Journal of Theoretical Accounting Research
ISSN: 0253-7214
Volume 22 Issue 1s Year 2026 Page 01-12

Bias, Fairness and the Machine in the Sustainable Digital Ecosystem: An Analytical Survey on Confronting Inequality in the Digital World

Prof.(Dr) SK Bose[1*], Mr. Deepak Bansal[2], Prof. (Dr) Yogita Sharma[3], Dr. Aditi Chaudhary[4]

[1*]School of Law, Manav Rachna University, skbose@mru.edu.in
[2]Assistant Professor, School of Law, Manav Rachna University
[3]School of Business, Manav Rachna University
[4]School of Law, Manav Rachna University

Abstract
The bursting growth of digital technologies has changed the way we classify, assess risk and allocate resources. Algorithmic models are increasingly operating as infrastructures of measurement in contemporary institutions. This research builds a theoretical accounting, derived accounting perspective to explore algorithmic bias and fairness in the sustainable digital ecosystems. Through analytical synthesis of interdisciplinary scholarship and documented empirical cases, the paper conceives algorithmic systems as extensions of accounting infrastructures with functions of recognition, measurement and classification. It argues that the sorts of biases built into the processes of data selection, proxy-based measurement and optimization represent forms of representational distortion like misstatement in accounting systems. Integrating together the distributive justice, capability theory and accountability scholarship, the study shows that the issue of algorithmic fairness is not a technical problem, but essentially a governance and institutional design problem. Comparative analysis on regulatory developments shows emerging dependency on risk-based classification, disclosure requirements and audit like oversight, and ongoing lack of developed areas of enforceability and contestability. The paper concludes that sustainable algorithmic governance would need structured accountability mechanisms that embed auditability, participatory oversight and contextual regulatory integration which would extend measurement, audit, and sustainability accounting theory into the field of algorithmic decision-making.

Keywords: Algorithmic bias, Algorithmic accountability, Auditability, Sustainable digital ecosystems, Regulatory governance.

1. Introduction

The integration of algorithmic systems into decision-making infrastructures has been progressing rapidly to change the architecture of measurement, classification and resource allocation of contemporary societies. From credit scoring and risk assessment to hiring analytics and welfare distribution, the functions performed by machine learning models can be seen to be similar and, in many cases, extensions of the traditional accounting process. These systems measure behaviour, attribute levels of risk, classify individuals and organize access to economic and social opportunity. While usually framed as neutral and efficiency-boosting instruments, algorithmic systems are powerful mechanisms of representation and representational institutionality’s. As has been noted by accounting scholarship for many years now, systems of measurement do not passive reflect reality, they actively create it (Hines, 1988).

Nevertheless, it is now evidenced by a growing literature in the interdisciplinary field that algorithmic systems tend to reproduce and amplify structural inequalities that exist. Empirical studies of the effects of algorithmic systems in areas such as criminal justice, healthcare, or labour markets have uncovered enduring patterns of bias in the output of such algorithms (Chouldechova, 2017; Obermeyer et al., 2019). These findings call into question the question of naturalism, as in computational systems, as well as bringing the normative assumption(s) of the design of algorithms to the table for discussion. From an accounting perspective such distortions increase basic questions with respect to accountability, transparency and fairness in digitally mediated structures of governance. If accounting can be thought of as a system for making activities visible and calculable then biased algorithms are distorted accounting practices which can undermine distributive justice as well as institutional legitimacy.

Theoretical accounting research has long focused on the idea that accountability is not only a technical exercise but a social and political process and it is influenced by the relations of power (Roberts, 1991). Contemporary algorithmic governance makes these concerns worse as it locks in to institutional decision frameworks which harbour opaque models with little to no avenues for contestation. Messner (2009) points to suggested inherent bounds of accountability mechanism as there are constraints on transparency and dialogue. Audit systems can become ritual forms of verification and hide, instead of uncover, substantive risks (Power, 1997). Applied in the case of algorithmic systems, because of this insight, fairness audits and compliance frameworks risk becoming symbolic exercises if they are not preceded by deeper theoretical reflection. Accordingly algorithmic bias is not limited to a narrow technical problem of correction and has to be situated in more general debates about accountability, representation and sustainability in the accounting systems.

In the case of sustainable digital ecosystems, the interests of algorithmic governance are pronounced. Sustainability accounting scholarship highlights the need for assessment of economic activity that is not limited to its monetary aspect, but also its consequences to the social and environmental condition (Gray et al. 1995). A digital system that systematically disadvantages specific groups of people by means of biased predictive modelling can hardly be considered sustainable in any sense. Corporate social reporting research also underscores the dangers of reputational and legitimacy risks associated with institutional practices that are perceived to be unjust (Bebbington et al., 2008). Algorithmic systems used in financial services, public administration and for corporate governance are therefore of great importance to the trust and social stability of the institutions themselves in the long term.

Recent research in the ethics of machine learning has observed that there are a range of sources of bias that are built throughout the algorithmic life cycle, including historical bias/ data selection bias, representation bias, proxy bias (Mehrabi et al., 2021). These types of biases are in many cases a reflection of existing social inequalities that is baked in training data and propagated within optimization processes. Such biases, from an accounting perspective are similar to the systematic misstatement of measurements of certain privilege over others. The normative question is therefore not just, whether algorithms are accurate or not, but whether they are fair and accountable in the context of the institutions. As put forward by Floridi et al, (2018), in order for ethical frameworks for AI, it should include principles of fairness, transparency, and human oversight. Yet, unless these principles rest on a coherent structure of established accountability theory, the regulatory responses will be in danger of being conceptually incoherent.

Against this background, the present work joins the international discussion about algorithmic systems should be thought of as an extension of accountancy infrastructural in sustainable digital ecosystems. The major objective of this research is to create a theoretical accounting structure and gain an understanding of algorithmic bias and fairness. Specifically, the aim of the study is (1) to examine the functionality of the algorithmic decision systems as mechanism of measurement and representation; (2) to assess, to what extent, these systems reproduce the structural inequalities; and (3) to propose an accountability-oriented model for introducing fairness to digital system's governance structure. By identifying the presence of algorithmic bias in the theoretical traditions of accounting and accountability research, the study makes a contribution to the growing space of theoretical accounting research in emergent digital contexts.

The main hypothesis of this study is that algorithmic bias is a form of distorted representation of accounting that subverts institutional accountability and sustainability without proper governance and oversight mechanisms. Drawing on accountability theory (Roberts, 1991), audit theory (Power 1997) and sustainability accounting scholarship (Gray et al., 1995), the paper argues that justice in algorithmic systems is not possible though technical modification and that it is necessary to integrate structured accountability frameworks. This position is in response to sociotechnical scholarship that has warned us to be wary of computational approaches and supports demands for systemic institutional change (Selbst et al., 2019).

By referring to the research studies on algorithmic fairness (Chouldechova, 2017; Mehrabi et al., 2021 , ethical frameworks for artificial intelligence (Floridi et al., 2018) and the critical frameworks of accounting (Hines, 1988; Roberts, 1991), the study positions itself between the research studies of digital governance and theoretical accounting research. In so doing, it opens horizons of analytical analysis in accounting scholarship into the domain of algorithmic systems, or systems in which machines are understood not in terms of their positions as technical artefacts, but as institutional actors in and embedded in calculative regimes. The following sections expand on the theory that underlies this, look at empirical examples of the ways bias may be manifested and the model of governance that we propose in an attempt to promote fairness and sustainability within digital accounting infrastructures.

2. Review of Literature

The recent literature on the numerous ways in which algorithms can be biased and how accountability and transparency have failed to meet the challenge of ensuring algorithmic fairness are a symptom of growing recognition that digital systems are not neutral computational devices but sociotechnical systems embedded in the institutional practices that shape institutional outcomes. Early empirical interventions of studies of algorithmic fairness have shown that ML systems perpetuate demographic disparities in a way that is systematic to the use of historical, skewed datasets to train models. Buolamwini and Gebru (2018) in their seminal paper on commercial gender classification systems showed that there are significant differences in intersectional error rate, in particular between dark-skinned women and everyone else. Their results revealed that facial recognition technologies were found to have an almost perfect accuracy of near 100% for light-skinned males while much higher error rates were found in marginalized demographic groups. This study has become fundamental in demonstrating how the representation bias that is embedded in the training data can result into structurally unequal outcomes.

At the same time as differences of technical performance, scholarships have tended to focus on structural and institutional aspects of algorithmic governance in recent years. Ananny and Crawford (2018) challenge the current assumption that one can only have accountability with transparency in algorithmic systems only. They imply that advocacy for transparency of algorithms tends to oversimplify the complexity of sociotechnical systems and do so on the assumption that visibility is a necessary factor that leads to understanding and/or control. Their deliberation allows that the obscuring of underlying power relationships can be achieved by transparency regimes whereby they do not take into consideration institutional settings, proprietary limitations, and epistemic asymmetries among developers and populations that are being impacted. This critique has important implications for models of governance that are based primarily on the principle of disclosure as the primary mechanism of accountability.

The shortcomings of frameworks based on transparency link directly to wider theoretical arguments on accountability. Bovens (2007) conceptualizes accountability as a mechanism involving a relationship between actor and forum where accountability is owed that the actor is obliged to explain and justify his or her conduct before a forum that has the power to question and assess that conduct. This framework makes a distinction between informational transparency and substantive accountability in that it refers to the importance of answerability and enforceability. When used to think about algorithmic systems this model of Bovens introduces an inherent tension, for many automated decision systems have not got clearly identified actors who can in any meaningful way justify outcomes to those affected, thus making traditional accountability mechanisms problematic. As a result, accountability gets diffused over developers, deploying institutions and regulatory bodies with the result that there are structural gaps in oversight.

The confluence of these lines of research are pointing to a change in the literature from purely technical definitions of fairness, to institutional and governance-oriented standards. Buolamwini and Gebru (2018) from measurable disparities in system outputs; Ananny and Crawford (2018) questions the limits of transparency as a remedy and then Bovens (2007) offers a normative framework to assessing if governance structure facilitates enough to ensure answerability and redress. Together, these works are building towards an emergent understanding on the role of algorithmic systems as being one that is imbedded in larger systems of accountability rather than standing as isolated technical artefacts.

Moreover, the ugly concept of intersectionality that is tacitly enabled by empirical studies of bias makes lone-axis studies of fairness more difficult. Buolamwini and Gebru's (2018) findings prove that the issue of errors disparities is not well explained by studying the issues of race or gender separately. It is rather compound identities, which lead to qualitatively different patterns of exclusion. This is an important insight that questions traditionally aggregated demographic based metrics-based auditing practices to suggest in the need for more granular accountability mechanisms that have the capacity to capture multidimensional inequities.

Simultaneously, Ananny an Crawford (2018) problematizes transparency when called in reference to performative nature of any regimes of disclosure. Transparency may be symbolical to suggest obedience without tackling the imbalance issues at the basic structure in terms of power and knowledge. This criticism has been consistent to accountability theory insistence on the need to have meaningful oversight that would require access to information instead of institutional arrangements that would facilitate contestation and corrective action (Bovens, 2007). In the case of algorithmic governance, for example, they may include independent bodies for auditing, mechanisms for review by regulatory bodies or oversight structures that are participatory.

Taken together, these contributions are indicative of an important conceptual development in literature. Early empirical demonstrations of bias (Buolamwini & Gebru, 2018) proved the existence of systematic discrepancies in the output of algorithms. Subsequent theoretical critiques of transparency (Ananny & Crawford, 2018) contended that there is doubt as to whether the dominant solutions to governance are adequate to address these disparities. Accountability theory (Bovens, 2007) offers an institutional framework for assessing institutional response and capacity for answerability and enforcement. This multilayered approach elaborates the fact that algorithmic fairness is not just a question of how fair a prediction model is or how universally bias mitigation techniques could be applied. Instead, it is essentially a matter of the organization of governance in digital ecosystems, and what institutional framework to provide accountability with.

Digital systems can make structural inequalities worse if they are not made in these ways, pretending as they are mere technology and not covert ways of reconciling them. The challenge is therefore how to find ways of reconciling technical interventions with institutional accountability mechanisms that can allow for fairness, answerability and legitimacy in algorithmic decision-making environments.

3. Theoretical Understanding of Bias, Fairness, and the Machine

3.1 Conceptualizing Algorithmic Bias

Algorithmic bias can be defined as the ability to obtain results that wrongly give particular individuals or groups a advantage or disadvantage and are systematic and repeatable. Unlike human prejudice: prejudice is the result of conscious or unconscious cognitive processes, whereas algorithmic bias is the product of the interactions between technical design choices and social structures that are embedded in the lifecycle of machine learning systems. These include the makeup of training data and representativeness, feature selection, objective function specification and deployment context.

From an accounting point of view, algorithmic systems are measurement and classification systems that frame economic and institutional decision-making. As has long been argued in academic study of accounting, systems of measurement neither merely reflect reality but help to create it (Hines, 1988). When algorithmic models allocate risk scores, classify applicants or predict behaviour, they are a kind of calculating machines similar to accounting systems that make individuals and activities visible in quantified form. Consequently, we can view algorithmic bias having the same kind of distortion in representational accuracy as systematic misstatements within the financial reporting or performance measurement systems.

The modern taxonomy of algorithmic bias draws distinctions between a number of overlapping categories. Historical bias occurs when a training data represents the existing structural inequalities. Representation bias is when some groups are not well represented in data sets, which means that the model will not be able to perform well. Measurement bias is caused by the use of proxy variables which imperfectly measure the construct of interest while relating to protected characteristics. Aggregation bias occurs when a single model is applied across different heterogeneous populations with different relationships. Evaluation bias is when benchmark datasets being used for validation are not a representative representation of the real world (Mehrabi et al. 2021).

Viewed from a theoretical accounting perspective, such categories look similar to failures in data integrity, inappropriate aggregate practices and flawed benchmarking standards. Just as the accounting system requires consistent standards of recognition, measurement and disclosure to make them comparable and fair, the algorithmic system requires structured governance systems to make sure that they are not distorted in terms of representation.

3.2 Theories of Fairness for Algorithmic Systems

The notion of fairness in algorithmic decision making has a normative nature and has some parallels with discussions that already exist in the accounting theory on the issue of distributive justice, accountability and stakeholder equity. Rawlsian distributive justice as it is based on the difference principle, involves the structuring of inequalities so as to maximize the lot of the least advantaged. Applied to algorithmic systems this principle is opposed to the application of financial scoring or allocation system which would systematically disadvantage vulnerable populations from such processes unless they can be shown to be demonstrably justified by the greater institutional good. In the current accounting language, it throws up questions of equitable distribution of resources and equitable presentation in frameworks for risk modelling.

Capability approaches popularized by Sen (1999) and Nussbaum (2011) steer attention from the distribution of resources and instead look to the extension of substantive freedoms. Within algorithmic governance, this view is based on whether digital measuring systems are enhancing or inhibiting people's access to employment, credit, healthcare and education. From an accounting's perspective, what capability-oriented fairness stipulates is an ambitious need that calculative systems don't, by accident, reduce economic participation by unjust, or biased, classification practices.

Critical race theory and feminist technoscience puts the structural dimensions of inequality within data practices. Benjamin (2019) shows how apparently neutral technological systems invest racial hierarchy and feminist data scholarship shows how power works through classification and categorization processes. Algorithmic fairness, therefore, requires an evaluation of the socio-economic assumptions that are contained in data infrastructures.

Procedural justice theory brings in a governance dimension in that it focuses on transparency, participation, and contestability. Accountability scholarship identifies accountability as an interaction between actors that require them to explain their conduct to a forum that has the ability to evaluate and punish (Bovens, 2007). Applied to algorithmic systems, fairness needs to not only refer to equitable outcomes, but also to institutional structures that allow for explaining, overseeing and redressing outcomes. This dimension is closely aligned to the theory of audit and regulatory governance.

Utilitarian methods do focus on aggregate welfare maximization and are still a powerful force in optimizing industry practices. However, tensions are created by the objectives of efficiency versus the goals of minority protection. Kantian deontological ethics goes further in insisting that the individual person must not be treated as a data point in the optimization process. Together, these frameworks in Table1, show that algorithmic fairness is not limited to statistical measures but needs to be implemented as part of wider accountability mechanisms for digital accounting infrastructures.

Table 1: Theoretical Frameworks for Algorithmic Fairness

Theoretical Framework	Core Principle	Fairness Criterion	Accounting Application
Rawlsian Distributive Justice	Maximize advantage of worst-off	Equality of outcomes	Credit scoring, resource allocation reporting
Capability Approach	Expansion of substantive freedoms	Enhancement of capabilities	Access to financial services, employment screening

Critical Race Theory	Structural analysis of power	Anti-subordination	Risk modelling, financial classification systems
Feminist Technoscience	Power in data and design	Intersectional equity	Hiring analytics, performance measurement
Procedural Justice	Transparent and participatory governance	Accountability and contestability	Audit systems, regulatory oversight

3.3 The Machine in the Sustainable Ecosystem

Placing the machine within the sustainable digital ecosystem entails the recognition of algorithmic systems as part of the institutional structures of accounting and governance. Rather than isolated technical artefacts, these systems are calculative infrastructures for the enactment of financial allocation, risk and organizational accountability. Environmental sustainability issues are not limited to energy consumption, but are also related to sustainability accountancy and carbon disclosure. The computational intensity of large-scale models raises some implications on environment reporting frameworks and measurement practices of ESGs. As organizations rely more and more on AI systems to integrate into the organization's operations, the carbon footprint of digital infrastructures becomes an issue of corporate accountability and transparent reporting.

Social sustainability has an overlay with accountability structures, too. Algorithmic systems rely on high amounts of data labour and institutional control mechanisms. Whereas governance frameworks are weak with consequent accountability gaps, there is a potential for externalising social costs on vulnerable populations. From a theoretical accounting perspective, sustainable digital ecosystems need to have structures in place that internalise these externalities through the reporting structures and oversight.

Economically, algorithmic governance has effects on sharing financial opportunities and market power. Automated decision systems have the potential to contribute to asymmetries in access to credit, employment and capital markets. Without strong accountability structures, these systems risk destroying the long-term institutional legitimacy. Thus, sustainable algorithmic governance must incorporate fairness principles as part of the logic of accounting, auditing and regulatory infrastructures to ensure that digital measuring systems are stimulated within greater societal objectives. In Figure1 sustainable algorithmic governance would need continuity of fairness with accounting, auditing and regulatory infrastructures to ensure digital measurement systems as subordinated to broader societal objectives.

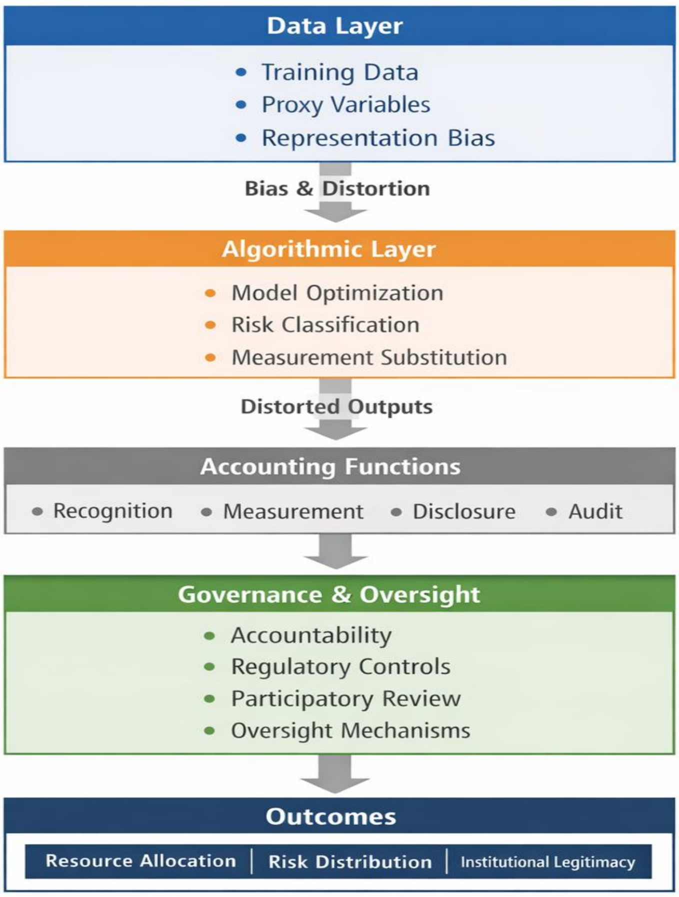

Figure 1: Algorithmic Systems as Accounting Infrastructure Model

3.4 Theoretical Contribution to Accounting Research

This research is a contribution to theoretical accounting research taking fundamental notion from the area of accounting into another area of algorithmic governance. Traditional accounting theory has for a long time examined the role of systems of recognition, measurement and disclosure in constructing organizational and economic reality (Hines, 1988; Hopwood, 1987). However, most of the research action currently taking place in the accounting literature has been represented by the financial reporting systems, the performance metrics, and the disclosure of sustainability in the corporate context. It has paid comparatively little attention to the role of algorithmic decision systems that have become in more institutional domains the de facto infrastructural of measurement and classification.

First, the paper promotes the theory of measurement by an understanding of the algorithmic bias as a form of representational distortion just like systematic misstatement. By revealing the significance and effect of variables such as proxy variables, criteria of optimization as well as the aggregation practice in structuring distribution results, the study moves beyond financial quantification presented in accounting measurement to computational risk classification and prediction modelling.

Second, the study helps to strengthen the accountability theory by looking to the relational context of answerability and contestability (Roberts, 1991; Bovens, 2007). Algorithmic infrastructures help distribute responsibility amongst approaches comprising of developers, institutions, and regulators, and have the effect of complicating conventional accountability mechanisms. This paper describes the ways in which digital systems create accountability gaps; and offers governance integration to correct this.

Third, the paper adds some audit theory: it reconceptualises fairness auditing as an emergent form of algorithmic assurance. The information case presented for the purpose of the study draws based on Power's (1997) analysis of audit expansion, makes the case of bias audits becoming subject to the risk of symbolisms in the absence of their being embedded in structured regimes of oversight. This amounts to an extension of audit scholarship to computational environments in which systemic opacity creates problems for traditional models of verification and assurance.

Finally, the paper brings in a social justice and environmental thinking to digital governance through the lens of sustainability accounting. Drugs and Pharmaceuticals Prioritizing equity in algorithmic decision-makingfairness by algorithmcommon designs devise allocation strategies that typically rely on a preexisting distribution of resources. Sustainable computing environments, algorithms and drugs By emphasizing both environmental consequences of computational fairness and the distributive consequences of algorithmic allocation, the research places digital infrastructures in broader frameworks of sustainability reporting and ESG accountability. Taken together, these sources re-think the understanding of algorithmic systems as institutional accounting infrastructures and form a conceptual basis with which future studies of digital measurement, governance, and legitimacy may be conducted.

4. Analysis of Data Development and Cases of Clash with Inequality

4.1 The Data Pipeline and Structural Inequality

Datasets used to train machine learning models are extracted from societies that have long-running structural inequalities of race, gender, class, caste, disability and geography. When these datasets are conceived of as objective representations as opposed to those that are social constructions, algorithmic systems are able to reproduce and legitimize existing hierarchies. Buolamwini and Gebru (2018) provided seminal empirical evidence of this kind of dynamic exhibiting that commercial facial recognition systems had substantially higher error rates for dark-skinned women than for light-skinned men due to underrepresentation in training datasets.

From an accounting point of view, the data pipeline works similarly to the processes of data recognition and classification in financial reporting systems. Training sets indicate what is being recognized, how is it being categorized, and how is it being subsequently evaluated. Representation bias therefore is similar to systematic mismeasurement within the accounting systems where the measurement data is incomplete or skewed, thus generating distorted output. Measurement bias caused by proxy variables also follows in a similar way as with the use of imperfect indicators in performance measurement frameworks. When proxy variables are used in place of substantive constructs, they can contain the encoding of structural inequalities, while being technically neutral.

Geographic data production concentration makes representational distortions even worse. Datasets which have been produced mainly in North American and Western European settings may fail to fully represent social and economic realities in other settings, and models produced in this way may underperform in diverse institutional environments. In accounting terms such distortions are the inappropriate benchmarking standards not taking into account variation in the context. Consequently, bias within the data pipeline cannot be explained just as technical error, but broken down as representational integrity broken down within digital measurement infrastructures.

4.2 Documented Instances of Algorithmic Discrimination

Institutional research is documented for algorithmic discrimination in many different areas of institutions like systematically failing risk modelling and resource allocation. In criminal justice, risk assessment tools like COMPAS caused much controversy because of racial disparity in false positive rates. Chouldechova (2017) showed statistical incompatibility of a range of fairness criteria if the base rates vary between the groups, which points to the normative nature of algorithmic evaluation.

From an accounting point of view, this case shows in Table 2, how systems for classified risk can lead to inequitable results where the standards of measure are not consistent with distributional goals.

In the domain of healthcare, Obermeyer et al. (2019) demonstrated that many currently used algorithms used the healthcare expenditure as a proxy for need to identify high-risk patients. Due to unequal access reflected by expenditure, Black patients were systematically assigned lower risk scores than equally sick white patients. This failure of the "proxy" is similar to errors of measurement substitution in accounting systems, in which there is a tendency to focus on financiality observable conditions that mask substantive conditions. The upshot of all this was an ill design of measurement with an incorrect allocation of healthcare resources.

Financial services are other examples of biased algorithmic classification. Automated credit & mortgage pricing systems have been found to be differential across demographic groups despite controlling for traditional modes of creditworthiness. Such patterns show the power of algorithmic scoring as a financial representation of structuring financial access, i.e. access to capital. When classification criteria reflect historical inequalities, models of risk are used as discriminatory accounting mechanisms which impact economic opportunity. Similarly, the reproductive practices of algorithmic hiring systems that have been trained on previously gender-skewed employment data have duplicated patterns of occupational segregation. These systems translate past institutional biases into predictive metrics with 'reinforcement of the institutionalized exclusion.' In each of the domains, algorithmic discrimination reflects distortions of calculative infrastructures that are akin to failures in processes of recognition, valuation, and disclosure in accounting systems.

Table 2: Documented Instances of Algorithmic Discrimination

Domain	System/Algorithm	Nature of Bias	Accounting Implication
Criminal Justice	COMPAS	Disparate false positive rates	Risk classification distortion
Healthcare	Care Management Algorithm	Expenditure proxy bias	Measurement substitution failure
Employment	Hiring Algorithm	Historical gender bias	Biased performance modelling
Computer Vision	Facial Recognition	Dataset underrepresentation	Recognition and classification error
Financial Services	Credit/Mortgage Scoring	Differential pricing outcomes	Discriminatory financial risk modelling

4.3 Confrontation Strategies and Institutional Responses

Recognition of the problem of algorithmic harm has spurred technical and institutional intervention to solve this concern by identifying ways of improving fairness and accountability. These interventions work similarly as internal controls mechanisms in the accounting systems by attempting to discover the distortion in the representation before they would be put in use.

However, if one is to rely only on technical adjustment, the risk of failing to address inequality sources that have a primarily structural character may be very high. Selbst et al (2019) warn against "solutionism" that metrics of fairness cannot address the problems of deeper socio-institutional imbalances that are imbedded within data generation processes. In accounting terms, rectification of computational outcomes without an interdiction on consideration of recognition standards and governance structures can treat symptoms and not attempt to get to the root of the issues.

Institutional responses have thus increased to include algorithmic impact assessment and guidelines in governance and oversight mechanisms evaluating the functioning of automated decision systems prior to implementation. Such measures are similar in some respects to ex ante audit and compliance frameworks that are meant to ensure that accountability is preserved. By incorporating assessments of fairness into formal regulatory frameworks and institutional governance systems, these strategies attempt to provide a way of incorporating and embedding ethical considerations into digital accounting infrastructures. Collectively, such approaches are indicative of a move away from reactive bias correction towards the structural integration of principles of governance. Due to the growing role of algorithmic systems as central elements of institutional regimes of measurement, fairness must be integrated into the design, deployment and oversight processes of algorithmic systems rather than solved through ex post technical adjustments.

5. Global Trends, Comparison, Case Studies, and Critical Analysis

5.1 Regulatory Landscape: A Comparative Assessment

The regulatory response to algorithmic bias across the world has been marked by considerable heterogeneity based on varying legal traditions, political priorities and state institutional capacities. The European Union has taken the lead as a regulatory force by having the Artificial Intelligence Act which creates a risk-based classification system for AI applications. And, high-risk systems deployed in areas such as employment, credit scoring, law enforcement, and education are subject to mandatory restrictions regarding transparency, human oversight, data governance, and non-discrimination. From an accounting point of view, this risk-based approach is similar to structured classification frameworks in financial reporting, in which activities are classified based on materiality and risk exposure, which is also how the intensity of governance is connected to assessed impact. Transparency and audit requirements provided by the Act serve the same way that disclosure and internal control mechanisms are intended to ensure accountability.

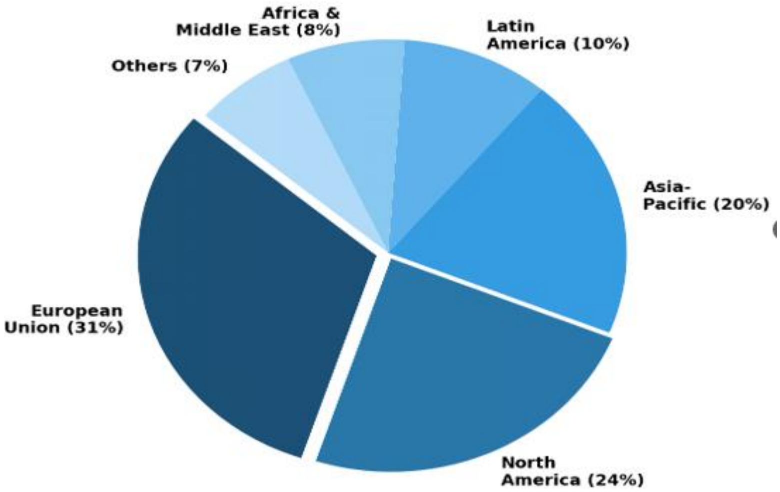

Figure 2: Compiled by the Author from OECD AI Policy Observatory (2024) and Stanford HAI AI Index Report (2024)

Compiled by the Author from: 1. Authority and Beneficiary Compulsions in Artificial Intelligence, Compulsions in Artificial Intelligence Per National Standards and Regulations, Migration Compulsions and Restrictions on Artificial Intelligence, Regulation of Artificial Intelligence Impacts of a Digital Divide in Artificial Intelligence (2021) 2. Artificial Intelligence Status and Trends, Report by Stanley M. Schneider, Artificial Intelligence Accelerator, Project on Technology, International Data Corporation (IDC). 3. Artificial Intelligence and E-Governance Trends in Digital Latrine Use Per National Survey and Research Efforts from Stanford University and University of California, The United States has taken a more patchwork approach on this through executive guidance with sector-specific regulation and litigation over regulation instead of wide legislation. Recent federal directions on safe and trustworthy AI set some safety requirements for fairness and safety, but are heavily dependent on how the agency performs them. Sectoral measures, e.g. bias audit requirements for automated decision tools at the employment level, are incremental steps on the way to a structured oversight. However, the lack of a coherent regulatory scheme results in patchy enforcement with similar elements of governance in the disclosure and audit requirements across sectors in financial regulation.

In the Asia-Pacific region policy approaches to algorithmic governance vary a great deal. China's regulation model for generative AI focuses on state control of AI and content control, which includes parts that are similar to centralized control mechanisms. By contrast, India's approach has been based largely on policy frameworks and guidance in the form of principles-based approaches, such as national artificial intelligence strategies. While the Digital Personal Data Protection Act (2023) is an important step towards having a statutory agenda on data governance, the fundamental concern of the Act focuses on the privacy aspect rather than being algorithmically fair. From an accountability point of view, lack of special audit and reporting requirements for high-risk algorithmic systems opens the door for governance gaps, especially in the public sector and financial applications.

Emerging economies in Latin America and Africa have their own set of regulatory issues, perhaps not least among these is the lack of institutional capacity and dependence on outside developed technologies that might not accurately reflect local social and economic contexts. Brazil's proposed legislation on AI, being consistent with its data protection regime, is a step forward for a rights-based form of governance. Similarly in Table 3, continental AI strategies in Africa appreciate structural vulnerabilities (dealing with data extraction and representational bias). Though issues of implementation and lack of resources make it difficult to develop strong accountability infrastructures. Across jurisdictions regulatory diversity is the result of different conceptions of risk classification, disclosure obligations and enforcement authority, highlighting the importance of harmonized principles of governance that could incorporate fairness within digital accounting systems.

Table 3: Comparative Analysis of AI Fairness Regulatory Frameworks

Jurisdiction	Primary Instrument	Approach	Enforcement	Governance Implication
European Union	AI Act (2024)	Risk-based regulation	Binding penalties	Structured audit and disclosure model
United States	Executive & sector laws	Fragmented oversight	Agency-based enforcement	Sectoral accountability gaps
China	Generative AI Measures	State-directed governance	Centralized enforcement	Strong supervision, limited contestability
India	Policy frameworks; DPDP Act	Principles-based	Limited binding force	Disclosure and audit deficit
Brazil	Proposed AI Bill	Rights-based model	Pending adoption	Emerging accountability integration
African Union	Continental AI Strategy	Development-focused	Aspirational	Capacity-dependent oversight

5.2 Case Study Analysis

5.2.1 COMPAS and the Fairness Impossibility Problem

The COMPAS recidivism prediction tool is the widely studied example of algorithmic bias in the criminal justice system. Academic evaluation showed that there were considerable differences in error rates among racial divisions, leading to a debate on what the standard of fairness should be. Chouldechova (2017) showed that if the base rates are different in different groups, it is mathematically impossible to simultaneously meet different fairness requirements, such as calibration and equal error rates. This result shows that evaluation of fairness entails normative choices as to which statistical standards must be prioritized. From the point of view of accounting this dispute is like conflicts between competing measurement standards in financial accounting, where different principles of valuation may give divergent representations. Risk scoring systems, whether they are financial or otherwise, lock into normative judgment in what appears to be a technical measuring system.

5.2.2 Healthcare Allocation & The Proxy Problem

In terms of healthcare resource allocation, Obermeyer et al. (2019) have found that a popular algorithm used healthcare expenditure as an estimate of patient need. Because expenditure was based on unequal access as opposed to underlying health status, Black patients were systematically given lower risk scores than other white patients who were equally ill. This case is an example of proxy bias - involving the use of observable financial indicators in the place of substantive conditions. In terms of accounting, such substitution is similar to the use of imperfect measurement proxies which obscure realities. The resultant misclassification resulted in unequal access to management programs of care and illustrates how faulty measurement design leads to the distortion of resource allocation within systems of institutional care.

5.2.3 The Dutch SyRI System and State Surveillance

The system, which was created in the Netherlands to detect welfare fraud, encompassed the Dutch System Risk Indication (SyRI) in which data from multiple sources of the government were combined to produce risk scores. The case reveals an overall lack of procedural accountability in this area, a lack of contestability, and a lack of disclosure of the criteria against which risk is assessed. From a governance perspective, SyRI demonstrates how open systems for automated decision making are often impossible without adequate audit and control systems, thus undermining the public's trust.

5.3 Critical Analysis

The quest for an algorithmic fairness in sustainable digital ecosystems leads to multiple tensions that are inter-linked. The first is the issue of efficiency and equity. Predictive accuracy or cost minimization are often the goals of algorithmic systems, which are both conflicting and may not promote fairness with respect to demographics. Resolving this tension requires normative judgment about the relative importance given to optimization of performance and distributive fairness - judgments similar to those faced in the standard-setting debates in accounting with respect to issues of measurement and materiality.

The second tension is of transparency or complexity. As the scale and opacity of machine learning models grow, the auditability of AI models decreases. Explainability techniques yield some insight and few result in complete interpretability. Under the theoretical accounting perspective, this poses certain problems involving the adequacy of disclosure-based governance where underlying models are not amenable to meaningful scrutiny and verification.

The third tension is between global standards and local situations. Fairness measures devised in particular institutional contexts may not reflect structural inequalities in other contexts. Having a good governance means thus having adaptive accountable systems that are able to strike a balance between standardization and context sensitivity. Embedding fairness into digital accounting infra-structures requires technical improvement but also long-term institutional management based on accountability theory.

6. Future Challenges and Limitations

6.1 The Foundation Model Challenge

The explosion of foundation models, generative language model of large scale, present challenges of algorithmic fairness and accountability of unprecedented nature. These models are trained on massive and often opaque corpora of text on the internet encoded with extant social biases, structural inequalities and representational distortions. Unlike the task-specific systems, the foundation models are general-purpose computational infrastructure which can be integrated across all kinds of institutional contexts, from financial services to governance to risk assessment. Given the complexity of their scale and architectural nature, the conventional audit mechanisms based on the relationships of training data on the model parameters and the downstream outputs are not immediately observable and traceable. From a theoretical accounting perspective, this opacity renders transparent and contestable measurement system - a core element of a good accounting system - impossible.
Moreover, in foundation models, emergent capabilities and emerging behavioral patterns may not be evident when the foundation models are being developed and validated but come to the fore when foundation models are deployed. Such unpredictability presents a problem for the standard ex ante compliance models based on pre-deployment testing and static audit procedures. Governing these systems therefore demands new and more comprehensive paradigms of oversight that include continuous monitoring, independent review of audits and adaptive governance frameworks that have the flexibility to respond to emerging risks. Embedding fairness within foundation model ecosystems requires the need for embedding accountability principles in continued operational controls instead of one-dimensional and outcome-based assessments. Absent such institutionalized structures of oversight, models of foundations run the risk of operating as opaque calculative infrastructures, whose distributive consequences are not sufficiently explored and poorly scrutinized.

6.2 The Sustainability Paradox of Computational Fairness

A tension of structures exists between the computational resources to get algorithmic fairness and the environmental sustainability of those computations. Fairness interventions, e.g. bias auditing, adversarial debiasing, training multiple model variants to access fairness-accuracy trade-offs etc. are also associated with high computational costs. In the context in which the volume of energy used and carbon emissions already caused by AI systems are already a cause of concern around sustainability, efforts to make computations fairer may also paradoxically create bigger environmental burdens of AI systems. This is a source of tension, a reason why goals of distributive justice and environmental responsibility are to be balanced in digital governance systems.

6.3 Limitations to Technocentric Approaches

An unending constraint in the algorithmic fairness discussion is the move to technocentrism, which is the belief that bias is inherently a technical issue that can be solved by improving algorithms, providing better datasets and more sophisticated auditors. While technical interventions are needed to address algorithmic bias, they are not enough to address the structural inequalities that produce algorithmic bias. As Selbst et al. (2019) noted, fairness interventions that focus solely on mathematical properties of algorithms but neglect the social context of their use run the risk of delivering solutions that on the technical level, fairness is considered, but on the social level of the deployment of an algorithm, it is not. A hiring algorithm that achieves statistical parity but inequality is perpetuated if the labor market per se is structured by discriminatory practices, which determine the pool of applicants, the definition of job qualifications and the evaluation of job performance.

6.4 Data Sovereignty and Digital Colonialism

The political economy of knowledge production and consumption of data at a global scale raises questions about digital colonialism, where data is mined from populations living in the Global South, and used to train artificial intelligence systems, which are controlled by corporations - in the Global North with the benefits going disproportionately to rich countries and the harms to marginalized communities. Couldry and Mejias (2019) characterized this dynamic as "data colonialism" arguing that the appropriation of human data represents a new form of colonial extraction. Addressing algorithmic fairness in the context of sustainability therefore means addressing data sovereignty and making sure communities maintain meaningful control over the collection, use and governance of data about them.

6.5 The Democratic Deficit in Algorithmic Rule

Finally, the governance of algorithmic systems is faced with a massive democratic deficit. The design and implementation of algorithms that have a powerful impact on life opportunities, from accessing credit and employment to criminal justice sentencing and healthcare, are usually determined behind closed doors by private companies with little public accountability. Affected communities tend to have little meaningful input into the design of systems that determine their lives and mechanisms for challenging algorithmic decisions do not exist in most jurisdictions. Addressing this democratic deficit requires the development of participatory framework models of governance, community oversight mechanisms and regulating frameworks that ensure meaningful deliberation of the public about the values embedded in algorithmic systems.
The difficulty of democratic control is coupled with the technical difficulty of algorithmic systems, which leads to asymmetries of information between the regime and the population concerned. Even well-intentioned transparency requirements may not be enough if

transparency demands too much information about technical matters to be judged meaningfully by non-specialist audiences. This raises the prospect of the need for genuine democratic accountability which requires not only transparency, but the interpretive infrastructure derived from such transparency, including independent auditing bodies, public interest technologists and community-based organizations with the technical capacity to evaluate algorithmic systems on behalf of affected populations.

6.6 Intersectionality and the Limits of Single-Axis Analysis

A further limitation on existing approaches to algorithmic fairness is that they often analyse bias along single demographic axes, looking at racial bias and gender bias separately rather than looking at intersections between the two. Crenshaw's original understandings of intersectionality require an acceptance that people occupying multiple marginalised positions, e.g. Black women or disabled persons from lower castes, may experience forms of algorithmic harm qualitatively different from those experienced along any one axis. Buolamwini and Gebru's (2018) finding that facial recognition systems performed worst for dark-skinned women is an example of this intersectional dynamic as the disparity could not be narrowed down to either just race or gender.

7. Conclusion

Based on this analytical examination of the issues of bias, fairness and machine driven inequality within sustainable digital ecosystems the study comes to the conclusion that algorithmic bias is not an accident in machines but a structural feature of digital measurement systems. Algorithmic models are akin to calculative infrastructures of categorising, assessing and distributing resources akin to accounting systems. And when these infrastructures are based on historically skewed data, and proxy variables, that is based on opaque optimization criteria, reproduce and magnify existing social inequalities. The theoretical frameworks examined - including those of distributive justice, capability approaches, critical structural perspectives - emphasized that algorithmic fairness is in essence normative and it cannot be reduced to purely technical correction.

The empirical evidence surveyed in the criminal justice field, the healthcare sector, the employment sector, the financial services industry and welfare administration exposes consistent patterns of representational distortion and unfair risk classification. From the point of view of theoretical accounting, these instances involve the failure of recognition, measurement and accountability mechanisms which are embedded within digital governance systems. Although there have been increasing efforts by the global regulatory environment to take action on fairness issues, enforcement efforts have been uneven and fragmented across jurisdictions. Regulatory responses have been mostly about transparency and compliance, but many have underdeveloped audit and oversight structures that are able to ensure sustained accountability.

The case analyses carried out further show that algorithmic bias is a frequent outcome of proxy mismeasurement, inappropriate benchmarking, and procedural deficiencies. The attempts to reduce such harms, either via judicial action, regulatory change, or fairness auditing, are efforts on the road to resumption of accountability in digital infrastructures. Looking towards the future, there are several interrelated challenges which guide the course of sustainable algorithmic governance. The obscurity and level of basis of the foundation models make the conventional approaches of auditors more complicated. The computational challenges associated with fairness interventions raise questions of environmental sustainability which touch on questions of sustainability accounting frameworks. Technocentric solutions are at risk of not attending to socio-institutional context, and there are global asymmetries in terms of control of data to ensure governance capacity and then accountability gaps. Additionally, no involvement of stakeholders in the process of designing algorithms speaks of a broader lack of democratic answerability.

This work therefore ventures that algorithmic fairness (i.e. sustainable algorithmic fairness) demands an integrated mode of governance featuring a tri-partite principle of (1) structured accountability mechanisms including auditability and disclosure in algorithmic systems (2) regulatory pluralism that engages with the contextual variation in different contexts but ensures a consistency in engagements with the non-discrimination and institutional responsibility and (3) participatory oversight frameworks that enhances contestability and stakeholder inclusion in digital measurement regimes.

Sustainable digital transformation is impossible without issues of representation and accountability. As algorithmic systems become more of an extension of accounting systems, ensuring their fairness requires engagement of relating normative commitments in measuring design, governance architecture and oversight practices. Advancing theoretical (and experimental) accounting research in the digital era thus implicates taking machines as not only technical artefacts but as institutional actors, whose calculative practices determine effects of control, of distribution and of societal legitimacy.

References

[1] Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society, 20(3), 973–989. https://doi.org/10.1177/1461444816676645

[2] Benjamin, R. (2019). Race after technology: Abolitionist tools for the new Jim code. Polity Press.

[3] Binns, R. (2018). Fairness in machine learning:

Lessons from political philosophy. In Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency (pp. 149–159). https://doi.org/10.1145/3287560.3287583
[4] Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. European Law Journal, 13(4), 447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x
[5] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency (pp. 77–91).
[6] Burchell, S., Clubb, C., Hopwood, A. G., Hughes, J., & Nahapiet, J. (1980). The roles of accounting in organizations and society. Accounting, Organizations and Society, 5(1), 5–27. https://doi.org/10.1016/0361-3682(80)90017-3
[7] Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. Big Data, 5(2), 153–163. https://doi.org/10.1089/big.2016.0047
[8] Citron, D. K., & Pasquale, F. (2014). The scored society: Due process for automated predictions. Washington Law Review, 89(1), 1–33.
[9] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People: An ethical framework for a good AI society. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5
[10] Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. ACM Transactions on Information Systems, 14(3), 330–347. https://doi.org/10.1145/230538.230561
[11] Gray, R., Kouhy, R., & Lavers, S. (1995). Corporate social and environmental reporting: A review of the literature and a longitudinal study. Accounting, Auditing & Accountability Journal, 8(2), 47–77. https://doi.org/10.1108/09513579510146996
[12] Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. Minds and Machines, 30(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8
[13] Hines, R. D. (1988). Financial accounting: In communicating reality, we construct reality. Accounting, Organizations and Society, 13(3), 251–261. https://doi.org/10.1016/0361-3682(88)90003-7
[14] Hopwood, A. G. (1987). The archaeology of accounting systems. Accounting, Organizations and Society, 12(3), 207–234. https://doi.org/10.1016/0361-3682(87)90038-9
[15] Irani, L. (2015). The cultural work of microwork. New Media & Society, 17(5), 720–739. https://doi.org/10.1177/1461444813511926
[16] Kitchin, R. (2017). Thinking critically about and researching algorithms. Information, Communication & Society, 20(1), 14–29. https://doi.org/10.1080/1369118X.2016.1154087
[17] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1–35. https://doi.org/10.1145/3457607
[18] Messner, M. (2009). The limits of accountability. Accounting, Organizations and Society, 34(8), 918–938. https://doi.org/10.1016/j.aos.2009.07.003
[19] Miller, P., & Power, M. (2013). Accounting, organizing, and economizing: Connecting accounting research and organization theory. The Academy of Management Annals, 7(1), 557–605. https://doi.org/10.5465/19416520.2013.783668
[20] Nussbaum, M. C. (2011). Creating capabilities: The human development approach. Harvard University Press.
[21] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342
[22] Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press.
[23] Power, M. (1997). The audit society: Rituals of verification. Oxford University Press.
[24] Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 469–481). https://doi.org/10.1145/3351095.3372828
[25] Roberts, J. (1991). The possibilities of accountability. Accounting, Organizations and Society, 16(4), 355–368. https://doi.org/10.1016/0361-3682(91)90027-C
[26] Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., & Aroyo, L. M. (2021). "Everyone wants to do the model work, not the data work": Data cascades in high-stakes AI. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1–15). https://doi.org/10.1145/3411764.3445518
[27] Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 59–68). https://doi.org/10.1145/3287560.3287598
[28] Sen, A. (1999). Development as freedom. Oxford University Press.
[29] Tinker, T., Merino, B., & Neimark, M. (1982). The normative origins of positive theories: Ideology and accounting thought. Accounting, Organizations and Society, 7(2), 167–200.