Journal of Advanced Computer Science & Technology, 13 (1) (2026) 66-73

Journal of Advanced Computer Science & Technology

Website: www.sciencepubco.com/index.php/JACST
https://doi.org/10.14419/dvgv3160
Research paper

The Trust Paradox in AI-Driven Hiring: Socio-Cultural Sensitivity As A Mediator of Fairness

Dr. Ruchi Saxena *

Assistant Professor, Lucknow Public College of Professional Studies, Lucknow, India
**Corresponding author E-mail: ruchisaxena123@gmail.com*

Received: 31-07-2026, Accepted: 01-08-2026, Published: 23-08-2026

Abstract

As Artificial Intelligence technology becomes part of recruitment processes, it is typically discussed as a remedy for human judgment and the handling of administrative tasks. However, we do not yet understand the social consequences of the use of these technologies at the pace at which they are accepted and used, which is creating a “trust paradox”. However, AI is being sold as objective, but often it reinforces inherent inequity and diminishes candidate trust. In this chapter, the problem that advanced algorithmic recruitment is not creating institutional or public trust is examined, and it's argued that technical approaches to ‘fairness’ shortcomings are not sufficient if they are not aligned with the varied, lived experiences of the global workforce.
A Universal Metric is not necessarily appropriate for the local context.
Standardized success measures, frequently tied to Western norms of workplace success, are typically used to train existing AI hiring systems on large datasets. Rather than this “universal candidate profile” - they basically try to do this - these models do not consider the fact that the meaning of “merit”, "potential" or “professionalism” is firmly rooted in cultural and socio-economic context.
This section will take up the conflict between algorithmic uniformity and socio-cultural diversity. For example, the ‘gaps in employment’ or ‘non-traditional educational path’ disadvantages some data sets but doesn't for others where it is more a reflection of common cultural experience, such as precarious jobs or non-conventional educational journeys. First, it is not only a technical bias, but a failure to demonstrate cultural literacy in the design of the underlying AI models. These systems continue to reproduce inequalities, particularly by rejecting candidates who don't resemble the dominant culture in the training system, by defining “fit” as a rigid monolith.
Theoretical Framework: Socio-Cultural Sensitivity as mediator
This middle part asserts that to be understood as fair and, therefore, credible, AI should strive for a shift towards “socio-cultural sensitivity”. A framework that infuses socio-cultural processes as a necessary middleman in algorithmic decisions, using concepts from organizational behavior and cultural psychology, will be suggested.
Cultural caliper of Merit: I will talk about how the AI architecture can be changed to embrace local conceptions of success. Systems should use multi-objective optimization of systems, based on the labor market constraints and values of the local market, instead of a single objective.
This section will draw on my previous research into ‘green nudges’ and behavioral alignment to consider how trust is a result of the alignment between institutional values and candidate expectations. This perceived procedural justice arises as an AI tool displays an understanding of the cultural level that matches the candidate's community norms.
The chapter will look at how socio-cultural sensitivity can be conveyed with Explainable AI (XAI) as a bridge. An ethically robust system should explain the decision in a culturally resonant way – meaning providing explanations for why a decision was made that also reflects the candidate's context in his or her locality at work.
The world's implications and navigating the North-South Divide.
Ethical issues surrounding AI in recruitment vary from country to country. In this section, it will be contrasted how AI is being used in markets when regulations are stringent (Global North) with those where regulations are emerging in the market (Global South). I will define hiring algorithms as a type of algorithmic colonialism if they are “exported” from one cultural context without adequate adaptation and thus imposed on another. I’ll argue that, when measuring the regulatory-equity gap, we can propose that TNC's should apply a strategy of “contextual compliance” whereby AI systems are notified for both mathematical neutrality and regulation according to the socio-cultural values and human rights of the area where they are used.
The purpose of this chapter is to advance the Handbook of Business Ethics and Values in a Globalized World by moving beyond the dichotomy of “bias vs. fairness” to consider the issue of “contextual legitimacy”.
In my contribution, I have two things I want to mention:
Most of all, it provides a framework for HR practitioners to assess the “cultural fit” of AI platforms. Second, it creates a theoretical connecting “string” for scholars to walk across the digital-dialectic-technology divide, to grasp the fact that trust in technology is not only a technical issue but also a social one. I want to propose, in contrast, an alternative vision of an ethical HRM, in which the “human” of Human Resources is brought back, where local value systems are explicitly, actively, and responsibly incorporated into our future

Copyright © Dr. Ruchi Saxena. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

automated HR. In the end, AI-induced fairness is nothing more than a hollow charade, unless it can humble itself to acknowledge its cultural shortcomings in its logic.

Keywords: *contextual legitimacy, Explainable AI, Green nudges, socio-cultural sensitivity, Trust Paradox*

1. Introduction: The Algorithmic Recruitment Revolution

The introduction of Artificial Intelligence (AI) in the domain of recruiting and selection is one of the most significant changes in the evolution of Human Resource Management (HRM) in the modern world. Companies around the world are moving away from manual and traditional hiring processes towards automated systems that offer unparalleled efficiency, speed, and unbiased decision-making. Firms hope to find the right ‘top talent' using predictive analytics, machine learning algorithms, and automated screening tools, while reducing the possible bias of intuitive human judgment. But there has been an ethical dilemma that's come to light because of the swift digitalization of the hiring process – the "Trust Paradox.

The paradox of the "neutral" algorithm and the "real" workforce stems from the basic incongruity between technology and society. This paradox of “neutral” algorithm and “real” workforce is directly related to the basic incongruity existing between technology and society. AI has been touted as a tool to reduce bias, but it is often referred to as a ‘black box' that reflects historical systemic inequalities, creating hired decisions that are opaque, unaccountable, and possibly discriminatory. This makes for an impasse in this relationship for candidates. People make an application for jobs with the assumption that the institutions are fair. If done in such a way that it involves opaque algorithms that discriminate against employment gaps, against non-traditional backgrounds and/or against unconventional educational pathways, the employer's perceived legitimacy is undermined.

This paradox is fuelled further by emerging markets, especially in the global south. Many AI models, which have been trained using data collected from Western professional norms, are sometimes "context-blind". They do not know how to identify the idiosyncrasies of other labour markets or the capacity for greater socioeconomic stresses, linguistic diversity, and patchy schooling experiences that are commonplace. Such widespread adoption of standardized methods in different cultural contexts can be broadly classified as "colonial algebra-utilization" because it denigrates local merit based on its lack of resonance with the "data sets" within the Global North.

Utilizing the following lens, this chapter challenges the Trust Paradox by claiming that insofar as it is a matter of technological optimization, fairness cannot be guaranteed. Ethical recruitment in a globalized society requires a profound understanding of socio-cultural complexities to emphasize the design of AI systems. Ethical recruitment in a globalized society calls for deep socio-cultural sensibility to call out the AI-driven system design. By adopting the Stimulus-Organism-Response (SOR) approach that has previously been used for understanding user trust of digital nudging, this chapter suggests that socio-cultural sensitivity is an important mediator construct. It changes how people feel about the output of algorithms from mistrust to trust. Through Reception-PT3's investigation into how automated recruitment technologies operate in different socio-cultural contexts in which they are implemented, this chapter offers attempts to move beyond the technological 'remedies' to create a culturally nuanced, human-centred, and ethical approach to recruitment practices. It will be followed by a discussion about the theory behind algorithmic trust, the specific blinks of the eye that have occurred with context-naive interventions, and a blueprint for socio-cultural intelligence in the future of global recruitment.

2. Literature Review: Algorithmic Recruitment & Fairness, and The Sociological Gap

Human Resource Management (HRM) has undergone dramatic transformation in the digital age, with HR tools becoming increasingly sophisticated, from tools like predictive models of candidate suitability, through automated screening of CVs, to tools that predict an employee's likely attrition rate, now tattooed within the organizational decision-making process. While the rise of automated recruiting is thought to be a response to the need to stop entrenched human cognitive biases, there is a growing body of research indicating that automated recruiting systems might perpetuate systemic inequality, yet masquerade as being mathematically neutral. This article explores the need for a more nuanced discussion on three central points: the defects of the black-box recruitment algorithm, the merits of Explainable AI (XAI) solutions on the ground, and the social and cultural obstacles to the use of globalized Western models in globalized labor markets.

2.1. The promise and peril of algorithmic recruitment

The major use cases of predictive analytics in HRM are used to help increase efficiency, save money, and increase the quality of hire. In contrast, recent studies have pointed out that ensemble models like Random Forest and XGBoost have emerged as the preferred prediction models of choice in recent years due to their predictive power and, in many cases, their accuracy. But there's another tension in the literature: the more sophisticated the models are, the more interpretable they are also. There's also a tension in the literature: the more predictive the model, the more opaque it will be.

If there are gaps in accountability, then “black-box” models are used in high-stakes, person-centred decisions – for example, in recruitment, promotion and termination. These risks have been addressed by the regulatory framework, which includes the European Union's Artificial Intelligence Act and New York City's Local Law 144 requiring enhanced transparency and human oversight of automated employment decision tools. However, as Raghavan suggests, in many cases the algorithmic vendor's attempts to claim the product is unbiased are not backed by any sort of validation, and it's up to the HR practitioner to "take the provider's word for it" to know that the algorithm is fair.

2.2. Beyond the black box: XAI as an operational imperative

Recently, the literature on Explainable AI (XAI) has highlighted the importance of going beyond black-box AI and achieving transparency in these predictions. Two widely used post-hoc explanation methods exist in current research in this area: Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME). LIME gives instance-level locally faithful explanations, whereas SHAP (in particular the Kernel SHAP variant) gives an explanation of the feature based on a theoretically sound perspective, and does so in a consistent way across the board.

One thing that stands out as a rich part of this discourse is the reconfiguration of the positioning of XAI. Early articles tended to portray XAI as a ‘forensic’ or ‘audit’ tool, following the decision to explain or check a decision. Recent empirical research, however, suggests that a shift is needed to an "operational tool" model in which XAI is already applied in the "active recruitment workflow". Indeed, if implemented into the pipeline of decision-making, these methods are able to go past the risk score and offer recruiters insights into the factors

of the candidate's evaluation, including work–life balance factors, tenure, and job satisfaction indicators. This transparency is essential for trust calibration and enables HR managers to make decisions with a human touch while respecting regulatory environments and ethical principles.

2.3. The socio-cultural disconnect in globalized markets

While methods based on computational technology such as XAI serve as a good baseline for the issue of fairness, it's not enough if the metrics used to qualify for the meritocratic scheme are culturally insensitive. An often-stated goal of the algorithmic bias literature has been to explain the threat of structural inequality being built into training data, but it has tended to make assumptions about the homogeneity of the data on which these models apply.

This is an especially significant research gap when thinking of the rollout of these tools in novel, different market settings. This disconnection has been explored in recent years in relation to the "Trust Paradox" in "behavioural green nudges". Trust is a prerequisite for the user to share data and rely on algorithmic recommendations in this context, but over-trust, or improperly calibrated trust of algorithms and automated systems, can put users at risk of distributive unfairness, data privacy losses, and algorithmic bias.

When standardized recruitment models are "exported" to the Global South without adaptation, they often encounter a lack of socio-cultural literacy. For instance, the presence of "gaps in employment" or the lack of traditional education or training can be valid indicators on the labour market in contexts where they were historically evaluated but might not be relevant in other areas, or because of linguistic diversity, language does not correlate with employment as it does in Western contexts, or because family-based socioeconomic structures are quite different. These cultural mismatches make these systems “algorithmic colonies” or mechanisms that erase local talent, neglecting or disregarding “local experiences” that are not normatively deemed as holding the same value as the “Western sense.”

2.4. Synthesis and research gap

The literature clearly indicates that technical transparency (in the form of XAI) is a necessary but not sufficient condition for ethical HR, leaving this topic open to continued discussion. The topic of technical transparency is still open to discussion, as the literature has clearly established that the approach of transparency (in the form of XAI) is an integral element of the concept of ethical HR, but only a part of it. Consistently, there is a "Micro-Trust" gap, which refers to the subjective trust that the user has in a particular tool, and a "Macro-Trust" gap, which focuses on the broader level of social acceptance and ethical justification for automated governance.

There is an absence of a unified body of literature that connects technical transparency to cultural sensitivity. The great majority of HR ethics models have stuck to two extremes: either they have been more concerned with the mathematical aspects of fairness, or with the philosophical approach of organizational ethics. To achieve this, this chapter aims to propose that socio-cultural sensitivity acts as a mediator between the two extremes of ethical and unethical recruitment with the help of AI in order to make AI-driven recruitment truly ethical. This work brings together the technical adroitness of SHAP/LIME and the sociological strength of trust calibration and proposes a new approach for studying HRM ethics by developing a framework of “culturally calibrated” technological systems that would honour the agency and dignity of the “pool of candidates in the world.”

3. Theoretical Framework: Socio-Cultural Sensitivity as a Mediator in The Trust Paradox

To understand this, it is important to look past a technology-adopted-as-it-would-be-absolutely-efficient mindset. The first step to that is to draw your attention beyond the input-output model of technology adoption. This chapter uses an adapted Stimulus-Organism-Response (SOR) model to explain how the algorithmic recruitment 'Stimuli' react with ‘Organisms' (the candidates), and why Socio-cultural sensitivity is the key mediator that determines their responses in terms of trust and engagement.

This figure visually maps how socio-cultural sensitivity acts as a mediator between the technical stimulus (AI) and the candidate’s response (trust).

(Visual Description for your illustrator or for captioning):

- Stimulus (Algorithmic Recruitment): A box containing "Black-Box Algorithms," "Standardized Merit Metrics," and "Opaque Data Processing."
- Mediator (Socio-Cultural Sensitivity): A central bridge labelled "Socio-Cultural Sensitivity," containing sub-layers of "Cultural Calibration" and "XAI Explanation."
- Organism (Candidate Internal State): A circular node labelled "Cognitive Appraisal," containing "Perceived Procedural Justice" and "Trust Calibration."
- Response (Outcomes): Two diverging arrows leading to "Institutional Legitimacy & Trust" (if mediation is successful) or "Systemic Alienation & Disengagement" (if mediation is absent).

Figure 1. The Stimulus-Organism-Response (SOR) Model Applied to AI-Driven Recruitment

3.1. The stimulus: algorithmic recruitment as a black box

Environment is a stimulus in the SOR model. The stimulus here is the AI-powered recruitment interface – it could be a platform, or the evaluation tool, or perhaps a computer-based, automated screening process. In today's reality, this often translates to a stimulus that has little transparency and applies “normative” standards. The system is developed by far-away developers who put in their own favoured descriptions of preferred, measurable features.

Such systems are "Black Box," imparting a feeling of depersonalization. If the interface requires candidates to interact with it in a certain, pre-established way by showing professional communication, it is telling each candidate a clear message: the system doesn't accept you; you do not possess a socio-cultural identity. This stimulus can violate their notion of procedure, since the rationale behind the "score" they earn is a mystery.

3.2. The organism: the candidate's internal state and trust calibration

The "Organism" is the psychological state in the individual that is occurring within him. Here is the source of the Trust Paradox. Organizational trust studies indicate that people evaluate levels of trust on two main factors: ability (can it do the work?); benevolence (does it have the intervener's interest in mind?).
An "Organism" has cognitive dissonance when the candidate communicates with an AI environment that does not respect him or her from a cultural point of view. But there may be some lack of benevolence – they may recognize "It is fast. It is high-tech" and think "Not so!" If a candidate fails to score a Western way of communicating using an algorithm based on Western communication, they go from being optimistic, expecting that they will be selected, to a feeling of systemic alienation if they are not. This creates the "Trust Paradox" situation: The more trusting the system is "neutral," the less the candidate trusts the system to treat his/her application fairly.

3.3. The mediator: socio-cultural sensitivity

A key argument presented in this chapter is that the term "Socio-Cultural Sensitivity" serves as the mediator in this SOR model. It filters the first stimulus - the AI - to see if the organism - the candidate - feels this is a threat or a "true evaluator".
In "Mediating Interpretation," incorporating socio-cultural sensitivity brings changes to an AI system's interpretation of input data. One such scenario is where it becomes aware that in some cultures, for example, it is more important to show humility than to express themselves straightforwardly and assertively, a characteristic which is often a major attribute of Western-trained AI. In doing this, the stimulus converts from one of exclusion to one of contextual recognition.
The higher the cultural awareness of an AI tool reflects a community's norms, the greater the perceived procedural justice is. If providing an explanation by an XAI system (e.g., "This score is based on how much you value situations in which you work together in a team in this position") touches a candidate's professional self-conception, the dissonance is handled. The candidate makes his or her own reality evident in a system that reflects this awareness.

3.4. The response: legitimacy and engagement

The behavioural and attitudinal result is the last element: the "Response". The reaction is one of scepticism, disengagement, and probably an insistence on not accessing the system again — or at least in the future — without having a socio-cultural sensitivity.
On the other hand, the mediator socially sensitive is present, and the answer is towards institutional validity. The candidate sees the firm as fair, since the firm's technology seems to "speak" to the wide variety of human capital the firm is trying to attract. This model helps to clarify how trust is a relational product, crossed by the level of "accuracy" achieved technically, but also by the capacity of the system to "see" the candidate in his cultural and professional context.

4. The Global Disconnect: Northern AI Models vs. Southern Contexts

The ethical challenges posed by AI adoption in recruitment processes are not all the same; they are mediated by, and often exacerbated by, government or regulatory laws, economics, or culture. A major failing in the way the algorithmic management of HR functions is rolled out today is that a model based on data from the Global North (mainly North America and Western Europe) can be simply exported and applied to the Global South (mainly the developing rest of the world) without any change.
The section makes the case that this 'off-the-shelf' application of hiring algorithms is an instance of some 'algorithmic colonialism' whereby the cultural and socioeconomic differences of emerging markets are transcoded to fit what seems like the 'normative' profile of Western success.
Table 1 visually maps how socio-cultural sensitivity acts as a mediator between the technical stimulus (AI) and the candidate's response (trust).

- Stimulus (Algorithmic Recruitment): A box containing "Black-Box Algorithms," "Standardized Merit Metrics," and "Opaque Data Processing."
- Mediator (Socio-Cultural Sensitivity): A central bridge labelled "Socio-Cultural Sensitivity," containing sub-layers of "Cultural Calibration" and "XAI Explanation."
- Organism (Candidate Internal State): A circular node labelled "Cognitive Appraisal," containing "Perceived Procedural Justice" and "Trust Calibration."
- Response (Outcomes): Two diverging arrows leading to "Institutional Legitimacy & Trust" (if mediation is successful) or "Systemic Alienation & Disengagement" (if mediation is absent).

Table 1: Comparative Analysis: Context-Blind vs. Culturally Calibrated Recruitment Models

Feature/Dimension	Context-Blind (Northern-Centric) Model	Culturally Calibrated Model
Merit Definition	Monolithic; prioritizes standardized career paths.	Pluralistic; accounts for regional labour market norms.
Employment Gaps	Flagged as "Risk" / "Lack of Stability."	Contextualized; identifies external economic volatility.
Data Grounding	Western-centric historical training datasets.	Locally grounded datasets incorporating regional nuances.
XAI Explanations	Generic, feature-based importance scores.	Narrative-based; resonant with local professional values.
Primary Goal	Mathematical neutrality/Standardization.	Procedural justice/Contextual legitimacy.

Note: This table illustrates the shift from standardized, often exclusionary algorithmic processes to responsive, ethical HR practices.

4.1. The fallacy of universal meritocracy

Algorithms for hiring are just as much an expression of value as they are used for recruiting. These models learn historical trends of success across professionals when they are trained on big data sets. But these patterns are very much rooted in the unique institutional story of the country of origin. For instance, in an economy that is rich in revenue, such as the U.S., and has a tradition of e-business adoption and practice, a high-fidelity indicator of success may be "consistent, full-time employment over ten years. This might seem like a reasonable approximation to the termini in one specific circumstance, but it is a discriminatory barrier in the case of emerging economy candidates, where underemployment, job volatility, and informal work prevail structurally and are not necessarily career gaps.

With these 'standardized' models, AI perpetuates one notion of merit. Candidates who have not followed the "Western style" career route are more likely to be "high-risk" or "low potential," for various reasons: not enough digital infrastructure, career path volatility due to the region's economy, and different cultural perspectives on trajectories of the career. This is no simple mistake, but rather the failure of a fair design of an ethical and sociocultural sensitivity required to carry out an equitable assessment. These black-box processes lack accountability, because they cannot be questioned as to 'why' they exclude a talented person, and as such, are automating what arguments I have developed in my previous research: the process of excluding talented members of society because they do not resemble the training dataset.

4.2. Algorithmic colonialism and the regulatory-equity gap

Algorithmic colonialism refers to the use of AI tools and technology in the Global South without making significant efforts to understand the social dynamics of the area, in effect, using the area to process data using foreign-built intelligence. This is complicated further by an enormous 'regulatory-equity gap'.

In the Global North, AI frameworks like the AI Act of the European Union or the measure of NYC's Local Law 144 require significant accountability measures, such as transparency, bias auditing, and human oversight. These rules force the organisations to justify their hiring rationale. But in many developing markets, these protections are in the early stages or not at all. If there is no requirement for bias auditing, multi-national companies could be employing higher risk/higher risk audited hiring tools in the Global South than in their home country. This division has enabled the evolution of two distinct methods of human capital management: one that is what they together recommend to strict ethical review, and one that works in the dark but may perpetuate historic inequalities under the guise of "technological advancement.

4.3. Moving toward "contextual compliance"

To overcome the Trust Paradox in today's globalized society, organizations need to move away from the standardizing premise and take the 'contextual compliance' approach. This philosophy also takes into account that AI-powered recruitment processes cannot be assessed for mathematical neutrality when mathematical neutrality is an elusive objective to become the ultimate reality, but for adherence to human rights and values embedded in the local socio-cultural landscape where they are used.

Contextual compliance involves making three changes:

1) Localized Data Grounding: The models should be trained and/or adapted with local data that reflect the realities of the local labour market. In an economy with cyclical industry fluctuation, a "high-risk" indication for an "employment gap" in a "stable" economy is not necessarily a "high-risk" indication in an "employment gap" in a "fluctuating economy".
2) Cultural Calibration of Explanations: As discussed in our theoretical framework, the explanations that XAI systems give should be culturally resonant. An explanation delivered in an aggressive or non-partisan manner within a certain culture will be rejected, as it will not lead to trust and hence to professional job selection.
3) Local Oversight Panels: Multinational companies should develop local advisory boards made up of HR professionals, sociologists, and community leaders that can assess the effect of algorithmic hiring tools. The panels can serve as the 'human-in-the-loop' needed to keep the AI a useful gadget and not a machine that blithely dispenses disciplinary decisions.

To sum up, the use of AI in global hiring is much more than just a technological process—it is a full-fledged socio-political way of proceeding. If these tools are not addressing socio-cultural sensitivity in the design and implementation phases, then we endanger the very inequalities that these technologies were supposed to resolve and will consequently exacerbate. Tackling and understanding the gap between 'north' and 'south' demands the most important responsibility in the era of globalized AI for every HR endeavour that bundles itself up in ethics.

5. Socio-Cultural Sensitivity: Defining the Mediator (Cultural Calibration & XAI)

The "Trust Paradox" (where an AI model gives a "Math verdict" and a "Soc explanation" is missing) is a problem of meaning, and the solution is changing our way of thinking about the "mediator" in our Stimulus-Organism-Response (SOR) model by studying the "improved" mapping issue.

The "Trust Paradox" (where an AI model gives a "Math verdict" but a "Soc explanation" is missing) is a problem of meaning, and the solution is to look at the "improved" mapping issue in our Stimulus-Organism-Response (SOR) model. Being 'socio-cultural sensitive' is not an optional extra with algorithmic recruiting; it is the necessary add-on to the algorithmic output that transforms it into action and makes it politically acceptable in our globalized job markets. In this part, the authors contend that Explainable AI (XAI) frameworks offer the technical "language" for this mediation, which can be implemented through Kernel SHAP and LIME, but needs to be "culturally calibrated" to be effective.

This table demonstrates how an HR manager translates technical XAI output (SHAP/LIME) using socio-cultural sensitivity.

Table 2: HR Decision-Support Matrix: Interpreting Algorithmic Logic Through Socio-Cultural Sensitivity

Algorithmic Signal (XAI Attribute)	Interpretation: Western/Stable Context	Interpretation: Emerging/Volatile Context	Strategic HR Action
Career Velocity	High velocity indicates high potential/ambition.	High velocity may indicate unsustainable job-hopping due to market instability.	Adjust weighting; prioritize qualitative achievements over quantitative metrics.
Institutional Prestige	Prestige of elite universities used as a proxy for capability.	Prestige may reflect privilege; focus on non-traditional learning/skill acquisition.	Shift evaluation from credentialism to competency-based screening.
Tenure/Stability	Long tenure signals commitment and loyalty.	Long tenure may reflect lack of exposure to diverse professional environments.	Balance tenure with breadth of experience and adaptive capacity.
Communication Style	Direct, assertive communication preferred.	High-context communication (humility/indirectness) may be valued.	Re-calibrate "Soft Skill" scoring to honor diverse interaction norms.

Note: This matrix serves as a policy tool for HR managers using Explainable AI in diverse, globalized labor markets.

5.1. The technical foundation: SHAP and LIME as interpretable bridges

We need to first normalize to a new stance: the idea that post-hoc interpretability tools are not just compliance auditing tools but tools for the calibration of trust. In my previous work, I showed that using Random Forest attrition classifiers, one can often get a false signal from native feature importance (Gini/MDI), enough to weight some high-cardinality features too much (age, tenure, etc.).
In contrast, Kernel SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) give a much more accurate characterisation of the "rules" being used by the model. For example, when analysing employees who presented for the first time for a concern, these techniques invariably highlighted three factors, namely 'overtime', 'job satisfaction' and 'monthly income' as the stable and salient factors of concern. Both SHAP and LIME converge independently of one another to the same factors, and they both employ entirely different mathematical machinery: one based on game-theoretic coalition averaging, the other on local surrogate regression, which provides a "corroborating baseline" for HR managers.
The raw SHAP and LIME outputs, on the other hand, are a mathematical stimulus (either a bar chart or a bee swarm chart). This "socio-cultural mediator" is the HR manager's interpretation of this stimulus. A high SHAP value for "tenure" is not necessarily indicative of a meaningful relationship. In an up-and-down employment environment, it could be seen as an alternative to loyalty. With a volatile structural reality in emerging markets, long tenure might be a sign of small-mindedness or an inability to deal with a changing space. Adopting the socio-cultural sensitivity approach to LIME interpretation is thus the site of its technical output.

5.2. Cultural calibration: moving from transparency to resonance

Here, a process of what I call the "cultural calibration" of algorithmic explanations refers to a conscious "realignment" process in ways that ensure that their predictions reflect actual socioeconomic dynamics and professional values on the local labour market. To make this happen, HR must transcend the "why" of technology and get into the "why" in culture.
Think about a recruitment system that filters resumes using an artificial intelligence system based on "career progression velocity. One of the problems with a context-blind model is that it could use a uniform velocity metric for the entire world. The HR manager could do this with a culturally calibrated model, which would take into account the socio-cultural sensitivity. Because early career professional learning is sometimes slowed by the bureaucracy found in institutions, it is important to make the explanation explicit that this is because of some external limitation on the candidate rather than some internal shortage of candidates.
Furthermore, the system can use feature attribution (SHAP) to offer a prompt to the manager: "Career velocity" is a feature that is heavily weighted for career growth jobs; making sense of this within the local hiring context of [Region X]? This relationship elevates the AI beyond operational restrictions to become a strategic decision support tool. It does not assume that the data is actual but rather a sign to be interpreted as inherent facts and phenomena in the local reality.

5.3. Operationalizing sensitivity: the HR manager as translator

To apply these insights to practice, there is a need to structure the role of HR manager as a "translator" of algorithmic logic in organizations. This requires a modification in how XAI is applied in the HR pipelines:
XAI tools need to be embedded into the recruitment dashboard, rather than as a separate forensic document which can only be accessed once after a recruitment. The SHAP values should be available at decision time so that the recruiter can communicate the rationale of the model.
The technical output (SHAP values) should be automatically translated into bilingual narratives based on the company's cultural values. Instead of showing "YearsAtCompany = +0.03", the application should show "The model attributes the candidate's job stability indicator aligned with our preference for job length in our organization."
Human in the Loop Feedback: When a recruiter enters a recommendation in place of the algorithm, because they are more familiar with the socio-cultural aspects of the decision, the system should record this "override" as a data point. With time, the model can learn to refine its global importance weightings, taking into account these proven local inputs, thus establishing a self-correcting loop which allows for both technical inputs and human wisdom.

5.4. Mitigating the trust paradox

This is an approach to the Trust Paradox per se. The issue with the paradox is that candidates and managers feel like they are being faced with a logic they are not familiar with and not one that can or should change. With culturally calibrated XAI comes "procedural justice." The fact that a candidate is rejected, but the system has been able to explain and convey that the candidate was rejected because of this, in a way that allows the candidate respect for the professional context, and also the HR manager, through an interpretable framework, to verify that logic makes sense, changes the interaction. Makes the hiring process more of a "conversation" than a "black-box calculation.
In the end, fairness mediated by socio-cultural sensitivity suggests that "moral dimension of technology depends on its context." HR managers can make sure the "human" side of Human Resource Management is in the lead when it comes to the digital future with Kernel SHAP and LIME, which can unveil the inner workings of the models, and with the use of socio-cultural sensitivity, which would interpret the structure and expose the workings of the model.

6. Policy and Practical Agenda: Operationalizing Ethical AI in HRM; Ethical AI in HRM will be Operationalized at The Level of a Policy and a Practical Data-Based Agenda

More sensitive, interpretable computer-based decision support tools for selection necessitate dedicated HR policy restructuring that goes beyond the "neutral" automated recruitment. AI procurement has become more than just a technical transaction—it's a socio-technical intervention. In order to overcome the limitations of the Trust Paradox and reduce the dangers of "algorithmic colonialism," practitioners should follow three pathways of action, namely, governance, cultural calibration, and ethical auditing.

6.1. Transitioning to human-centered governance

The number one policy call is to decentralise AI from the role of the decision-maker. It is important that a "Human-in-the-Loop" (HITL) protocol becomes part of the organizational structure if it is to be adopted instead of being merely performative.
There should be a clarification of AI tools in organizational policy from being "Gatekeepers" to being "Decision Support Systems" (DSS), not “Automated Decision Systems” (ADS). This classification results in the adoption of different accountability rules and regulations. Should an AI tool flag a candidate and classify them as "high risk" or "unsuitable", the HR manager will have to be the one providing evidence that they are not, based on the AI's XAI explanations.
Structured Human Discretion: HR departments need to establish "Discretionary Protocols" that ask HR recruiters not only to give a reason for trumping the algorithmic suggestion, but actually do so. This is a loop which doesn't punish the algorithm, because it was "wrong," but is recording the human context that has passed, which is also valuable data to aid future cultural calibration.

6.2. Protocol for cultural calibration

To prevent a "West v. rest" mentality that would risk applying Western standards to the marketplace of the global corporate world, organizations need to adopt a "Protocol for Cultural Calibration. This protocol moves from a global standardisation to a regional contextualisation.
Localized Feature Weighting: Businesses that are using global recruitment solutions should explore the validity of recruiter model features in the geographic region they are targeting as compared to the source region. If the validity conflicts, then the policy should determine that an automatic weight change should be made on the features.
Every explanation from XAI frameworks such as SHAP, LIME, etc., should go through a "Cultural Translator" layer in the model. This layer helps ensure that the concepts (fit, potential, stability) in the description of the language are within the professional vernacular currently used in the local area. For example, in a market such as communal collaboration among people and others, soft skills players account for relationships and brainstorming, while explaining the importance of career velocity-players focus on self-performance.

6.3. Formalizing institutional ethics oversight

Traditional HR boards are not well-versed in running such a specialized operation as algorithmic recruitment.
Interdisciplinary Ethics Boards: HR and IT personnel should be joined in the committees by experts in related fields such as sociology, labour economics, and, locally, in the legal aspects of the team. The responsibility of these boards should be to review the "Trust Calibration" semi-annually to ensure that candidates feel that this process is inclusive and equitable and that the results reflect the firm's diversity and inclusion (D&I) objectives.
Candidate Recourse Mechanisms: Policy should include a clear and easily accessible path for candidates to challenge algorithmic decisions. As a certainty, this isn't simply a regulatory need (as in the EU AI Act); it's an ethical requirement for trust. Wherever there is a possibility of a candidate believing an algorithm has approached their career path incorrectly, the organization must ensure there is a review process where they get some human input that is transparent regarding what initiatives have been taken.

6.4. Ethical auditing beyond mathematical neutrality

HR audit procedures are too narrowly focused on “bias detection,” typically measured as the difference in the selection rate across demographic groups by doing a simple mathematical calculation. This is necessary but not enough.
Contextual Auditing: There is a need for organizations to shift towards ‘Contextual Audits’ of the systemic impacts of the model. It means the question: Is the model systematically excluding nontraditional career options that are widely used in this particular economic context? Does the model promote a "normative" job role which pushes out qualified local talent?
Multinational companies should have a zero-tolerance policy of publishing an annual transparency report on the use of AI in recruiting. Such reports should not rely on the subtlety of "proprietary technology" to avoid having to address clear questions about the variables considered, the method of interpretation of the model logic (e.g., SHAP/ LIME), and what steps have been taken in making sure that regional cultural resonances have been properly achieved.

6.5. A summary of the policy agenda

Integrating AI into global recruitment is a dynamic negotiation of technology's efficiencies and respect for human dignity. These policies represent a paradigm shift from algorithmic imposition to algorithmic partnership by collaborating with people within the algorithms, calibrating algorithms with culture, and auditing dialogically in context. It is not an AI uproar; it's just 'control with humility and socio-cultural intelligence', showing respect to the diverse workforce it exists for. There is a solution to the Trust Paradox—but it requires the organization to seek to make fairness something that can be made and fine-tuned to the human ingredients of each market in which it does business.

7. Conclusion: Toward an Ethical AI Future in HRM

AI has become a defining moment in the world of HR. The proposition that AI can improve efficiency and minimise human error is indisputable, but as the preceding chapter has described, AI can be deployed without paying proper attention to the socio-cultural context of the global workforce, thus creating a "Trust Paradox. That's a paradox, however—the neutral nature of the technology doesn't equal the fairness needed in a healthy organization when it comes to hiring practices: Algorithmic opacity also undermines the legitimacy needed to develop healthy hiring practices.
By examining the importance of socio-cultural sensitivity as the required connector between stimulus, organism, and response, we have shown how important socio-cultural sensitivity is for the trust-builder. Using algorithmic "culturally calibrated decision support" instead of defining it as "algorithm as gatekeeper" can help to resolve the tension between standardization and human diversity in organizations. Explainable AI (XAI) solutions like Kernel SHAP and LIME can make this transition possible by offering the technical infrastructure needed to operationalize these algorithms. These tools take the conversation beyond compliance and enable HR managers to ask culturally

positive questions of the data, rather than the background one that could spot a compliance problem, giving them more actionable, more interpretable, and more information-rich insights to benefit from.

Finally, the ideal of an ethical HRM is not an automated process but smart management of the interplay between human and artificial intelligence. To dismantle "algorithm colonialism," hiring models and processes of the North must learn to adapt to hiring models and processes of the South in a proactive manner, and local human oversight needs to be empowered. The direction ahead requires that fairness be not a static mathematical parameter but a dynamic value that must be creatively shaped with socio-cultural intelligence. This complexity can be harnessed to create recruitment ecosystems that are as fair and respectful as they are effective, thereby ensuring that the age of AI has a positive impact on the candidates available to organizations around the world.

References

[1] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint*.

[2] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*.

[3] Molnar, C. (2020). *Interpretable machine learning: A guide for making black box models explainable*.

[4] Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. https://doi.org/10.1145/3351095.3372828.

[5] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/2939672.2939778.

[6] Rudin, C. (2019). Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead. *Nature Machine Intelligence*. https://doi.org/10.1038/s42256-019-0048-x.

[7] Saxena, R. (2026). The trust paradox in green nudges in the Indian context: Mediating the efficacy of artificial intelligence-driven interventions with socio-cultural sensitivity. *Asian Journal of Advances in Research*. https://doi.org/10.56557/ajoair/2026/v9i1563.

[8] Saxena, R. (2026). Explainable AI for human resource decision-making: A SHAP- and LIME-based framework for interpretable employee attrition prediction. [Unpublished manuscript/Internal paper].

[9] Strumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*. https://doi.org/10.1007/s10115-013-0679-x.

[10] Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.