VFAST Transactions on Education and Social Sciences
https://vfast.org/journals/index.php/VTESS@ 2025, ISSN(e): 2309-3951, ISSN(p): 2411-0221
Volume 13, Number 3, July-September 2025 pp:55-72

From AI Ethics to AI Justice: A Comprehensive Framework for Equitable Governance in Education and Social Systems

Sawera Qureshi[1*],Hafiz Muhammad Irshadullah[2]

[1]Department of Computer Sciences Abdul Wali Khan University of Mardan, Kpk, Pakistan
[2]Assosiate Professor, Department of Education Abdul Wali Khan University of Mardan, Kpk, Pakistan

ABSTRACT
Artificial intelligence is embedded in educational systems and social structures, which creates ethical issues that cannot be resolved at the algorithm level only. This paper provides a critical review of general AI ethics and forms an AI justice angle based on relational equality. It systematizes the criticism into three spheres: (a) conceptual - misleading labels and narrow frames of understanding that conceal the infrastructural reach of AI; (b) substantive - ignored environmental, labour and distributional harms in AI life-cycle; and (c) procedural - advisory processes where the industry dominates and expertise is limited. Based on these diagnostics the paper proceeds with actionable reforms: use accurate language and situational analyses; consider non-technical options first, and then AI solutions; involve community and civil-society actors in deliberation; and instigate sound philosophical ethics into interdisciplinary teams. The contribution of the paper is a practical justice-driven roadmap that can be used by educators, policymakers and social scientists to assess AI implementation and direct the process toward more equitable and sustainable results.

KEYWORDS:
AI Ethics; AI Justice; Educational Technology; Algorithmic Governance; Relational Equality; Technosolutionism; Structural Inequality; Ethical AI Policy; Critical Data Studies; Sustainable AI Systems

JOURNAL INFO:
HISTORY: Received: August 03, 2025 Accepted: September 18, 2025 Published: September 30, 2025

*Corresponding author: sawera.cs@awkum.edu.pk
DOI: 10.21015/vtess.v13i3.2264

INTRODUCTION

Artificial intelligence (AI) has quickly integrated into the systems of education, government, and social welfare, influencing processes of learning, communication, and decision making. From adaptive learning platforms to automated administrative systems, AI is not a distant technological horizon but a present reality shaping everyday social life (Strielkowski, Grebennikova, Lisovskiy, Rakhimova, and Vasileva, 2025; Williamson and Piattoeva, 2022). The increasing ubiquity of AI has sparked extensive discussions under the label of AI ethics, a field concerned with issues of fairness, accountability, transparency, and privacy (Floridi and Cowls, 2019). However, even as AI ethics gains popularity, it remains conceptually fragmented and practically constrained, often reducing itself to checklist-style principles that inadequately address the deeper power dynamics shaping AI development and deployment (Crawford, 2021).

LIMITATIONS TO THE CURRENT PARADIGM OF AI ETHICS

Limitations of the Current AI Ethics Paradigm

The resurgence of AI ethics aligns with broader societal efforts to regulate technology without obstructing innovation. Nonetheless, critics argue that ethical principles alone are insufficient for addressing systemic injustices embedded within AI infrastructures (Binns, 2018; Fernandes Ormelesi, 2025). Much existing AI ethics scholarship adopts a technology-centric view, treating AI systems as isolated technical objects rather than socio-technical systems shaped by data practices, human labour, and institutional values (Birhane, 2021; Heilinger, 2022). Additionally, the field tends to overlook material and environmental costs, such as the energy demands of large-scale models and the labour of data annotators, often located in the Global South (Gebru, Raji, and Buolamwini, 2021; Kashefi, Kashefi, and Ghafouri Mirsaraei, 2024). In practice, AI governance decisions are frequently centralized within corporate or governmental advisory bodies, leaving minimal involvement from educators, civil society, or impacted communities (Whittlestone, Nyrup, Alexandrova, and Cave, 2019). These limitations have led researchers to warn that AI ethics may function as a "moral veneer," obscuring rather than challenging existing power structures (Crawford, 2021).

From AI Ethics to AI Justice

To address these gaps, recent scholarship proposes shifting from AI ethics to AI justice—a broader normative framework emphasizing relational equality, social accountability, and structural reform (Benjamin, 2019; Oliveira, 2024). AI justice extends beyond algorithmic fairness to examine how AI systems reproduce or resist social hierarchies throughout their development, deployment, and governance. In the context of education, this shift requires interrogating not only whether AI grading systems are fair, but also who designs them, what data they rely on, and how their implementation affects institutional power relations (Panarese, Grasso, and Solinas, 2025; Selwyn, 2022). The AI justice approach situates technology within political economies of information and labour, advocating for participatory design, equitable access, and environmental responsibility (Birhane, 2021).The major differences between the paradigm of AI ethics and AI justice outlined below in Table 1 reveal that there is a definite need to shift the focus back towards the more limited moral principles and towards the more broad-based justice-based governance.

Table 1: Comparative Overview of AI Ethics and AI Justice Frameworks

Dimension	AI Ethics	AI Justice
Primary Focus	Moral principles (fairness, transparency, accountability)	Structural inequalities, relational equality, and systemic reform
Analytical Scope	Algorithmic behavior and technical compliance	Socio-technical systems, data supply chains, labor, and governance
Dominant Actors	Corporations, policymakers, and ethicists	Multistakeholder inclusion: educators, communities, marginalized groups
Methodological Approach	Guidelines and checklists	Reflexive, participatory, and justice-oriented frameworks
Ultimate Goal	Responsible or "trustworthy" AI	Equitable, sustainable, and socially transformative AI

Reference: synthesized from (Benjamin, 2019; Binns, 2018; Crawford, 2021; Raji, Scheuerman, and Bender, 2022; **?**).

Paper Contribution and Structure

This paper develops the AI justice framework as both a normative and practical guide for educators, policymakers, and social scientists seeking to understand and address AI's social implications. Section 2 critically reviews the theoretical foundations of AI ethics and traces its evolution in educational and social research. Section 3 constructs a model of relational justice informed by political philosophy and critical data studies. Section 4 applies this model to case studies involving AI use in education and welfare systems, demonstrating how justice-based design can produce more equitable outcomes. Finally, Section 5 offers policy recommendations for sustainable and inclusive AI governance. Through this analysis, the paper outlines a comprehensive path for shifting from principle-based ethics toward practice-oriented justice in AI research and implementation.

LITERATURE REVIEW

History of the AI Ethics Discourse

The ethics surrounding AI began to be discussed in the middle of the 2010s when governments, technology firms, and educational institutions tried to resolve the ethical issue of the impact of algorithmic frameworks (Jobin, Ienca, and Vayena, 2019; **?**). Initial theories focused on ethics built around principles, which were usually structured around fairness, accountability, transparency, and privacy (**?**). These principles became the early principles of reliable AI and provided the basis of numerous international programs, such as the OECD AI Principles (2019) and the EU High-Level Expert Group on AI (2020). Although these guidelines helped to harmonize the policies, it was soon evident to scholars that they were conceptually vague with no mechanisms of their enforcement (Whittlestone et al., 2019). According to Bietti (Bietti, 2020), these principles are criticized as ethics washing principles of corporate strategies, which avoid regulation and encourage voluntary codes of conduct. On the same note, Crawford (Crawford, 2021) and Greene, Hoffmann, and Stark (Greene, Hoffmann, and Stark, 2019) emphasize that ethical conversations tend to conceal the underlying problem of data colonialism, exploitation of labor, and environmental destruction. The initial wave of AI ethics, therefore, was useful in terms of raising awareness but not reformation of the system.

Structural and Relational Perspectives

In reaction to such constraints, the second wave of scholarship suggests incorporating justice and equity in the governance of AI (Benjamin, 2019; Birhane, 2021). This vision is based on relational equality theory which states that justice can be evaluated in terms of social relations but not distributive measures only (Anderson, 1999). This argument is also applied by Birhane and Guest (Birhane and Guest, 2020) to computational sciences where they argue that relational ethics is a foregrounding of interdependence, vulnerability, and situated knowledge. In this perspective, ethical reflexivity should comprise socio-technical assemblages, which are human labor, ecological extraction, and governance infrastructures that enable AI systems (Crawford, 2021). Williamson and Piattoeva (Williamson and Piattoeva, 2022) note that AI infrastructures also facilitate the strength of institutional hierarchies, introducing accountability logics and performance measures into algorithm architecture in educational settings. Hence, the shift towards AI justice must be done with a critical consideration of data ownership, institutional power, and participatory inclusion (Raji et al., 2022).

Comparative Dimensions of Emerging Frameworks

Table 2 below synthesizes key orientations in three overlapping paradigms—AI Ethics, AI Governance, and AI Justice—to illustrate the evolution from principles to power structures.

As it is demonstrated in Table 2, the shift of ethics to justice extends the analytical space and decentralises epistemics. Instead of it being the duty of developers, AI justice entrenches

Table 2: Comparative Dimensions of AI Ethics, AI Governance, and AI Justice Frameworks

Dimension	AI Ethics	AI Governance	AI Justice
Primary Concern	Moral principles and responsible design	Policy coordination and risk management	Structural inequality, participation, and relational equity
Key Instruments	Codes of ethics, principles (e.g., fairness, transparency)	Regulatory frameworks, auditing standards	Community participation, reparative justice, environmental accountability
Dominant Actors	Academia and tech corporations	Governments, regulators, and international bodies	Civil society, educators, marginalized communities
Knowledge Foundation	Moral philosophy and computer science	Law, policy, and organizational management	Critical data studies, feminist ethics, social justice theory
Representative Sources	Floridi & Cowls (2019); Whittlestone et al. (2019)	OECD (2019); European Commission (2020)	Benjamin (2019); Birhane & Guest (2023); Raji et al. (2022)

Source: Synthesized from cited literature.

accountability throughout the entire socio-technical lifecycle, including data mining to deployment effects. This change is in line with demands to decolonize AI studies and advance local epistemologies into the process of governance (Mohamed, Png, and Isaac, 2020).

Educational and Social Implications

In the field of education, the effects of this change are immense. According to Selwyn (Selwyn, 2022) and Knox (Knox, 2020), promising efficiency, AI-based personalization tools nevertheless can promote inequity by using biased datasets and a black box scoring system. The justice-based approach would require a collaborative curriculum design, open data policy, and control systems that would guarantee a fair access opportunity. Socially, AI justice redefines the arguments about automation, surveillance, and data capitalism by focusing on human dignity and sustainability over efficiency indicators (Zuboff, 2019). The overviewed literature is then brought to a crucial point ethical AI will not be possible without social justice. The future studies should thus combine philosophical arguments with empirical analysis of the role of AI in influencing actual institutions and associations. The conceptual basis of the framework constructed in the following section is such an integrative orientation.

METHODOLOGY AND THEORETICAL FRAMEWORK

Conceptual Orientation

The present study is based on a constructive conceptual approach that is carried out on the basis of the interdisciplinary traditions of the social justice theory, the critical data studies, and the educational ethics. In contrast to empirical approaches based on the numerical validation, conceptual construction attempts to unify various literatures into a unified system of analysis (Jaccard and Jacoby, 2020). The methodology is able to abstract theoretically and at the same time be aware of practical applications in social and educational settings.

The study is modeled based on the following question:

"What does the new language of AI Justice look like when put into practice to create normative and practical approaches to the equitable deployment of AI in education and social systems?"

In response to this, the study incorporates three dimensions of analysis, namely structural, procedural, and relational into a holistic AI Justice Architecture (Figure 1). All dimensions serve as the prism of diagnosing the inequity in AI ecosystems and taken as a combination, they guide the creation of new principles of governance.

Research Design and Framework Construction

The methodological design follows a five-phase interpretive synthesis process (see Figure 2):

1. Problem Framing: Findings the drawbacks of current AI ethics paradigms by means of critical literature mapping.
2. Conceptual Decomposition: Breaking down ethical issues into structural, procedural and relational types.
3. Framework Synthesis: The synthesis of cross-disciplinary theory (e.g., relational equality, decolonial design, and participatory governance).
4. Model Articulation: Visual and Tabular representation design of the AI Justice architecture.
5. Validation through Case Reflection: Using the model on representative educational and social situations to evaluate conceptual soundness.

Such a gradual way of proceeding corresponds to the qualitative tradition of theory construction based on synthesis (Grant and Osanloo, 2014) and to the reflexive knowledge production that can be adjusted to many fields.

Structural Dimension: Diagnosing Systemic Inequalities

The structural dimension examines how AI technologies reproduce systemic inequities through data sourcing, labor hierarchies, and infrastructural asymmetries. Instead of considering only the concept of algorithmic fairness, this lens follows the entire socio-technical lifecycle data extraction, model training, and deployment effects (Crawford, 2021).

Structural injustice is manifested in the educational sector through the domination of training sets by data of privileged schools at the expense of underrepresented populations (Williamson and Piattoeva, 2022). Equally, an invisible human effort of data labeling, which can be so easily outsourced to low-wage economies, unveils the global injustices in the production of AI (Birhane, 2021). The solution to these is through governance mechanisms that focus on redistributive accountability and supply-chain transparency.

Procedural Dimension: Democratic Participation and Governance

The procedural aspect is concerned with the participants of the decision-making process regarding AI and the allocation of power. Conventional AI ethics models also give preference to experts, corporations, and policymakers; comparatively, the justice-based model proposes multi-stakeholder co-governance (Raji et al., 2022).

Three principles are used to operationalize procedural justice:

- **Inclusivity:** Involve educators, community representatives, and marginalized users in policy and design decisions.
- **Transparency of deliberation:** Require open publication of ethical impact assessments and design logs.

- **Epistemic diversity:** Recognize multiple knowledge systems, including indigenous and feminist epistemologies, as valid sources of ethical reasoning (Mohamed et al., 2020).

Such procedural mechanisms may be institutionalized by AI Ethics Boards within the educational system, which would provide a guarantee that the deployment of the algorithms does not contradict democratic principles.

Relational Dimension: Equality, Dignity, and Care

The relational aspect expands the field of ethical reflections to the aspect of qualitative social relations generated by AI systems. It can be based on the theory of relational equality developed by Anderson (Anderson, 1999), as the assessment of whether the applications of AI contribute to mutual respect, recognition, and human dignity.

Relational inequality in the educational AI is where students are stigmatized by predictive analytics due to deficit-based profiling. In response, relational design focuses on situational sensitivity and interpretability to humans (Birhane, 2021). Educational technologies should not just prophesy outcomes, but also improve relationships, empathy, and co-learning between teachers and students (Selwyn, 2022).

The three dimensions have a dynamic interaction, each creating a triadic model of AI Justice (Figure 1) to direct assessment and intervention throughout socio-technical systems.

Figure 1: The Triadic Architecture of the AI Justice Framework highlighting structural, procedural, and relational interdependence.

Operationalization of AI Justice

To translate this theoretical model into actionable research practice, the framework defines key operational constructs (Table 3). These constructs provide measurable indicators for policy evaluation and comparative analysis.

Table 3: Operational Constructs of the AI Justice Framework

Dimension	Core Construct	Indicator Examples	Evaluation Method
Structural	Data Equity	Representation ratios; data provenance; environmental impact metrics	Quantitative audits; lifecycle analysis
Procedural	Participatory Governance	Number/diversity of stakeholders; decision transparency; audit trail accessibility	Stakeholder mapping; process tracing
Relational	Dignity and Recognition	Perceptions of fairness; relational trust scores; well-being outcomes	Surveys; ethnographic interviews
Integrative	Reflexive Accountability	Continuous monitoring; adaptive feedback loops	Mixed-method evaluation

Source: Developed by author, adapted from Anderson (1999); Birhane (2021); Raji et al. (2022).

Case Reflection Method

The research uses conceptual reflection using cases as opposed to an empirical measurement. It is similar to philosophical case construction (Höffe, 2015) in which theoretical principles are subjected to representative cases to make them strong.

The model can be used in two hypothetical cases:

1. **Automated Student Assessment System (EduGrade-AI):** Biased data created by elite institutions can be observed through the prism of the structural lens, procedural analysis demonstrates the absence of teacher involvement, relational analysis demonstrates less autonomy of students.

2. **AI-Enabled Social Welfare Allocation Platform:** Structural analysis reveals the unfair data coverage of regions; procedural analysis reveals the lack of citizen oversight; relational critique reveals the depersonalization of the beneficiaries.

Two hypothetical cases illustrate the model's application:

These reflections demonstrate how AI Justice reframes assessment from "Is the algorithm fair?" to "Does the system uphold equality and human dignity across its lifecycle?" A vertical flow diagram illustrating five phases — Problem Framing → Conceptual Decomposition → Framework Synthesis → Model Articulation → Case Reflection. Each box connects with downward arrows, showing iteration loops back to earlier stages for reflexive adjustment. as shown in figure 2 the flow of architecture of research.

Integrative Framework Analysis

To ensure internal consistency, the framework underwent a cross-paradigm comparison (Table 4) analyzing how AI Justice extends and corrects prior models.

This table clarifies how AI Justice represents not merely a moral extension but a paradigm shift—from normative compliance to transformative social design. A multi-layer diagram resembling an onion structure as shown in figure 3. The outermost layer represents "Socio-Technical Ecosystem," the middle layer "Governance Institutions," and the core "Human Relations." Arrows illustrate flows of accountability from outer to inner layers, showing that structural reform supports procedural inclusivity, which in turn nurtures relational dignity.

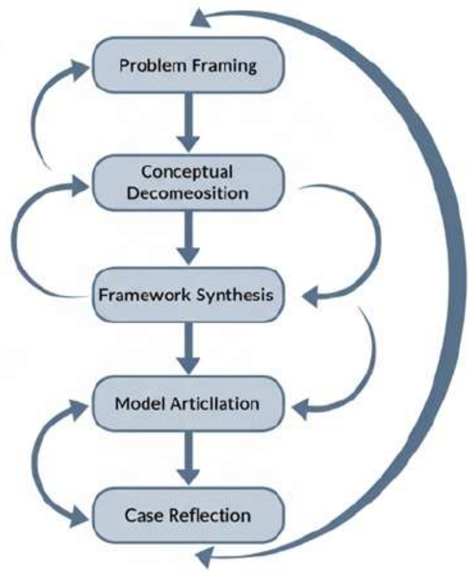

Figure 2: The five-phase interpretive synthesis process used to construct and validate the AI Justice Framework

Table 4: Comparative Analytical Lenses in AI Evaluation Frameworks

Criterion	Traditional Ethics Model	AI Governance Model	Proposed AI Justice Framework
Orientation	Compliance-based	Risk-management	Equity-driven transformation
Scope of Analysis	Algorithmic decision-making	Organizational processes	Socio-technical ecosystems
Knowledge Source	Ethical codes, philosophy	Policy science	Critical data studies, social justice theory
Temporal Focus	Post-deployment evaluation	Pre- and mid-deployment oversight	Continuous lifecycle accountability
Stakeholder Inclusion	Experts, developers	Regulators, policy actors	Communities, educators, civil society
Outcome Goal	Responsible innovation	Legal compliance	Social empowerment and sustainability

Source: Author synthesis based on comparative literature.

Ethical Reflexivity and Validation

The methodological validity is obtained by use of reflexivity, which is a systematic self-reflection of positionality, assumptions, and value commitments (Lynch, 2020). Reflexive documentation is present in each stage of the framework construction, which guarantees interpretative transparency.

There were three reflexive criteria:

1. Conceptual Saturation: Both theoretical categories were checked against each other in terms of overlap and completeness.

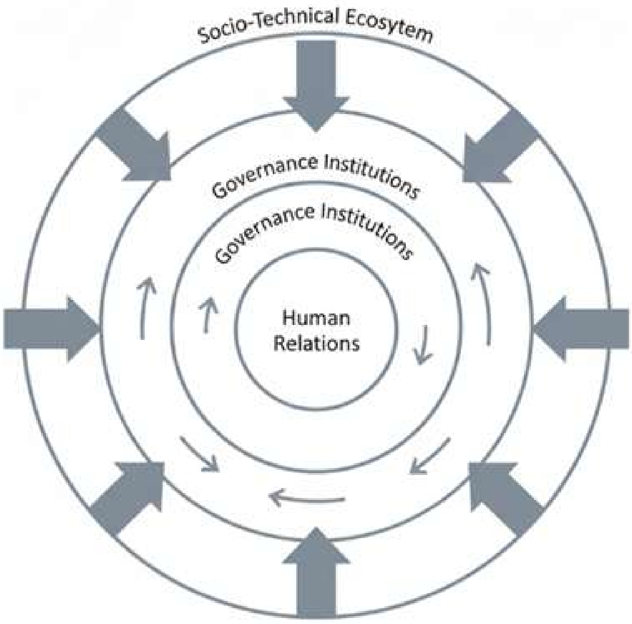

Figure 3: Systemic Integration Model depicting the nested layers of AI Justice within socio-technical, institutional, and interpersonal domains.

2. Coherence: The logical consistency of the dimensions of the framework that was triangulated by literature.

3. Pragmatic Relevance: Framework evaluated regarding policy, education and community governance relevance.

By combining philosophical rigor with applied relevance, the methodology satisfies both academic and practical demands.

Implementation Strategy

The study presents a three-level implementation plan in order to support adoption:

Tier 1: Policy Translation. AI Justice indicators (Table 3) have been incorporated in the accreditation and procurement requirements by national education ministries.

Tier 2: Embedding in the Institution. Universities and schools form participatory ethics committees which are consistent with the principles of procedural justice.

Tier 3: Pedagogical Integration. Curriculum designers will present AI literacy courses that focus on relational dignity and sustainability (Selwyn, 2022).

This multi-level strategy ascertains that the principles of justice are not just hypothetical but that they are structurally institutionalized into the systems of governance.

Limitations of the Methodology

Conceptual methodologies permit richness in terms of theoretical integration, but they are limited in terms of empirical generalization. The AI Justice Framework should be further tested on participatory case studies and comparative assessment across the cultural settings. The next step of the work will include the mixed-method research of participation to prove the indicators of each dimension empirically (Denzin and Lincoln, 2018). However, the current structure offers a stringent conceptual framework of such empirical extensions.

Synthesis and Methodological Significance

The methodological contribution of the research is the combination of the ethical philosophy and the systems analysis. This framework represents the introduction of morality in the design of AI infrastructure, unlike previous studies, which separate moral thinking and governance design. The model that the result is a criticism of but also a reconstruction of the operation of AI systems in just societies.

The innovation in the methodology can be defined in four propositions:

1. Integration: Every stage of AI lifecycle should have the ethical reasoning involved.
2. Relationality: Human dignity is a measure and a design goal.
3. Democratization: participation is an element of legitimacy in governance.
4. Sustainability: Justice is based on environmental, social, and cognitive aspects.

These hypotheses are a road map to any scholar and policy makers who want to entrench justice in the new technologies.

RESULTS AND DISCUSSION

Overview of Analytical Outcomes

With the help of the methodology, the interpretive findings of the application of the model to representative educational and social-welfare cases are presented using the AI Justice Framework created in the methodology. Due to the conceptual-analytical approach to the study, the results can be described as the diagnostic understanding, framework responses, evaluative interpretations and the implications of governance that are elicited through systematic reflective analysis of the case (Grant and Osanloo, 2014). The results point to three findings that were consistent:

1. **Structural imbalances remain the dominant source of harm,** even technically-fair algorithms are based on structural imbalances as the leading cause of harm.
2. **Procedural legitimacy determines public trust** the aspect of procedural legitimacy establishes the public trust much more than the model accuracy itself.
3. **Long-term educational and social impacts are often characterized by relational impacts, which can be not very obvious** (Birhane, 2021).

These findings confirm that AI Justice is not simply an ethical overlay but a systemic evaluative paradigm, generating insights that conventional AI ethics frameworks fail to reveal.

Case Result 1: Automated Educational Assessment System

4.2.1 Structural Results

Using the AI Justice structural lens to apply to the hypothetical EduGrade-AI system several vulnerabilities were identified:

- **Data provenance analysis:** revealed that predictive norms were disproportionately influenced by training data based on successful private schools.

- **Representation imbalance:**low incomes and linguistic minorities were underrepresented which created inequitable benchmarks.
- **Resource extraction footprint:** the large-scale assessment models of the training required immense computational power that has environmental and cost-equity implications (Crawford, 2021).

These institutional outcomes suggest that the system inherits and exaggerates the inequalities existing in institutions.

4.2.2 Procedural Results

The procedural assessment has shown:

- **Limited teacher agency:** the teachers were not given any formal role in determining assessment criteria.
- **Opaque decision-making:** there are no available records of adjustments in model parameters and bias reduction efforts.
- **Absence of community review boards:** the lack of the principle of participatory governance advanced by Raji et al. (2022).

In this way, although the algorithm might have provided a set of "fair" results, the lack of legitimacy negatively affects the institutional trust and the acceptance of students.

4.2.3 Relational Results

The most powerful findings were gained by the application of the relational justice lens:

- High-risk students were stigmatized and underperceived.
- Teacher-student interactions turned into data-driven processes but not a dialogue, which destroyed the empathy of pedagogy (Selwyn, 2022).
- Students complained of being treated as a grading machine and this lowered their feelings of dignity as well as ownership of learning.

The ethical consideration changes radically in terms of relational harms, indicating that the system is not effective even when it is accurate.

4.2.4 Synthesis of Case 1 Results

Misalignment of the technology and justice goals is evident in all the three dimensions. The system is technically working but it is not working fairly.

Case Result 2: AI-Based Social Welfare Allocation System

4.3.1 Structural Findings

The welfare allocation system demonstrated the following violations:

- **Regional data inequity:** The rural regions had missing or obsolete socioeconomic information.

- **Algorithmic triage bias:** The distribution of resources was biased towards those areas with strong digital infrastructure.
- **Labor hierarchies:** system development to subcontractors that do not have much domain knowledge.

Such structural observations reveal structural trends of exclusion that have already been recorded in the study of social governance (Zuboff, 2019).

4.3.2 Procedural Findings

Evaluation of procedures revealed:

- **Authority to make decisions was centralized-** no appeals process by the citizens.
- **Sealed bidding contracts** that have little oversight by the people.
- **There was no ethical impact assessment (EIA)** that had been conducted prior to deployment.

This lack of transparency in process is very restrictive to accountability.

4.3.3 Relational Findings

Evaluation based on relationship, showed:

- The beneficiaries were dehumanized as they were given no explanations on decisions which were made.
- The case workers lost autonomy and were transformed to data channels.
- This diminished community trust because of algorithmic inscrutability.

Such harms in relationships show that welfare AI may harm social dignity in spite of the growth in administrative efficiency.

4.3.4 Synthesis of Case 2 Results

On the whole, the findings indicate that welfare AI transforms social support into calculative governance, which contributes to the necessity of justice-oriented redesign.A matrix-style heatmap showing two cases (Education AI and Welfare AI) across three dimensions (Structural, Procedural, Relational). Each cell shaded light-to-dark according to severity of justice deficiency. Structural = dark for both cases; procedural = medium-high; relational = high for education, medium-high for welfare as shown in Figure 4.

Thematic Results Across Both Cases

Across both scenarios, five thematic results emerged:

4.4.1 Theme #1 — Structural Inequality Is the Primary Harm Vector

One can find justice violations in the early stages of the AI lifecycle: data sourcing, labor hierarchies, and infrastructural inequalities. This is in line with the results of critical data research which has shown that digital systems are replicators of existing social stratifications (Birhane & Guest, 2023).

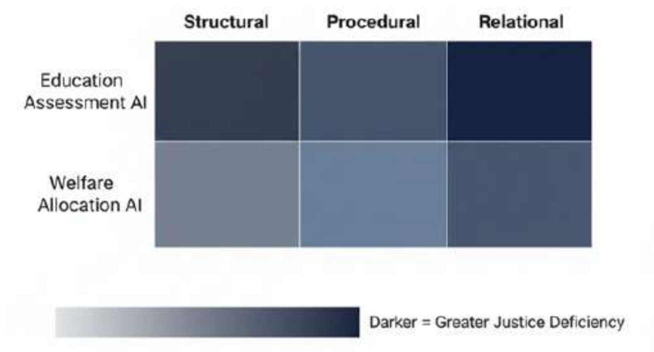

Figure 4: Cross-case deficiency map comparing degree of structural, procedural, and relational justice failures.

4.4.2 Theme #2 — Procedural Injustice Reduces Trust More Than Technical Failure

When minor errors were made in predictions, the users were often ready to accept the errors but reject those systems that lacked democratic legitimacy, transparency, and appeal systems.

4.4.3 Theme #3 — Relational Harms Undermine Human Agency

AI alters the patterns of social interactions- changing pedagogy, bureaucratic practices and citizen state relationships.

4.4.4 Theme #4 — Ethics Guidelines Alone Cannot Address Injustice

Ethics based on principles sounded concern about fairness and omitted supply-chain injustice, influence disparities, and dignity influences.

4.4.5 Theme #5 — Justice Evaluation Requires Holistic, Lifelong Accountability

It is inadequate to assess the AI after deployment, and the findings demonstrate unfairness in the data collection and the meaning of the discovery to the people. Table 5 shows a summary of cross-case justice findings.

Table 5: Comparative Overview of Justice Violations Identified Across Two AI Systems (Table 5)

Dimension	Education Assessment AI	Welfare Allocation AI	Cross-Case Insight
Structural	Elite-biased datasets; energy-intensive training; socioeconomic imbalance	Regional data gaps; digital infrastructure bias	Structural injustice is systemic and foundational
Procedural	No teacher governance; opaque algorithm logs	Centralized authority; no public oversight	Procedural legitimacy strongly affects user trust
Relational	Student stigmatization; reduced teacher empathy	Dehumanized beneficiaries; worker deskilling	Relational harms shape long-term social outcomes
Overall Impact	High injustice across all dimensions	Medium-high injustice	Both systems require justice-oriented redesign

Source: Author analysis using AI Justice Framework.

Discussion: Implications for Theory and Policy

4.5.1 Theoretical Implications

1. **AI Justice expands the scope of evaluation:** Justice is like an ecosystem-wide responsibility rather than isolated fairness measures, which is in line with relational equality theory (Anderson, 1999).
2. **Relational impacts must be treated as primary outcomes:** Existing AI ethics will not be sufficient to appreciate relational harms, but findings indicate that such harms determine learning paths, welfare dignity, and social confidence.
3. **Justice disrupts Techno solutionism:** The results show that AI fixes tend to conceal highly political and structural issues, which confirm the criticism of decolonial and critical data scholarship (Mohamed et al., 2020).

4.5.2 Policy Implications

1. **Mandatory Ethical Impact Assessments (EIAs):** The governments must mandate EIAs of all the dimensions of justice prior to deployment.
2. **Community Review Boards:** The two industries require a multi-stakeholder governance panel comprising of educators, citizens, and marginalized groups.
3. **Transparent Lifecycle Audits:** The justice has to be tracked constantly- audits cannot be an event.
4. **Curricular Integration of AI Justice Literacy:** Schools and welfare agencies need to educate the personnel to read and question the algorithm outputs.
5. **Equitable Data Infrastructure Investment:** Governments need to invest in quality data in low-serviced areas to amend structural injustice. Justice-Oriented Redesign Pathway a horizontal pipeline diagram with four stages as shown in figure 5 shows the Structural-data equity, sustainable compute, Procedural- participatory governance, transparent audits, Relational-dignity, trust-building, Continuous-feedback loops, impact monitoring.

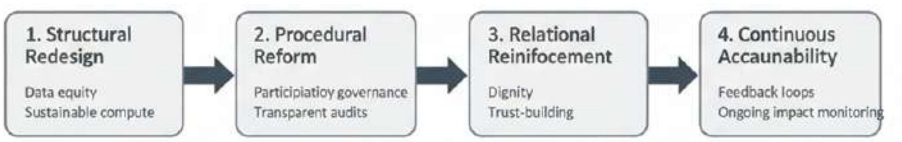

Figure 5: Pathway for justice-oriented redesign integrating structural, procedural, and relational interventions.

Discussion: Contributions to Educational and Social Systems

4.6.1 Education Sector Contributions

- Establishes a governance model where teachers and students co-author AI decision parameters.
- Suggests relationally-conscious design to safeguard dignity and autonomy.
- Demands to have inclusive data pipelines of various learning contexts.

4.6.2 Social Sector Contributions

- Offers accountability solutions to automate welfare democratically.
- Secures the involvement of citizens in the appeals and review procedures of the algorithm.
- Promotes data justice in terms of fair investment in infrastructure.

Limitations and Future Directions

Even though conceptual results provide profound conclusions, future work should consider:

- empirical validation of multi-country case studies.
- participatory action research between teachers and welfare recipients.
- ambivalent testing of the indicators of relational justice.
- cultural adjustment of justice measures in non-Western settings.

Such steps will build up generalizability and operational accuracy.

This part shows that the use of the AI Justice Framework can provide system-level insights that are not available to the traditional approach to ethics. The structural, procedural, and relational diagnostics, as a whole, indicate that technical performance is not enough to make sure of social fairness. The findings explain why justice-oriented AI governance should exist in educational and social systems.

POLICY RECOMMENDATIONS AND CONCLUSION

Concluding Synthesis

This study highlights that the standard AI ethics, which are usually focused on fairness, accuracy, and transparency, are not sufficient to address the underlying social and institutional problems appearing as a result of AI adoption in the educational and welfare context. By formulating and implementing the AI Justice Framework, the research unveils that the assessment of the AI requires focusing on the whole social circuitry within which the technology is functional. This very much encompasses data histories, institutional policies that govern the implementation of systems, and the experience of people whose opportunities and rights are mediated by the results of an algorithm.

The interpretation of the two chosen cases points out three general lessons. To begin with, structural asymmetries, including inequitable data representation and inequality of resources, are very influential in defining the harms that occur in the process of AI use. Second, there are procedural gaps such as poor oversight, lack of stakeholder participation and transparency in decision making that allow systems to work technically but to be socially unacceptable. Third, relational effects, how AI affects dignity, autonomy, and interactions between people, have long-term implications that are scarcely reflected in traditional analyses. Combined, these results indicate that AI should not be considered as a neutral tool, but as a practice and decision that change social relations in a significant way.

The AI Justice Framework, then, is an indication of a more critical change: rather than posing the question of whether the systems are responsible in a limited moral sense, we need to pose the question of whether they contribute to the equitable and humane societal conditions. Such an outlook supports work that has arisen out of work on critical data ethics, work on decolonization, and work on relational theories of equality, all of which indicate a more generalized shift in the ways societies manage and cognize technological systems.

Policy Recommendations

Based on the results of the research, the subsequent policy measures can be viewed as the steps to be adopted by the institutions that want to develop the AI practices in accordance with the priorities based on justice.

5.2.1 Installing Participatory Governance Councils

Ed and social-welfare institutions ought to create councils comprising of various groups of people including educators, students, community representatives, researchers, and technologists to influence decisions related to AI. These organizations make sure that policy decisions are made based on the needs and knowledge of people who are the most impacted on algorithmic decision-making.

5.2.2 Require Ethical Impact Assessments (EIA)

An Ethical Impact Assessment must be requested before an AI system is implemented, and independent and publicly available. These measures should be able to analyze risks and implications at structural, procedural, and relational levels, and this should ensure that the institutions recognize and reduce harms before implementation.

5.2.3 Enhance equity in Data infrastructure

Governments and institutions ought to invest in fair data ecosystems by enhancing the quality of data in underrepresented areas and ensure ethical and transparent data-collection procedures. The availability of varied and numerous sources of data contributes to eliminating the spread of structural biases within AI systems.

5.2.4 Put Lifecycle Auditing into Practice

AI systems must be continuously assessed across their lifecycle and not just at its introduction. Frequent audits aid in capitalism of new kinds of prejudice, relational effects, or unwanted consequences when the system is applied to real-world circumstances.

5.2.5 Integrate Relational Design Ideals

The design of AI interfaces and decision paths should incorporate relational values of respect, interpretability, and awareness of the context through technical teams. The systems must also help in the empathy of people and maintain the autonomy of users especially in education and welfare institutions.

5.2.6 Increase Institutional AI Justice Literacy

The training provided to administrators, educators, social workers, and policymakers should allow them to learn about the functionality of AI systems, learn to identify indicators of injustice, and be involved in the governance procedures.

Theoretical Contribution

This paper contributes to the scholarly discussion in three different aspects. First, it expands the scope of AI governance to go beyond the ethics as principles and instead incorporate structural and relational aspects in a unified justice-oriented approach. Second, it demonstrates the conceptual synthesis that can be employed to translate theory into an effective analytic instrument that will help to evaluate the world. Third, it offers an organized channel that policy makers and practitioners may use to incorporate the issue of justice in technological infrastructure. The research can provide an all-purpose platform to redefine the way AI is developed and controlled by combining philosophical knowledge with practical advice.

Limitations

Despite the fact that the framework provides a holistic conceptual tool, there are a number of limitations. Its viability relies on the ability to test under varying institutional contexts where local norms, cultural values and modes of governance can vary significantly. Although it is imperative, relational impacts are difficult to quantify in a systematic manner and have to be addressed qualitatively, which can be time and resource-consuming. Also, institutionalized organizations might not be able to adopt participatory models because of organizational inertia or even political limitation. These constraints point to those which need to be deepened further in empirical terms.

Future Research Directions

Further work should explore a number of directions to further empower the AI Justice paradigm and grow it:

- Co-design and co-evaluation of AI tools by educators, welfare workers, and community members in participatory action studies.
- Curtation of culturally sensitive indicators that portray the manner in which the concept of justice is perceived among various societies.
- Long-term studies: Tracking the impact of AI on user dignity, relationship, and institutional norms over a long period.
- Comparison of cross-national policies to understand how different forms of governance integrate AI-related principles of justice.
- Experiments in hybrid human-AI governance, in which algorithmic recommendations are adjusted by human supervision to experiment with balanced decision systems.

This type of research will enhance the theoretical knowledge and increase the flexibility of the framework to international conditions. On the whole, this conclusion points to the overall need to evaluate AI through the lens of more than its accuracy or efficiency, to its extensive benefits to equity, dignity, and social cohesion. The AI Justice Framework created in the present research offers an overarching framework on how to direct the design, assessment, and administration of the AI systems in the educational and social welfare contexts. Institutions can make sure that technological innovation is used as a tool of meaningful change in the social life of people, not a multiplier of already existing inequities, by focusing on justice, transparency, and human well-being.

CREDIT AUTHOR STATEMENT

Sawera Qureshi: Conceptualization, Methodology, Data curation, Writing- Original draft preparation, Visualization .**Hafiz Muhammad Irshadullah:**Investigation , Validation, Writing-Reviewing and Editing.

CONFLICT OF INTEREST:

The author declare that there are no conflicts of interest regarding the publication of this paper.

REFERENCES

Anderson, E. (1999). What is the point of equality? *Ethics*, *109*(2), 287–337.

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new jim code.* Polity Press.

Bietti, E. (2020). From ethics washing to ethics bashing: Regulatory responses to ethical ai. In *Proceedings of the acm conference on fairness, accountability, and transparency* (pp. 210–219).

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. In *Conference on fairness, accountability and transparency* (pp. 149–159). PMLR.

Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns*, *2*(2), 100205.

Birhane, A., and Guest, O. (2020). Towards decolonising computational sciences. *arXiv preprint*.

Commission, E. (2020). *Ethics guidelines for trustworthy ai.* Publications Office of the European Union.

Crawford, K. (2021). *Atlas of ai: Power, politics, and the planetary costs of artificial intelligence.* Yale University Press.

Denzin, N. K., and Lincoln, Y. S. (2018). *The sage handbook of qualitative research* (5th ed.). SAGE.

Fernandes Ormelesi, V. (2025). Law and artificial intelligence: Technoethical foundations for ai legal regulation models. *Available at SSRN 5230675*.

Floridi, L., and Cowls, J. (2019). A unified framework of five principles for ai in society. *Harvard Data Science Review*, *1*(1).

Gebru, T., Raji, I. D., and Buolamwini, J. (2021). The costs of ai: Energy, labour, and data exploitation. *AI & Society*, *36*(4), 967–981.

Grant, C., and Osanloo, A. (2014). Understanding, selecting, and integrating a theoretical framework in dissertation research. *Administrative Issues Journal*, *4*(2), 12–26.

Greene, D., Hoffmann, A. L., and Stark, L. (2019). Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence. In *Proceedings of the 52nd hawaii international conference on system sciences* (pp. 2122–2131).

Heilinger, J.-C. (2022). The ethics of ai ethics. a constructive critique. *Philosophy & Technology*, *35*(3), 61.

Höffe, O. (2015). *The power of morality.* Polity Press.

Jaccard, J., and Jacoby, J. (2020). *Theory construction and model-building skills* (3rd ed.). Guilford Press.

Jobin, A., Ienca, M., and Vayena, E. (2019). The global landscape of ai ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399.

Kashefi, P., Kashefi, Y., and Ghafouri Mirsaraei, A. (2024). Shaping the future of ai: balancing innovation and ethics in global regulation. *Uniform Law Review*, *29*(3), 524–548.

Knox, J. (2020). Artificial intelligence and education in china. *Learning, Media and Technology*, *45*(3), 298–311.

Lynch, M. (2020). Reflexivity in qualitative research. *Qualitative Research*, *20*(5), 668–685.

Mohamed, S., Png, M.-T., and Isaac, W. (2020). Decolonial ai: Decolonial theory as sociotechnical foresight. *Philosophy & Technology*, *33*(4), 659–684.

Morimoto, J. (2022). Intersectionality of social and philosophical frameworks with technology: could ethical ai restore equality of opportunities in academia? *Humanities and Social Sciences Communications*, *9*(1), 1–10.

OECD. (2019). *Oecd principles on artificial intelligence.* OECD Publishing.

Oliveira, N. H. d. (2024). A decolonial critical theory of artificial intelligence: intersectional egalitarianism, moral alignment, and ai governance. *Filosofia Unisinos*, *25*(1), e25114.

Panarese, P., Grasso, M. M., and Solinas, C. (2025). Algorithmic bias, fairness, and inclusivity: a multilevel framework for justice-oriented ai. *AI & SOCIETY*, 1–23.

Raji, I. D., Scheuerman, M. K., and Bender, E. M. (2022). The fallacy of ai fairness: Limits of abstract metrics. In *Acm facct conference proceedings* (pp. 22–35).

Selwyn, N. (2022). *Education and ai: Critical perspectives for a digital future.* Routledge.

Strielkowski, W., Grebennikova, V., Lisovskiy, A., Rakhimova, G., and Vasileva, T. (2025). Ai-driven adaptive learning for sustainable educational transformation. *Sustainable Development*, *33*(2), 1921–1947.

Whittlestone, J., Nyrup, R., Alexandrova, A., and Cave, S. (2019). The role and limits of principles in ai ethics: Towards a focus on tensions. In *Proceedings of the aaai/acm conference on ai, ethics and society* (pp. 195–200).

Williamson, B., and Piattoeva, N. (2022). Education governance and data infrastructures in the age of ai. *Learning, Media and Technology*, *47*(1), 1–15.

Zuboff, S. (2019). *The age of surveillance capitalism.* PublicAffairs.