

# Algorithmic Fairness in AI-Driven Project Management for Post-Conflict and Developing Regions: A Critical Review and Framework for Context-Aware AI

Saleem Ahmed Al-Azazi<sup>1</sup>, Aawag Mohsen Alawag<sup>2\*</sup>, Amin Saif<sup>3</sup>

<sup>1</sup> Faculty of Engineering and Information Technology, Taiz University, Taiz, Yemen

<sup>2</sup> Faculty of Civil Engineering, Universiti Teknologi MARA (UiTM), Shah Alam 40450 Selangor, Malaysia

<sup>3</sup> Faculty of Engineering and Information Technology, Taiz University, Taiz, Yemen

\*Corresponding author: aawagmohsen@uitm.edu.my

Received: 25 August 2025 / Accepted: 25 November 2025 / Published online: 31 March 2026

---

## Abstract

Artificial Intelligence (AI) is poised to revolutionize construction project management, offering unprecedented gains in efficiency, scheduling, and risk assessment. However, deploying AI uncritically in the data-poor and socially fragile environments of post-conflict and developing regions creates profound ethical risks. Standard technical solutions for algorithmic fairness, which require large, clean datasets, are fundamentally misaligned with these realities. This misalignment risks automating historical inequalities and eroding public trust in reconstruction efforts. This critical review synthesizes a multidisciplinary body of literature to map the intersection of AI applications in construction with the specific ethical challenges of these vulnerable settings, including pre-existing, technical, and emergent biases. Synthesizing these findings, we propose a "Context-Aware AI Fairness Framework," a holistic, lifecycle approach structured around four pillars: (1) Foundational Scoping & Participatory Design, (2) Data Governance & Bias-Aware Data Management, (3) Contextualized Model Development & Mitigation, and (4) Human-in-the-Loop Deployment & Continuous Monitoring. The paper concludes by arguing that the prevailing fixation on "big data" is a critical limitation and calls for a new research direction focused on developing robust AI systems that can effectively reason with "small data" and integrate rich qualitative inputs, thereby ensuring that AI serves as a tool for equitable and sustainable development rather than a driver of a new digital divide.

**Keywords:** Artificial Intelligence (AI), Algorithmic Fairness, Post-Conflict Reconstruction, Ethical AI, Data Scarcity

---

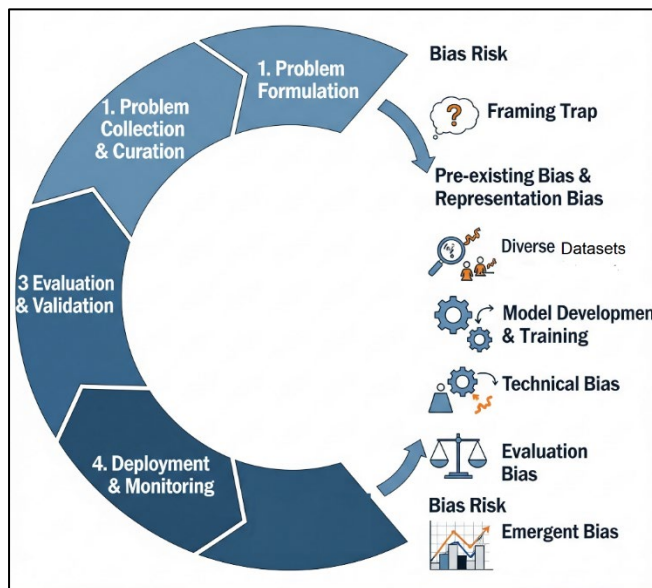
## 1. Introduction

This review critically synthesizes a multidisciplinary body of literature to build the argument for a context-aware approach to algorithmic fairness, moving beyond decontextualized technical fixes (Selbst et al., 2019). The analysis is structured to first deconstruct the foundational element of any AI system—the data—by exposing its inherent, context-specific biases (Alvarez et al., 2024; Mittelstadt et al., 2016). It then critically appraises the standard technical solutions for bias mitigation, arguing their inadequacy for the data-poor environments typical of post-conflict and developing regions (Goknil et al., 2023; Plevris et al., 2025). Finally, it reviews the broader ethical frameworks and documented implementation challenges (Brand, 2023; Tjebane et al., 2022) that, together, demonstrate the necessity for a new, holistic model. While numerous AI fairness frameworks exist, they often focus on technical bias mitigation (Alvarez et al., 2024) or are designed for data-rich, stable institutional environments in the Global North. This paper argues that such frameworks are insufficient for post-conflict and developing regions, which face unique challenges of data scarcity, institutional

fragility, and profound trust deficits. The proposed 'Context-Aware AI Fairness Framework' addresses this gap by integrating participatory governance and socio-political analysis directly into the AI lifecycle.

### 1.1 The Myth of Neutral Data: Sources of Bias in Construction Datasets

The foundational premise of data-driven decision-making, a core component of modern approaches like Industry 4.0, which rely on robotics and data sharing (Alawag, Liew, et al., 2024), is the quality and neutrality of the underlying data. However, a significant body of research asserts that data is never truly neutral but is instead a value-laden product of social and technical processes (Alvarez et al., 2024; Mittelstadt et al., 2016). In the context of construction project management, particularly in developing and post-conflict regions, the data used to train AI systems is susceptible to multiple forms of bias that can undermine algorithmic fairness. These biases can be broadly categorized as pre-existing, technical, and emergent (Mittelstadt et al., 2016). As illustrated in **Figure 1**, these biases are not isolated incidents but can be introduced at every stage of the AI project management lifecycle, from initial problem formulation to final deployment and monitoring.



**Figure 1.** Sources of Bias in the AI Project Management Lifecycle (adapted from (Alvarez et al., 2024; Mittelstadt et al., 2016).

Figure 1 points out that bias can enter an AI project at different stages. This idea builds on the work of Alvarez et al. (2024) and Mittelstadt et al. (2016). What matters is that each stage carries its own risks, and if they are ignored, fairness is easily lost.

The analysis that follows looks at these risks stage by stage. The key message is simple: choices that look like routine technical steps can, in reality, carry heavy ethical consequences.

**Pre-existing bias** is perhaps the most difficult to deal with. These biases come from social and historical inequalities that already shape the data (Alvarez et al., 2024). In post-conflict or developing settings, construction records often reflect older patterns of exclusion. Some communities, whether ethnic or social groups, were consistently left out of housing projects, road networks, or other public resources (Cochran, 2011; Harrowell et al., 2018). Training an AI model on such records risks repeating the same injustice, this time under the cover of technical objectivity (Selbst et al., 2019). The issue grows worse when the so-called “ground truth” is itself unreliable. Data such as project delays or cost overruns may look objective, but in practice they are often influenced by subjective reporting and uneven monitoring (Alvarez et al., 2024).

Technical bias is another challenge. It often comes from the way data is collected and organized. Tools like Building Information Modeling (BIM) were created to provide shared and consistent information (Alawag et

al., 2023). However, in many developing contexts BIM use is still limited (Alakhali et al., 2024). Within the AEC industry more generally, missing records, inconsistent formats, and incomplete entries are common problems (Aleke et al., 1833; Goknil et al., 2023). On top of this, much information remains as “dark data”—unstructured and unused. This problem is especially visible in data-poor environments, where the result is sparse and unbalanced datasets (Goknil et al., 2023; Plevris et al., 2025). This type of data is used to train AI, which results in skewed models that frequently do not adequately represent underrepresented groups or unique project circumstances (Egwim et al., 2023). Using proxy variables for sensitive attributes is another way that technical bias manifests itself. Even seemingly innocuous data points, like the location of a project or financial indicators, can have a strong correlation with social class or ethnicity. Such proxies may inadvertently result in discriminatory outcomes when they are employed in risk assessment or resource allocation models (Cochran, 2011; Mittelstadt et al., 2016).

After an AI system is implemented, bias may also surface. These are known as emergent biases, and they frequently arise as a result of feedback loops that reinforce the initial bias (Mittelstadt et al., 2016). Projects in areas that were previously disregarded, for instance, are more likely to fail if a scheduling tool consistently allocates fewer resources there due to skewed risk scores. New information that appears to “prove” the skewed prediction is then produced by their failure. These cycles can exacerbate resentment, exacerbate social divisions, and directly impede efforts toward equitable and inclusive reconstruction in post-conflict settings where trust is already brittle (Brown et al., 2019; Harrowell et al., 2018).

## 1.2 A Critical Appraisal of Bias Mitigation Algorithms: A Toolkit Unfit for Context

Researchers in the “fair AI” field have built a set of methods to deal with algorithmic bias. These usually fall into three groups: pre-processing, in-processing, and post-processing (Alvarez et al., 2024). On paper, they seem useful. But once applied in post-conflict or data-poor environments, their impact is limited and sometimes even counterproductive.

Pre-processing tries to fix the problem at the data stage by rebalancing or cleaning the dataset (Mittelstadt et al., 2016). In practice, this is extremely difficult when the available data is already small or inconsistent. Making too many changes can erase the little valuable information that exists, so the model ends up both less accurate and not any fairer (Goknil et al., 2023).

In-processing methods attempt to “bake fairness” into the training process (Alvarez et al., 2024). These require advanced computing power and technical expertise, which are often unavailable in regions caught on the weaker side of the global AI divide (Anzolin et al., 2012; Liu, 2023). Even more challenging, they usually depend on choosing a single fairness metric. That choice is never neutral—it means accepting one definition of fairness over another, which is an ethical decision, not just a technical one (Alvarez et al., 2024; Selbst et al., 2019).

Post-processing works differently: it adjusts the model’s outputs after training (Mittelstadt et al., 2016). But critics argue this is more of a “band-aid” than a real solution. It hides the problem instead of addressing the root causes, and it tends to break down when project conditions change quickly (Alvarez et al., 2024).

When viewed together, these methods reflect a kind of techno-solutionism. They treat fairness as if it were only a mathematical issue, ignoring the social and political realities that shape data and outcomes. This is a serious gap, since real fairness often depends on collaboration with stakeholders, clear governance, and constant monitoring—factors that purely technical fixes cannot deliver (Alawag, Alqahtani, et al., 2024; Musarat et al., 2022). At the heart of the problem is the fact that poor data quality and the contested nature of fairness itself cannot be solved by algorithms or statistics alone (Selbst et al., 2019).

## 1.3 Beyond Technical Fixes: Ethical and Governance Frameworks for AI

It is now widely accepted that technical fixes alone cannot solve the problem of bias in AI. Because of this, many researchers and policymakers are turning their attention to ethical and governance frameworks that focus on the public interest. The shift here is important: instead of treating fairness as a mathematical puzzle, the discussion moves toward principles such as accountability, transparency, human rights, and oversight. In fact, there is a growing global agreement around these ideas under the banner of “Responsible AI” (Brand, 2023; Isagah et al., 2024). For the humanitarian and development fields, these principles are not abstract theories.

They are daily necessities. Without them, organizations cannot earn the trust of the communities they work with (Brown et al., 2019).

A good example is algorithmic accountability. This is often put into practice through tools like Algorithmic Impact Assessments (AIAs), which are used to check possible risks before a system is deployed (Brand, 2023; Walsh et al., 2019). Transparency is another key piece. Some governments, for example, now keep public “algorithm registers” so that people can see when and how AI is used in public services (Brand, 2023).

But perhaps the most important principle is participation. AI cannot just be designed from the top down. Communities affected by these systems need to be directly involved in shaping them. In other words, fairness must be defined with them, not for them (Brown et al., 2019). This lesson is especially true in post-conflict reconstruction, where history shows that projects designed without local voices almost always fail (Anderlini et al., 2004).

There is also a major ethical concern around privacy and consent. Using data from vulnerable people is never a neutral act. If handled poorly, it can cause real harm. That is why strong, context-sensitive protections are essential (Wheeler, 2014).

#### 1.4 The Implementation Reality: Structural Barriers to AI Adoption

The big promises of AI often collide with the hard realities of implementation in developing countries. For example, a resource allocation tool trained on biased historical data might deprioritize aid to a marginalized ethnic group, as seen in post-conflict reconstruction scenarios where historical records themselves reflect patterns of exclusion (Harrowell et al., 2018). Similarly, a predictive model for project delays might fail spectacularly when deployed in a region with weak institutions and unreliable data, a challenge highlighted in post-genocide Rwanda (Tjebane et al., 2022; World Bank, 2024). Evidence from the AEC sector and other fields shows a recurring pattern: structural, organizational, and political obstacles make the idea of a simple “technology transfer” unrealistic. The barriers are many. High upfront costs, which are especially difficult for small and medium-sized enterprises, lack of government backing, and weak stakeholder coordination in post-disaster situations are just a few (Alawag, Liew, et al., 2024; Alakhali et al., 2024; Baarimah et al., 2024; Saad et al., 2025).

At the core of the problem is what has been called the “AI divide.” This refers to the uneven distribution of infrastructure, funding, and human expertise between the Global North and the Global South (Anzolin et al., 2012; Berg et al., 2024). There is also a significant skills gap. Adoption in the AEC sector is frequently hampered by a lack of local professionals with advanced technology training, which is a problem in many countries (Plevris et al., 2025; Shi et al., 2020; Tjebane et al., 2022).

Institutional and organizational flaws also come into play. These include limited state capacity, institutional resistance to change, and exorbitant costs that SMEs cannot afford. Time and again, such barriers are highlighted in both post-disaster recovery efforts and in broader technology adoption in developing contexts (Baarimah et al., 2024; Saad et al., 2025). Rwanda after the genocide is a telling example: even basic data collection for planning was severely constrained by weak institutions and political realities (Tjebane et al., 2022; World Bank, 2024).

It is also important to recognize that adopting technology is never free of politics. The case of post-conflict Nepal makes this clear. Here, unresolved political tensions and the capture of resources by elites turned reconstruction projects—designed with good intentions—into tools for local power struggles (Harrowell et al., 2018). This body of evidence conclusively shows that a purely techno-solutionist approach is destined to fail. Successful and ethical AI implementation requires addressing not only technical problems but also the underlying structural, institutional, and political realities of the context.

To consolidate the findings of this critical review, **Table 1** provides a summary that maps specific AI applications in construction to their most salient bias risks and the contextual challenges that complicate their

mitigation. This mapping synthesizes the core argument of the literature review and sets the stage for the framework proposed in the subsequent sections.

**Table 1.** Mapping AI Applications to Bias Risks and Mitigation Challenges in Context

| <b>AI Application in Construction</b> | <b>Primary Bias Risks &amp; Sources</b>  | <b>Mitigation Challenges in Post-Conflict &amp; Developing Contexts</b>   |
|---------------------------------------|--|---|
| Risk Assessment                       | Pre-existing Bias: Models trained on historical data may learn to associate certain geographic areas or demographic groups with higher risk due to past patterns of systemic neglect and discrimination, not objective risk factors (Cochran, 2011; Selbst et al., 2019).      | Data Scarcity & Poor Quality: Lack of reliable, representative historical data makes it impossible to train accurate or fair models. Using what little data exists can amplify existing biases (Goknil et al., 2023; Plevris et al., 2025).   |
| Resource Allocation                   | Technical Bias: Using seemingly neutral proxies (e.g., proximity to services, property records) that correlate with ethnicity or socioeconomic status can lead to discriminatory allocation of funds and materials (Cochran, 2011; Mittelstadt et al., 2016).                  | Institutional Fragility: Weak state capacity and the absence of robust data governance frameworks make it difficult to manage and audit complex allocation algorithms, increasing the risk of elite capture or corruption (A JOURNEY FROM POST-CONFLICT, 2024), (Harrowell et al., 2018). |
| Project Scheduling                    | Emergent Bias: AI schedulers optimizing for "efficiency" based on flawed data can create feedback loops, deprioritizing projects in marginalized areas, which then suffer more delays, "proving" the initial biased assumption (Brown et al., 2019; Mittelstadt et al., 2016). | High Cost & Skills Gap: In-processing mitigation techniques are impractical to implement because they are computationally costly and require specialized expertise that is frequently unavailable locally due to the "AI divide" (Anzolin et al., 2012; Tjebane et al., 2022).            |
| Safety Monitoring                     | Technical Bias: Disparities in safety monitoring and protection may result from computer vision models trained on globally sourced datasets performing poorly in identifying local workers or site conditions (Alvarez et al., 2024; Shi et al., 2020).                        | Lack of Legitimacy & Trust: In communities where people have little faith in authority, the use of opaque monitoring technologies may be interpreted as surveillance, and this could undermine the legitimacy of the project and community support (Brown et al., 2019; Wheeler, 2014).   |

## 2. Research Method

### 2.1 Research Design

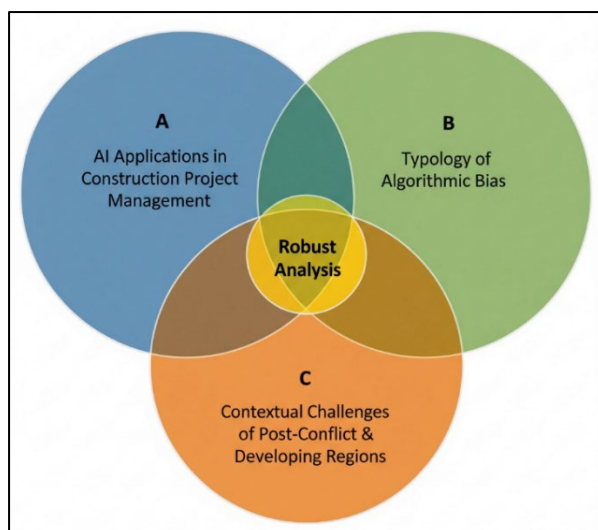
A critical review design was used in this investigation. Since the goal of the research is to go beyond a simple summary of the body of existing literature, a critical review was chosen as the most suitable methodology. Rather, it aims to methodically locate, integrate, and analyze the disparate, interdisciplinary corpus of work at the nexus of post-conflict development, project management, and artificial intelligence (AI). This design aims to assess and critique the conceptual boundaries and underlying assumptions of existing research in order to facilitate the development of a new analytical framework to fill in identified gaps (Khan, 2025). The main goal of this design is to create a logical map that critically links AI applications in the construction industry to their potential for unethical behavior in the particular circumstances of developing and post-conflict areas.

## 2.2 Data Source and Selection Criteria

A carefully selected collection of academic and policy literature served as the "data" for this critical review. In order to ensure thorough coverage of the engineering, computer science, and social science literature, a purposive literature search was carried out across a number of significant academic databases, including Scopus, IEEE Xplore, ACM Digital Library, and Google Scholar, in order to establish this evidence base. The search strategy combined several strings of keywords designed to capture the core themes of the research, such as: ("Artificial Intelligence" OR "algorithmic bias" OR "AI ethics") AND ("construction project management" OR "AEC industry") AND ("post-conflict reconstruction" OR "developing countries" OR "humanitarian sector"). The inclusion criteria for sources were: (1) direct relevance to at least two of the core research themes (AI, project management, post-conflict/developing contexts); (2) publication in a peer-reviewed journal or as a formal report from a recognized international organization (e.g., World Bank, United Nations) or governmental body; and (3) publication date largely within the last decade to ensure contemporary relevance, with foundational older texts included where necessary. This selection process resulted in a final corpus of 33 high-impact sources that form the foundation for the thematic synthesis.

## 2.3 Data Analysis

The analytical method employed was thematic synthesis. This process involved several distinct stages aligned with a critical review methodology. First, a familiarization phase where all 33 selected sources were read in detail to identify initial concepts. Second, a systematic coding phase was conducted, where text segments from the literature were tagged and categorized based on the three dimensions of the analytical framework (Figure 2): (A) AI Applications in Construction, (B) Typology of Algorithmic Bias, and (C) Contextual Challenges. Third, these codes were collated, compared, and refined to generate broader descriptive themes (e.g., 'data scarcity as a bias amplifier,' 'institutional fragility as an implementation barrier'). Finally, these themes were synthesized to build the critical argument, culminating in the development of the 'Context-Aware AI Fairness Framework' presented in the results. This structured process ensures that the framework is directly grounded in the gaps and intersections identified across the multidisciplinary literature. The core of this process involved structuring the analysis around a three-dimensional analytical framework, developed specifically for this study and visualized in Figure 2. The figure below illustrates the critical review methodology employed in this paper. A robust analysis of the research problem exists only at the intersection of (A) AI applications in construction project management, (B) the socio-technical nature of algorithmic bias, and (C) the unique contextual challenges of post-conflict and developing regions.



**Figure 2.** The Three-Dimensional Analytical Framework. Source: Author's elaboration

This figure illustrates the critical review methodology employed in this paper. A robust analysis of the research problem exists only at the intersection of (A) AI applications in construction project management, (B) the socio-technical nature of algorithmic bias, and (C) the unique contextual challenges of post-conflict and developing regions.

The analysis proceeded by systematically coding the selected literature according to the three dimensions of this framework:

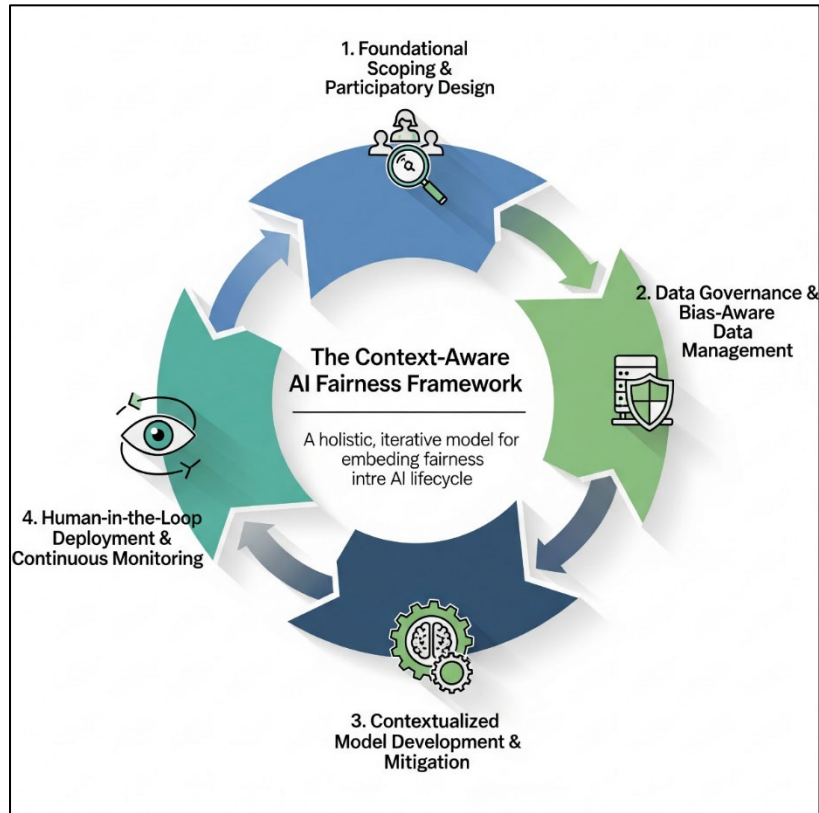
1. **Functional Applications of AI:** Categorizing AI applications based on their primary project management function, such as Scheduling and Sequencing (Khan, 2025), Risk Assessment (Aleke et al., 1833; Shi et al., 2020), and Resource Allocation (Bah Esseme et al., 2025).
2. **Typology of Algorithmic Bias:** Classifying identified risks according to a synthesized typology of bias, including Pre-existing Bias (Anderlini et al., 2004), Technical Bias (Cochran, 2011), and Emergent Bias (Berg et al., 2024).
3. **Contextual Challenges:** Mapping the findings to the specific challenges of post-conflict and developing regions as documented in the literature.

The development of the context-aware framework in this paper is directly informed by the nuanced critique made possible by this structured thematic synthesis, which offers a rigorous methodological foundation to move the discourse beyond a generic "promise and peril" narrative.

### 3. Results and Discussion

The literature review reveals a glaring discrepancy between the complex realities of post-conflict and developing environments and the technical approaches typically suggested for algorithmic fairness. On paper, standard "techno-solutionist" approaches might seem good, but they frequently fall short in environments characterized by inequality, a lack of data, and a lack of trust (Harrowell et al., 2018; Selbst et al., 2019).

The study's primary result, a novel conceptual model known as the Context-Aware AI Fairness Framework, is derived from this synthesis. This framework adopts a more comprehensive approach than approaches that solely concentrate on correcting bias in algorithms. It emphasizes participation, good data governance, context-sensitive modeling, and rigorous human oversight, treating fairness as a process that permeates the whole AI lifecycle. The goal is pragmatic. In addition to researchers, the framework is intended to support international organizations and policymakers who are attempting to implement AI in delicate environments. The goal is to ensure that these systems are socially and ethically sustainable in addition to being technically dependable. An illustration of this holistic, iterative model's operation is provided in Figure 3.



**Figure 3.** The Context-Aware AI Fairness Framework. Source: Author's elaboration

An iterative, comprehensive approach to integrating equity into the AI lifecycle for project management in delicate situations. The framework places a strong emphasis on an ongoing cycle of governance, participation, contextual adaptation, and supervision.

The framework is structured around four foundational pillars, each detailed below.

**Pillar 1: Foundational Scoping & Participatory Design** must be incorporated into the project from the start and cannot be addressed after the fact. In order to identify current disparities and power imbalances, this stage begins with a thorough examination of the local context, including its history, politics, and social structures (Anderlini et al., 2004; Harrowell et al., 2018). This directly counters the risk of pre-existing bias, where a tool might unknowingly adopt historical patterns of ethnic marginalization. For practitioners, this means allocating budget and time for ethnographic studies and stakeholder workshops before any code is written, treating this as a non-negotiable project milestone. Participatory design techniques that actively involve local communities must be incorporated into the process to prevent top-down approaches. In this case, people are partners in establishing the objectives of the system and determining what constitutes a "fair" result rather than passive recipients (Brown et al., 2019). Gaining the trust—also known as the "social license to operate"—that projects in vulnerable environments require is crucial (Brand, 2023; Brown et al., 2019). Before any major deployment, high-stakes systems should also go through an Algorithmic Impact Assessment (AIA) to identify risks to human rights and social equity in advance (Brand, 2023; Walsh et al., 2019).

**Pillar 2: Data Governance & Bias-Aware Data Management** should never be seen as a neutral input. It is shaped by social contexts, and that means it must be handled with care. A strong data governance process is required, including ethical rules for how information is collected, stored, and protected, with special attention to vulnerable groups (Wheeler, 2014). In places where data is scarce, the focus should be on collecting better, more representative information rather than relying on incomplete records (Goknil et al., 2023). This pillar directly addresses the challenge of 'small data' and technical bias. Practically, this guides project managers to

invest in participatory data collection (as detailed in the case study in Box 1) rather than attempting to apply complex bias mitigation algorithms to sparse, unreliable datasets. Practitioners also need to question the assumption that historical data represents “ground truth.” Datasets should be documented in detail, for example, through datasheets, to increase transparency and accountability (Alvarez et al., 2024).

**Pillar 3: Contextualized Model Development & Mitigation** in fragile, low-trust environments, simpler models are often the better choice. Interpretability should come before complexity, since people must be able to understand and question the system if accountability is to mean anything (Brown et al., 2019; Mittelstadt et al., 2016). In high-trust-deficit environments, like post-conflict zones, a simple, explainable model that local stakeholders can understand is more sustainable than a complex 'black box' model. This implies a practical shift for technical teams, prioritizing simpler models (e.g., logistic regression, decision trees) that facilitate human-in-the-loop oversight. The choice of fairness metrics should also be transparent and explained, with an open discussion of trade-offs, so that decisions align with the goals defined by the community (Alvarez et al., 2024; Selbst et al., 2019). Bias mitigation methods can still be applied, but they must be used with caution, especially when data is sparse or noisy, and their performance checked against outcomes in the real world (Alvarez et al., 2024; Goknil et al., 2023).

**Pillar 4: Human-in-the-Loop Deployment & Continuous Monitoring** is designed to ensure sustainable accountability. For all critical project decisions, AI systems must serve as decision-support tools, not autonomous agents. A qualified human professional must remain "in the loop," with the authority to scrutinize, override, and assume ultimate responsibility for the final decision (Brown et al., 2019; Walsh et al., 2019). This pillar is the primary safeguard against emergent bias and automation bias. For practitioners, this means designing system interfaces that present AI recommendations as support (e.g., 'Risk Flag: 75%'), not as a final directive (e.g., 'Action: Deny Funding'), and establishing clear, accessible appeal mechanisms for affected communities. Clear and accessible mechanisms for contestability and redress must be established, allowing individuals to appeal algorithmic outcomes (Brand, 2023; Mittelstadt et al., 2016). Finally, systems must be continuously monitored for performance degradation and emergent biases, with regular, independent audits ensuring ongoing compliance (Alvarez et al., 2024).

To move the framework from a conceptual model to an actionable guide, **Table 2** provides a detailed breakdown of its operational components.

**Table 2.** Operationalizing the Context-Aware AI Fairness Framework

| <b>Pillar</b>                                    | <b>Key Activities</b>  | <b>Suggested Tools &amp; Methods</b>   | <b>Primary Responsible Actors</b>                                     |
|--|--|--|---|
| 1. Foundational Scoping & Participatory Design   | Conduct socio-political context analysis (Harrowell et al., 2018). Co-define fairness goals with communities (Brown et al., 2019). Proactively assess potential harms (Brand, 2023).                       | Ethnographic Studies, Stakeholder Workshops (Brown et al., 2019), Algorithmic Impact Assessments (AIAs) (Brand, 2023; Walsh et al., 2019).                             | Project Leads, Donors, Social Scientists, Community Representatives.  |
| 2. Data Governance & Bias-Aware Data Management  | Establish ethical protocols for vulnerable populations (Wheeler, 2014). Critically evaluate data sources (Alvarez et al., 2024). Document dataset contents and provenance.                                 | Data Governance Policies, Source Criticism, Datasheets for Datasets (Alvarez et al., 2024), Data Privacy Impact Assessments (DPIAs).                                   | Data Stewards, Project Managers, Legal/Compliance Officers.           |
| 3. Contextualized Model Development & Mitigation | Prioritize interpretable models in low-trust environments (Brown et al., 2019). Justify the choice of fairness metrics transparently (Selbst et al., 2019). Validate models on local, representative data. | Interpretable Models, Model Cards (Mitchell et al., 2019), as cited in (Alvarez et al., 2024), Contextual Validation, Causal Fairness Analysis (Alvarez et al., 2024). | Data Scientists, AI Developers, and Multi-Stakeholder Advisory Board. |

|   |  |  |  |
|---|--|--|--|
| 4. Human-in-the-Loop Deployment & Continuous Monitoring | Ensure meaningful human oversight for critical decisions (Walsh et al., 2019). Establish appeal and redress mechanisms (Mittelstadt et al., 2016). Continuously audit for bias (Alvarez et al., 2024). | Human-in-the-Loop (HITL) Protocols, Public Algorithm Registers (Brand, 2023), Regular Fairness Audits. | Project Managers, End-Users, Independent Auditors, Community Liaisons. |
|---|--|--|--|

**Box 1: The Framework in Practice – A Hypothetical Case Study**

An international NGO plans to use an AI tool to prioritize the allocation of funds for rebuilding homes in a post-conflict district with two ethnic groups who have a history of tension.

- **Pillar 1 (Scoping):** Instead of immediately deploying a tool, the NGO first conducts stakeholder workshops with leaders and members from both ethnic groups. They learn that historical property records are biased, heavily favoring one group (a form of pre-existing bias). The community collectively defines "fairness" not just as equal distribution, but as prioritizing the most vulnerable households, irrespective of ethnicity (a participatory definition of fairness). They conduct an AIA, which identifies the risk that using biased land registry data could reignite ethnic conflict.
- **Pillar 2 (Data Governance):** The NGO decides against using the biased official land registry data as the primary training source. Instead, they invest in a new, participatory data collection effort, working with community representatives to verify household vulnerability on the ground. This new dataset is carefully documented using a datasheet, noting its collection methods and limitations.
- **Pillar 3 (Model Development):** The technical team chooses a simple, interpretable regression model over a complex "black box" to predict household vulnerability based on the new data. The model's logic is explainable to community leaders. They explicitly decide not to use ethnicity as a variable but test the model to ensure it does not have a disparate impact on either group using the community-defined fairness metric.
- **Pillar 4 (Deployment):** The AI tool provides a ranked list of vulnerable households but does not make the final allocation decision. A committee of NGO officers and trusted local representatives makes the final funding decisions, using the AI's recommendations as one of several inputs (HITL). A public notice board in the community center lists the allocation criteria and provides contact for a local liaison officer whom families can approach with questions or appeals, ensuring contestability. The NGO commits to a six-month review to audit the outcomes and gather community feedback.

3.1 Discussion

Translating this framework into practice requires significant organizational and strategic shifts, where emerging technologies like blockchain can offer new mechanisms for enhancing transparency, collaboration, and trust (Bazel et al., 2023), which are central to the framework's success (Alawag, Alaloul, et al., 2024). For implementing organizations, the primary challenge is one of governance and culture, not just technology. It demands cultivating an internal culture that prioritizes ethical AI, which begins with leadership championing a vision aligned with human rights and sustainable development (Brand, 2023). This necessitates significant investment in capacity building to address the acute shortage of requisite skills in developing regions (Plevris et al., 2025; Tjebane et al., 2022). Governance must be operationalized through formal structures, such as creating multi-stakeholder advisory bodies for independent oversight, a practice successfully implemented in cities like Barcelona (Brand, 2023).

For policymakers and international funding bodies, the framework implies a shift in funding priorities and regulatory oversight. Donors should allocate resources not only for technology procurement but also for the crucial foundational work of contextual analysis and participatory design (Anderlini et al., 2004). Furthermore, governance bodies should consider mandating ethical guardrails, such as AIAs, for any AI system used in publicly funded reconstruction projects, mirroring the risk-based approach of emerging regulations (Brand, 2023; Walsh et al., 2019).

Real engagement with stakeholders is what makes or breaks a project. It is through these interactions that

fairness takes shape in practice (Selbst et al., 2019). Local communities are best placed to spot the biases that outside experts often miss (Brown et al., 2019). And in fragile, post-conflict settings, participation is also the key to building trust (Anderlini et al., 2004).

Still, turning this idea into practice is anything but simple. Elites run the risk of taking control of the framework in corrupt systems (Harrowell et al., 2018). Additionally, it makes the assumption that states are capable of handling these procedures, which frequently isn't the case in institutions that are already fragile (A JOURNEY FROM POST-CONFLICT, 2024). Furthermore, it is extremely difficult to obtain genuine participation in traumatized and divided communities (Wheeler, 2014).

The value of the framework is not negated by these issues. Instead, they draw attention to a more significant reality: improvements in governance, more robust anti-corruption policies, and initiatives to reestablish social trust and reconciliation are all necessary for its success.

#### **4. Conclusion**

This critique has revealed the potential that the use of Artificial Intelligence in construction Project Management brings, as well as how such applications in this post-conflict and developing environments, without any critique at all, pose high ethical risks. Also, the mainstream paradigm for AI, which relies on big historical datasets, is fundamentally inappropriate for the data-poor, institutionally weak, and socially fractured realities of these environments. Technical forms of tackling equity are inadequate for handling biases rooted within historical and socio-political contexts. In answer to that, this paper offered the "Context-Aware AI Fairness Framework" as a more responsible way forward that reframes the discussion around participation, governance, and human oversight. The limitations and implementation challenges identified highlight several areas in need of future study. First, there is a pressing need for empirical research that tests and validates the proposed framework in real-world pilot projects. Action-research methodologies, where researchers collaborate directly with NGOs and communities during a reconstruction project, would be particularly valuable. Longitudinal case studies are required to track the challenges and outcomes of its implementation in messy, real-world conditions. Second, this review reinforces the call for a new research paradigm focused on developing AI methods that are effective in the 'small data' or data-poor contexts that prevail in much of the world. This involves advancing techniques such as transfer learning, one-shot learning, and causal inference that can learn from limited, incomplete, and noisy data, moving beyond the assumption that more data is always the answer. Third, as the future of sustainable development will rely on the integration of multiple advanced technologies, such as AI, IoT, and BIM (Baarimah et al., 2023), further interdisciplinary research is needed to create and validate practical toolkits for integrating rich qualitative inputs into AI decision-making systems. This requires developing novel methods for developing hybrid models that can formally combine quantitative project data with insights from participatory workshops, stakeholder interviews, and expert socio-political analysis directly into the decision-making loop. Finally, future work should explore the intersection of this framework with other emerging technologies, such as blockchain, which may offer novel mechanisms for enhancing the transparency, auditability, and stakeholder trust that are critical to the framework's success (Bazel et al., 2023; Alawag, Alaloul, et al., 2024). By pursuing this new research trajectory focusing on hybrid systems that effectively combine "small data" with qualitative understanding, the academic and practitioner communities can begin to shape an AI that empowers communities. It calls for greater interdisciplinary collaboration between computer scientists, social scientists, and humanities scholars to build systems that respect and respond to the complexities of the human condition. Embracing this challenge is essential to ensure that AI contributes to a more just and sustainable global future, even in the world's most challenging environments.

#### **Acknowledgments**

The authors would like to express their sincere gratitude to Taiz University for providing the supportive environment to conduct this research. Special thanks are also extended to the anonymous reviewers for their invaluable comments and suggestions, which helped improve the quality of this manuscript.

### **Declaration of Conflicting Interests**

The authors declare that there is no conflict of interest. All authors have reviewed and approved the final version of the manuscript for publication.

### **Author Contribution**

Saleem Ahmed Alazazi: Conceptualization, Methodology, Investigation, Data Curation, Writing – Original Draft, Visualization. Dr. Aawag Mohsen Alawag: Supervision, Conceptualization, Methodology, Writing – Review & Editing, Validation. Dr. Amin Saif: Supervision, Writing – Review & Editing, Project Administration, Resources.

### **Declaration of Generative AI in the Writing Process**

The authors declare that no generative AI or AI-assisted technologies were used in the writing process of this manuscript.

### **Data Availability**

Data sharing is not applicable to this article as no new datasets were generated or analysed during the current study. All information supporting this critical review is derived from published sources that are cited in the reference list.

### **Ethics Statement**

The authors declare that this research did not involve human or animal subjects. As this study is a critical review based exclusively on published literature, formal ethical approval and informed consent were not required.

### **References**

- Alakhali, A. K., Baarimah, A. O., Alkhadri, A. H., Gabir, A. A., Hasan, K., & Alawag, A. M. (2024). Unlocking the Potential of BIM for Transformation in the Kenyan Construction Industry. *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETISIS)*, 822–828.
- Alawag, A. M., Alaloul, W. S., Al-dhawi, B. N. S., Baarimah, A. O., Bazel, M. A., & Mushtaha, A. W. (2024). A Review and Bibliometric Analysis of Blockchain Adoption Within the Context of Smart Construction Projects. *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETISIS)*, 805–811.
- Alawag, A. M., Alaloul, W. S., Liew, M. S., Al-Awag, E. A., Baarimah, A. O., & Omer, M. M. (2023). Incorporating the Applications of Building Information Modeling (BIM) Technology in the Sustainable Retrofitting of Heritage Buildings: A Systematic Review. *2023 International Conference on Sustaining Heritage: Innovative and Digital Approaches (ICSH)*, 125–130.
- Alawag, A. M., Alqahtani, F. K., Alaloul, W. S., Liew, M. S., Baarimah, A. O., Al-Mekhlafi, A.-B. A., & Sherif, M. A. (2024). Developing framework for implementing total quality management (TQM) in sustainable

industrialized building system (IBS) in construction projects. *Sustainability*, 16(23), 10399.

Alawag, A. M., Liew, M. S., Alaloul, W. S., Baarimah, A. O., & Alsalaheen, M. (2024). Implementation of industrial revolution 4.0 toward successful practices of total quality management in the sustainable construction industry. *AIP Conference Proceedings*, 3041(1), 060003.

Aleke, C. U., Usang, W. O., Obi-Obuoha, A., Jolaosho, A. A., & Michael, O. O. (1833). Artificial Intelligence as a tool for enhancing Building Information Modeling (BIM). *World Journal of Advanced Research and Reviews*.—2024.—24 (02).—Pp, 1846.

Alvarez, J. M., Colmenarejo, A. B., Elobaid, A., Fabbriizzi, S., Fahimi, M., Ferrara, A., Ghodsi, S., Mougan, C., Papageorgiou, I., & Reyer, P. (2024). Policy advice and best practices on bias and fairness in AI. *Ethics and Information Technology*, 26(2), 31.

Anderlini, S. N., & El-Bushra, J. (2004). Post conflict reconstruction. *Inclusive Security, Sustainable Peace: A Toolkit for Advocacy and Action*, 51–68.

Anzolin, G., Haraguchi, N., De Sousa, A. P. N., Savrasov, A., & Reis, J. (2012). *Bridging the AI Divide: Empowering Developing Countries through Manufacturing*.

Baarimah, A. O., Alaloul, W. S., Liew, M. S., Al-Aidrous, A. H. M. H., & Alawag, A. M. (2024). Post-disaster reconstruction projects: Insights from the construction industry in developing countries. *AIP Conference Proceedings*, 3041(1). doi: 10.1063/5.0194562/3269793

Baarimah, A. O., Gartoumi, K. I., Alaloul, W. S., Liew, M. S., Alawag, A. M., & Bahamid, R. A. (2023). Applications of Heritage Building Information Modeling (HBIM): A Bibliometric Review and Future Trends. *2023 International Conference on Sustaining Heritage: Innovative and Digital Approaches (ICSH)*, 104–111.

Bah Esseme, A., Faniyan, A., & Farayola, G. (2025). AI in Project Management: Enhancing Efficiency. *Decis. Mak. Risk Manag. J. Artif. Intell. Mach. Learn. Data Sci*, 3, 2107–2114.

Bazel, M. A., Baarimah, A. O., Mohammed, F., Alaloul, W. S., Alawag, A. M., & Alzubi, K. M. (2023). Blockchain Technology in Construction Industry: Current State and Future Trends through Bibliometric Analysis. *2023 4th International Conference on Data Analytics for Business and Industry (ICDABI)*, 114–123.

Berg, J., Snene, M., & Velasco, L. (2024). *Mind the AI Divide: Shaping a Global Perspective on the Future of Work*. United Nations. [https://www.ilo.org/publications/major-publications/mind ...](https://www.ilo.org/publications/major-publications/mind...)

Brand, D. J. (2023). Ensuring the ethical development and use of AI in local governance. *Journal of Ethics and Legal Technologies*, 5(JELT-Volume 5 Issue 2), 95–110.

Brown, A., Chouldechova, A., Putnam-Hornstein, E., Tobin, A., & Vaithianathan, R. (2019). Toward algorithmic accountability in public services: A qualitative study of affected community perspectives on algorithmic decision-making in child welfare services. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12.

Cochran, M. (2011). Fairness in Disparity: Challenging the Application of Disparate Impact Theory in Fair Housing Claims Against Insurers. *George Mason University Civil Rights Law Journal (CRLJ)*, 21(2).

Egwim, C. N., Alaka, H., Demir, E., Balogun, H., Olu-Ajayi, R., Sulaimon, I., Wusu, G., Yusuf, W., & Muideen, A. A. (2023). Artificial intelligence in the construction industry: A systematic review of the entire

construction value chain lifecycle. *Energies*, 17(1), 182.

Goknil, A., Nguyen, P., Sen, S., Politaki, D., Niavis, H., Pedersen, K. J., Suyuthi, A., Anand, A., & Ziegenbein, A. (2023). A systematic review of data quality in CPS and IoT for industry 4.0. *ACM Computing Surveys*, 55(14s), 1–38.

Harrowell, E., & Özerdem, A. (2018). The politics of the post-conflict and post-disaster nexus in Nepal. *Conflict, Security & Development*, 18(3), 181–205.

Isagah, T., & Ben Dhaou, S. (2024). *Responsible and inclusive urban AI: Opportunities and challenges for advancing sustainable development goals*.

Khan, M. F. (2025). *Artificial Intelligence in Project Scheduling Management: A Systematic Literature Review*.

Liu, Y. (2023). Artificial intelligence: Revolutionary potential and huge uncertainties. *Digital Progress and Trends Report*, 85–106.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 2053951716679679.

Musarat, M. A., Alaloul, W. S., Altaf, M., Alzubi, K. M., Irfan, M., & Alawag, A. M. (2022). Supply Chain Management in Green Construction: A Systematic Review. *2022 International Conference on Data Analytics for Business and Industry (ICDABI)*, 381–386.

Plevris, V., & Hosamo, H. (2025). Responsible AI in structural engineering: a framework for ethical use. *Frontiers in Built Environment*, 11, 1612575.

Saad, S., Rasheed, K., Ammad, S., Hasnain, M., Ullah, H., Qureshi, A. H., Alawag, A. M., Altaf, M., & Sadiq, T. (2025). Offsite modular construction adoption in developing countries: Partial least square approach for sustainable future. *Ain Shams Engineering Journal*, 16(1), 103228.

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68.

Shi, Z. R., Wang, C., & Fang, F. (2020). Artificial intelligence for social good: A survey. *ArXiv Preprint ArXiv:2001.01818*.

Tjebane, M. M., Musonda, I., & Okoro, C. (2022). Organisational factors of artificial intelligence adoption in the South African construction industry. *Frontiers in Built Environment*, 8, 823998.

Walsh, T., Levy, N., Bell, G., Elliott, A., Maclaurin, J., Mareels, I., & Wood, F. (2019). The effective and ethical development of artificial intelligence: an opportunity to improve our wellbeing. Australian Council of Learned Academia (ACOLA). Wheeler, K. (2014). Post-Conflict Reconstruction Lessons In Technology.

World Bank. (2024). A Journey from Post-Conflict Reconstruction to Sustainable Peace and Development: Three Decades of *Trusted Partnership with Bosnia and Herzegovina*. World Bank Group.