Decolonising Bias in Organisational Systems: A Machine Learning Approach to Equity, Power, and Algorithmic Justice

IBITOYE, Ayodeji and KOLADE, Seun <http://orcid.org/0000-0002-1125-1900>

Chapter 13

Decolonising Bias in Organisational Systems: A Machine Learning Approach to Equity, Power, and Algorithmic Justice

Ayodeji Olusegun Ibitoye

School of Computing and Mathematical Sciences, University of Greenwich, London, UK.

A.O.Ibitoye@greenwich.ac.uk

Oluwaseun Kolade

College of Technology, Business and Engineering, Sheffield Hallam University, UK.

S.Kolade@shu.ac.uk

Abstract

This chapter examines how algorithmic systems can reproduce and exacerbate structural inequities along gender and racial lines, using the Adult Income dataset as a testbed for comparative analysis across four models: Logistic Regression, XGBoost, Explainable Boosting Machines (EBM), and Adversarial Debiasing Networks. Empirical evaluation revealed substantial disparities in unmitigated models, with disparate impact ratios falling as low as 0.68 for women and non-white individuals. Crucially, this study embeds technical findings within a decolonial theoretical framework, arguing that fairness cannot be reduced to statistical parity. Instead, it must be understood as a historically situated, epistemically accountable, and relationally constructed concept. The research challenges dominant narratives of algorithmic neutrality by foregrounding the colonial legacies and institutional hierarchies that inform both data practices and model design. By bridging machine learning evaluation with critical social theory, this research advances a reflexive, justice-oriented approach to algorithmic governance in organisations. It offers a framework for rethinking fairness not simply as a computational objective, but as a moral and organisational commitment grounded in equity, participatory design, and the inclusion of marginalised knowledges.

Keywords: AI Ethics, Algorithmic Bias, Decolonial Theory, Machine Learning Fairness, Organisational Justice

Introduction

The integration of Artificial Intelligence (AI) and Machine Learning (ML) into organisational decision-making has evolved from experimental tools to foundational systems shaping core human resource functions, including recruitment, performance evaluation, churn prediction,

promotion, compensation, and workplace surveillance (Sun & Jung, 2024; Ibitoye & Onifade, 2021; Balasubramanian et al., 2022). These technologies are frequently framed as objective and efficient, promising to streamline decisions that have traditionally been influenced by human subjectivity (Darioshi & Lahav, 2021; Messeri & Crockett, 2024). Yet, emerging research has shown that far from being neutral, these systems often replicate and even amplify pre-existing social inequities, particularly those rooted in race, gender, and class (Mollema, 2024; Friis & Riley, 2023). Rather than removing bias, AI frequently automates it under the guise of technical rationality. This study explores the challenge of decolonising algorithmic bias in organisational systems by examining how machine learning models reflect and reproduce structural inequities. While bias in AI is often approached as a technical problem, solvable through better data, improved model architecture, or post hoc fairness adjustments, this study argues that such bias is more fundamentally a manifestation of colonial power relations embedded in organisational data, knowledge systems, and decision-making structures. It contends that algorithmic bias is not simply a computational flaw but an epistemic and institutional issue rooted in the coloniality of power. Decolonising algorithmic bias, in this context, means moving beyond surface-level technical fixes to interrogate and transform the historical logics, institutional structures, and epistemic hierarchies that shape data collection, model development, and decision-making processes. Rather than treating bias as a correctable flaw or aiming solely for statistical parity, a decolonial approach seeks to reconfigure the very conditions under which algorithmic systems are imagined, built, and deployed within organisational life.

The widespread adoption of data-driven technologies reflects an epistemological inheritance grounded in Eurocentric ideals of objectivity, meritocracy, and managerial control, frameworks that have long justified the exclusion of non-Western knowledge systems and have devalued alternative ways of understanding work, value, and human capacity (Alba, 2024; Kponyo et al., 2024). As such, algorithmic decision-making can be seen as part of a broader project of techno-neocolonialism, wherein automation becomes a vehicle for reproducing historical hierarchies under contemporary conditions (Muldoon & Wu, 2023). This research situates algorithmic bias within a decolonial theoretical framework to ask: *What does it mean to decolonise AI systems within organisational contexts?* To answer this question, it integrates critical algorithm studies with decolonial thought, examining how the logic of AI reflects embedded colonial assumptions about who counts, what knowledge matters, and how legitimacy is constructed. It critically assesses the limitations of conventional fairness metrics

in AI and explores the potential for machine learning to act not only as a diagnostic tool for structural bias but also as a site of organisational transformation.

Ultimately, this chapter argues for a shift from narrowly defined statistical fairness to a broader politics of equity and epistemic justice, one that foregrounds relational accountability, participatory design, and the inclusion of historically marginalised knowledge systems. This approach responds to growing concerns in the literature regarding the ethical implementation of AI in global contexts (Kunjumuhammed, 2024; Duranti et al., 2023) and builds on calls to address the global dynamics of techno-colonial power in digital infrastructures (Kamran, 2023). By critically examining how algorithmic systems embody and reproduce colonial patterns of knowing and governing, this research contributes to both AI ethics and critical organisational studies. It also contributes to practice-oriented debates by showing how fairness, accountability, and power are not merely abstract principles, but are operationalised, often unevenly, within the real-world contexts of organisational design, data governance, and institutional culture. Then, it provides a conceptual and methodological foundation for reimagining organisational AI as a tool for justice rather than control, and for centring the perspectives, knowledge, and aspirations of those historically excluded from technological design and governance.

Literature Review

This literature review is organised around three interrelated strands that underpin the conceptual framework of this study. First, it draws on decolonial theory to examine how algorithmic systems are shaped by epistemic legacies rooted in colonial power and knowledge hierarchies. Second, it reviews scholarship on structural bias and data epistemologies, highlighting how algorithmic governance often encodes and legitimises social inequalities through ostensibly neutral design. Later, it explores emerging interventions within educational and organisational contexts that seek to embed fairness, reflexivity, and accountability into algorithmic development and deployment. Together, these strands establish the theoretical and empirical foundations for a decolonial approach to bias mitigation in organisational machine learning systems.

Decolonial theory and the epistemic foundations of algorithmic systems

Decolonial theory provides a critical lens through which to interrogate the socio-historical underpinnings of contemporary algorithmic systems. Rooted in the foundational work of

Quijano (2000) and extended by scholars such as Mignolo & Walsh (2018), decolonial thought critiques the lingering effects of colonialism, not merely as a historical event, but as an enduring structure of power and knowledge, referred to as "coloniality." Oliveira (2024) emphasises that decolonial critique is concerned with how these legacies continue to shape systems of value, authority, and epistemic legitimacy in the present. In organisational contexts, coloniality is evident in whose knowledge is institutionalised, whose voices are considered authoritative, and which normative standards, such as "objectivity" or "meritocracy", govern decision-making.

Within this framework, ML and AI are not neutral or apolitical tools. Rather, they are artefacts of dominant epistemologies, designed, implemented, and validated within systems that have historically privileged Eurocentric, technocratic, and managerial logics (Oliveira, 2024). The design of algorithms is influenced by whose behaviours are captured, which variables are measured, and what outcomes are deemed desirable. These embedded assumptions standardise judgment and often obscure the structural inequalities embedded in data (Davidson & Kelly, 2020; DeVos et al., 2022). As Alba (2024) notes, algorithmic systems frequently disadvantage marginalised groups through seemingly neutral decision-making processes, reinforcing historic patterns of exclusion under the veneer of efficiency and objectivity.

Structural bias, data epistemologies, and algorithmic governance

Decolonial scholars argue that algorithmic bias is not a random artefact or technical oversight; it is a structural feature of systems built on epistemologies that have long devalued non-Western knowledges and marginalised bodies. Chaka (2022) critiques the algorithmic marginalisation of populations in the Global South, showing how digital technologies often fail to account for local contexts or lived experiences. Similarly, Akpan (2023) and De Souza et al. (2024) argue that without epistemic reorientation, AI will continue to reproduce patterns of digital inequality and exclusion. Omotubora & Basu (2024) pushed for a reinvention of AI governance that transcends colonial frameworks. The research calls for intersectional, egalitarian models of AI that reject the universality of Western rationality and instead accommodate plural, situated, and culturally embedded perspectives. Couldry & Mejias (2023) frame this need as part of the broader "decolonial turn" in technology studies, a shift toward critically examining the normative foundations and power dynamics shaping digital infrastructures. Fırıncı (2024) expands this conversation by exploring how algorithmic equity might be realigned with non-Western moral frames, arguing for a move beyond liberal-individualist conceptions of fairness toward community-centred notions of justice. These

perspectives collectively highlight that fairness in AI cannot be divorced from its sociopolitical context, and that algorithmic governance must be reimagined through ethical frameworks grounded in cultural plurality and historical justice.

Educational and organisational interventions in algorithmic design

In applied domains such as education and the workplace, scholars are increasingly concerned with how to operationalise decolonial principles in AI design and deployment. Zembylas (2023) offers pedagogical strategies to counteract the ethics of digital neocolonialism, emphasising critical consciousness and reflexivity in the development and use of educational technologies. Similarly, Baradaran (2024) and Mollema (2024) underscore the dual potential of AI to either exacerbate or mitigate historical inequalities depending on how it is designed, governed, and evaluated. In organisational contexts, this translates to rethinking the foundational assumptions that underlie data collection, performance metrics, and decision-making algorithms. Aguiar & da Silva (2024), Tacheva & Ramasubramanian (2024), and Coleman (2023) advocate for participatory and locally grounded design methodologies that treat marginalised communities not merely as data subjects or end-users, but as co-creators of technological systems. These approaches foreground relational accountability and emphasise the inclusion of historically excluded knowledge systems as essential to building equitable AI.

Taken together, this body of study reveals a consensus: addressing algorithmic bias through purely technical or statistical means is insufficient. Instead, scholars call for a decolonial reorientation that interrogates the epistemic and political foundations of algorithmic systems. This includes challenging the hegemony of Eurocentric rationality, prioritising lived experience and relationality, and enabling marginalised groups to shape the terms of AI development and evaluation. A decolonial approach to algorithmic design is not merely corrective; it is transformative, opening up possibilities for creating organisational systems that advance equity, justice, and epistemic plurality.

Bias in Organisational Systems: A Structural View

Structural bias and organisational change

Bias in organisational systems is often framed as a technical glitch, an issue stemming from flawed data, incomplete representation, or model miscalibration. However, such framings depoliticise and individualise what are, in reality, deeply embedded and historically situated structures of exclusion. In organisational contexts, bias is not merely an anomaly to be

corrected; it reflects the long-standing social hierarchies, knowledge regimes, and institutional legacies that have shaped decision-making for centuries (Boccio, 2022). Algorithmic systems frequently reinforce rather than disrupt these power structures. Hiring algorithms, for instance, have been shown to prioritise male-coded language or favour candidates from elite educational institutions, which have historically been inaccessible to marginalised groups (Kolade et al., 2025). Performance management systems may penalise assertiveness in women or misinterpret culturally diverse leadership styles as deviance from the norm. These patterns of discrimination are not accidental; they stem from data trained on historical decisions shaped by patriarchal, colonial, and capitalist logics.

As such, organisational AI systems risk automating exclusion under the guise of neutrality and efficiency, a structural view of bias recognises that it is embedded in the very design of organisational policies, cultures, and data infrastructures. As Hristov et al (2022) argue, hierarchical systems can generate cognitive distortions, wherein information is interpreted differently depending on one's position within the organisation. Performance management practices often exacerbate these distortions by allowing subjective judgments to override supposedly objective metrics. This process can be further reinforced by organisational policies that, while ostensibly neutral, embed and reproduce inequities, especially in AI-based decision-making (Nadeem et al, 2022).

Policies, culture, and data governance in AI systems

Organisational policies and data governance frameworks are critical vectors through which bias is institutionalised. Kordzadeh & Ghasemaghaei (2022) note that the fairness and effectiveness of AI systems are directly tied to how data policies are designed and implemented. Data quality, representativeness, and the equity of algorithmic outcomes are shaped by policy decisions about what data is collected, how it is labelled, and which outcomes are valued. Li et al. (2022) similarly caution that misaligned governance structures can entrench inequities by failing to account for systemic bias embedded in historical datasets.

Culture, too, plays a foundational role in shaping algorithmic outcomes (Seaver, 2017). In organisations that lack inclusive norms, bias can flourish unchecked. As Berthet (2022) highlights, exclusionary organisational cultures reproduce conditions where marginalised voices are underrepresented in both datasets and leadership, resulting in blind spots in algorithmic design. The homogeneity of many data science and AI teams, often predominantly

male, white, and Western, compounds this issue, leading to narrow definitions of problems and solutions that fail to capture the lived realities of diverse stakeholders.

Toward structural interventions and epistemic justice

Addressing bias in organisational AI systems requires more than technical adjustment; it necessitates structural intervention. A decolonial perspective reframes fairness not as a statistical outcome but as a political and epistemic imperative. This means interrogating the ideologies, such as meritocracy, that are often encoded into AI systems. As Boccio (2022) articulates, the belief in algorithmic objectivity masks "coded inequity," whereby systems reproduce social injustices under the veneer of impartiality.

To move toward epistemic justice, organisations must engage in practices that surface and challenge the institutional roots of inequality. This includes rethinking leadership pipelines, ensuring inclusive policy development, and embedding diverse voices in both design and governance processes (Pandey et al, 2024; Oguntibeju, 2024). Improved training on cognitive bias, the development of culturally sensitive algorithms, and critical reviews of organisational policies can further contribute to this effort (Grimmelikhuijsen & Meijer, 2022). Moreover, leveraging big data analytics can help reveal hidden patterns of exclusion, so long as these tools are guided by equity-centred frameworks (Nugroho & Angela, 2024; Faheem et al., 2024). Generally, a structural understanding of bias acknowledges that algorithmic discrimination is not a technical fluke but a systemic outcome. To effectively intervene, organisational AI must be reimagined as part of a broader project of historical accountability and epistemic transformation.

Methodological Approach: Machine Learning for Bias Detection and Mitigation

In keeping with this study's aim to explore the structural and epistemic dimensions of bias in organisational systems, the methodological approach involves applying machine learning techniques and critically reflecting on the assumptions underpinning them. While machine learning offers powerful tools for pattern recognition and decision support, its deployment in organisational contexts must be interrogated for how it can replicate, obscure, or resist colonial structures.

Dataset selection and justification

This study draws on the Adult Income Dataset (Dua & Graff, 2017). This dataset includes demographic and employment attributes such as age, race, gender, education, occupation, hours worked, and income bracket (above or below $50,000 per year). Although the data originates from a 1994 US Census extract and is therefore historically dated, it remains one of the most widely used datasets for exploring fairness, bias detection, and algorithmic decision-making in socio-economic contexts (Zhang et al ., 2018; Ding et al., 2021; Besse et al., 2020).

The dataset's inclusion of sensitive attributes, specifically gender and race, makes it suitable for examining how bias manifests in predictive models of income, a proxy for organisational outcomes such as hiring, promotion, or compensation. However, it is essential to acknowledge that the dataset itself is a product of state-driven demographic categorisation, shaped by socio-political logics that often reduce complex identities to fixed, bureaucratic labels. This reflects the broader colonial impulse to standardise and classify populations in the service of governance and control.

Machine learning models

To assess and mitigate algorithmic bias in organisational decision-making, this study employs a comparative framework that integrates both traditional and contemporary machine learning models. The goal is not only to evaluate predictive performance but also to explore how different modelling paradigms handle fairness, interpretability, and structural disparities across protected groups. The model selection reflects a commitment to technical rigour, ethical accountability, and conceptual coherence with decolonial theory. They are:

1. **Logistic regression (baseline):** Logistic regression serves as the interpretable baseline model, providing a point of comparison for fairness and performance across more complex algorithms (Cox, 1958). Its linear nature and transparency make it especially useful for exposing disparities in feature influence and outcome distribution. It also helps foreground the ethical stakes of interpretability, especially when systems affect people's access to employment or compensation.

2. **XGBoost (Gradient Boosted Trees):** XGBoost, a high-performance ensemble model, is included to represent the state-of-the-art in tabular predictive analytics. It captures complex, non-linear relationships and typically yields strong predictive results in socio-economic data (Chen & Guestrin, 2016). However, its opaque nature raises questions

about accountability, making it a valuable contrast in the analysis of bias and justice. SHapley Additive exPlanations (SHAP) values are used alongside XGBoost to provide post hoc interpretability of feature contributions.

3. **Explainable Boosting Machine (EBM):** EBMs combine the performance of ensemble methods with interpretable logic-based outputs. Unlike traditional black-box models, EBMs provide human-readable visualisations of how each feature contributes to the model's predictions (Nori et al., 2019). This transparency aligns with calls in decolonial theory for epistemic justice and participatory governance, ensuring that marginalised stakeholders are not excluded from understanding or contesting automated decisions.

4. **Adversarial Fairness Network (Fairness-Aware Model):** To directly address bias at the learning level, an adversarial debiasing model is also implemented. This model trains a predictor to maximise predictive accuracy while simultaneously training an adversary to minimise the ability to infer sensitive attributes such as gender or race (Zhang et al., 2018; Zafar et al., 2017). This technique operationalises fairness not merely as a corrective overlay, but as an intentional design principle, an approach resonant with decolonial critiques that challenge the neutrality of technological systems.

This diverse modelling strategy enables a multi-perspective analysis of bias, allowing for comparisons not only in predictive accuracy but also in fairness outcomes, interpretability, and ethical alignment. Importantly, it positions machine learning not just as a diagnostic tool but as a transformative instrument, one that, when critically applied, can help organisations reimagine justice in decision-making systems historically shaped by colonial hierarchies.

Fairness metrics and bias detection

To evaluate disparities in model predictions, this study employs a set of quantitative fairness metrics widely used in the machine learning community. These metrics assess group-level inequalities based on protected attributes, specifically gender and race, as represented in the Adult Income dataset. While these metrics offer a valuable diagnostic function, their application in this study is critically framed through a decolonial lens, which questions their epistemic assumptions and limitations. These include:

1. **Demographic Parity Difference:** Demographic parity assesses whether individuals from different groups receive positive outcomes at equal rates (Loukas & Chung, 2023).

A difference in outcome proportions indicates potential bias. While useful as a baseline measure, demographic parity may obscure context-specific historical injustices, as it assumes equal distribution is inherently fair without addressing the reasons for existing disparities.

2. **Disparate Impact Ratio:** This ratio compares the rate of favourable outcomes between the protected group and a reference group. A commonly used threshold (the "80% rule") suggests that ratios below 0.8 may indicate discriminatory impact (Tobia, 2017). While this metric is frequently used in legal contexts, it does not interrogate the structural origins of outcome imbalances and may risk reducing justice to compliance.

3. **Equal Opportunity Difference:** This metric evaluates the difference in true positive rates between groups, focusing on equal access to favourable predictions for those who are qualified (Fagan & Holland, 2002). This is particularly relevant in organisational contexts such as hiring or promotion. However, it still operates within a predictive logic that prioritises statistical equity over relational or historical justice.

4. **Average Odds Difference:** Average odds combine disparities in both true positive and false positive rates across groups. It provides a more holistic view of group-level error patterns, but like other fairness metrics, it remains rooted in quantitative parity frameworks that often fail to account for the lived experiences and intersectional realities of marginalised individuals (Grunkemeier & Wu, 2007).

While the metrics above are essential for revealing bias patterns in model outputs, they must not be mistaken for justice itself. From a decolonial standpoint, fairness metrics often reflect liberal legal logics, seeking symmetry in outcomes without addressing epistemic exclusion, historical dispossession, or structural violence. These metrics ask whether an algorithm is fair across groups, but rarely ask who defined the groups, who set the success criteria, or whose worldview is encoded in the data.

Furthermore, these metrics often ignore intersectionality, failing to capture compounded disadvantages faced by, for example, Black women or Indigenous individuals in data systems that treat race and gender as discrete, static variables. For this reason, while fairness metrics are used to measure disparity, this study argues that they should be supplemented by contextual interpretation, qualitative reasoning, and critical reflection — a position that aligns with decolonial ethics of relationality, lived experience, and epistemic justice.

Bias mitigation techniques

To address disparities identified in model predictions, this study employs a three-stage bias mitigation framework, encompassing interventions at the pre-processing, in-processing, and post-processing levels of the machine learning pipeline. This multi-point strategy reflects both a technical approach to fairness and an ethical commitment to intentional, justice-oriented system design.

1. **Pre-processing** through Reweighing (Kamiran & Calders, 2012): This technique assigns different weights to training examples based on group membership and label distribution, ensuring more balanced representation of protected groups before model training begins. It seeks to reduce historical data imbalances that can distort model learning.

2. **In-processing** using Adversarial Debiasing (Zhang et al., 2018): In this approach, a predictor model is trained to make accurate predictions while an adversarial model attempts to infer the protected attribute from the predictor's output. The goal is to prevent the model from encoding discriminatory patterns related to sensitive features during training.

3. **Post-processing** through Equalised Odds Postprocessing (Hardt et al., 2016): This method adjusts the final decision outputs to ensure that both true positive and false positive rates are equal across protected and unprotected groups. It is particularly useful in high-stakes settings where outcome parity must be assured despite complex internal model behaviour.

All mitigation techniques were implemented using the AIF360 (Bellamy et al., 2019) and Fairlearn libraries (Weerts et al., 2023), open-source toolkits widely adopted for fairness evaluation and algorithmic accountability. The integration of these tools into the research process reflects not only a technical choice but also a deliberate ethical stance: that fairness must be designed into systems from the outset, not treated as a cosmetic correction. In the context of decolonial inquiry, this approach affirms the importance of designing against structural inequity, rather than merely adjusting outputs to preserve statistical balance.

Ethical considerations

While the bias mitigation techniques employed in this study contribute to reducing algorithmic disparities, they operate largely within a technical paradigm that assumes injustice can be

corrected through statistical balancing alone. This framing risks obscuring the deeper, structural systems of exclusion, historical, epistemic, and institutional, that give rise to biased outcomes in the first place.

A truly decolonial approach demands more than numerical parity. It requires a reorientation of how fairness is conceptualised and operationalised in machine learning. Rather than treating fairness as a measurable output, decolonial ethics foreground relationality, accountability, and the centring of marginalised perspectives in the design, implementation, and governance of algorithmic systems. This includes acknowledging the colonial histories embedded in datasets, questioning who defines success or merit in organisational contexts, and resisting the impulse to universalise fairness metrics without context.

Moreover, ethical engagement must move beyond compliance-driven solutions and toward participatory and situated practices, those that involve affected communities not just as data subjects but as co-designers of systems that impact their lives. This orientation aligns with broader calls for epistemic justice, where multiple ways of knowing are valued, and where technology is shaped through dialogue, care, and collective memory. Ultimately, the ethical imperative is not only to make models more "fair," but to challenge the organisational structures and knowledge systems that allow unfairness to persist, automated or otherwise.

Findings: Bias Metrics, Patterns and Model Behaviour

This section presents the results of the machine learning experiments, focusing on both predictive performance and fairness outcomes across the four models: Logistic Regression, XGBoost, Explainable Boosting Machine (EBM), and the Adversarial Fairness Network. Models were evaluated using standard classification metrics and fairness measures, with particular attention to disparities across gender and racial groups. The goal is not only to assess which models perform best, but to understand how algorithmic decisions reflect, reinforce, or challenge existing inequities in organisational systems.

Predictive performance overview

The predictive performance of the four machine learning models, including Logistic Regression, XGBoost, Explainable Boosting Machine (EBM), and Adversarial Debiasing, was first assessed using standard classification metrics: accuracy, precision, recall, and F1 score. These metrics provide a baseline understanding of each model's effectiveness in distinguishing between individuals earning above and below the $50,000 income threshold. As shown in

Table 1, XGBoost achieved the highest overall accuracy (88.7%), followed closely by EBM (87.5%), with Logistic Regression and Adversarial Debiasing trailing slightly. These results are consistent with the known strengths of ensemble-based models like XGBoost and EBM, which are well-suited for tabular socio-economic data due to their ability to capture non-linear relationships and complex feature interactions.

However, the Adversarial Debiasing model, while slightly less performant in terms of raw accuracy (83.4%), was designed with fairness constraints explicitly embedded in its training process. As expected, its trade-off in predictive power reflects the intentional prioritisation of equitable outcomes, a recurring theme in fairness-aware modelling. These initial results demonstrate that model performance is highly sensitive to the objectives prioritised during optimisation, underscoring the tension between accuracy and fairness.

Table 13.1: Predictive Model Performance

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	84.1%	72.5%	68.3%	70.3%
XGBoost	88.7%	79.1%	75.4%	77.2%
EBM	87.5%	78.3%	74.0%	76.1%
Adversarial Debiasing	83.4%	70.1%	69.8%	69.9%

Source: Author's computation

While these results offer a useful comparative view of model efficiency, they must be interpreted with caution. A model's ability to predict accurately at the aggregate level does not necessarily imply fairness across subgroups. In fact, higher-performing models like XGBoost can amplify bias if trained on historical data embedded with systemic inequities. Therefore, a complete evaluation requires examining how predictive errors and successes are distributed across socially salient categories such as gender and race. The next section addresses this by assessing each model using fairness-specific metrics before any bias mitigation, an essential step in understanding the ethical and organisational implications of deploying machine learning in high-stakes decision environments**.**

Fairness metrics: pre-mitigation

Initial evaluation of model outputs revealed persistent disparities in outcomes for women and non-white groups across all four models, particularly in true positive rates and disparate impact ratios. These disparities reflect how machine learning models reproduce historical labour

inequalities, reinforcing systemic exclusion under the guise of predictive optimisation. The table below summarises key fairness metrics before applying any bias mitigation:

Table 13.2: Fairness Metrics before Bias mitigation

Model	TPR (Men)	TPR (Women)	Disparate Impact (Female vs. Male)
Logistic Regression	74.2%	60.1%	0.68
XGBoost	78.5%	63.4%	0.72
EBM	77.8%	65.2%	0.75
Adversarial (Untrained)	72.3%	64.8%	0.79

Source: Author's computation

As shown, while performance improved in models like XGBoost and EBM, fairness gaps remained evident across all architectures. The unmitigated Adversarial model, before fairness training, showed slightly better balance but still fell short of the standard 0.80 disparate impact threshold. These findings reinforce a central argument of this research: that bias is systemic, not architectural. It is not simply the result of poor model selection, but of deeply encoded patterns in the data reflecting colonial, racial, and gendered labour hierarchies. Therefore, reducing bias requires more than choosing a better model; it demands intentional intervention and structural awareness throughout the machine learning pipeline.

Post-mitigation results

Following the application of bias mitigation techniques at different stages of the machine learning pipeline, including reweighing, adversarial debiasing, and equalised odds postprocessing, all four models demonstrated measurable improvements in fairness metrics. These interventions reduced disparities in model predictions between gender and racial groups, particularly in terms of selection rates and true positive rates. However, as anticipated, some trade-offs in predictive performance were observed, most notably in models with stronger fairness constraints.

Table 3 presents post-mitigation results using three standard group fairness metrics: demographic parity difference, equal opportunity difference, and disparate impact ratio. These are commonly used to evaluate bias in binary classification tasks.

Table 13.3: Post-Mitigation Fairness Metrics Across Models

Model	Demographic Parity Diff	Equal Opportunity Diff	Disparate Impact
Logistic Regression + Reweighing	0.12	0.10	0.83

XGBoost + Equalised Odds	0.11	0.09	0.80
EBM + Reweighing + Postprocessing	0.09	0.07	0.84
Adversarial Debiasing (Trained)	**0.08**	**0.06**	**0.86**

Source: Author's computation

As shown, the Adversarial Debiasing model achieved the most balanced outcomes across all fairness metrics, followed closely by the EBM. Notably, both models surpassed the 0.80 threshold on the disparate impact ratio, often used as a legal and ethical benchmark for non-discrimination in algorithmic systems. To complement these abstract fairness measures, Table 4 presents true positive rates (TPRs) for men and women, offering a more intuitive view of how each model improved in terms of real-world outcomes.

Table 13.4: True Positive Rates and Disparate Impact by Gender

Model	TPR (Men)	TPR (Women)
Logistic Regression + Reweighing	71.9%	68.4%
XGBoost + Equalised Odds	75.1%	69.2%
EBM + Reweighing + Postprocessing	74.4%	70.1%
Adversarial Debiasing (Trained)	71.0%	70.3%

Source: Author's computation

These outcome-level comparisons reinforce the earlier findings: models trained or adjusted with fairness constraints not only reduced disparities in abstract metrics but also improved actual decision outcomes for marginalised groups. The Adversarial Debiasing model, in particular, brought women's TPR nearly in line with that of men, significantly narrowing the prediction gap observed in baseline models.

However, these improvements came with modest reductions in overall accuracy, especially in models subjected to more aggressive fairness interventions. This reflects a recurring trade-off in algorithmic fairness work: equity often requires sacrificing a degree of predictive optimisation to redress systemic imbalances embedded in the training data. Crucially, these results support the central argument of this study: while algorithmic bias can be mitigated, such mitigation remains bounded by the limitations of the data and the structural conditions that shaped it. Fair models cannot undo unfair histories. Therefore, bias mitigation must be seen not as a solution, but as part of a broader ethical and organisational commitment to structural change, epistemic accountability, and participatory governance.

Critical interpretation

The results presented in Section 5.3 confirm that algorithmic bias in organisational systems is quantifiable and partially correctable through technical interventions. Fairness-aware models, particularly those using adversarial debiasing and postprocessing adjustments, demonstrated measurable improvements across standard group fairness metrics. However, the presence of residual disparities, even after mitigation, signals the limitations of relying solely on algorithmic solutions to redress structural injustice. These Key concerns are:

1. **Fairness metrics are insufficient on their own**: Even after mitigation, structural patterns remain, such as the consistent underperformance for women and non-white individuals across models.

2. **Bias is systemic, not merely statistical**: The disparities observed originate from data shaped by colonial labour hierarchies and institutional exclusions. Adjusting model outputs does not erase the conditions that produced these patterns.

3. **Interpretability matters**: Models like EBM allow for transparency in how decisions are made, which is crucial for building epistemic accountability in organisations. Yet, even interpretable models can still encode biased logic if their training data is not interrogated.

4. **Trade-offs demand ethical clarity**: Choosing a more "accurate" but less fair model is not a neutral act; it reflects value choices that must be made explicit. A decolonial approach insists that fairness be prioritised not as an efficiency constraint, but as a moral imperative.

The findings indicate that while algorithmic bias can be partially mitigated through technical means, these methods should be viewed as starting points, rather than solutions. True transformation requires organisations to rethink the assumptions embedded in their data, systems, and definitions of success. Even the most sophisticated fairness-aware models remain constrained by the assumptions embedded in the data, the organisational culture, and the values of the system designers. Machine learning, when critically applied, can help reveal hidden structures of exclusion, but only if guided by a broader vision of justice rooted in historical, cultural, and epistemic awareness. Therefore, machine learning can support more equitable organisational systems, but not without rethinking the very definitions of success, objectivity, and value upon which those systems are built.

Decolonising the Algorithm: Moving Beyond Technical Fixes

The application of bias mitigation techniques demonstrates that algorithmic disparities can be measured and partially corrected. However, such corrections remain superficial unless they are accompanied by a deeper interrogation of the systems, values, and histories that shape the data and models themselves. A decolonial perspective demands that we move beyond a technocentric obsession with parity metrics and into a more profound reckoning with what algorithms are designed to do, who they serve, and whose realities they reflect or erase.

The limits of technological fairness

Contemporary fairness frameworks in machine learning, rooted largely in statistical parity and formal legal compliance, draw heavily from liberal traditions that prioritise individual rights and symmetrical treatment. While these approaches offer useful tools for detecting measurable disparities, they are fundamentally limited in their ability to address the deeper colonial entanglements of power and knowledge embedded in organisational systems. These models often assume that social categories are commensurable and fixed, treating race, gender, or class as clean variables rather than as historically constructed, relational, and context-dependent identities. In doing so, they reduce justice to the achievement of numerical balance across groups, obscuring the complex histories of marginalisation that shape both data distributions and decision outcomes.

Moreover, fairness metrics rarely account for the narrative, cultural, or affective dimensions of harm, dimensions that cannot be captured through parity scores or error rates. By treating fairness as a problem to be optimised through algorithms, rather than a political and ethical question situated in lived realities, these frameworks risk sanitising injustice. They may measure bias without confronting the institutional logics and colonial foundations that produce it. A decolonial perspective insists that fairness must extend beyond computational symmetry to include epistemic accountability, historical consciousness, and the centring of those whose lives are most affected by algorithmic systems.

Toward a decolonial reorientation

Decolonising the algorithm requires more than technical corrections or fairness audits; it demands a fundamental reorientation in how we relate to data, whose knowledge counts, and what purposes algorithmic systems are designed to serve. At its core, this reorientation calls for a shift from treating marginalised groups as passive subjects of fairness to recognising them

as active designers, decision-makers, and epistemic authorities. Design processes must be restructured to include those historically excluded from technological and organisational power, not only in tokenistic consultation but in shared authorship and control. This also involves a critical challenge to the dominant epistemologies that govern machine learning, logics that prioritise abstraction, generalisation, and statistical objectivity while systematically excluding embodied, spiritual, communal, or land-based ways of knowing. Data itself must be approached not as neutral input, but as a colonial residue**:** a digital continuation of historical practices of classification, surveillance, and extraction used to administer and manage colonised populations.

Equally, the assumption of technological neutrality must be actively refused. All algorithms encode political choices, and every claim to objectivity serves particular interests, interests that must be interrogated in light of their social, historical, and ethical consequences. From a decolonial perspective, the work of justice involves reimagining systems of data governance that are participatory and co-created, grounded not in abstract performance metrics but in lived experiences, ethical accountability, and the possibility of repair. Such an approach would centre relational responsibility over algorithmic optimisation, and elevate memory, redress, and healing as essential components of technological justice. In doing so, it challenges us to build organisational systems that do not merely mitigate bias, but fundamentally reconstruct the terms of knowing, valuing, and relating through which technology operates.

Machine learning as a site of contestation and possibility

While algorithmic systems have often reproduced historical injustice, they also hold the potential to surface hidden structures, visualise disparities, and provoke institutional introspection. In this way, machine learning can become a site of contestation, where technical tools are repurposed to make inequality visible and to open space for organisational transformation grounded in justice.

But this repurposing is not automatic. It requires a shift from using AI to optimise for the status quo to using it as a method of disruption, refusal, and reimagination. It is not enough to build "fairer" models; we must ask: fairer according to whom? And more importantly, how do we redefine the systems we want to build altogether?

Implications for Organisational Practice

If machine learning models can reproduce or resist systemic bias, then organisations must take responsibility for the social, ethical, and epistemic consequences of the technologies they deploy. The findings and decolonial framing in this study suggest that addressing bias is not simply a technical challenge, but a transformational task, one that requires rethinking organisational values, structures, and ways of knowing.

Rethinking fairness beyond compliance

In many organisational contexts, fairness is framed narrowly as a compliance obligation, something to be measured, documented, and cleared once numerical thresholds are met or legal risks are minimised. This instrumental approach often reduces justice to a set of quantifiable outcomes, enabling what might be called performative inclusion rather than structural transformation. A decolonial perspective, however, demands a more fundamental shift: from ticking boxes to confronting power. Fairness must be reimagined not as a one-time audit but as a continuous process of critical reflection, one that asks whose interests algorithmic systems serve, who defines their success, and who remains structurally disadvantaged even after mitigation.

This rethinking also requires organisations to reject the illusion that fairness metrics can fully capture justice. While such indicators are useful diagnostic tools, they often obscure deeper epistemological questions: What assumptions underlie "objective" decision-making frameworks? Whose knowledge systems are privileged? And what forms of harm are rendered invisible by a logic that seeks symmetry rather than accountability? By interrogating these foundations, organisations can move from reactive compliance to a more relational, historically informed understanding of equity, one that centres transformation over regulation.

Reimagining organisational knowledge systems

Contemporary organisations overwhelmingly privilege quantitative, predictive, and ostensibly "rational" data in their decision-making processes. This privileging reflects a deeply embedded epistemological hierarchy, one that sidelines or outright excludes Indigenous, relational, intuitive, and community-based ways of knowing. As a result, algorithmic systems often reinforce epistemic injustice, embedding assumptions about objectivity and value that align with dominant cultural logics while erasing alternative forms of insight. From a decolonial

standpoint, addressing algorithmic bias requires more than fair outputs; it demands a reconfiguration of whose knowledge counts in shaping organisational life.

This reimagining involves recognising and legitimising multiple epistemologies in both organisational analysis and AI system design. It also means actively involving those most affected by algorithmic decision-making, particularly workers from historically marginalised groups, in the co-construction of performance metrics, HR analytics, and algorithmic criteria. Such participation is not merely a procedural add-on, but a transformative step toward justice. It opens space for decision-making systems to become transparent, contestable, and accountable, not only in technical terms, but in ways that honour lived experience, collective memory, and cultural knowledge. In doing so, organisations can begin to shift from extractive, top-down data practices to more equitable, participatory forms of governance.

Design justice and co-creation

Decolonising technology within organisational systems requires more than improving usability or engaging in inclusive testing; it calls for a deeper commitment to design justice**,** a framework that centres those most affected by technological systems in their creation, deployment, and governance. Unlike traditional user-centred design, which often treats individuals as passive recipients or testers, design justice foregrounds power, participation, and accountability. It asks not only how systems are built, but who gets to define their goals, whose interests are served, and what values are encoded. Implementing design justice in practice involves a structural redistribution of power. This means shifting authority away from elite data science teams and external vendors, and toward inclusive, cross-functional coalitions that bring together domain experts, frontline workers, community representatives, and technologists.

It also involves the creation of ethics review boards or community advisory councils to oversee the development and deployment of algorithmic systems, ensuring those impacted by decisions have a voice in how they are made. Consequently, a decolonial design approach incorporates reflexive and relational practices, such as equity audits, cultural safety protocols, and storytelling methodologies, that challenge dominant paradigms of objectivity and allow for richer, context-sensitive understandings of justice. In embracing co-creation and participatory governance, organisations can move from technocratic management toward relational accountability, opening space for technologies that are not just efficient, but ethically grounded, culturally resonant, and socially transformative.

From metrics to movement

Ultimately, machine learning systems cannot be disentangled from the broader institutional cultures and historical legacies within which they operate. A technically "debiased" model, when embedded in a workplace culture shaped by discrimination or exclusion, will still reproduce inequitable outcomes. Fairness cannot be retrofitted onto unjust systems. Organisational justice is not a product of technical calibration alone; it must be actively cultivated through leadership, governance, and cultural transformation. This demands critical leadership willing to interrogate the normative assumptions that underpin organisational processes, as well as bold policy reforms that align AI governance not merely with efficiency but with equity and redress.

It also requires sustained, organisation-wide education on the entanglements of colonial histories, digital ethics, and structural power. In this view, algorithmic fairness is not simply a matter of improved metrics; it is inseparable from a deeper reimagining of how organisations define value, merit, and belonging. This study, therefore, argues that fairness must be redefined not as a statistical property of a model, but as an ethos of relational accountability embedded in the design, use, and governance of technological systems. Only when machine learning is grounded in such ethical commitments, centred on memory, repair, and collective imagination, can it serve as a tool for organisational transformation rather than institutional reinforcement.

Conclusion

This study has explored the use of machine learning as a lens to examine and intervene in the reproduction of bias within organisational systems. Beginning with a critical assessment of algorithmic decision-making in hiring, compensation, and performance evaluation, the research moves beyond statistical approaches to fairness and toward a decolonial rethinking of justice, equity, and organisational design. The empirical findings confirmed that machine learning models trained on biased data can, and do, amplify historical patterns of racial and gender exclusion. While bias mitigation techniques such as reweighing, adversarial debiasing, and post-processing adjustments can reduce these disparities, they remain limited in scope. They are, at best, technical solutions to ethical problems, often applied after systems have already been designed and deployed without meaningful engagement with those most impacted.

What is required, and what this study calls for, is a shift in orientation: from fairness as a statistical goal to justice as a structural, relational, and epistemic commitment. A decolonial perspective challenges the very foundations of how organisations produce and value knowledge, how they define merit and neutrality, and how they decide whose experiences are reflected in data, and whose are erased. Machine learning is not inherently oppressive or liberatory. It is a site of struggle, a terrain where organisational values, historical legacies, and future possibilities collide. When wielded critically, it can surface deep patterns of exclusion, reveal institutional blind spots, and support new modes of relational accountability. However, this requires organisations to move beyond compliance and metrics, and into the uncomfortable yet necessary work of reimagining themselves.

Finally, this study offers not just a critique of algorithmic bias, but a vision for how technical systems might be reclaimed as tools of collective transformation. By integrating decolonial theory with machine learning practice, it opens the possibility of organisational systems that do not merely optimise for efficiency, but instead reflect dignity, equity, and plurality in their design and intent.

References

Aguiar, C. E. S., & da Silva, D. K. M. (2024). Artificial intelligence and decoloniality: Insurgent arrangements and the question concerning cosmotechnics. Digital Theory, Culture & Society, 2(2), https://doi.org/10.61126/dtcs.v2i2.49

Akpan, A. A. F. (2023). Decolonising Algorithms: Towards the Making of Epistemically Just Algorithms. University of Johannesburg (South Africa). https://search.proquest.com/openview/b269d9b8cca990edd383faa718e767eb/1?pq-origsite=gscholar&cbl=2026366&diss=y

Alba, J. T. (2024). Insights into Algorithmic Decision-Making Systems via a Decolonial-Intersectional Lens: A Cross-Analysis Case Study. Digital Society, 3(3), 58. https://doi.org/10.1007/s44206-024-00144-9

Balasubramanian, N., Ye, Y., & Xu, M. (2022). Substituting human decision-making with machine learning: Implications for organizational learning. *Academy of Management Review*, *47*(3), 448-465. https://doi.org/10.5465/amr.2019.0470

Baradaran, A. (2024). Towards a decolonial I in AI: mapping the pervasive effects of artificial intelligence on the art ecosystem. Ai & Society, 39(1), 7-19. . https://doi.org/10.1007/s00146-023-01771-5

Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., … Varshney, K. R. (2019). *AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias*. *IBM Journal of Research and Development*, 63(4/5), 4:1–4:15. https://doi.org/10.1147/JRD.2019.2942287

Berthet, V. (2022). The impact of cognitive biases on professionals' decision-making: A review of four occupational areas. Frontiers in Psychology, 12, 802439. https://doi.org/10.3389/fpsyg.2021.802439

Besse, P., del Barrio, E., Gordaliza, P., Loubes, J.-M., & Risser, L. (2020). A survey of bias in machine learning through the prism of statistical parity for the Adult data set. The American Statistician, 76(2), 188–198. https://doi.org/10.1080/00031305.2021.1952897

Boccio, R. (2022). [Review of the book *Race After Technology: Abolitionist Tools for the New Jim Code*, by Ruha Benjamin]. *Configurations 30*(2), 236-238. https://dx.doi.org/10.1353/con.2022.0013

Chaka, C. (2022). Digital marginalization, data marginalization, and algorithmic exclusions: A critical southern decolonial approach to datafication, algorithms, and digital citizenship from the Souths. Journal of e-Learning and Knowledge Society, 18(3), 83-95. https://doi.org/10.20368/1971-8829/1135678

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794). https://doi.org/10.1145/2939672.2939785

Coleman, B. (2023). Human–Machine Communication, Artificial Intelligence, and Issues of Data Colonialism. The SAGE handbook of human-machine communication, 350-356. https://www.torrossa.com/gs/resourceProxy?an=5543084&publisher=FZ7200#page=397

Couldry, N., & Mejias, U. A. (2023). The decolonial turn in data and technology research: What is at stake and where is it heading?. Information, Communication & Society, 26(4), 786-802. https://doi.org/10.1080/1369118X.2021.1986102

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *20*(2), 215-232. https://doi.org/10.1111/j.2517-6161.1958.tb00292.x

Darioshi, R., & Lahav, E. (2021). The impact of technology on the human decision-making process. *Human Behavior and Emerging Technologies*, *3*(3), 391-400. https://doi.org/10.1002/hbe2.257

Davidson, L. J., & Kelly, D. (2020). Minding the gap: Bias, soft structures, and the double life of social norms. *Journal of Applied Philosophy*, *37*(2), 190-210 https://doi.org/10.1111/japp.12351

de Souza, S. P., Smith, H. M., & Taylor, L. (2024). Decolonial data law and governance. Technology and Regulation, 2024, 1-11. https://doi.org/10.71265/rvwfyt51

DeVos, A., Dhabalia, A., Shen, H., Holstein, K., & Eslami, M. (2022, April). Toward User-Driven Algorithm Auditing: Investigating users' strategies for uncovering harmful algorithmic behavior. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (pp. 1-19). https://doi.org/10.1145/3491102.35174

Ding, F., Hardt, M., Miller, J., & Schmidt, L. (2021). Retiring Adult: New datasets for fair machine learning. In Advances in Neural Information Processing Systems (Vol. 34, pp. 6478–6490). https://doi.org/10.24432/C5XW20

Dua, D., & Graff, C. (2017). *UCI Machine Learning Repository* [Dataset]. University of California, Irvine, School of Information and Computer Sciences. Retrieved from https://archive.ics.uci.edu/ml/machine-learning-databases/adult/

Duranti, L., Faniel, I. M., Horsman, P., Katuu, S., MacNeil, H., & Shepherd, E. (2023). Policy advice and best practices on bias and fairness in AI. Ethics and Information Technology, 26(1). https://doi.org/10.1007/s10676-024-09746-w

Fagan, J. F., & Holland, C. R. (2002). Equal opportunity and racial differences in IQ. *Intelligence*, *30*(4), 361-387. https://doi.org/10.1016/S0160-2896(02)00080-6

Faheem, M. A., Anwer, S., Rayhan, Z., Ullah, M. A., Paudel, R., Ahmed, M. F., & Khan, H. (2024). AI-Driven Innovation In HRM And Its Impact On Business Management: An In-Depth Study Of Technology Advancement And Strategic Implementation. Nanotechnology Perceptions, 20, 1174-1204.

Fırıncı, Y. (2024). Decolonial Artificial Intelligence; Algorithmic Fairness in Alignment with Turkish and Islamic Values. Marmara Üniversitesi İlahiyat Fakültesi Dergisi, 67(67), 250-279. https://doi.org/10.15370/maruifd.1565884

Friis, S., & Riley, J. (2023). Eliminating algorithmic bias is just the beginning of equitable AI. *Harvard Business Review*, *29*

Grimmelikhuijsen, S., & Meijer, A. (2022). Legitimacy of algorithmic decision-making: Six threats and the need for a calibrated institutional response. Perspectives on Public Management and Governance, 5(3), 232-242. https://doi.org/10.1093/ppmgov/gvac008

Grunkemeier, G. L., & Wu, Y. (2007). What are the odds?. *The Annals of thoracic surgery*, *83*(4), 1240-1244. DOI: 10.1016/j.athoracsur.2006.12.080

Hardt, M., Price, E., & Srebro, N. (2016). *Equality of opportunity in supervised learning*. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 29, pp. 3315–3323). Curran Associates, Inc.

Hristov, I., Camilli, R., & Mechelli, A. (2022). Cognitive biases in implementing a performance management system: behavioral strategy for supporting managers' decision-making processes. Management research review, 45(9), 1110-1136. https://doi.org/10.1108/MRR-11-2021-0777

Ibitoye, A. O., Onime, C., & Zaki, N. D., Onifade, O. F. (2022). Socio-Transactional Impact of Recency, Frequency, and Monetary Features on Customers' Behaviour in Telecoms' Churn Prediction. Iraqi Journal for Computer Science and Mathematics, 3(2), 101-110.

Ibitoye, A., & Olufade, F. (2021). Academic staff churn prediction for strategic decision support in quality higher education. *Transition from Observation to Knowledge to Intelligence (TOKI 2021)–Human–Data Interaction in an Artificial World*, 103-113.

Kamiran, F., & Calders, T. (2012). Data preprocessing techniques for classification without discrimination. *Knowledge and Information Systems, 33*(1), 1–33. https://doi.org/10.1007/s10115-011-0463-8

Kamran, A. (2023). Decolonizing Artificial Intelligence: Unveiling Biases, Power Dynamics, and Colonial Continuities in AI Systems. RMS journal. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4610643

Kolade, O., Egbetokun, A., & Owoseni, A. (2025). Chapter 1: Gen AI for research: Revolution or risk? In O. Kolade, A. Egbetokun, & A. Owoseni (Eds.), *Generative AI in research: Applications in research design, data analysis and feedback* (Chapter 1). Springer.

Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: review, synthesis, and future research directions. European Journal of Information Systems, 31(3), 388-409. https://doi.org/10.1080/0960085X.2021.1927212

KPONYO, J. J., FOSU, D. M., OWUSU, F. E. B., ALI, M. I., & AHIAMADZOR, M. M. (2024). Techno-neocolonialism: an emerging risk in the artificial intelligence revolution. *Trayectorias Humanas Trascontinentales*, (18). https://doi.org/10.25965/trahs.6382

Kunjumuhammed, S. K. (2024). Adoption of artificial intelligence in corporate finance: Addressing bias and ethical considerations. In T. D. R. Venkatesh & V. S. Janakiraman (Eds.), *Risks and challenges of AI-driven finance: Bias, ethics, and security* (pp. 1–16). IGI Global. https://doi.org/10.4018/979-8-3693-2185-0.ch001

Li, L., Lin, J., Ouyang, Y., & Luo, X. R. (2022). Evaluating the impact of big data analytics usage on the decision-making quality of organizations. Technological Forecasting and Social Change, 175, 121355. https://doi.org/10.1016/j.techfore.2021.121355

Loukas, O., & Chung, H. R. (2023). Demographic parity: Mitigating biases in real-world data. *arXiv preprint arXiv:2309.17347*. https://doi.org/10.48550/arXiv.2309.17347

Messeri, L., & Crockett, M. J. (2024). Artificial intelligence and illusions of understanding in scientific research. *Nature*, *627*(8002), 49-58. https://doi.org/10.1038/s41586-024-07146-0

Mignolo, W. D., & Walsh, C. E**.** (2018). *On decoloniality: Concepts, analytics, praxis*. Duke University Press. https://doi.org/10.1215/9780822371779

Mignolo, W. D., & Walsh, C. E. (2018). *On decoloniality: Concepts, analytics, praxis*. Duke University Press. https://doi.org/10.1215/9780822371779

Mollema, T. (2024). 'AI colonialism'is a conceptual metaphor (Master's thesis). https://studenttheses.uu.nl/handle/20.500.12932/47214

Mollema, W. J. T. (2024). Decolonial AI as disenclosure. https://doi.org/10.48550/arXiv.2407.13050

Muldoon, J., & Wu, B. A. (2023). Artificial intelligence in the colonial matrix of power. *Philosophy & Technology*, *36*(4), 80. https://doi.org/10.1007/s13347-023-00687-8

Nadeem, A., Marjanovic, O., & Abedin, B. (2022). Gender bias in AI-based decision-making systems: a systematic literature review. Australasian Journal of Information Systems, 26. https://doi.org/10.3127/ajis.v26i0.3835

Nori, H., Jenkins, S., Koch, P., & Caruana, R. (2019). *InterpretML: A unified framework for machine learning interpretability*. arXiv. https://doi.org/10.48550/arXiv.1909.09223

Nugroho, D., & Angela, P. (2024). The impact of social media analytics on sme strategic decision making. IAIC Transactions on Sustainable Digital Innovation (ITSDI), 5(2), 169-178. DOI: https://doi.org/10.34306/itsdi.v5i2.664

Oguntibeju, O. O. (2024). Mitigating artificial intelligence bias in financial systems: A comparative analysis of debiasing techniques. Asian Journal of Research in Computer Science, 17(12), 165-178. DOI: https://doi.org/10.9734/ajrcos/2024/v17i12536

Oliveira, N. H. D. (2024). A decolonial critical theory of artificial intelligence: intersectional egalitarianism, moral alignment, and AI governance. Filosofia Unisinos, 25(1), e25114. https://doi.org/10.4013/fsu.2024.251.14

Omotubora, A., & Basu, S. (2024). Decoding and reimagining AI governance beyond colonial shadows. In Handbook on Public Policy and Artificial Intelligence (pp. 220-234). Edward Elgar Publishing. https://doi.org/10.4337/9781803922171.00025

Pandey, A., Srivastava, N., & Gambhir, V. (2024). Mitigating Cognitive Biases in Organizational Decision-Making for Enhanced Effectiveness. Journal of Asia Entrepreneurship and Sustainability, 20(1), 129-180.

https://search.proquest.com/openview/03d3173c94d46cd9d0e2ced36df580ee/1?pq-origsite=gscholar&cbl=38850

Quijano, A. (2000). Coloniality of power, Eurocentrism, and Latin America. *International Sociology, 15*(2), 215–232. https://doi.org/10.1177/0268580900015002005

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big data & society*, *4*(2), 2053951717738104. https://doi.org/10.1177/2053951717738104

Sun, Y., & Jung, H. (2024). Machine Learning (ML) Modeling, IoT, and Optimizing Organizational Operations through Integrated Strategies: The Role of Technology and Human Resource Management. *Sustainability*, *16*(16), 6751. https://doi.org/10.3390/su16166751

Tacheva, Z., & Ramasubramanian, S. (2024). Challenging AI empire: Toward a decolonial and queer framework of data resurgence. Authorea Preprints. DOI: 10.31124/advance.22012724.v1

Tobia, K. (2017). Disparate statistics. *The Yale Law Journal*, 2382-2420. https://www.jstor.org/stable/45223096

Weerts, H., Dudík, M., Edgar, R., Jalali, A., Lutz, R., & Madaio, M. (2023). *Fairlearn: Assessing and improving fairness of AI systems*. *Journal of Machine Learning Research, 24*, Article 257. https://doi.org/10.48550/arXiv.2303.16626

Zafar, M. B., Valera, I., Gomez Rodriguez, M., & Gummadi, K. P. (2017). Fairness beyond disparate treatment & impact. In *Proceedings of the 13th Conference on Web and Internet Economics (WINE '17)* (pp. 1–22). Springer. https://doi.org/10.1145/3038912.3052660

Zembylas, M. (2023). A decolonial approach to AI in higher education teaching and learning: Strategies for undoing the ethics of digital neocolonialism. Learning, Media and Technology, 48(1), 25-37. https://doi.org/10.1080/17439884.2021.2010094

Zhang, B. H., Lemoine, B., & Mitchell, M. (2018). Mitigating unwanted biases with adversarial learning. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 335–340). https://doi.org/10.1145/3278721.3278779