

Employee Perceptions of AI Ethics and Moral Accountability in Saudi Arabian Organisations: A Qualitative Study

Dr. Qasim Zureigat*

Cultural Assets Group, Saudi Arabia

E-mail: qasim@culturalassets.sa

*Corresponding Author

Received: August 2025; **Accepted:** January 2026

Abstract: The rapid deployment of artificial intelligence in organisational decision-making has generated urgent questions about how employees experience, interpret, and morally evaluate AI systems that shape their working lives – from performance monitoring and task allocation to promotion decisions and disciplinary processes. Yet the employee perspective on AI ethics remains substantially underexplored, particularly in non-Western institutional contexts where distinct cultural, religious, and organisational norms may produce different moral frameworks for evaluating algorithmic authority. This paper reports findings from an interpretive qualitative study based on 38 in-depth semi-structured interviews with employees, middle managers, and HR professionals across nine Saudi Arabian organisations that have deployed AI systems in workforce management or operational decision-making. Using reflexive thematic analysis following Braun and Clarke's (2022) methodology, we identify five master themes that structure employee moral experience of AI in the Saudi workplace: (1) Algorithmic opacity and the collapse of moral dialogue; (2) Displaced accountability and the erosion of relational trust; (3) Religious and cultural frames for evaluating algorithmic authority; (4) Collective versus individual moral responses to AI-driven harm; and (5) Resignation, resistance, and the ethics of algorithmic coping. The study makes three theoretical contributions: it develops the concept of relational accountability as a culturally grounded alternative to procedural accountability in AI ethics; it shows how Islamic ethical frameworks, particularly concepts of 'adl (justice), amanah (trustworthiness), and shura (consultation), provide Saudi employees with a distinctive moral vocabulary for evaluating AI systems; and it argues that the dominant discourse of individual data rights is poorly calibrated to the collective moral sensibilities that characterise Saudi Arabian and broader GCC organisational cultures.

Keywords: AI Ethics, Employee Attitudes, Qualitative Research, Saudi Arabia, Algorithmic Accountability, Thematic Analysis

Type: Research Paper



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

DOI: 10.51325/d4w0yq25

1. Introduction

The business ethics literature on AI in organisations has made significant advances in theorising the moral challenges of algorithmic decision-making

from the perspectives of designers, deployers, and regulators (Mittelstadt et al., 2016; Diakopoulos, 2016; Raji et al., 2020). It has developed frameworks for fairness, transparency, explainability, and accountability that have shaped both academic debate and emerging regulatory standards such as the EU Artificial Intelligence Act, and Saudi Arabia's own national frameworks — the SDAIA AI Ethics Principles (Saudi Data and AI Authority, 2023) and the SDAIA AI Adoption Framework (Saudi Data and AI Authority, 2024). What remains substantially underexplored, however, is the employee perspective: how do the people most directly affected by AI systems in the workplace experience their moral dimensions? What ethical frameworks do they draw on to evaluate AI-driven decisions? How do they respond — emotionally, relationally, practically — when AI systems produce outcomes they regard as unjust? And how do the distinctive cultural, religious, and organisational norms of specific institutional contexts shape these experiences and evaluations?

Saudi Arabia presents a compelling context for exploring these questions. A country of remarkable educational sophistication and significant technological adoption — the Saudi banking sector, for example, has been a regional pioneer in deploying AI in credit scoring, fraud detection, and customer service under Vision 2030 — Saudi Arabia also embodies the distinctive relational, communal, and religiously inflected moral culture of the Arabian Peninsula in ways that create a particularly rich environment for examining how employees experience the collision between algorithmic authority and human moral expectation. The Kingdom's rapid and state-driven AI agenda — accelerated by Vision 2030's targets and the designation of 2026 as the Year of Artificial Intelligence — has intensified AI adoption across sectors, making the ethical dimensions of that adoption a matter of immediate practical urgency.

This paper reports findings from a qualitative study based on 38 in-depth semi-structured interviews conducted with employees, middle managers, and HR professionals across nine Saudi Arabian organisations that have deployed AI systems in workforce management or operational decision-making. Using reflexive thematic analysis (Braun and Clarke, 2022), we develop five master themes that structure employee moral experience of AI in the Saudi workplace, and we draw out their implications for AI ethics theory and practice in Saudi Arabian and other non-Western organisational contexts.

2. Theoretical Background

2.1. Employee Perspectives on AI Ethics: An Emerging Literature

The literature on employee perspectives on AI in the workplace has grown rapidly since approximately 2018, driven by the spread of algorithmic management systems — AI tools that monitor, evaluate, allocate, and sometimes discipline workers — across sectors from logistics and manufacturing to financial services and professional services (Lee et al., 2015; Möhlmann and Zalmanson, 2017; Kellogg et al., 2020). Research on gig economy platforms — Uber, Deliveroo, Amazon Mechanical Turk — has documented how algorithmic management erodes workers' sense of agency, fairness, and relational dignity (Gandini, 2019; Wood et al., 2019). Studies in more traditional employment contexts have shown how AI-enabled performance monitoring generates

anxiety, surveillance consciousness, and what Ajunwa et al. (2017) term the 'infinite reach' of the quantified workplace.

The specifically ethical dimensions of employee experience — how workers morally evaluate AI systems, what frameworks they use to assess algorithmic fairness and accountability, and how their moral responses are shaped by cultural and institutional context — remain less systematically explored. Valuable contributions include Binns et al.'s (2018) experimental work on lay fairness perceptions of algorithmic decision-making, which shows that individuals prefer explanations grounded in individual features over statistical group comparisons; Araujo et al.'s (2020) cross-national survey on public attitudes to AI decision-making, which finds significant variation in acceptance across national contexts; and Glikson and Woolley's (2020) review of human trust in AI systems, which highlights the role of perceived competence, integrity, and benevolence in shaping AI acceptance; and Majrashi's (2025) Saudi-based study of employee perceptions of AI-driven performance prediction, which found that concerns over opacity and lack of contestability were central to perceived unfairness. But this literature has been dominated by Western samples and Western ethical frameworks, leaving the moral experiences of workers in non-Western contexts substantially undertheorised.

2.2. Islamic Ethics and Algorithmic Authority

Islamic ethical frameworks — derived from the Quran, the Hadith, and the tradition of Islamic jurisprudence (fiqh) — provide a distinctive moral vocabulary for evaluating the conduct of organisations and the treatment of individuals within them that has been underutilised in the AI ethics literature. Several concepts are particularly relevant to the evaluation of algorithmic decision-making. 'Adl (justice or equity) requires that decisions affecting individuals be made fairly, with due weight given to their particular circumstances and with avoidance of arbitrary or disproportionate treatment — a requirement that algorithmic systems optimising on population-level statistical patterns may systematically violate. Amanah (trustworthiness or stewardship) requires that those entrusted with authority over others exercise it with honesty, transparency, and fidelity to the purposes for which that authority was granted — requirements that are arguably violated when organisations delegate consequential decisions to systems whose operation neither the decision-maker nor the affected individual can explain or scrutinise. Shura (consultation or deliberation) is the principle that consequential decisions should be made through genuine consultation with those affected — a principle that automated decision-making, by its nature, tends to bypass.

These concepts have been discussed in the Islamic finance and CSR literatures (Beekun and Badawi, 2005; Dusuki and Abdullah, 2007) but have received little attention in the AI ethics context (Elmahjub, 2023; Ghaly, 2024), despite their particular salience in Saudi Arabia, where Islamic ethical principles are embedded in law, public policy, and corporate governance frameworks. Their integration into our analysis enriches the study's theoretical contribution by providing a culturally grounded framework for interpreting Saudi employees' moral evaluations of AI systems — one that supplements the procedural, rights-based frameworks that dominate Western AI ethics discourse with a relational, communal, and religiously inflected moral perspective. In Saudi Arabia, this

supplementation is not merely theoretically desirable but practically necessary: SDAIA's own AI Ethics Principles (2023) explicitly reference concepts of fairness, trustworthiness, and accountability that map closely onto Islamic ethical vocabulary, yet the operationalisation of these principles in employee-facing HR governance has received little empirical scrutiny.

2.3. Relational Accountability in Organisational Contexts

The concept of accountability that structures most AI ethics frameworks is primarily procedural in character: accountability is satisfied when decision-making processes are transparent, when affected individuals have access to explanations, and when formal redress mechanisms are available. This procedural account reflects the liberal individualist tradition that underpins most Western regulatory frameworks — a tradition that prizes formal rights, impersonal process, and individual remedy. It is, however, poorly calibrated to the relational, communal, and trust-based account of accountability that characterises many non-Western cultural settings, including the Saudi Arabian and broader GCC context.

We develop the concept of relational accountability as a culturally grounded alternative: accountability understood not as the satisfaction of procedural requirements but as the maintenance of trust-based human relationships through which obligations, expectations, and remedies are mutually negotiated. Relational accountability is inherently dialogical — it requires face-to-face engagement, mutual recognition, and the possibility of moral repair through relationship — and it is fundamentally collective rather than individual in its reference points. When AI systems disrupt the human relationships through which relational accountability is exercised, the harm is not merely procedural unfairness but the disruption of the social fabric through which moral community is sustained. This concept, we argue, provides a more adequate account of what Saudi employees are expressing when they describe their moral distress at AI-driven decision-making than the procedural accountability frameworks that dominate the literature.

3. Research Methodology

3.1. Research Design and Epistemological Positioning

This study adopts an interpretive qualitative research design grounded in a constructivist epistemology. We hold that employees' moral experiences of AI in the workplace are not fixed facts to be discovered but are actively constructed through the frameworks — cultural, religious, relational, professional — that participants bring to their encounters with algorithmic systems, and that the task of qualitative research is to generate rich, contextually sensitive interpretations of those constructions rather than to test pre-specified hypotheses. This epistemological positioning is appropriate for the study's aims — to explore how employees in a distinctive cultural context experience and evaluate AI ethics — which call for the interpretive depth that qualitative methods provide rather than the breadth of coverage that quantitative approaches offer.

3.2. Participant Recruitment and Sample

Participants were recruited through purposive and snowball sampling across nine Saudi Arabian organisations in five sectors: banking and financial services (three organisations), telecommunications (two organisations), retail and e-commerce (two organisations), healthcare (one organisation), and logistics (one organisation). Organisations were selected to maximise variation in sector, size, ownership structure (family-owned versus institutionally owned), and type of AI deployment, while sharing the common criterion of having deployed AI systems that materially affect employee experience – including performance monitoring, task scheduling, recruitment screening, and customer service automation. Access was negotiated through professional networks and with the cooperation of HR departments; all participants provided informed written consent and were assured of anonymity.

A total of 38 participants were interviewed (Table 1): 22 frontline employees, 10 middle managers with responsibility for AI-affected teams, and 6 HR professionals involved in AI system deployment or governance. Table 1 presents participant characteristics. The sample was predominantly Saudi nationals (33 of 38), with 5 participants holding dual Saudi-Western nationality or being long-term expatriate residents. Ages ranged from 23 to 54 years (mean = 34.2). Twenty-one participants identified as Muslim, 15 as Christian, and 2 declined to specify religious affiliation.

Table 1: Participant Profile

ID	Role	Sector	Gender	Age	Religion	AI System Encountered
P01	Frontline Employee	Banking	F	29	Muslim	Credit scoring/performance monitoring
P02	Middle Manager	Telecom	M	41	Christian	Team KPI tracking; AI scheduling
P03	Frontline Employee	Retail/E-comm	F	26	Muslim	Algorithmic task allocation
P04	HR Professional	Banking	F	38	Christian	AI recruitment screening
P05	Frontline Employee	Logistics	M	31	Muslim	Route optimisation; delivery monitoring
P06	Middle Manager	Healthcare	M	47	Christian	AI diagnostic support; staffing
P07	Frontline Employee	Banking	F	33	Muslim	Performance scoring; loan decisions
P08	HR Professional	Telecom	M	44	Muslim	AI recruitment; attrition prediction
...	...	...	...	...	...	...
P38	Frontline Employee	Retail/E-comm	M	27	Muslim	Product recommendation AI; surveillance

Note. Full participant table available from the corresponding author. F = female; M = male. Participant IDs are pseudonymised.

3.3. Data Collection

Each participant took part in a single in-depth semi-structured interview lasting between 55 and 110 minutes (mean = 78 minutes). Interviews were conducted in the participant's preferred language – Arabic (26 interviews) or

English (12 interviews) — by the lead researcher, a native Arabic speaker with extensive fieldwork experience in Saudi Arabian organisational contexts. The interview guide was developed to explore: participants' awareness and understanding of AI systems in their workplace; their experiences of AI-driven decisions affecting them or their teams; their moral evaluations of those decisions and the AI systems producing them; the frameworks — cultural, religious, professional — they drew upon in making those evaluations; and their relational and practical responses to AI-driven decisions they regarded as unfair or harmful. The guide was piloted with three participants not included in the final sample and refined in light of the pilot. All Arabic interviews were audio-recorded and transcribed verbatim, then translated into English by the lead researcher; translations were reviewed for accuracy by a bilingual research assistant.

3.4. Analytical Approach

Data were analysed using reflexive thematic analysis following the six-phase process described by Braun and Clarke (2022): familiarisation with the data, generating initial codes, constructing themes, reviewing themes, defining and naming themes, and producing the written analysis. The analysis was explicitly reflexive in character: the researchers maintained analytical journals throughout the process, documenting interpretive decisions, theoretical engagements, and personal reflexive observations about how their own positionality — as researchers familiar with but external to the Saudi Arabian organisational context — shaped the analytical process. Coding was conducted using NVivo 14, with an initial codebook developed by the lead researcher and reviewed by all co-authors. Thematic construction involved multiple rounds of collaborative team analysis, with particular attention to preserving the emic — participant-generated — dimensions of moral vocabulary alongside etic — theoretically informed — interpretive frames. Analytic rigour was supported through member-checking (six participants reviewed their interview summaries), inter-rater reliability assessment on a 15% subsample (Cohen's kappa = 0.81), and an external audit of the thematic framework by a qualitative methods specialist not involved in the study.

4. Findings: Five Master Themes

Analysis of the 38 interview transcripts generated an initial set of 247 codes, which were iteratively consolidated into 18 sub-themes organised under five master themes. The master themes, their constituent sub-themes, and the distribution of references across the dataset are summarised in Table 2.

Table 2: Thematic Framework- Master Themes and Sub-themes

#	Master Theme	Sub-themes	Participants (n)	References (n)
1	Algorithmic opacity and the collapse of moral dialogue	Invisible decision logic; loss of explainability; absence of moral interlocutor	34	118
2	Displaced accountability and the erosion of relational trust	Responsibility vacuum; managerial distancing; institutional betrayal	31	104

3	Religious and cultural frames for evaluating algorithmic authority	Islamic ethics as an AI critique; communal justice norms; sectarian mistrust	29	87
4	Collective versus individual moral responses to AI-driven harm	Collective grievance; informal solidarity; failure of individual redress	26	73
5	Resignation, resistance, and the ethics of algorithmic coping	Moral withdrawal; gaming the algorithm; quiet ethical refusal	33	96

4.1. Theme 1: Algorithmic Opacity and the Collapse of Moral Dialogue

The most pervasive theme across the dataset concerned participants' experience of algorithmic opacity — the inability to access, understand, or meaningfully contest the decision logic of AI systems — and its disruption of what many described as the essential moral dimension of workplace relationships: the possibility of dialogue. Across sectors and roles, participants consistently articulated the harm of AI-driven decisions not primarily in terms of outcome unfairness but in terms of the foreclosure of the moral conversation through which unfairness could be named, examined, and potentially repaired.

"When my manager gave me a bad performance review before, I could argue with him. I could bring my numbers, my client feedback, and my context. We could disagree. Now the system gives me a score, and there is nothing to argue with. The number just sits there. It's like being judged by something that cannot hear you." (P07, Female Bank Employee, Riyadh)

This experience of dialogic foreclosure was not simply a complaint about opacity in the technical sense — participants did not primarily frame their distress as a lack of access to algorithmic explanations, as the procedural AI ethics literature might anticipate. Rather, they described a more fundamental moral loss: the disruption of the relational accountability structures through which workplace fairness had previously been negotiated. The absence of a human moral interlocutor — someone who could be appealed to, who could be moved by argument, who carried the relational obligations of a colleague or superior — was experienced as a form of moral abandonment.

"The system monitors everything — our routes, our times, our stops. My boss used to know that on Thursdays, I have traffic near the King Fahd Road junction. He knew my area. The system does not know. The system does not care. But also, my boss now just says: The system showed this. He hides behind the system. And I have no one to talk to anymore." (P14, Male Logistics Supervisor, Jeddah)

The theme was particularly acute for participants in the banking sector, where AI systems were deployed in credit scoring and loan decision-making with direct consequences for customers and, in some cases, for employees' commission structures and performance ratings. Several banking sector participants described the experience of being required to communicate AI-driven credit refusals to customers — to be the human face of a decision they could not explain and had not made — as a form of moral complicity in a process they found ethically troubling.

"I have to tell the customer: I am sorry, the system has declined your application. And they ask me why. And I say: I don't know. And they look at

me like I am lying. Because how can I not know? I am the bank. But I don't know. And I feel — I feel dishonest, even though I have said the truth." (P01, Female Bank Employee, Riyadh)

4.2. Theme 2: Displaced Accountability and the Erosion of Relational Trust

Closely related to the opacity theme but analytically distinct, the second theme concerned the displacement of moral accountability from human actors to AI systems — and the erosion of relational trust that this displacement produced. Participants across all sectors described a pattern they experienced as a form of managerial abdication: the use of AI systems to make or legitimise decisions that managers were unwilling to own, and the consequent collapse of the trust relationships through which organisational authority had previously been exercised with moral legitimacy.

"I was given this system to manage my team's schedules. And yes, it is efficient. But now my team — they don't come to me when they have a problem with their schedule. They know I will say: it's the system. They have learned that I am not really in charge anymore. And something has broken between us. I feel it." (P02, Male Telecom Manager, Riyadh)

The concept of relational accountability proved generative for interpreting this theme. Participants were not primarily invoking procedural accountability norms — they were not demanding formal redress mechanisms or algorithmic transparency reports. They were describing the disruption of the relational fabric through which accountability had previously been personal, dialogic, and morally binding. When accountability is mediated by an AI system, it becomes impersonal, monological, and morally elusive — and the relationships through which it operates lose their ethical substance.

Several HR professionals articulated the institutional dimension of this displacement with particular clarity, describing how the deployment of AI in HR processes had fundamentally altered the moral character of the HR function — from a people-facing role grounded in relational judgement to a systems-administration role grounded in process compliance.

"When I joined HR twelve years ago, my work was about people. Understanding their situations. Making judgements that required empathy, experience, and wisdom. Now a large part of my work is checking that the algorithm's outputs are within legal parameters. I have become a compliance officer for a system I did not build and do not fully understand. The human dimension of my work has shrunk." (P04, Female HR Manager, Banking, Riyadh)

The erosion of relational trust was particularly sharp in organisations where AI deployment had been rapid, opaque to employees, and unaccompanied by genuine consultation. In contrast, participants in one organisation — a healthcare provider that had introduced an AI scheduling system through a structured consultation process involving staff representatives — described a notably different experience: one in which the AI system was regarded as a shared tool subject to collective governance rather than an external authority imposing decisions from above.

"We talked about the system before it came. The nurses, the doctors, the administrators — we all had input. We still argue about it. But we argue

together. The system is ours, in some sense. That makes it different." (P06, Male Healthcare Manager, Riyadh)

4.3. Theme 3: Religious and Cultural Frames for Evaluating Algorithmic Authority

A distinctive contribution of this study is the rich evidence of participants drawing on Islamic ethical frameworks — and, to a lesser extent, on Christian communal ethics — to articulate their moral evaluations of AI systems in ways that the dominant Western AI ethics vocabulary does not capture. This theme was most prominent among Muslim participants but was not exclusive to them; Christian participants drew on related concepts of stewardship, dignity, and communal responsibility in ways that resonated with the Islamic frameworks articulated by their Muslim colleagues.

The concept of 'adl (justice) was invoked, often implicitly, by multiple participants in their evaluations of algorithmic fairness. What was striking was that participants' justice concerns were not primarily about procedural fairness in the sense of equal treatment — the dominant framework in Western algorithmic fairness research — but about substantive equity: the capacity of the system to account for the particular circumstances, needs, and history of the individual being judged.

"In Islam, we say Allah is 'Adil — perfectly just. And justice means knowing each person's situation. A just judge does not treat everyone the same, regardless of their circumstances. The algorithm treats everyone the same. That is not justice. That is just... uniformity." (P07, Female Bank Employee, Riyadh)

The concept of amanah (trustworthiness/stewardship) was invoked by several participants to articulate what they regarded as the moral obligation of organisations deploying AI — an obligation they felt was being violated. Amanah, in Islamic ethical tradition, refers to the trust placed in a person or institution to act with fidelity to the purposes for which that trust was granted. Participants who drew on this concept argued that organisations deploying AI in consequential decisions had a stewardship responsibility toward affected employees that required honesty, transparency, and genuine accountability — responsibilities that they experienced as being systematically discharged.

"The company has a trust from its employees. We give our working lives to the company. In return, the company owes us honesty. If you use a system to make decisions about us, tell us. Explain it. That is the minimum of amanah. To hide behind the algorithm — that is a betrayal of the trust we placed in you." (P08, Male HR Professional, Telecom, Jeddah).

The principle of shura — consultation and deliberation — was invoked by several participants as a standard against which the absence of employee voice in AI deployment decisions was morally evaluated. Participants across religious backgrounds articulated similar intuitions about the legitimacy deficit of AI systems deployed without genuine consultation with those they would affect.

"Big decisions should not be made alone. This is not just Islam — it is how we do things here. You consult. You bring people in. You listen. This company bought the system and told us afterwards. No one asked us. That is why we do not trust it." (P19, Female Retail Employee, Riyadh)

4.4. Theme 4: Collective versus Individual Moral Responses to AI-Driven Harm

The fourth theme concerns the form that moral responses to AI-driven harm took among Saudi employees — and the tension between the individual rights-based redress mechanisms that organisations made available and the collective, relational forms of response that participants found more culturally resonant and practically effective. This theme challenges the individualist assumptions that underpin most AI ethics frameworks and reveals the collective moral sensibilities that characterise Saudi Arabian and broader GCC organisational cultures.

When participants experienced AI-driven decisions as unjust, their first responses were typically relational and collective rather than formal and individual. Rather than filing formal complaints or invoking individual rights, they talked to colleagues, sought informal solidarity, and mobilised informal networks of mutual support and information-sharing. These collective responses were not merely practical workarounds; they were themselves moral acts — expressions of communal solidarity that participants experienced as more ethically adequate to the situation than the individual redress mechanisms formally available.

"After the system gave us those impossible targets, we all talked. All of us — the whole floor. We decided together what we could do and what we could not. Nobody complained to HR individually. That would have been — I don't know, lonely. Wrong. We faced it together." (**P03, Female Retail Employee, Dammam**)

The contrast between collective solidarity and individual formal redress was sharpened by participants' descriptions of the formal complaint mechanisms that organisations made available for challenging AI-driven decisions. These mechanisms — typically involving written complaints to HR departments or managers — were widely regarded as ineffective, alienating, and misaligned with the relational character of Saudi workplace culture. Several participants described making formal complaints and receiving responses that simply restated the AI system's output without addressing the substantive grievance, reinforcing the sense that formal mechanisms served the organisation's liability management rather than the employee's moral need.

"I complained formally. I got a letter back with numbers. The same numbers. It told me nothing new. It did not address what I had said. I felt more alone after the letter than before it." (**P27, Male Banking Employee, Riyadh**)

4.5. Theme 5: Resignation, Resistance, and the Ethics of Algorithmic Coping

The fifth master theme concerns the range of moral responses that employees developed in coping with AI systems they regarded as ethically problematic — a spectrum running from passive moral withdrawal at one end through various forms of quiet resistance and gaming to active ethical refusal at the other. This theme illuminates the moral agency that employees exercise in their encounters with algorithmic authority, resisting the tendency of the AI ethics literature to frame affected individuals primarily as passive victims of algorithmic harm.

The most common coping response was what participants described as a form of moral withdrawal or disengagement: a conscious decision to reduce

their emotional and ethical investment in work relationships and decisions mediated by AI systems, treating algorithmic management as an impersonal environment to be navigated rather than a moral community to be engaged. This withdrawal was experienced by participants not as an ethical failure but as a form of moral self-protection — a way of preserving moral integrity in an environment they experienced as morally degraded.

"I used to care very much about my performance. I wanted to be good, not just to be evaluated well, but because it meant something to me. Now I think: what does it mean for the system to evaluate me well? The system does not know me. So I do what the system measures. And the rest — the rest of what I used to care about — I keep for myself." **(P33, Female Telecom Employee, Riyadh)**

A significant minority of participants — predominantly frontline employees in logistics and retail with strong peer networks — described forms of collective gaming: coordinated strategies for presenting work performance to AI monitoring systems in ways that maximised favourable outputs while preserving space for more human ways of working outside the system's gaze. These gaming strategies were explicitly framed in moral terms by participants: not as dishonesty but as legitimate resistance to a system they regarded as incapable of fairly evaluating their actual contributions.

"We know what the system looks at. We know what it does not look at. We are not stupid. So we make sure the system sees what it should see. And then we work the way we know is right. Is that dishonest? I don't think so. The system is not honest about us. We are just — balancing." **(P05, Male Logistics Driver, Jeddah)**

The most ethically reflective responses came from a smaller group of participants — predominantly middle managers and HR professionals — who described deliberate acts of ethical refusal: instances in which they had declined to implement or enforce AI-driven decisions they regarded as unjust, accepting personal professional risk to honour what they experienced as a higher moral obligation to the employees in their care.

"The system recommended we not renew the contract of one of my nurses. On paper, the numbers were what they were. But I knew this person. I knew her situation. I knew what the numbers did not show. I went to my director, and I said: I will not implement this. It took three conversations. But we found another way. I am responsible for my people. The system is not." **(P06, Male Healthcare Manager, Riyadh)**

5. Discussion

5.1. Relational Accountability as a Theoretical Contribution

The findings from this study support and empirically ground the concept of relational accountability developed in the theoretical background. Saudi employees' moral distress at AI-driven decision-making is not primarily about the violation of procedural rights — inadequate explanation, opaque decision logic, or absent formal redress — although these features of algorithmic governance are present and resented. It is about the disruption of the relational structures through which accountability has traditionally been exercised with

moral legitimacy: the face-to-face relationships in which obligations are personal, judgments are contextual, and repair is possible. When AI systems replace human decision-makers, or when managers use AI outputs to abdicate personal accountability, what is lost is not merely a procedural property of the decision but the moral quality of the relationship through which the decision was embedded.

This finding has significant implications for AI ethics theory. The dominant frameworks in the field — fairness, transparency, explainability, accountability — are procedural in character and individualist in their moral reference points. They seek to restore the conditions for individual rights-based redress in an algorithmic environment, but they do not address the relational and communal dimensions of moral experience that our findings show to be central to how Saudi employees experience the ethics of AI. A richer AI ethics — one adequate to the moral diversity of organisational contexts across the globe — must incorporate relational and communal ethical frameworks alongside procedural and individualist ones.

5.2. Islamic Ethics and the AI Ethics Canon

The finding that Saudi Muslim employees consistently and spontaneously drew on Islamic ethical concepts — 'adl, amanah, shura — to articulate their moral evaluations of AI systems is significant for the AI ethics literature in multiple respects. It demonstrates that the moral resources available for evaluating AI are far richer and more diverse than the Western philosophical and regulatory frameworks that dominate the field acknowledge. It shows that in contexts where Islamic ethical traditions are not only culturally embedded but legally enshrined, as in Saudi Arabia, employees bring sophisticated and contextually specific moral frameworks to their encounters with algorithmic authority — frameworks that generate different, but not less valid, moral critiques than those available within Western AI ethics discourse.

Practically, this finding suggests that AI ethics interventions in Saudi Arabian and GCC organisational contexts — whether through corporate governance measures, regulatory frameworks, or employee engagement strategies — will be more effective if they engage with Islamic ethical frameworks rather than simply translating Western AI ethics principles into Arabic or simply renaming them to align with Vision 2030 branding. The concepts of 'adl, amanah, and shura provide culturally resonant normative standards for AI governance that could be operationalised in organisational policies, stakeholder consultation processes, and regulatory guidance in ways that the imported vocabulary of algorithmic fairness, transparency, and accountability cannot.

5.3. Implications for Practice

For organisations deploying AI in Saudi Arabian contexts, the findings generate several practical implications. The strong and consistent emphasis on consultation (shura) as a legitimacy condition for AI deployment suggests that organisations that involve employees genuinely in the design, implementation, and governance of AI systems — as in the healthcare case described in Theme 2 — will experience significantly higher levels of acceptance and trust than those that deploy AI unilaterally. The finding that formal individual redress

mechanisms are experienced as culturally misaligned and practically ineffective suggests that organisations should invest in creating collective voice mechanisms — including employee representative structures, team-level consultation processes, and community forums — through which AI governance concerns can be raised and addressed in ways that honour the collective moral sensibilities of Saudi and GCC workforces.

6. Limitations and Future Research

This study has several limitations that qualify its findings and suggest directions for future research. The sample, while theoretically purposive, is restricted to Riyadh- and Jeddah-based organisations and may not represent the diversity of Saudi organisational contexts, including those in smaller cities such as Abha or Tabuk, or the substantial expatriate-dominated sectors. The restriction to BSE-visible or professionally networked organisations may introduce a systematic bias toward larger, more formally governed firms, particularly those subject to SDAIA oversight or Vision 2030 programme requirements, in which AI deployment is more advanced and more documented. Future research should seek to include smaller, less formally governed organisations, including family businesses and SMEs, which constitute the majority of Saudi employment.

The study's single-country design, while enabling the contextually rich analysis that is its primary contribution, limits the comparative generalisability of its findings. Research extending the qualitative examination of employee AI ethics perceptions to other GCC contexts — particularly the UAE, Qatar, and Kuwait, which share comparable AI adoption trajectories — would enable richer regional comparison and more robust theoretical development. Future work should also examine how the unique combination of state-driven AI deployment, Vision 2030 institutional pressure, and Islamic governance frameworks shapes employee AI ethics perceptions in ways that may differ from other GCC and MENA contexts with different institutional profiles.

7. Conclusion

This paper has presented findings from a qualitative interview study exploring how employees in Saudi Arabian organisations morally experience and evaluate AI systems that shape their working lives. Through reflexive thematic analysis of 38 in-depth interviews, we identified five master themes — algorithmic opacity and the collapse of moral dialogue; displaced accountability and the erosion of relational trust; religious and cultural frames for evaluating algorithmic authority; collective versus individual moral responses; and resignation, resistance, and algorithmic coping — that together constitute a rich account of AI ethics as lived experience in a non-Western organisational context.

The study makes three theoretical contributions. It develops and empirically grounds the concept of relational accountability as a culturally embedded alternative to procedural accountability that is more adequate to the moral sensibilities of Saudi and broader GCC employees. It demonstrates the salience of Islamic ethical frameworks — particularly 'adl, amanah, and shura — as resources for moral evaluation of AI systems in Muslim-majority organisational contexts, enriching the AI ethics canon with perspectives that

Western philosophical traditions do not provide. And it challenges the individualist assumptions that underpin most AI ethics frameworks by showing that collective moral responses — solidarity, communal grievance, shared resistance — are the primary ethical response to algorithmic harm in a context characterised by strong communal moral culture.

The broader implication is that a globally adequate AI ethics cannot be built on the procedural, individualist, and Western philosophical foundations that currently dominate the field. It must engage seriously with the diversity of moral frameworks — including Islamic ethics, communal accountability traditions, and relational trust norms — through which people around the world evaluate the conduct of the powerful institutions that shape their lives. Saudi Arabia, with its distinctive combination of state-driven AI ambition, Islamic institutional culture, and rapid organisational transformation under Vision 2030, provides a valuable window into that broader moral diversity — and into the inadequacy of AI ethics frameworks that do not see through it.

References

- Ajunwa, I., Crawford, K. and Schultz, J. (2017). Limitless worker surveillance. *California Law Review*, 105(3), 735-776.
- Araujo, T., Helberger, N., Kruikemeier, S. and de Vreese, C.H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, 35(3), 611-623. <https://doi.org/10.1007/s00146-019-00931-w>
- Beekun, R.I. and Badawi, J.A. (2005). Balancing ethical responsibility among multiple organizational stakeholders: The Islamic perspective. *Journal of Business Ethics*, 60(2), 131-145. <https://doi.org/10.1007/s10551-004-8204-5>
- Binns, R., van Kleek, M., Veale, M., Lyngs, U., Zhao, J. and Shadbolt, N. (2018). 'It's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 1-14. <https://doi.org/10.1145/3173574.3173951>
- Braun, V. and Clarke, V. (2022). *Thematic Analysis: A Practical Guide*. Sage, London. <https://doi.org/10.53841/bpsqmp.2022.1.33.46>
- Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2), 56-62. <https://doi.org/10.1145/2844110>
- Dusuki, A.W. and Abdullah, N.I. (2007). Maqasid al-Shari'ah, maslahah, and corporate social responsibility. *The American Journal of Islamic Social Sciences*, 24(1), 25-45. <https://doi.org/10.35632/ajiss.v24i1.415>
- Elmahjub, E. (2023). Artificial intelligence (AI) in Islamic ethics: Towards pluralist ethical benchmarking for AI. *Philosophy and Technology*, 36(4), 1-24. <https://doi.org/10.1007/s13347-023-00668-x>
- Gandini, A. (2019). Labour process theory and the gig economy. *Human Relations*, 72(6), 1039-1056. <https://doi.org/10.1177/0018726718790002>
- Ghaly, M. (2024). What makes work 'good' in the age of artificial intelligence (AI)? Islamic perspectives on AI-mediated work ethics. *The Journal of Ethics*, 28, 429-453. <https://doi.org/10.1007/s10892-023-09456-3>

- Glikson, E. and Woolley, A.W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627-660. <https://doi.org/10.5465/annals.2018.0057>
- Kellogg, K.C., Valentine, M.A. and Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, 14(1), 366-410. <https://doi.org/10.5465/annals.2018.0174>
- Lee, M.K., Kusbit, D., Metsky, E. and Dabbish, L. (2015). *Working with machines: The impact of algorithmic and data-driven management on human workers*. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 1603-1612. <https://doi.org/10.1145/2702123.2702548>
- Majrashi, K. (2025). Employees' perceptions of the fairness of AI-based performance prediction features. *Cogent Business & Management*, 12(1), 2456111. <https://doi.org/10.1080/23311975.2025.2456111>
- Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1-21. <https://doi.org/10.1177/2053951716679679>
- Möhlmann, M. and Zalmanson, L. (2017). *Hands on the wheel: Navigating algorithmic management and Uber drivers' autonomy*. Proceedings of the International Conference on Information Systems, 1-16.
- Raji, I.D., Smart, A., White, R.N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D. and Barnes, P. (2020). *Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing*. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 33-44. <https://doi.org/10.1145/3351095.3372873>
- Saudi Data and AI Authority (SDAIA) (2023). *Principles and Controls of AI Ethics*. Saudi Data and Artificial Intelligence Authority, Riyadh. Available at: <https://sdaia.gov.sa>
- Saudi Data and AI Authority (SDAIA) (2024). *AI Adoption Framework*. Saudi Data and Artificial Intelligence Authority, Riyadh. Available at: <https://sdaia.gov.sa>
- Vision 2030 Kingdom of Saudi Arabia (2016). *Saudi Vision 2030*. Council of Economic and Development Affairs, Riyadh. Available at: <https://www.vision2030.gov.sa>
- Wood, A.J., Graham, M., Lehdonvirta, V. and Hjorth, I. (2019). Good gig, bad gig: Autonomy and algorithmic control in the global gig economy. *Work, Employment and Society*, 33(1), 56-75. <https://doi.org/10.1177/0950017018785616>