Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

When The Algorithm Does Not Know Kafala: Why Western AI Fairness Frameworks Fail in GRC Dual-Market Talent Decisions

Ms. Kriti Sarna

Abstract

Western algorithmic fairness frameworks, including the EU AI Act, NIST AI Risk Management Framework, and the academic fairness literature grounded in disparate impact and demographic parity, were designed for single-segment labor markets with anti-discrimination legislation, active regulatory enforcement, and majority-citizen workforces. The Gulf Cooperation Council (GCC) is none of these things. GCC labor markets are structurally bifurcated between state-mandated national citizen segments, governed by Saudization and Emiratization quota regimes, and a numerically dominant expatriate workforce bound by the kafala sponsorship system with limited legal mobility and differentiated institutional protection. This article argues that Western AI fairness frameworks fail in GCC talent decision contexts not merely in degree but in kind: they misidentify the relevant protected dimensions, cannot resolve the normative paradox created by legally mandated preferential hiring, and rest on regulatory enforcement assumptions with no equivalent in GCC data governance architecture. Drawing on institutional theory, Segmented Labour Market theory, and emerging GCC-specific empirical evidence, the article develops a three-dimensional critique of Western framework transfer, identifies five structural failure points, and proposes a GCC-contextualised fairness governance model organized around nationality-aware bias auditing, dual-segment impact assessment, and normative spillover management. The article contributes to the growing non-Western turn in organizational AI ethics and offers actionable guidance for HR practitioners, policymakers, and researchers operating at the intersection of AI transformation and workforce nationalization mandates.

Keywords: algorithmic fairness; GCC labor markets; dual-market economies; Saudization; Emiratization; kafala; institutional theory; Segmented Labour Market theory; AI governance; human resource management JEL Classification: M51, M12, O33, J71, K31

1. Introduction

In 2018, Amazon scrapped an AI recruitment tool after discovering it systematically downgraded resumes from women. The case became the canonical reference point for algorithmic fairness in talent decisions, catalysing a literature that now spans hundreds of empirical studies, international standards frameworks, and regulatory instruments, culminating in the EU AI Act's explicit classification of employment-related AI as high-risk. The conceptual infrastructure this literature produced, grounded in concepts such as disparate impact, demographic parity, equalized odds, and explainability, has become the global default for thinking about what fair algorithmic talent decisions should look like.

There is, however, a foundational problem with this infrastructure. It was built in, and for, institutional contexts that do not exist in large parts of the world where AI-driven talent decisions are being made at scale. The Gulf Cooperation Council (GCC) states, comprising Saudi Arabia, the United Arab Emirates, Qatar, Kuwait, Bahrain, and Oman, collectively employ more than 35 million workers, the majority of whom are migrant expatriates operating under a sponsorship architecture known as the kafala system. These are not peripheral labor markets. The GCC hosts some of the world's most capitalised aviation, banking, and logistics firms, and they are adopting AI-driven HR tools at speed. Yet the fairness frameworks being deployed in or exported to these markets were designed for demographic and institutional configurations that are structurally absent in the Gulf.

This article makes a straightforward but consequential argument: Western AI fairness frameworks do not merely need contextual adaptation in GCC talent markets. They fail in kind. The failure is not one of calibration but of architecture. The frameworks assume protected characteristic categories that misidentify the primary axes of bias in GCC talent decisions. They cannot resolve the normative paradox created by legally mandated preferential hiring of nationals under Saudization and Emiratization quota systems. They rest on regulatory enforcement mechanisms that have no institutional equivalent in GCC data governance. And they treat the workforce as a single segment when the GCC operates a structurally bifurcated dual-market system in which two legally distinct categories of workers receive qualitatively different institutional protections.

The article proceeds as follows. Section 2 reviews the Western algorithmic fairness literature and its dominant frameworks. Section 3 maps the structural characteristics of GCC dual-market labor systems. Section 4 develops a three-dimensional critique of Western framework transfer, identifying five specific failure points. Section 5 proposes a GCC-contextualised fairness governance model. Section 6 examines the role of EU AI Act normative spillover as a partial driver of fairness adoption in the GCC despite the absence of direct jurisdictional reach. Section 7 discusses theoretical contributions and practical implications. Section 8 concludes.

2. Western Algorithmic Fairness in HRM: Dominant Frameworks and Assumptions

2.1 The Academic Fairness Literature

The concept of algorithmic fairness in talent decisions draws on two parallel research traditions: the organizational justice literature, which provides the normative vocabulary of procedural, distributive, and interactional fairness (Colquitt, 2001; Greenberg, 1987); and the computer science fairness literature, which provides the technical vocabulary of bias measurement and mitigation. The synthesis of these traditions in HRM has generated a rich body of work examining how AI systems in recruitment, promotion, performance evaluation, and succession planning can encode and amplify historical inequality (Köchling and Wehner, 2020; Langer and König, 2023; Vassilopoulou et al., 2024).

The empirical record of AI bias in talent decisions is now well established. Resume screening algorithms trained on historical hire data reproduce the demographic characteristics of past hires. Computer vision-based interview assessment systems penalise non-dominant communication styles and accents. Predictive attrition models encode socioeconomic correlates of voluntary exit that disproportionately affect minority workers. Performance evaluation algorithms trained on manager ratings inherit the biases of those ratings.

In each case, the bias mechanism is the same: historical training data embeds the distributional inequalities of past human decisions, and the algorithm treats those inequalities as signal rather than noise (Köchling and Wehner, 2020; Raghavan et al., 2020; Dubey and Vachher, 2025).

The corrective frameworks the literature proposes operate at three levels. At the technical level, fairness-aware algorithm design, diverse training data, and pre-processing and post-processing bias mitigation. At the organizational level, human oversight protocols, algorithmic auditing, and governance accountability structures. At the regulatory level, mandatory risk assessments, transparency disclosures, and enforcement mechanisms with material financial consequences for non-compliance. These three levels correspond, broadly, to the architecture of the EU AI Act.

2.2 The EU AI Act and Its Employment Provisions

The EU AI Act (Regulation EU 2024/1689), which entered into force on 1 August 2024, represents the most comprehensive regulatory instrument for AI governance in force anywhere in the world. Its classification of AI systems used in employment decisions as high-risk is unambiguous and broad in scope. The high-risk classification covers AI systems used to place targeted job advertisements, screen or filter applications, evaluate or rank candidates, assess responses in screening interviews, make decisions about promotion, termination, work allocation or contract terms, and monitor and evaluate the performance and behaviour of workers (European Commission, 2024). Critically, even where a human manager remains formally responsible for the final decision, an AI system may still be high-risk if the AI system's output heavily influences which candidates advance or how workers are evaluated (DLA Piper, 2026).

High-risk system obligations under the Act include mandatory conformity assessments before deployment, technical documentation, data governance requirements ensuring training data is sufficiently representative of the populations on whom the system will operate, human oversight mechanisms, transparency disclosures to affected individuals, and post-market monitoring for discriminatory outcomes. The penalty ceiling for non-compliance with high-risk system obligations reaches up to 15 million euros or 3% of global annual turnover, whichever is higher.

The Act does not directly apply outside the European Union, though its extraterritorial provisions extend to non-EU providers whose AI systems are placed on the EU market or whose outputs are used in the EU. For the GCC, the Act functions not as binding law but as a normative template, shaping the expectations of multinational firms, international HR professional bodies, and academic research communities about what responsible AI in talent decisions should involve. This normative spillover dynamic is examined in Section 6.

2.3 The Embedded Assumptions

For the purposes of this article, three embedded assumptions of Western fairness frameworks are critical to identify, because each assumption is violated in the GCC context.

The first is the protected characteristic assumption. Western frameworks organise fairness analysis around the protected characteristics encoded in anti-discrimination legislation: gender, race, ethnicity, age, disability, religion. These categories reflect the historical patterns of discrimination that motivated legislative intervention in North American and Western European labor markets. The fairness metrics developed in the computer science literature, including demographic parity, equalized odds, and individual fairness, are operationalised in terms of these categories.

The second is the single-segment workforce assumption. Western frameworks assume that all workers exist within the same legal and institutional framework and are therefore entitled to the same procedural protections and fairness guarantees. The notion that a fairness audit should assess differential outcomes across legally distinct worker categories with different institutional rights is entirely absent from the mainstream literature.

The third is the regulatory enforcement assumption. Western fairness frameworks are embedded within, and derive much of their practical force from, regulatory systems with active enforcement bodies, material financial penalties, and citizens' rights of legal recourse against discriminatory algorithmic decisions. The normative authority of frameworks like the EU AI Act rests on the threat of enforcement. In jurisdictions where that enforcement architecture does not exist, the frameworks become voluntary, and voluntarism in fairness governance tends, as the literature on ethics-washing demonstrates, to produce declarations rather than practices.

3. The GCC Dual-Market Labor System: Structural Characteristics

3.1 The Kafala Architecture

The kafala, or sponsorship, system defines the legal relationship between foreign workers and their local sponsor, who is typically their employer. The system ties a non-national's legal residency to their sponsor/employer and bestows substantial power to employers over migrant workers through a complex of recruitment, labour, and immigration regulations as well as social dynamics (Kalush and Saraswathi, 2024). The kafala is not a single law but an interlaced architecture of regulations across multiple jurisdictions. Its core operational logic is simple: without employer sponsorship, a migrant worker has no legal right to reside in the country, and departure from employment typically requires employer consent.

The institutional consequences for labor market functioning are profound. Migrant workers under kafala face severely constrained labor mobility: switching employers requires sponsor approval, and absconding from employment can result in deportation. This creates asymmetric bargaining power between employer and worker that is not ameliorated by collective bargaining rights, since migrant workers in GCC states generally have no right to organize. Racial discrimination and gender-based violence have been documented as endemic in kafala-governed employment relationships, particularly for domestic workers and low-skilled laborers (ADHRB, 2025).

The demographic scale of the kafala system is the context within which AI talent decisions operate. In Qatar, nationals represent less than 15% of the total resident population. In the UAE, nationals comprise approximately 11% of the population. Even in Saudi Arabia, which has the largest national citizen population of any GCC state, expatriates constitute a majority of the private sector workforce. This demographic reality, combined with the legal architecture of kafala, means that the workforce on which AI-driven talent decisions operate is predominantly composed of workers with structurally reduced institutional protections.

3.2 Nationalization Quota Systems: Saudization and Emiratization

Operating in parallel with the kafala system, and creating the central normative paradox examined in this article, are the workforce nationalization programs of GCC states. Saudi Arabia's Nitaqat system, known as Saudization, requires every private-sector employer to fill a set percentage of its jobs with Saudi citizens, with quotas varying by sector and employer size, and non-compliance resulting in denial of visa issuance rights, denial of access to government services, and disqualification from public contract bidding. The UAE's Emiratization program operates similarly, with the banking sector targeted at 45% Emiratisation and sector-specific mandates across healthcare, hospitality, and logistics. The accounting profession in Saudi Arabia faces a quota of 40% set in October 2025, rising by 10 percentage points annually to reach 70% by 2028 (House of Saud, 2026).

These are not aspirational diversity targets. They are legally mandated preferential hiring regimes with material compliance consequences. Firms that fall below their Nitaqat tier face immediate operational restrictions. The compliance incentive is strong and continuous. An HR department in a Saudi private firm is legally required to preference Saudi nationals in hiring decisions. The question of whether an AI recruitment system that scores Saudi nationals higher than expatriate candidates with equivalent credentials is enacting fairness or discrimination has no clear answer within any existing Western fairness framework.

3.3 Nationality as the Primary Fairness Axis

Research specifically testing AI recruitment models in Saudi Arabian conditions has found that standard AI recruitment tools trained on historical hiring data revealed significant biases in gender and nationality-based hiring, making them inappropriate for the Saudi Arabian diverse hiring environment (Albaroudi et al., 2024, as cited in the HITHIRE study). The study categorized nationalities into Arab and non-Arab groups for fairness analysis, a categorization that has no equivalent in any Western fairness taxonomy.

This empirical finding is theoretically predictable. An AI system trained on historical hire data from a GCC private firm will learn that national employees have higher retention rates (because kafala makes exit costly for expatriates but also because nationals receive preferential contract terms), that national employees are promoted more frequently (because nationalization targets require visible national representation at senior levels), and that national employees appear in training completion records more prominently (because firms prioritize training investment in employees with longer expected tenure). Each of these patterns is real in the training data. None of them reflects unambiguous productive ability. All of them will be learned as predictive features.

4. Five Structural Failure Points of Western Framework Transfer

4.1 Failure Point One: Misidentification of Protected Dimensions

Western fairness frameworks are organized around protected characteristics drawn from anti-discrimination law: gender, race, ethnicity, religion, age, disability. These categories do not map onto the primary axes of bias risk in GCC talent decisions. In a GCC private sector organization making AI-assisted hiring decisions, the most consequential demographic division is not race in the Western sense but nationality status, specifically national citizen versus expatriate, and within the expatriate population, national origin in ways that interact with kafala tier, contract type, and Saudization quota position.

Standard bias auditing protocols, which test for differential outcomes across gender and race categories, will fail to detect the most significant fairness risks in this context. An AI system that produces identical gender parity outcomes while systematically disadvantaging South Asian expatriates relative to Arab expatriates, or that advantages Arab expatriates relative to national citizens in ways that inadvertently undermine Saudization compliance, will pass a standard Western bias audit and simultaneously produce the most consequential unfair outcomes in the GCC context.

The corrective implication is not merely to add nationality to the list of protected characteristics. Nationality in the GCC context is not a characteristic alongside others but the organizing axis of the entire labor market institutional architecture. Treating it as an additional demographic variable misses the structural point: the two-tier system is not an incidental feature of GCC workplaces but the legal foundation on which all employment relationships are built.

4.2 Failure Point Two: The Nationalization Paradox

Western fairness frameworks operate on the premise that differential treatment on the basis of protected characteristics is morally and legally prohibited. This premise is directly contradicted by the legal structure of GCC nationalization mandates. Saudization and Emiratization require differential treatment of nationals and expatriates in hiring, promotion, and retention decisions. This differential treatment is not a tolerated exception to fairness norms: it is the legal obligation.

No existing algorithmic fairness framework provides a mechanism for adjudicating this conflict. Demographic parity, the most commonly applied group fairness metric, would require that national and expatriate applicants be selected at equal rates. But Nitaqat requires exactly the opposite: that national applicants be selected at higher rates than their representation in the applicant pool would produce under neutral conditions. An AI system achieving demographic parity across the national/expatriate dimension would be simultaneously enacting algorithmic fairness in the Western sense and violating Saudi labor law.

This is not a technical challenge resolvable through parameter adjustment. It is a normative conflict that requires a new conceptual framework. The GCC context demands a distinction between two types of differential treatment that Western fairness literature conflates: differentially harmful treatment that disadvantages protected groups contrary to their legitimate interests, and differentially mandated treatment that operates contrary to expatriate preferences but within a legally authorised national policy framework. The design of AI systems for GCC talent decisions requires a framework that can distinguish these two types without either collapsing the distinction or pretending that mandated differential treatment has no fairness consequences for the affected expatriate population.

4.3 Failure Point Three: Single-Segment Workforce Assumption

Segmented Labour Market theory, developed by Piore (1979), proposes that labor markets are structurally divided into primary and secondary segments operating with distinct rules, reward structures, and mobility patterns. In the GCC, this segmentation is not a sociological tendency but a legal reality. National citizens and kafala-bound expatriates do not occupy the same legal institutional space. They have different rights to labor mobility, different rights to collective representation, different rights to legal recourse against employer decisions, and different levels of access to state labor market institutions.

Western fairness frameworks assume a single institutional space in which all workers are legally entitled to the same procedural protections. A fairness audit conducted under this assumption will assess whether the AI system produces differential outcomes across demographic groups but will not assess whether the system's design respects or violates the differential institutional rights of two legally distinct worker populations. The secondary segment worker who is screened out of a promotion by an AI system operating in a Western institutional context has recourse to anti-discrimination law, labor tribunals, and potentially class action litigation. The kafala-bound expatriate screened out of the same promotion in a GCC context has none of these protections and faces the additional constraint that challenging the decision risks employer retaliation in a system where the employer controls their right to remain in the country.

The implication is not merely that fairness audits in GCC contexts should assess differential outcomes across national and expatriate groups. It is that the ethical stakes of unfair algorithmic decisions are asymmetrically distributed in a way that Western frameworks, which treat all affected parties as legally equal, cannot register. The harm of an unfair algorithmic decision is compounded for kafala-bound workers by their structural inability to contest it.

4.4 Failure Point Four: Regulatory Enforcement Architecture

Western algorithmic fairness frameworks derive their practical force from regulatory enforcement systems. The EU AI Act creates financial penalties with material consequences for large firms: up to 35 million euros or 7% of global annual turnover for prohibited practices, up to 15 million euros or 3% of global annual turnover for high-risk system non-compliance. These figures are not primarily important as deterrents: they are important because they make compliance with fairness requirements a financial priority for boards and executive committees, embedding fairness governance into organizational risk management.

The GCC regulatory environment contains no equivalent enforcement architecture for AI fairness in talent decisions. The UAE has no horizontal AI law regulating private-sector recruitment AI (Recruitly, 2026). Saudi Arabia's PDPL does not contain algorithmic accountability provisions equivalent to GDPR Article 22 or the EU AI Act's high-risk system obligations. GCC data protection laws are recent, still-evolving, and do not address the employment-specific AI governance questions that the EU AI Act treats as a primary concern. The absence of enforcement architecture does not mean that fairness governance is impossible in GCC organizations. It means that the mechanism by which Western frameworks convert normative commitments into organizational practices, namely regulatory coercion, is not available.

In the absence of coercive mechanisms, institutional theory predicts that adoption of fairness practices will be driven by mimetic and normative isomorphism rather than legal compliance (DiMaggio and Powell, 1983). Organizations will adopt fairness controls to the extent that their multinational peers have done so (mimetic pressure) and to the extent that their HR leadership is exposed to international professional standards through bodies like SHRM and CIPD (normative pressure). Both of these mechanisms are present in GCC markets, particularly in sectors with high multinational firm penetration such as banking and aviation, but they are significantly weaker drivers of adoption than regulatory coercion. The result is systematic under-adoption of fairness controls relative to the level warranted by the bias risks present in these markets.

4.5 Failure Point Five: The Ethics-Washing Risk in Weak Regulatory Environments

International evidence warns that algorithms can reproduce historical inequities when their training data contain embedded biases, and that voluntary ethics codes often lapse into ethics-washing when enforcement incentives are weak (Green, 2022, cited in the GCC AI governance literature). Fragmented regulations in the GCC amplify these risks (Al-Shehabi and colleagues, 2024). Ethics-washing typically arises when non-binding instruments that laud ethical principles but impose no sanctions stand in for hard law or independent oversight.

The risk is particularly acute in GCC contexts because the adoption of Western fairness framework terminology, deploying concepts such as bias auditing, explainability, and demographic parity in corporate AI governance documentation, creates the appearance of fairness compliance without addressing the structural features of GCC labor markets that make those frameworks inadequate. A GCC firm that adopts an EU AI Act-style bias audit, tests for gender and race parity in its recruitment AI, finds no significant disparities, and declares its system fair, has performed a fairness audit that is simultaneously formally rigorous and substantively irrelevant to the most significant fairness risks in its operating context.

5. Toward a GCC-Contextualised Algorithmic Fairness Governance Model

The critique developed in Section 4 is not a counsel of paralysis. The fact that Western frameworks fail in GCC contexts does not imply that fairness governance in GCC talent decisions is impossible or unnecessary. It implies that a different framework architecture is required, one built from the structural characteristics of GCC labor markets rather than imported from institutional contexts where those characteristics do not apply.

The following model proposes three organizing dimensions for GCC-contextualised algorithmic fairness governance in talent decisions.

5.1 Dimension One: Nationality-Aware Bias Auditing

GCC bias auditing for AI talent systems must be reorganized around nationality status as the primary protected dimension, with national/expatriate status as the fundamental axis and national origin within the expatriate population as a secondary axis of analysis. This requires two substantial departures from standard Western audit practice.

First, it requires distinguishing between what may be termed compliance-driven nationality effects and unjustified nationality effects in algorithmic outputs. A recruitment AI that advantages Saudi nationals over expatriates in a proportion consistent with Nitaqat requirements is enacting a compliance-driven nationality effect. The same system that advantages Arab expatriates over South Asian expatriates with equivalent credentials is enacting an unjustified nationality effect. Standard demographic parity metrics treat both as equivalent sources of unfairness and prescribe equalization of selection rates across all groups. A GCC-contextualised audit must treat them as categorically distinct and apply different corrective logics to each.

Second, it requires auditing for proxy discrimination through nationality-correlated variables. AI systems trained on GCC workforce data will encounter variables such as contract type (fixed-term versus open-ended), visa category, salary band, and training participation rate that are strongly correlated with national/expatriate status without being nationality itself. Systems that are nationality-neutral in their explicit feature set may nonetheless encode effective nationality discrimination through these proxies. GCC-contextualised audit protocols must test for proxy effects explicitly, using counterfactual fairness approaches that assess whether outcomes would differ if nationality-correlated variables were held constant.

5.2 Dimension Two: Dual-Segment Impact Assessment

Western fairness impact assessments evaluate differential outcomes across demographic groups within a single institutional framework. GCC impact assessments must evaluate differential outcomes across two legally distinct worker populations whose institutional positions are not equivalent and whose exposure to harm from unfair algorithmic decisions differs qualitatively, not merely quantitatively.

A dual-segment impact assessment should evaluate three categories of impact for each AI talent decision system. The first is within-segment differential impact, assessing whether the system produces differential outcomes within the national segment (for example, by gender or region) and within the expatriate segment (for example, by national origin or visa category). The second is cross-segment differential impact, assessing whether the system's aggregate effects on each segment are proportionate to the differential institutional protections of each segment, rather than simply equal. The third is harm compounding assessment, evaluating whether unfavorable algorithmic decisions for kafala-bound expatriates are likely to trigger cascading harms through the kafala architecture, such as inability to switch employers, visa cancellation risk, or forced departure from the country.

The inclusion of harm compounding assessment represents a significant departure from Western impact assessment practice. Standard disparate impact analysis treats a discriminatory algorithmic decision as the terminal harm event. In GCC contexts, an unfavorable algorithmic decision for a kafala-bound worker may be the beginning of a harm cascade whose full extent extends far beyond the workplace.

5.3 Dimension Three: Normative Spillover Management

In the absence of direct regulatory coercion, GCC organizations operating in sectors with high international exposure face normative pressure to adopt fairness practices through mimetic and normative isomorphism channels. This normative spillover from international frameworks, particularly the EU AI Act, creates both opportunities and risks for GCC fairness governance.

The opportunity is that multinational firms operating both in EU jurisdictions and in the GCC face internal consistency pressure to extend their EU compliance frameworks to non-EU entities. This creates a private governance mechanism that can drive meaningful fairness practice adoption in the GCC's most internationalised sectors, particularly aviation and international banking, without waiting for GCC-specific regulatory development.

The risk is that the normative spillover mechanism imports Western framework architectures without adaptation, producing the ethics-washing dynamic identified in Section 4.5. Organizations that respond to normative spillover by adopting EU AI Act-aligned bias audit protocols will be seen to have responded to international fairness expectations while leaving the most consequential GCC-specific fairness risks unaddressed.

Normative spillover management therefore requires organizations to perform two simultaneous and distinct governance tasks: satisfying the Western-derived fairness expectations of international partners and professional bodies, while implementing the GCC-contextualised audit and impact assessment practices described in Dimensions One and Two. This dual-track approach is more resource-intensive than simply adopting a single framework, but it is the only approach that addresses the full spectrum of fairness risks in GCC dual-market talent decisions.

6. The EU AI Act as Normative Architecture Without Jurisdictional Reach

The EU AI Act entered into force on 1 August 2024, with employment-related high-risk obligations becoming applicable on 2 August 2026, subject to the Digital Omnibus simplification package that proposed deferral of certain high-risk deadlines to December 2027 (DLA Piper, 2026). Even in its deferred form, the Act creates the most comprehensive binding regulatory architecture for AI in talent decisions yet produced. Its employment provisions explicitly cover AI systems used in recruitment, selection, promotion, performance evaluation, work allocation, monitoring, and termination, with the clarification that material influence over employment outcomes is sufficient to trigger high-risk classification regardless of formal decision authority (European Commission DLA Piper guidelines, 2026).

The Act has no binding force in GCC jurisdictions. But its influence operates through at least four channels that are materially relevant to GCC fairness governance.

The first is the subsidiary compliance channel. GCC subsidiaries of EU-headquartered firms whose parent entities must comply with EU AI Act obligations face internal governance pressure to apply comparable standards across their global operations. A European bank operating in both Frankfurt and Dubai cannot realistically maintain EU-compliant bias auditing in its Frankfurt operations and no bias auditing in its Dubai operations without creating governance asymmetries that create reputational and operational risks.

The second is the vendor compliance channel. AI talent decision tools deployed in GCC organizations are frequently produced by vendors headquartered in or selling to EU markets. These vendors face EU AI Act compliance obligations that require them to design their products for high-risk use cases with conformity assessments, data governance documentation, and bias testing built in. GCC clients of these vendors therefore receive products that have been designed to Western fairness standards, creating a de facto adoption of Western framework architecture through the technology supply chain.

The third is the professional standards channel. HR professional bodies operating in the GCC, particularly SHRM MENA and CIPD MENA, draw their professional standards frameworks primarily from their parent organizations' international standards, which are increasingly being updated to reflect EU AI Act and similar frameworks. HR directors who are CIPD or SHRM-qualified, or who are pursuing postgraduate qualifications in HRM, are exposed to these standards and carry them into organizational practice through normative professional socialization.

The fourth is the investor and ratings channel. GCC firms listed on international exchanges or seeking international investment are subject to ESG rating frameworks that increasingly incorporate AI governance indicators. The Global Reporting Initiative, the UN Principles for Responsible Investment, and emerging AI-specific ESG metrics all reflect Western fairness framework assumptions. GCC firms seeking international capital face external pressure to adopt fairness governance practices that satisfy these metrics, even without direct regulatory obligation.

The cumulative effect of these four channels is that the EU AI Act functions as normative architecture without jurisdictional reach in GCC markets: it shapes the fairness governance expectations that GCC organizations face from international partners, vendors, professional bodies, and investors, without creating the enforcement mechanisms that would make those expectations materially binding. This is precisely the condition under which institutional theory predicts ethics-washing to be most likely, and it is therefore the condition under which the GCC-contextualised framework developed in Section 5 is most urgently needed.

7. Theoretical Contributions and Practical Implications

7.1 Theoretical Contributions

This article makes three theoretical contributions to the management and business economics literature.

The first contribution is conceptual and critical. By demonstrating that Western algorithmic fairness frameworks fail in kind rather than merely in degree in GCC talent market contexts, the article advances a structurally grounded critique of framework transfer assumptions in international management research. The literature on AI ethics and governance has been, to date, largely produced in and for Western institutional contexts, with non-Western applications treated as extensions requiring local adaptation rather than as contexts requiring fundamentally different conceptual architectures. This article argues for the latter position and develops it in the specific case of GCC dual-market labor systems, but the argument generalises to any employment context where structurally bifurcated labor markets, state-mandated preferential hiring regimes, or asymmetric institutional protection architectures are present.

The second contribution is the application of Segmented Labour Market theory to algorithmic fairness governance. Piore's (1979) framework has been applied extensively to understand labor market inequality but has not previously been applied to the fairness of algorithmic decision systems operating within segmented markets. This article demonstrates that SLM theory provides a powerful diagnostic tool for identifying the failure modes of single-segment fairness frameworks when applied to bifurcated labor markets, and for specifying the dual-segment impact assessment logic that a GCC-contextualised framework requires.

The third contribution is the concept of the nationalization paradox as a distinct category of algorithmic fairness problem. No existing fairness framework, technical or normative, provides a mechanism for resolving the conflict between demographic parity requirements and legally mandated preferential hiring regimes. This article names and defines this conflict, demonstrates its structural irresolvability within Western fairness architectures, and proposes a compliance-driven versus unjustified differential treatment distinction as the conceptual foundation for a GCC-contextualised resolution.

7.2 Practical Implications for HR Practitioners

For CHROs and HR directors in GCC private sector organizations deploying AI-driven talent decision tools, the article's practical implications are organized around four priorities.

The first priority is audit scope redesign. Existing bias auditing protocols should be reviewed and redesigned to incorporate nationality status as the primary protected dimension, with explicit testing for nationality proxy discrimination through correlated variables such as contract type, visa category, and compensation band. Audits conducted exclusively on gender and race categories, however carefully executed, do not address the primary fairness risks in GCC talent AI systems.

The second priority is compliance-driven effect mapping. For each AI talent decision system, HR practitioners should produce a compliance-driven effect map that explicitly documents which algorithmic nationality effects are legally required by Saudization or Emiratization mandates, which are neutral with respect to nationalization requirements, and which are neither required nor justified. Only the third category represents unjustified algorithmic bias in the GCC context: the first category is legal compliance, and conflating it with the third, as Western fairness frameworks do by default, prevents accurate risk identification.

The third priority is harm cascade assessment for secondary-segment workers. For AI systems making or influencing promotion, termination, and performance decisions, HR practitioners should assess whether unfavorable algorithmic outputs for kafala-bound employees are likely to trigger kafala-specific harm cascades. Where such risks are identified, mandatory human override requirements should be implemented for algorithmic decisions affecting kafala-bound workers, with documented rationale for each override.

The fourth priority is dual-track governance documentation. GCC organizations facing normative spillover pressure from international partners and professional bodies should maintain governance documentation organized in two tracks: a Western-framework-aligned track for international stakeholder communication, and a GCC-contextualised track for internal fairness governance. The first track satisfies international expectations; the second track addresses the actual fairness risks of the organization's operating context. Maintaining only the first track constitutes ethics-washing; maintaining only the second track creates communication gaps with international stakeholders. Both are needed.

7.3 Implications for Policymakers

For GCC policymakers and regulatory bodies, the article has two primary implications.

The first is the urgent need for algorithmic accountability provisions in GCC data protection law. Saudi Arabia's PDPL and the UAE's Federal Data Protection Law provide foundational data governance frameworks but contain no provisions addressing algorithmic decision-making in employment contexts. The EU AI Act's employment-specific high-risk classification, human oversight requirements, and algorithmic explainability obligations represent a regulatory standard that GCC jurisdictions should consider adopting in contextualised form. Critically, any GCC-specific algorithmic accountability framework must grapple with the nationalization paradox explicitly, providing legal guidance on how AI systems should be designed and audited in the context of mandatory Saudization and Emiratization requirements.

The second is the need for GCC-specific algorithmic fairness standards development through regional bodies. The GCC Standardization Organization (GSO) is the appropriate institutional vehicle for developing regional AI fairness standards that reflect the structural characteristics of GCC labor markets rather than importing Western standards without adaptation. Such standards could provide the normative infrastructure for a GCC-specific coercive institutional pressure mechanism, reducing the current dependence on mimetic and normative isomorphism as the primary drivers of fairness adoption.

8. Limitations and Future Research

This article has several limitations that future research should address.

The framework developed here is primarily conceptual and critical, drawing on theoretical synthesis and the limited empirical evidence available from GCC AI talent decision contexts. Empirical validation of the five failure points and the three-dimensional fairness model is urgently needed. Specifically, empirical work should test whether nationality-aware bias auditing of GCC recruitment AI systems detects differential effects that gender-and-race-only audits miss; whether the nationalization paradox produces measurable ethical compliance conflicts in organizations simultaneously subject to Nitaqat requirements and international fairness standards; and whether dual-segment harm cascade assessment captures harm patterns not registered by standard impact assessment protocols.

A second limitation is the article's focus on Saudi Arabia and the UAE as the primary GCC cases. Qatar, Kuwait, Bahrain, and Oman each have distinct nationalization architectures, labor market compositions, and regulatory environments. A comparative analysis across all six GCC states would reveal important variation in the severity and form of the failure points identified here. Qatar's kafala reform claims, for instance, create a different institutional landscape than Saudi Arabia's unchanged kafala architecture, and this difference has implications for harm cascade assessment.

A third limitation is the article's primary focus on the employer-side of algorithmic fairness. The worker-side experience of AI-driven talent decisions in GCC contexts, including how kafala-bound workers cognitively process, contest, or are precluded from contesting unfair algorithmic outcomes, is an important research domain that requires qualitative investigation. In-depth interviews with expatriate workers who have experienced AI-mediated rejection decisions in GCC hiring processes would provide evidence about the harm cascade mechanism at the individual level.

Future research should also examine the dynamic effects of GCC regulatory development on fairness adoption. As Saudi Arabia's PDPL matures, as UAE data governance frameworks develop, and as the EU AI Act's normative spillover effects intensify through the vendor compliance and investor channels identified in Section 6, the institutional pressure landscape for GCC algorithmic fairness governance will change. Longitudinal research tracking adoption of fairness controls across these regulatory transitions would illuminate how coercive pressure mechanisms develop in emerging market contexts.

9. Conclusion

The transfer of Western algorithmic fairness frameworks to GCC talent decision contexts is not a matter of adding nationality to a checklist of protected characteristics. It is a matter of confronting the structural incompatibility between frameworks designed for single-segment, enforcement-backed, anti-discrimination-law-grounded institutional contexts and the actual institutional architecture of GCC dual-market labor systems. The kafala system, the nationalization quota regimes of Saudization and Emiratization, the nascent and enforcement-light state of GCC data governance, and the structural vulnerability of the secondary-segment workforce together create a fairness governance challenge that Western frameworks are architecturally incapable of addressing.

The five failure points identified in this article, misidentification of protected dimensions, the nationalization paradox, the single-segment workforce assumption, the absence of enforcement architecture, and the ethics-washing risk in weak regulatory environments, are not incidental weaknesses of particular frameworks. They are consequences of the foundational assumptions embedded in all major Western algorithmic fairness frameworks as they currently stand. Addressing them requires not adaptation but reconstruction.

The three-dimensional GCC-contextualised fairness governance model proposed here, organized around nationality-aware bias auditing, dual-segment impact assessment, and normative spillover management, provides a starting architecture for that reconstruction. It is offered not as a finished framework but as a research agenda. The empirical work needed to validate, refine, and operationalise this model represents a significant opportunity for management and business economics research to contribute to one of the most consequential governance challenges of the current AI transformation wave in the world's most demographically distinctive labor markets.

References

1. Albaroudi, E., Mansouri, T. and Alameer, A. (2024). A comprehensive review of AI techniques for addressing algorithmic bias in job hiring. AI, 5(1), 383-404.
2. AlShehabi, O.H. (2021). Policing labour in empire: the modern origins of the kafala sponsorship system in the Gulf Arab States. British Journal of Middle Eastern Studies, 48(2), 291-310.
3. American Disability Rights and Human Rights (ADHRB). (2025). Trapped in the system: Women domestic workers and kafala in the Gulf states. ADHRB Publications.
4. Colquitt, J.A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. Journal of Applied Psychology, 86(3), 386-400.
5. DiMaggio, P.J. and Powell, W.W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. American Sociological Review, 48(2), 147-160.
6. DLA Piper. (2026, June). EU Commission publishes draft guidelines on high-risk AI in employment. DLA Piper GENIE Global Employment Law Developments.
7. Dubey, S. and Vachher, L. (2025). Stride towards fair AI: Addressing algorithmic prejudices in human resource analytics. International Journal of Human Resource Management [in press].
8. European Commission. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (AI Act). Official Journal of the European Union.
9. Greenberg, J. (1987). A taxonomy of organizational justice theories. Academy of Management Review, 12(1), 9-22.

10. Green, B. (2022). The contestation of tech ethics: A sociotechnical approach to technology ethics in practice. Journal of Social Computing, 2(3), 209-225.
11. House of Saud. (2026). Saudization (Nitaqat): Employment rules for foreign companies operating in Saudi Arabia. houseofsaud.com/business/saudization
12. Kalush, R. and Saraswathi, V. (2024). Reforms without rights: The GCC states' blinkered view of labour reforms. Gulf Research Center Explanatory Note No. 2 / 2024.
13. Kim, K. (2025). Strategic human resource management in the era of algorithmic technologies: Key insights and future research agenda. Human Resource Management. https://doi.org/10.1002/hrm.22268
14. Köchling, A. and Wehner, M.C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. Business Research, 13(3), 795-848.
15. Langer, M. and König, C.J. (2023). Introducing and testing the algorithmic aversion/appreciation questionnaire (3AQ). Journal of Personnel Psychology, 22(3), 131-142.
16. Migration Policy Institute. (2024, December). As the Gulf region seeks a pivot, reforms to its oft-criticized immigration policies remain a work in progress. migrationpolicy.org
17. Piore, M.J. (1979). Birds of passage: Migrant labor and industrial societies. Cambridge University Press.
18. Raghavan, M., Barocas, S., Kleinberg, J. and Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. Proceedings of the 2020 ACM Conference on Fairness, Accountability, and Transparency, 469-481.
19. Recruitly. (2026, May). AI Act for recruiters: Where the rules apply in 2026, country by country. recruitly.io/resources/ai-act-for-recruiters
20. Scott, W.R. (1995). Institutions and organizations. Sage.
21. Soleimani, M., Intezari, A., Arrowsmith, J., Pauleen, D.J. and Taskin, N. (2024). Reducing AI bias in recruitment and selection: An integrative grounded approach. Big Data and Society, 11(1).
22. Vassilopoulou, J., Doyle, N. and Özbilgin, M. (2024). Diversity and inclusion in the digital age: The role of algorithms in HRM. International Journal of Human Resource Management, 35(4), 589-614.
23. Faltara. (2026, May). Qatarization, Bahrainization, Omanization: What foreign companies operating beyond Saudi need to know in 2026. faltara.com/blogs/gcc-localization-comparison-2026

Appendix: Western versus GCC-Contextualised Algorithmic Fairness Governance

Dimension	Western Framework	GCC-Contextualised Framework
Primary protected dimension	Gender, race, ethnicity, age, disability, religion	Nationality status (national vs. expatriate); national origin within expatriate segment
Fairness metric baseline	Demographic parity; equalized odds; individual fairness	Compliance-adjusted parity: distinguishes mandated from unjustified differential treatment
Workforce assumption	Single-segment: all workers in same institutional framework	Dual-segment: nationals and kafala-bound expatriates have legally distinct protections
Harm assessment scope	Adverse algorithmic decision as terminal harm event	Harm cascade assessment: adverse decision as potential trigger of kafala-linked cascading harms
Enforcement mechanism	Regulatory coercion: material financial penalties for non-compliance	Normative spillover management: dual-track governance for international and internal audiences
Bias audit scope	Gender and race disparity testing against historical baseline	Nationality-aware auditing including proxy discrimination through correlated variables
Regulatory driver	EU AI Act; GDPR Article 22; national anti-discrimination law	PDPL (Saudi); UAE DPRL; emerging GSO standards; international normative pressure
Nationalization mandate handling	Not addressed: mandated differential treatment conflated with discriminatory treatment	Compliance-driven effect mapping: legal nationality effects distinguished from unjustified bias
Ethics-washing risk level	Moderate: enforcement architecture reduces but does not eliminate risk	High: weak coercive pressure in absence of GCC-specific regulatory enforcement

This comparison table illustrates the structural differences between Western-derived algorithmic fairness frameworks and the GCC-contextualised model proposed in this article across nine governance dimensions. Neither framework is universally superior: the Western framework reflects its institutional context accurately; the GCC-contextualised framework reflects the structurally distinct GCC dual-market context. The purpose of the comparison is to make visible the assumptions embedded in Western frameworks that fail when those frameworks are transferred to GCC operating environments.