGMs Purport and What It Is

The understanding of what constitutes minority status that AGMs purport is primarily physical. In other words, AGMs embody the understanding that having certain physical traits, such as being Black, female, etc., is what makes one a minority and distinguishes them from the dominant group [25]. This understanding of what constitutes a minority status is evident in the way AGMs are generated. Typically, AGMs are constructed by modifying existing data points to reflect physical traits such as being a woman or a person of color, etc., associated with minorities [6, 26].[9] To exemplify, Datagen, a synthetic data company, provides a list of characteristics, such as age, gender, and ethnicity, which its patrons can select from to generate artificial data [25]. In this way, a patron of Datagen could simply select "female," and artificial profiles with physical female traits (AGMs) would be generated. Upwage, a recruitment firm, adopted a similar method. The firm generated AGMs simply by mirroring the qualifications in the natural datasets and tweaking traits like gender and color. Let's say from the resume of a real candidate—Robert Paul, an AGM—Gloria Robert with the same qualifications but a different gender is generated. This process of generating AGMs suggests that AGMs embody the understanding that minority status is defined by physical attributes associated with minorities.

I, however, argue that what constitutes minority status is the experience of historical and systemic oppression. It is the lived experience of exclusion and marginalization that minorities have in common, and that which distinguishes them from dominant groups [27, 28]. While physical attributes such as color or gender may be used to identify minorities, they are not, in themselves, the defining or distinguishing feature of minority groups [29, 28]. To illustrate, White women and Black men share no physical traits, such as color and sex, in common, yet both demographics are considered minorities. Conversely, a black man and a white man share the attribute of being male, but only the former holds minority status. The first instance demonstrates that individuals can share minority status without sharing physical traits, while the second shows that individuals can share physical traits without necessarily sharing minority status. It is gender-based oppression that makes white women minorities, and racialized oppression that makes Black men minorities. For Black women, it is both gender and race-based, leading scholars to describe Black women as double minorities.[10] This points to the fact that physical traits are not the defining feature of minority status, but rather the experience of systemic oppression. Put differently, gender alone does not make women minorities; it is gendered oppression that does. Likewise, race alone does not make Black people minorities; it is racialized oppression that does.[11]

---

[9] As a reviewer helpfully observes, the creation of AGMs includes the modification of non-physical features, such as names or addresses. However, I focus on cases where it involves the modification of physical features for the sake of my argument.

[10] See [48, 49, 50] for more detailed discussions on the notion of double minority.

[11] I thank a reviewer for urging me to clarify this point.

3.2 AGMS and the Illusion of Representation

AGMs are not meaningfully representative of minorities. While AGMs exhibit physical traits typically associated with minorities, they are bereft of the experience of historical and systemic oppression, which constitutes minority status. Jacobsen [25] argues in these lines, as he notes that artificially generated data embodies a claim to representativeness that is ultimately shallow. He characterizes AGMs in light of Zizek's notion of "the Other deprived of Otherness," positing that the synthetic subject mimics the superficial features of an identity but is stripped of the substantive complexity that constitutes the identity. With AGMs, the features mimicked are being female or Black, etc., while what is stripped off is the lived experience of oppression [30].

I argue that AGMs are, in fact, more akin to privileged groups. They are not just the Other deprived of their Otherness, as Zizek and Jacobsen describe. They are an intentional substitution of the Other with their very opposite. This is so because they embody experiences of dominant groups marked by inclusion and access [31]. To illustrate, Upwage, the recruitment firm, embedded the AGMs it generated and deployed with qualifications and experiences that signal inclusion and access, typical of dominant groups. The outcome of such a process is a figure who, despite exhibiting physical traits associated with minorities, is more representative of privileged groups. This becomes clear when we recall that minority status is defined by the experience of systemic oppression, and groups that experience inclusion and access are designated as privileged groups. Therefore, AGMs, which do not embody the experience of minorities but rather that of dominant groups, despite exhibiting physical traits associated with minorities, align more closely with privileged groups.

Nevertheless, inadequate representation has been a point of critique of data representation. Scholars have long argued that data (natural datasets) fail to fully capture those they represent, as they flatten or oversimplify the complexities of individuals' lives.[12] This raises an important question: if both natural datasets and AGMs fall short in representing their subject of representation, what, then, sets AGMs apart in this regard?

I argue that AGMs add another layer to the issue of inadequate representation through data. While natural datasets oversimplify, they are collected from the real world and reflect some real-world contexts. AGMs, on the other hand, are not based on actual experiences and reflect an idealized version of the world. To illustrate this distinction, let us consider datasets in recruitment. Women are often underrepresented in natural datasets for managerial roles. These datasets oversimplify the experience of women by capturing the outcome and not the factors that contribute to these outcomes. The outcome captured by the datasets here is women's absence in managerial roles, and the contributing factor that is not explicitly captured is the systemic exclusion that keeps women from those roles. Nonetheless, the scarcity of women in the natural dataset in managerial roles signals the lived social reality of systemic exclusion that women endure. However, in datasets of managerial roles supplemented or replaced with AGMs, women will be well represented. These AGMs create an idealized world where women amply occupy managerial roles, effectively obscuring the systemic oppression that

[12] See [51, 52, 53, 41]

women endure in reality. Consequently, this creates the illusion of equal footing—the faux appearance or narrative that all demographics have equal opportunities. Thus, while both natural datasets and AGMs do not adequately represent their subject of representation, AGMs more radically depart from their subject of representation.

In this section, I have argued that AGMs are not meaningfully representative of minorities as they do not embody the lived experience of oppression, which I contend constitutes minority status. As I will show in the subsequent section, this has significant implications.

4 Are AGMs tools for normative intervention?

In this section, I will analyze the claim that AGMs address the issue of algorithmic bias and the attendant social inequality that disproportionately affects minorities.

4.1 Normative Intervention: The Holistic Approach

The normative intervention claim is valid only if AGMs holistically address algorithmic bias. A holistic approach to addressing algorithmic bias must involve two key components: the futuristic component, which is about preventing algorithms from replicating bias, and the reparative component, which is about acknowledging and redressing historic and systemic injustices that minorities endure [32]. These components are complementary and necessary to ensure that algorithmic bias is adequately addressed. Fulfilling the futuristic component while neglecting the reparative component risks entrenching the existing disadvantages that minorities face (This will be made clear shortly). While reparative measures can certainly be implemented by humans, algorithmic reparations are important given the mainstream deployment of algorithms in bottom-up processes. In recruitment, for instance, applicant tracking systems (ATS) filter out candidates who do not meet predetermined standards before any human review takes place, and in some cases, algorithms are even used to conduct preliminary interviews. If reparative considerations are not built into these stages of the process carried out by algorithms, large numbers of candidates from minority groups risk being excluded by the algorithm before they ever reach a human decision-maker.[13] Therefore, I argue that the claim that AGMs address the issue of algorithmic bias is valid only if the deployment of AGMs prevents algorithms from replicating existing biases and also enables algorithmic redress of historic and systemic injustices. To determine the validity of the normative intervention claim, in what follows, I will assess whether and to what extent AGMs facilitate the realization of the futuristic and reparative components.

[13] I thank a reviewer for urging me to expand on this point. See [32] for further discussions.

4.2 AGMs and the Futuristic Component

AGMs facilitate the algorithms' realization of the futuristic component. That is to say, they prevent algorithms from replicating biases. To support my claim here, I will first briefly explain how algorithms replicate societal bias and then show how the deployment of AGMs prevents algorithms from replicating bias.

Algorithms tend to replicate the bias reflected by the natural datasets that they are trained with. This is because algorithms, particularly machine learning systems, make predictions based on the training datasets. First, algorithms detect patterns in the datasets used for their training, then replicate those patterns in their predictions. Algorithms trained on natural datasets where privileged groups are typically overrepresented and minorities underrepresented tend to prioritize members of privileged groups in their recommendations. To put it differently, algorithms overfit to the characteristics of the overrepresented (privileged) group [5, 16]. Essentially, algorithms misinterpret the underrepresentation of minorities in the dataset as a signal of their unfitness or unsuitability for those roles. The system thus tends to favor members of privileged groups in their recommendations, replicating or even amplifying existing societal biases reflected in the data.

AGMs prevent algorithms from replicating bias. AGMs are deployed to balance datasets that are originally skewed toward privileged groups. In this way, they introduce diversity into the training datasets. This allows algorithms to associate competence with a wider range of groups, including minority groups. In effect, this makes the algorithm less likely to prioritize any groups in its recommendations.

To illustrate this, I will use the popular case of the now-defunct Amazon hiring algorithm. Amazon's hiring algorithm was trained on historical resume data collected over a ten-year period. Since technical roles have been historically male-dominated at Amazon, the majority of resumes in the training set came from men. Resultantly, the algorithm learnt to associate competence for technical roles with being male or male-dominant markers. This, in turn, led the system to prioritize male candidates in its recommendations for technical roles. The deployment of AGMs could improve the algorithm's performance. AGMs—synthetic profiles of female candidates with qualifications and experience in technical roles could be generated and deployed to balance the training datasets. In this way, the algorithm would be exposed to more diversity and learnt to also associate competence in technical roles with women. This, in turn, would prevent the algorithm from replicating bias against women in technical roles.

Additionally, both the study by Lee et al. [11] and Upwage's deployment of AGMs, detailed in section 2.2, highlight that AGMs can mitigate the replication of biases by algorithms.[14]

In sum, AGMs provide a safeguard against algorithms' replication of biases against minority groups. By doing so, they ensure that competent members are not unfairly disqualified by algorithms due to their identity. Let us now turn to assess whether AGMs facilitate the realization of the reparative component.

---

[14] See section 2.2 for detailed discussions of the study by Lee et al. [11] and real-world implementation of AGMs by Upwage.

4.3 AGMs and the Reparative Component

Reparations are necessary to disrupt the vicious circle of bias and inequality [33]. Without deliberate, effective reparation, bias and the attendant inequality resurface in new forms (This will become clearer shortly). Calls for reparations are grounded in the well-established understanding that historical and systemic oppression have ripple effects. Put simply, demands for reparations are justified by the widely recognized fact that exclusion in one sphere can hinder access and inclusion in others, thereby compounding disadvantage. For example, women's exclusion from science, technology, engineering, and mathematics (STEM) courses hinders their employment in technical roles.[15] Historic and systemic oppression thus provides the evidentiary basis for demands for reparations [34, 35].

Absent such evidence, the disparities between minorities and dominant groups may be construed as neutral outcomes, rather than the lingering effects of historic and systemic oppression. For example, if there is no evidence or recognition of women's exclusion from technical roles or the pathways leading to those roles (e.g., admission to STEM courses), their underrepresentation in technical roles will be interpreted as a neutral outcome rather than the product of systemic barriers, requiring redress.

AGMs undermine the capacity of algorithms to meet the reparative component. As I discussed previously, AGMs obscure evidence of historic and systemic oppression that minorities endure and create the illusion of equal footing.[16] Given this, algorithms will strictly interpret the disparity that members of minority groups may exhibit as a personal shortcoming rather than the outcome of historic and systemic oppression needing redress.

To illustrate this, consider a hiring algorithm trained with AGMs. The AGMs deployed, let's say, are synthetic profiles of women with exemplary managerial credentials, including prestigious educational qualifications and extensive managerial experience. The deployment of these AGMs will create the illusion of equal footing, from which the algorithm bases its recommendations. Consequently, any female candidate with a less-than-perfect resume, including gaps in employment or a less prestigious educational background, possibly due to historical and systemic oppression, will be classified by the algorithm as particularly unqualified. This female candidate will therefore not be recommended for the role.

Consequently, the deployment of AGMs further entrenches algorithmic bias and the attendant inequality insidiously. AGMs shift the exclusion of minorities as a result of group-based bias to their exclusion at an individualized level [36]. As can be gleaned from the illustration, disparities exhibited by minorities, stemming possibly from systemic barriers, will be interpreted by AI systems as signals that the individual (member of a minority group) is simply not good enough. In this way, bias and inequalities that disproportionately affect minorities remain, but are masked as neutral decisions about individual suitability [37, 32, 38, 39, 40]. This appearance of neutrality that AGMs portray may make the bias and inequality that minorities endure much harder to challenge.

[15] See [59, 54]

[16] See section 3.2 for detailed discussions.

Ultimately, the normative intervention claim is not valid. This is so because AGMs do not adequately address the issue of algorithmic bias. While AGMs facilitate algorithms' realization of the futuristic component, they compromise the reparative component. As is gleanable from our discussion, without the reparative component, algorithmic bias and the attendant inequality persist insidiously.

5 Further Comments

So far, I have argued that AGMs do not live up to the claims made about them, particularly in the context of algorithmic recruitment. As I showed, AGMs do not meaningfully represent minorities and also do not adequately address algorithmic bias, particularly in the context of algorithmic recruitment. This is so because they do not embody the experience of systemic oppression that constitutes minority status. Consequently, this undermines the realization of the reparative component of an adequate normative intervention.

Notably, there is a connection between how well minorities are represented and the possibility of satisfying the reparative component of a normative intervention. Representations of minorities that are bereft of the experience of historic and systemic oppression that defines minority status compromise algorithmic reparation and vice versa; representations of minorities that reflect these experiences can enable it. This is simply because the experience of historic and systemic oppression is the basis for algorithmic reparation.

Interestingly, the so-called bias in the data that Data fundamentalists seek to correct is evidence of systemic oppression that constitutes minority status and also the basis for algorithmic reparations. As we see with AGMs, artificially increasing the representation of minorities obscures the experience of historic and systemic oppression that constitutes minority status, and also grounds algorithmic reparations.

This signals a limitation in the framing of algorithmic bias put forward by Data fundamentalists. Their framing of algorithmic bias as an issue of data bias, indicated by the unrepresentative data, is overly simplistic and does not account for all the complexities involved. The underrepresentation of minorities in datasets is not merely a statistical anomaly. It is often the evidence of historic and systemic oppression that constitutes minority status and grounds algorithmic reparations. To treat this underrepresentation as merely a statistical anomaly to be corrected through means like AGMs overlooks its origins and significance. It thus follows that Data fundamentalists either disregard this reparative component of addressing algorithmic bias or underestimate the implications of their data-focused framing of algorithmic bias.

There is thus a need for a more robust framing of algorithmic bias that accounts for all complexities involved in the issue. Such a framing may require decentering data itself, or at the very least, resisting the tendency to view and treat unbalanced data as a mere statistical anomaly to be artificially corrected.

6 Conclusion

In this paper, I have argued and demonstrated that AGMs fall short in meeting the claims made about them, particularly in the context of algorithmic recruitment.

First, I analyzed the representation claim that holds that AGMs represent minority groups. I contended that AGMs do not represent minorities meaningfully. This is so because, as I showed, AGMs do not capture the experience of historic and systemic oppression, which I argue constitutes minority status.

Second, I analyzed the normative intervention claim that holds that AGMs address algorithmic bias, and the attendant inequality that disproportionately affects members of minority groups. I argued that AGMs do not adequately address algorithmic bias, particularly in the context of recruitment. I first asserted that any meaningful normative intervention must include two components: a futuristic component that ensures that algorithms do not replicate biases against minority groups, and a reparative component that ensures that algorithms recognize and redress historical and systemic injustices that minorities endure. However, as I demonstrated, while AGMs enable algorithms to meet the futuristic component, they compromise the reparative component. This, in turn, leads to the entrenchment of biases, maintaining the status quo.

Next, I turned to offer further reflections. I highlighted that Data fundamentalists' data-focused framing of algorithmic bias is overly simplistic. As I pointed out, the underrepresentation of minorities that Data fundamentalists view merely as a statistical anomaly that is to be corrected is evidence of systemic oppression that constitutes minority status and grounds algorithmic reparation. In light of this, I suggest that there is a need for a more robust framing of algorithmic bias that accounts for all the complexities. However, I do not make any suggestions for the sake of scope.

References

1. Marinucci, L., Mazzuca, C., Gangemi, A.: Exposing implicit biases and stereotypes in human and artificial intelligence: State of the art and challenges with a focus on gender. AI & Society **38**, 747–761 (2023). https://doi.org/10.1007/s00146-022-01474-3.

2. Dressel, J., Farid, H.: The accuracy, fairness, and limits of predicting recidivism. Science Advances **4**(1), 1-5 (2018). https://doi.org/10.1126/sciadv.aao5580.

3. Wang, A., Ramaswamy, V., Russakovsky, O.: Towards intersectionality in machine learning: Including more identities, handling underrepresentation, and performing evaluation. In: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT'22), pp. 1–14 (2022). https://doi.org/10.1145/3531146.3533101

4. Peng, A., Nushi, B., Kıcıman, E., Inkpen, K., Suri, S., Kamar, E.: What you see is what you get? The impact of representation criteria on human bias in hiring. In: Proceedings of the AAAI Conference on Human Computation and Crowdsourcing **7**(1), 125–134 (2019). https://ojs.aaai.org/index.php/HCOMP/article/view/5281/5133

5. Buolamwini, J., Gebru, T.: Gender Shades: Intersectional accuracy disparities in commercial gender classification. In: Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81, 77–91 (2018). https://proceedings.mlr.press/v81/buolamwini18a.html

6. Momha, M.: The synthetic mirror: Synthetic data at the age of agentic (2025). https://doi.org/10.48550/arXiv.2506.13818. Accessed 26 July 2025

7. Dilemgani, C.: Synthetic data generation: Techniques, best practices & tools. AIMultiple (2020). https://research.aimultiple.com/synthetic-data-generation/. Accessed 27 July 2025

8. Shin, H.C., Tenenholtz, N.A., Rogers, J.K., Schwarz, C.G., Senjem, M.L., Gunter, J.L., Andriole, K.P., Michalski, M.: Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: Göksel, O., Oguz, I., Gooya, A., & Burgos, N. (eds.) Simulation and Synthesis in Medical Imaging – Third International Workshop, SASHIMI 2018, Held in Conjunction with MICCAI 2018, LNCS, vol. 11037, pp. 1–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00536-8_1

9. Jaipuria, N., Zhang, X., Bhasin, R., Arafa, M., Chakravarty, P., Shrivastava, S., Manglani, S., Murali, V.N.: Deflating dataset bias using synthetic data augmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 772–773 (2020). https://arxiv.org/pdf/2004.13866

10. Savage, N.: Synthetic data could be better than real data. Nature (2023). https://doi.org/10.1038/d41586-023-01445-8.

11. Lee, D., Jaiswal, R.R., Byrne, A.: Utilising synthetic data from LLM for gender bias detection and mitigation in recruitment systems. In: Proceedings of the 2024 Conference on Human Centred Artificial Intelligence – Education and Practice (HCAIep '24), ACM, New York (2024). https://doi.org/10.1145/3701268.3701285

12. Selbst, A.D., Boyd, D., Friedler, S.A., Venkatasubramanian, S., Vertesi, J.: Fairness and abstraction in sociotechnical systems. In: Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 59–68 (2019). https://doi.org/10.1145/3287560.3287598

13. De Vries, T., Misra, I., Wang, C., van der Maaten, L.: Does object recognition work for everyone? In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 52–59 (2019). https://arxiv.org/pdf/1906.02659

14. Garg, N., Schiebinger, L., Jurafsky, D., Zou, J.: Word embeddings quantify 100 years of gender and ethnic stereotypes. Proceedings of the National Academy of Sciences **115**(16), E3635–E3644 (2018). https://doi.org/10.1073/pnas.1720347115

15. Gershgorn, D.: If AI is going to be the world's doctor, it needs better textbooks. Quartz (2022).https://qz.com/1367177/if-ai-is-going-to-be-the-worlds-doctor-it-needs-better-textbooks. Accessed 20 July 2025

16. Barocas, S., Hardt, M., Narayanan, A.: Fairness and Machine Learning: Limitations and Opportunities. MIT Press, Cambridge, MA (2023). https://fairmlbook.org/pdf/fairmlbook.pdf

17. Noble, S.U.: Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press (2018).

18. Bolukbasi, T., Chang, K.W., Zou, J.Y., Saligrama, V., Kalai, A.T.: Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In: Proceedings of the 30th International Conference on Neural Information Processing Systems, pp. 4349–4357 (2016). https://arxiv.org/pdf/1607.06520

19. Bauer, L., Bansal, M.: Identify, align, and integrate: Matching knowledge graphs to commonsense reasoning tasks (2021). https://arxiv.org/pdf/2104.10193

20. Long, L., Wang, R., Xiao, R., Zhao, J., Ding, X., Chen, G., Wang, H.: On LLMs-driven synthetic data generation, curation, and evaluation: A survey. (2024). https://arxiv.org/abs/2406.15126

21. Chang, A., Fontaine, M.C., Booth, S., Matarić, M.J., Nikolaidis, S.: Quality-Diversity Generative Sampling for Learning with Synthetic Data. In: Proceedings of the AAAI Conference on Artificial Intelligence **38**(18), 19805–19812 (2024). https://doi.org/10.1609/aaai.v38i18.2995

22. Markus, E., Venugopal, M., Asha, Tran, T.: Synthetic data generation: A comparative study. In: Proceedings of the International Conference on Artificial Intelligence, pp. 94–102 (2022). https://doi.org/10.1145/3548785.3548793

23. Patki, N., Wedge, R., Veeramachaneni, K.: The synthetic data vault. In: IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 399–410. IEEE, Montreal, QC (2016). https://doi.org/10.1109/DSAA.2016.49

24. Abay, N.C., Zhou, Y., Kantarcioglu, M., Thuraisingham, B., Savas, E.: Privacy Preserving Synthetic Data Release Using Deep Learning. In: Machine Learning and Knowledge Discovery in Databases, pp. 510–526 (2018). https://doi.org/10.1007/978-3-030-10925-7_31

25. Jacobsen, B.N.: Machine learning, synthetic data, and the politics of difference. Theory, Culture & Society **42**(3), 41-57 (2025). https://doi.org/10.1177/02632764241304687

26. Toh, S.-L., Park, J.: Fake it till you make it: Synthetic data and algorithmic bias. International Journal of Communication **19**(0), 1852–1858 (2025). https://ijoc.org/index.php/ijoc/article/view/24000/4991

27. Song, M.: Rethinking minority status and 'visibility'. Comparative Migration Studies **8**(5) 1-17 (2020). https://doi.org/10.1186/s40878-019-0162-2

28. Wirth, L.: The Problem with Minority Groups. In: Linton, R. (ed.) The Science of Man in the World Crisis, pp. 347–372. Columbia University Press, New York (1945).

29. Wilkinson, D.: Rethinking the Concept of 'Minority': A Task for Social Scientists and Practitioners. Journal of Sociology and Social Welfare **27**(1), 115–132 (2000).

30. Amaro, R.: The Black Technical Object: On Machine Learning and the Aspiration of Black Being. Sternberg Press, London (2022)

31. Lee, F., Hajisharif, S., Johnson, E.: The ontological politics of synthetic data: Normal-ities, outliers, and intersectional hallucinations. Big Data & Society **12**(2) (2025). https://doi.org/10.1177/20539517251318289

32. Davis, J.L., Williams, A., Yang, M.W.: Algorithmic reparation. Big Data & Society **8**(2) (2021). https://doi.org/10.1177/20539517211044808

33. Táíwò, O.O.: Reconsidering Reparations. Oxford University Press, New York (2022).

34. Spinner-Halev, J.: From historical to enduring injustice. Political Theory **35**(5), 574–597 (2007). https://doi.org/10.1177/0090591707304585

35. McCarthy, T.: Coming to Terms with Our Past, Part II: On the Morality and Politics of Reparations for Slavery. Political Theory **32**(6), 750–772 (2004). http://www.jstor.org/stable/4148144

36. Kleinberg, J., Mullainathan, S., Raghavan, M.: Inherent Trade-Offs in the Fair Determination of Risk Scores. (2016). https://arxiv.org/pdf/1609.05807

37. Kim, S., Lee, J., Oh, P.: Questioning artificial intelligence: How racial identity shapes the perceptions of algorithmic bias. International Journal of Communication **18**, 677–699 (2024). https://dr.ntu.edu.sg/server/api/core/bitstreams/ed9c0143-a83a-457e-837d-27f945992797/content

38. Green, B., Viljoen, S.: Algorithmic realism: Expanding the boundaries of algorithmic thought. In: Proceedings of the 2020 ACM Conference on Fairness, Accountability, and Transparency (FAT'20), pp. 19–31 (2020).https://doi.org/10.1145/3351095.3372840

39. Fazelpour, S., Lipton, Z.C.: Algorithmic Fairness from a Non-ideal Perspective. In: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, pp. 57–63 (2020). https://doi.org/10.1145/3375627.3375828

40. Benjamin, R.: Race After Technology: Abolitionist Tools for the New Jim Code. Polity Press, Medford (2019).

41. Crawford, K., Miltner, K., Gray, M.L.: Critiquing Big Data: Politics, ethics, epistemology. International Journal of Communication **8**, 1663–1672 (2014).

42. Veselovsky, V., Ribeiro, M.H., Arora, A., et al.: Generating faithful synthetic data with large language models: A case study in computational social science. (2023). https://arxiv.org/pdf/2305.15041

43. Nikolenko, S.I.: Synthetic Data for Deep Learning. Springer International Publishing, Cham, Switzerland (2021). https://arxiv.org/pdf/1909.11512

44. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016).

45. Krchova, I., Platzer, M., Tiwald, P.: Improving predictions on highly unbalanced data using open source synthetic data upsampling. (2025). https://arxiv.org/abs/2507.16419

46. Juwara, L., El-Hussuna, A., Emam, K.: An evaluation of synthetic data augmentation for mitigating covariate bias in health data. Patterns **5**(4), 1-19 (2024). https://doi.org/10.1016/j.patter.2024.100946

47. Bhanot, K., Qi, M., Erickson, J.S., Guyon, I., Bennett, K.P.: The Problem of Fairness in Synthetic Healthcare Data. Entropy **23**(9), 1165 (2021). https://doi.org/10.3390/e23091165

48. Greenman, E., Xie, Y.: Double Jeopardy? The Interaction of Gender and Race on Earnings in the U.S. Social Forces **86**(3), 1217–1244 (2008). https://doi.org/10.1353/sof.0.0008

49. Adler, M.A., Koelewijn-Strattner, G.J., Lengerman, J.J.: The intersection of race and gender among chemists: Assessing the impact of double minority status on income. Sociological Focus **28**(3), 245–259 (1995). http://www.jstor.org/stable/20831743

50. Sparrow, K.: Effects of the double minority (Black/Female) in a southern predominantly white university setting. Griot **11**(2), 8-14 (1992). https://www.proquest.com/docview/1297948839

51. Törnberg, P., Tornberg, A.: The limits of computation: A philosophical critique of contemporary Big Data research. Big Data & Society **5**(2) 1-12 (2018). https://doi.org/10.1177/2053951718811843

52. Kitchin, R.: The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. Sage Publications (2014).

53. Stevens, M., Wehrens, R., de Bont, A.: Conceptualizations of Big Data and their epistemological claims in healthcare: A discourse analysis. Big Data & Society **5**(2), 1-21 (2018). https://doi.org/10.1177/2053951718816727

54. Meoli, A., Piva, E., Righi, H.: Missing women in STEM occupations: The impact of university education on the gender gap in graduates' transition to work. Research Policy **53**(8), 1-16 (2024). https://doi.org/10.1016/j.respol.2024.105072

55. Dastin, J.: Insight - Amazon scraps secret AI recruiting tool that showed bias against women, https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/, (2018). Accessed 7 July 2025

56. Microsoft and OpenAI test synthetic data to train LLMs, as web data is "no longer good enough":r/ArtificialInteligence,https://www.reddit.com/r/ArtificialInteligence/comments/153wgc4/microsoft_and_openai_test_synthetic_data_to_train/?utm_. Accessed 2 August 2025

57. Synthetic data's fine line between reward and disaster, https://www.cio.com/article/3986687/synthetic-datas-fine-line-between-reward-and-disaster.html. Accessed 10 July 2025

58. Bias audits for AI in employment: how synthetic personas are shaping our approach | Upwage, https://www.upwage.com/blog/bias-audits-for-ai-in-employment?utm_. Accessed 7 July 2025

59. The Gender Gap in STEM | MIT Professional Education, https://professionalprograms.mit.edu/blog/leadership/the-gender-gap-in-stem/. Accessed 4 August 2025

60. Dobson, J.E.: Objective vision: Confusing the subject of computer vision. Social Text **41**(3), 35–55 (2023). https://doi.org/10.1215/01642472-10613653