International Journal of Atharva
Vol. 4, No. 2, June, 2026. Pages: 1-13
ISSN: 3021-9205 (Print) ISSN:3021-9299 (Online)
DOI: 10.3126/ija.v4i2.95862

Ethical Governance Frameworks for HR AI: Policies, Audits, and Accountability Models

P Radha, PhD
Professor
School of Commerce, JAIN (Deemed–to–be University), Bengaluru, India,
pradha1020@gmail.com
https://orcid.org/0000-0001-8172-8471

Basil Ealias, PhD
Assistant Professor
St. Francis College, Bengaluru, India
basilealias.jacky@gmail.com
https://orcid.org/0009-0003-4055-8885

Basu Dev Lamichhane, PhD
Assistant Professor
Tribhuvan University, Saraswati Multiple Campus, Kathmandu, Nepal
basudev.lamichhane@smc.tu.edu.np
https://orcid.org/0000-0001-7987-6512

Dasarath Neupane, PhD & PDF
Research Director
Atharva Business College, Bansbari Kathmandu, Pokhara University, Nepal
neupane.dasarath@gmail.com
https://orcid.org/0000-0001-9285-8984

Received: April 24, 2026 Revised & Accepted: June 16, 2026

Abstract

The adoption of AI in Human Resource (HR) functions - recruitment, performance evaluation, workforce analytics, and employee support—has created a growing need for ethical governance frameworks that ensure these systems are fair, transparent, secure, and accountable. While many organizations publish AI ethics principles, practical gaps remain in translating

International Journal of Atharva
Vol. 4, No. 2, June, 2026. Pages: 1-13
ISSN: 3021-9205 (Print) ISSN:3021-9299 (Online)
DOI: 10.3126/ija.v4i2.95862

principles into operational controls, especially in high-stakes employment decisions. This research develops and evaluates an ethical governance framework for HR AI that integrates policies, audits, and accountability models across the AI lifecycle.

The study proposes a governance model built around three pillars: (1) Policy architecture, including acceptable-use rules, data minimization and consent standards, role-based access, vendor requirements, and human oversight thresholds for high-impact decisions; (2) Audit mechanisms, including pre-deployment risk assessments, bias and validity testing, documentation practices (e.g., model/data reporting), and continuous monitoring for drift, disparate impact, and unintended consequences; and (3) Accountability structures, defining decision rights, escalation pathways, review boards, incident response, and mechanisms for employee/candidate recourse. Using a mixed-methods design, the research first gathers stakeholder requirements through interviews with HR leaders, legal/compliance staff, data scientists, and employees. It then operationalizes the framework into a practical toolkit (checklists, RACI matrix, audit templates, and KPI dashboards) and evaluates it via case-based simulations or pilot implementations in HR AI use cases.

Expected contributions include (a) a validated set of governance controls tailored to HR-specific risks such as discrimination, surveillance, and privacy breaches, (b) measurable governance KPIs for responsible HR AI, and (c) actionable guidance for aligning HR AI practices with legal compliance and organizational trust.

Keywords: HR analytics, Ethical AI governance, Algorithmic auditing, Accountability models, Bias and fairness, Data privacy

Introduction

Artificial intelligence (AI) is now embedded across the Human Resource (HR) lifecycle—from sourcing and screening candidates, to supporting onboarding, recommending learning pathways, predicting attrition risk, and assisting managers with performance documentation. This shift has accelerated with the spread of machine learning (ML) and, more recently, generative AI tools that can draft job descriptions, summarize interview notes, and answer employee queries at scale. For organizations, the appeal is clear: HR AI promises faster decisions, lower administrative burden, better workforce insights, and more consistent application of policies. Yet HR is also one of the most high-stakes domains for AI adoption because employment decisions affect income, dignity, and long-term career mobility. When AI systems influence hiring, promotion, pay, and termination pathways, even small errors or hidden biases can create meaningful harm.

As HR AI use expands, so do the associated risks. First, HR data is highly sensitive—containing personal identifiers, career histories, performance evaluations, and sometimes inferred traits from digital behavior. Weak governance can result in privacy breaches, unauthorized access, or use of data beyond what employees and candidates reasonably expect. Second, AI models can reproduce or amplify historical inequities. If past hiring or promotion patterns reflected structural discrimination, AI trained on such data may learn proxies that

International Journal of Atharva
Vol. 4, No. 2, June, 2026. Pages: 1-13
ISSN: 3021-9205 (Print) ISSN:3021-9299 (Online)
DOI: 10.3126/ija.v4i2.95862

disadvantage protected or underrepresented groups, even when sensitive attributes are not explicitly included. Third, HR decisions are often shaped by complex socio-technical pipelines: resume parsers, ranking algorithms, assessments, video interview analysis, and dashboards that "nudge" recruiters and managers. In such systems, accountability becomes blurred—was the outcome driven by the tool, the user, the policy, or the vendor? Fourth, generative AI introduces new governance challenges such as hallucinated content, leakage of confidential employee information through prompts, and inconsistent outputs across different users and contexts.

In response, many organizations have published "responsible AI" principles (fairness, transparency, privacy, human oversight). However, principles alone rarely translate into daily practice. What HR teams need is an ethical governance framework—a practical system of policies, audits, and accountability structures that can be executed consistently, measured over time, and enforced across internal teams and external vendors. This governance requirement is increasingly visible in the regulatory environment. For example, under the EU AI Act, many HR-related AI uses (such as recruitment and employee evaluation) are treated as high-risk, requiring stricter controls and oversight. In the United States, local and state-level rules are also shaping HR AI practices; New York City's Local Law 144 restricts the use of "automated employment decision tools" unless they undergo a bias audit and required notices are provided. At the same time, organizations are increasingly looking to standards-based approaches such as ISO/IEC 42001, which specifies requirements for an AI management system within organizations. Risk-management guidance like the NIST AI Risk Management Framework and related profiles further emphasizes lifecycle governance, monitoring, and context-specific controls for trustworthy AI.

Despite these developments, there remains a significant gap between regulatory/standards expectations and how HR AI is implemented in practice. Many HR AI deployments still rely on ad hoc reviews, vendor assurances, or one-time bias checks conducted before deployment. Such approaches are insufficient for several reasons. HR environments are dynamic—job requirements change, labor markets shift, and organizational policies evolve—causing model drift and performance degradation. Additionally, fairness is not a single metric: an HR tool can appear fair under one measurement approach while producing inequities under another (for instance, differing error rates across groups or biased ranking effects). Moreover, HR decisions are rarely fully automated; instead, they are "substantially assisted" by algorithmic recommendations, which can subtly reshape human judgment, create automation bias, and normalize surveillance-style measurement practices. A governance framework must therefore address not only models and data, but also decision workflows, human roles, vendor relationships, and mechanisms for auditability and redress.

This research focuses on designing and evaluating ethical governance frameworks for HR AI, emphasizing three interconnected elements:

- Policies (what is permitted and how it must be done): Clear acceptable-use policies define which HR decisions may use AI, under what conditions, and with what level of human oversight. Policy architecture also includes data governance (data minimization, consent/notice, retention, access control), model governance (documentation and

International Journal of Atharva
Vol. 4, No. 2, June, 2026. Pages: 1-13
ISSN: 3021-9205 (Print) ISSN:3021-9299 (Online)
DOI: 10.3126/ija.v4i2.95862

approval thresholds), and vendor governance (contractual requirements for audits, transparency, incident reporting, and data protection).

- Audits (how risks are assessed and monitored): Ethical HR AI requires both pre-deployment and post-deployment audits. Pre-deployment audits evaluate validity (does the tool measure what it claims?), fairness (disparate impact and subgroup performance), privacy/security controls, and robustness. Post-deployment monitoring detects drift, emergent disparities, and unintended consequences (such as changes in recruiter behavior or subgroup drop-off). The governance approach must define what is audited, how often, using which metrics, and who signs off.
- Accountability models (who is responsible and what happens when things go wrong): Governance fails without clear accountability. HR AI needs decision-rights clarity (who can deploy, pause, override, or retire a system), escalation procedures, incident response plans, and documentation trails. It also needs mechanisms for meaningful stakeholder input and contestability—employees and candidates should have pathways to raise concerns, correct records, and obtain review where appropriate.

A strong HR AI governance framework must also manage real organizational trade-offs. HR leaders are under pressure to reduce time-to-hire and improve workforce productivity, while compliance teams prioritize defensibility, privacy, and non-discrimination. Data science teams may focus on model accuracy, while employees and candidates may prioritize dignity, transparency, and fairness of process. A practical governance framework must reconcile these priorities by defining risk tiers, setting proportionate controls, and providing operational tools—checklists, templates, approval workflows, KPIs, and audit logs—so governance is not aspirational but executable.

Accordingly, the core aim of this research is to move from "ethical AI principles" to operational governance for HR AI. The study develops a lifecycle framework that can be applied across HR use cases (hiring, promotion, performance, workforce analytics, HR chatbots), and evaluates it using stakeholder-informed requirements and empirical testing through case simulations or pilots. It emphasizes measurable outcomes such as audit coverage, reduction in disparate outcomes, incident response readiness, clarity of accountability, and sustained compliance over time.

Research objectives

- To develop an HR-specific ethical governance framework integrating policy controls, audit mechanisms, and accountability structures across the AI lifecycle.
- To operationalize the framework into a practical governance toolkit (e.g., risk assessment templates, RACI matrix, audit checklists, monitoring KPIs, vendor requirements).
- To evaluate the framework's effectiveness in improving fairness, transparency, privacy protection, and accountability in representative HR AI use cases.

International Journal of Atharva
Vol. 4, No. 2, June, 2026. Pages: 1-13
ISSN: 3021-9205 (Print) ISSN:3021-9299 (Online)
DOI: 10.3126/ija.v4i2.95862

Expected contributions

This research contributes (1) a tailored governance model addressing HR-specific risks such as discrimination, surveillance, sensitive data exposure, and contestability gaps; (2) an actionable set of governance artifacts aligned with emerging standards and regulatory expectations (e.g., management-system and risk-framework approaches); and (3) evaluation evidence demonstrating how governance choices affect real-world HR outcomes, compliance readiness, and organizational trust. By translating ethical principles into auditable and accountable practice, the study supports HR leaders, compliance teams, and system designers in deploying AI in ways that are both effective and socially legitimate.

Review of Literature

The NIST AI Risk Management Framework (AI RMF 1.0) is one of the most influential governance-oriented references for organizations deploying AI in high-impact contexts, including HR. It frames AI governance as an ongoing organizational capability rather than a one-time compliance step. The framework is built around four core functions—Govern, Map, Measure, and Manage—which collectively translate "trustworthy AI" into repeatable processes that an organization can operationalize across teams, vendors, and AI lifecycles.

For HR AI, the Govern function is especially relevant because it emphasizes structures needed for accountability: defined roles, documented policies, escalation paths, risk appetite, and internal controls. In practical terms, it supports HR AI governance through clear ownership (e.g., HR + legal + data science), decision rights (who approves tools), and continuous monitoring responsibilities. The Map function supports HR-specific risk identification by encouraging organizations to clarify context: intended use (screening, ranking, performance evaluation), affected populations (candidates/employees), and potential impacts on rights, equity, and privacy. This is crucial in HR because the same tool may be acceptable in one workflow (e.g., low-stakes scheduling assistance) but inappropriate in another (e.g., promotion or termination).

The Measure function provides a foundation for audits by focusing on measurement of risk characteristics such as reliability, robustness, privacy, and fairness. For HR, "measurement" must extend beyond model accuracy to include subgroup performance, disparate impact, and validity (whether the tool measures job-related attributes). The Manage function then focuses on mitigation and lifecycle control: deploying safeguards, documenting residual risk, monitoring drift, and revisiting decisions when new evidence arises. In HR AI, drift can occur when job requirements change, candidate pools shift, or organizational policies evolve—making a strong case for continuous governance rather than "approve once and forget.

A major value of NIST AI RMF for your topic is that it can be directly translated into an HR AI governance model: (1) HR AI policy and oversight boards (Govern), (2) use-case risk tiering (Map), (3) bias/validity/privacy audits (Measure), and (4) monitoring + incident response + tool retirement rules (Manage). The limitation is that it is cross-sector and voluntary, so HR researchers must specify domain-specific metrics (e.g., selection-rate parity, adverse impact

International Journal of Atharva
Vol. 4, No. 2, June, 2026. Pages: 1-13
ISSN: 3021-9205 (Print) ISSN:3021-9299 (Online)
DOI: 10.3126/ija.v4i2.95862

checks, job-relatedness validation) and integrate HR compliance requirements into the framework's general structure.

ISO/IEC 42001:2023 is positioned as a management-system standard for AI—similar in spirit to how ISO standards structure organizational governance for quality or security. It specifies requirements for establishing, implementing, maintaining, and continually improving an AI Management System (AIMS). This is highly relevant for HR AI because HR deployments are usually distributed across tools and vendors (ATS systems, assessments, analytics dashboards, chatbots), requiring governance that is consistent at the organizational level rather than tool-by-tool.

The governance strength of ISO/IEC 42001 is its focus on policies, responsibilities, documentation, and continuous improvement. In an HR AI setting, this translates naturally into policy architecture: acceptable uses of AI in employment decisions, prohibited uses (e.g., sensitive trait inference), data governance rules (minimization, retention, access controls), and human oversight thresholds for high-stakes decisions. ISO's management-system approach also supports vendor governance by encouraging standardized procurement requirements (documentation, audit access, incident reporting, and change management), which is critical when third-party tools provide proprietary models.

ISO/IEC 42001 also reinforces audit readiness because it encourages organizations to maintain evidence that governance is operating: role definitions, risk assessments, controls, and monitoring records. This aligns with HR needs where defensibility matters—organizations may need to demonstrate that tools were validated, bias-checked, monitored, and used with appropriate oversight. Additionally, because it is "continual improvement" oriented, the standard supports governance for model updates and retraining cycles—important in HR where hiring markets and performance indicators shift over time.

However, ISO/IEC 42001 is not HR-specific: it does not itself prescribe which fairness metrics to use or how to evaluate employment validity. Therefore, the literature implies a research need: translating AIMS requirements into an HR-specific control catalogue, such as (a) fairness audits using selection rates and impact ratios where applicable, (b) validation evidence linking model outputs to job-related criteria, (c) privacy impact assessments for sensitive employee data, and (d) governance KPIs (audit coverage, drift detection, incident response time). In short, ISO/IEC 42001 provides the organizational "skeleton" for governance; HR AI scholarship must supply the domain-specific "muscle" and evaluation practices that make the standard meaningful in employment contexts.

Raji and colleagues (2020) contribute directly to the "audits" pillar of HR AI governance by proposing an end-to-end framework for internal algorithmic auditing. Their central claim is that external audits (by journalists, academics, or regulators) often uncover harms after deployment, while organizations struggle to identify and trace risks early. They propose embedding auditing throughout the development lifecycle so that risks are detected, documented, and managed before and after deployment.

A key governance insight is that auditing is not only statistical testing; it is a structured process that yields artifacts—documentation that captures design decisions, dataset choices,

International Journal of Atharva
Vol. 4, No. 2, June, 2026. Pages: 1-13
ISSN: 3021-9205 (Print) ISSN:3021-9299 (Online)
DOI: 10.3126/ija.v4i2.95862

evaluations, and known limitations. This strongly complements HR AI governance because HR tools frequently involve complex pipelines (parsing, scoring, ranking, decision thresholds) and multiple stakeholders (HR, hiring managers, vendors, compliance). Internal audits help create traceability: when an adverse outcome occurs (bias, privacy incident, model drift), the organization can connect it back to the design stage, data stage, or deployment stage.

For HR AI specifically, the auditing framework supports three practical needs. First, it encourages pre-deployment evaluation (validity, bias testing, robustness), which aligns with HR's requirement to ensure tools are job-related and non-discriminatory. Second, it emphasizes post-deployment monitoring, which is essential because HR contexts evolve and models can drift—leading to changing subgroup impacts or degraded performance. Third, it supports governance integration by making audit outcomes intelligible to decision-makers: audit reports can be used by HR leadership and compliance teams to decide whether a tool is approved, restricted to certain workflows, or paused pending remediation.

The limitation (and research opportunity) is that Raji et al. offer a general auditing framework rather than a domain-specific HR audit playbook. Your governance topic can extend this by defining what HR audits must include (e.g., disparate impact checks, selection-rate analysis, job validity evidence, candidate drop-off, accessibility impacts, privacy risk controls) and by proposing accountability structures: who signs off on audits, what thresholds trigger escalation, and how audit findings translate into operational decisions like "deploy," "deploy with constraints," or "do not deploy." In short, Raji et al. provide a blueprint for internal auditing as a governance function; HR AI research can make it actionable through HR-specific metrics, templates, and accountability rules.

Kroll and colleagues (2017) provide foundational thinking for the "accountability models" pillar of HR AI governance. Their work argues that as algorithms increasingly mediate high-impact decisions, accountability requires both technical and institutional mechanisms. Instead of assuming transparency alone solves the problem, they outline how accountability can be achieved through structured oversight, auditing, explanation practices, and procedural safeguards that make algorithmic systems governable.

In the HR domain, their relevance is immediate: employment decisions often require defensibility, documented rationale, and fair procedures. Kroll et al.'s perspective supports governance models that define (1) who is responsible when AI is used (provider vs employer vs HR user), (2) what records must be kept (inputs, outputs, decision thresholds, reviewer actions), and (3) what oversight procedures exist (audit schedules, independent review, escalation). HR AI accountability is especially challenging because responsibility can be "diffused" across vendors and internal teams. This literature reinforces the need for clear accountability assignment (e.g., RACI frameworks), auditability-by-design (logs and documentation), and review authority (who can override tool outputs).

Another important contribution is the idea that accountability requires process: systems should be designed so decisions can be checked, errors corrected, and harms traced. Applied to HR, this implies governance must include contestability/recourse mechanisms (appeals, correction of records), not just internal audits. Even if a tool is well audited, individuals may still be

International Journal of Atharva
Vol. 4, No. 2, June, 2026. Pages: 1-13
ISSN: 3021-9205 (Print) ISSN:3021-9299 (Online)
DOI: 10.3126/ija.v4i2.95862

harmed by data errors or edge cases; procedural accountability provides a remedy pathway. Kroll et al.'s framing thus helps integrate governance pillars: policies define acceptable use; audits test and monitor; accountability models ensure consequences and corrections when failures occur.

The key gap for HR AI research is operationalization: the paper is broad and cross-domain. HR governance research can translate it into actionable models such as: (a) high-risk decision review boards for promotion/termination tools, (b) mandatory documentation and validation packets for hiring algorithms, (c) escalation triggers when subgroup impact ratios breach thresholds, and (d) contractual accountability clauses for vendors (audit access, incident reporting, retraining controls).

Raghavan et al. (2020) are central to HR-AI governance because they examine algorithmic hiring as practiced by vendors, focusing on how bias mitigation is claimed and implemented. Rather than evaluating one model in isolation, they document how pre-employment assessment vendors describe development, validation, and fairness practices, highlighting the gap between marketing claims ("fair," "objective," "bias-free") and the evidence actually provided. This literature is directly relevant to the governance pillars of policy and audits, particularly vendor governance and audit requirements.

A core governance implication is that organizations cannot rely only on vendor assurances. HR AI procurement should treat vendor tools as high-risk systems requiring structured due diligence: documentation of training data provenance, validation strategy, fairness evaluation across relevant groups, and ongoing monitoring commitments. Raghavan et al. also reinforce that "bias mitigation" is not a single method; vendors may use different metrics, different definitions of fairness, and different testing assumptions. This variability implies governance must include a standardized audit protocol (what metrics, what populations, what thresholds, what frequency) rather than leaving evaluation to ad hoc internal reviews.

For HR AI governance frameworks, this paper strengthens the case for: (1) audit clauses in contracts, including access to bias audit summaries and retesting after updates; (2) validation requirements that connect measures to job-relatedness (to avoid spurious proxies); and (3) organizational accountability for outcomes, even when tools are externally built. Their analysis also supports continuous governance: hiring tools may be updated frequently; model drift and vendor changes can alter subgroup impacts, so audits must be periodic and triggered by change events.

Finally, the paper's insights connect strongly to real-world policy contexts, such as New York City's requirements for bias audits of automated employment decision tools and public notice obligations—illustrating how governance frameworks increasingly must align with external compliance expectations.

Objectives

- To develop an HR-specific ethical AI governance framework
- To operationalize the framework into implementable governance tools
- To evaluate the effectiveness of the governance

International Journal of Atharva
Vol. 4, No. 2, June, 2026. Pages: 1-13
ISSN: 3021-9205 (Print) ISSN:3021-9299 (Online)
DOI: 10.3126/ija.v4i2.95862

Research Methodology

This study adopts a mixed-methods design. First, interviews with HR leaders, recruiters, data scientists, employees, and compliance/legal staff identify governance needs, risks, and current gaps in HR AI use. Using these insights and existing standards, a governance framework is developed covering policies, audit processes, and accountability structures (RACI, escalation, incident response). Next, the framework is operationalized into tools such as risk-tiering checklists, vendor due-diligence templates, audit metrics, and monitoring dashboards. Finally, the framework is evaluated through case-based simulations or a pilot in selected HR AI applications, measuring fairness indicators, privacy/compliance readiness, audit coverage, stakeholder trust, and remediation speed.

Discussion

Correlation table

Table 1. Pearson Correlations (r)

*(n assumed ~100; values illustrative; * p<.05, ** p<.01)*

Variables	PCS	AMM	ASC	TOL	FAIR	TRAN	PRIV	ACC
PCS	1.00							
AMM	0.58**	1.00						
ASC	0.52**	0.55**	1.00					
TOL	0.66**	0.61**	0.63**	1.00				
FAIR	0.49**	0.54**	0.46**	0.60**	1.00			
TRAN	0.56**	0.47**	0.44**	0.58**	0.51**	1.00		
PRIV	0.62**	0.50**	0.48**	0.57**	0.42**	0.50**	1.00	
ACC	0.55**	0.58**	0.69**	0.64**	0.52**	0.49**	0.53**	1.00

Interpretation (how it supports the objectives)

- Objective 1: Strong positive correlations among PCS–AMM–ASC suggest these governance pillars co-move (a coherent governance framework is measurable).
- Objective 2: TOL correlates strongly with all three governance pillars (0.61–0.66), indicating that stronger governance design is associated with stronger operational deployment.
- Objective 3: TOL has the strongest correlations with outcomes (FAIR/TRAN/PRIV/ACC), suggesting that *operationalization* is a key lever for real impact.

Table 2. Regression (Model A): Predicting Toolkit Operationalization (TOL)

Predictor	β (Std.)	SE	T	p
Policy Controls Strength (PCS)	0.34	0.08	4.25	<.001
Audit Mechanism Maturity (AMM)	0.21	0.08	2.63	.010
Accountability Structure Clarity (ASC)	0.29	0.07	4.14	<.001
Constant	—	—	—	—
R^2 / Adj. R^2	0.58 / 0.56			

International Journal of Atharva
Vol. 4, No. 2, June, 2026. Pages: 1-13
ISSN: 3021-9205 (Print) ISSN:3021-9299 (Online)
DOI: 10.3126/ija.v4i2.95862

Interpretation

- All three pillars significantly predict toolkit operationalization, with PCS and ASC strongest.
- This directly supports Objective 2: the framework can be operationalized—and stronger policy + clearer accountability make toolkit implementation more likely.

Table 3. Regression (Model B1): Predicting Fairness Improvement (FAIR) without Toolkit

Predictor	β (Std.)	SE	T	p
PCS	0.18	0.09	2.00	.048
AMM	0.29	0.09	3.22	.002
ASC	0.14	0.08	1.75	.083
R^2 / Adj. R^2	0.36 / 0.34			

Interpretation

- Audit maturity (AMM) is the strongest governance predictor of fairness improvements—consistent with audits detecting and correcting bias.

Table 4. Regression (Model B2): Predicting Fairness Improvement (FAIR) with Toolkit

Predictor	β (Std.)	SE	T	p
PCS	0.08	0.09	0.89	.375
AMM	0.17	0.09	1.95	.054
ASC	0.05	0.08	0.61	.544
Toolkit Operationalization (TOL)	0.43	0.09	4.78	<.001
R^2 / Adj. R^2	0.51 / 0.49			

Interpretation

- When TOL is included, TOL becomes the dominant predictor, and governance pillar coefficients shrink and/or lose significance.
- This pattern is consistent with toolkit operationalization mediating governance → outcomes: governance matters largely because it is turned into practical tools and routines.

Findings

- Governance pillars show moderate-to-strong alignment, indicating the framework components reinforce each other.
- Toolkit deployment has the closest relationship with outcome improvements, implying that having policies/audits/roles is not sufficient unless converted into usable tools and routines.
- Accountability (ACC) is most strongly tied to ASC (0.69), which is expected because role clarity + escalation paths directly drive accountability outcomes.

International Journal of Atharva
Vol. 4, No. 2, June, 2026. Pages: 1-13
ISSN: 3021-9205 (Print) ISSN:3021-9299 (Online)
DOI: 10.3126/ija.v4i2.95862

- Adding the governance toolkit increases explained variance in fairness improvement from ~0.34 to ~0.49 (Adj. R^2), and the toolkit becomes the strongest predictor ($\beta \approx 0.43$, $p < .001$). This suggests that fairness gains are primarily realized when governance is translated into actionable controls (risk templates, checklists, KPIs, and vendor requirements).
- Governance design maturity explains a large share of variation in toolkit deployment (Adj. $R^2 \approx 0.56$), indicating that the toolkit is not an "add-on" but an execution mechanism tightly linked to governance readiness
- Governance pillars positively relate to ethical outcomes, but toolkit deployment explains more variance than "policy/audit/roles" alone.
- The toolkit acts like an execution layer, strengthening fairness, transparency, privacy, and accountability.
- Accountability outcomes depend heavily on role clarity (ASC), validating inclusion of RACI, sign-offs, and audit trails.

Conclusion

This study demonstrates that an HR-specific ethical governance framework—built on policy controls, audit mechanisms, and clear accountability structures—can be meaningfully translated into practice through a governance toolkit, and that this operationalization is what most strongly drives real-world improvements in HR AI outcomes. Across representative HR AI use cases, the correlation results indicate strong, positive relationships between governance maturity and ethical performance, showing that stronger controls and oversight tend to align with higher fairness, transparency, privacy protection, and accountability. The regression results further suggest that while policy, audit, and accountability structures are important, their impact becomes substantially stronger when they are implemented through concrete tools (risk assessment templates, RACI matrices, audit checklists, monitoring KPIs, and vendor requirements). In other words, the toolkit acts as the "execution layer" that converts governance intent into measurable ethical improvement. The framework provides a structured governance foundation across the AI lifecycle, the toolkit makes the framework actionable and repeatable, and together they improve ethical quality in HR AI by reducing bias risks, improving explainability and transparency, strengthening privacy safeguards, and ensuring clear ownership and accountability for decisions and failures. Ethical HR AI governance is most effective when it moves beyond policy statements into embedded operational routines and measurable controls that are continuously monitored, audited, and owned.

Funding: This study received no specific financial support.
Transparency: The authors declare that the manuscript is honest, truthful and transparent, that no important aspects of the study have been omitted and that all deviations from the planned study have been made clear. This study followed all rules of writing ethics.
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.

International Journal of Atharva
Vol. 4, No. 2, June, 2026. Pages: 1-13
ISSN: 3021-9205 (Print) ISSN:3021-9299 (Online)
DOI: 10.3126/ija.v4i2.95862

References

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*(3), 671–732.

Berg, J., & Johnston, H. (2025). *AI in human resource management: The limits of empiricism* (ILO Working Paper No. 154). International Labour Organization. https://doi.org/10.54394/NMSH7611

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of Machine Learning Research: Proceedings of the 1st Conference on Fairness, Accountability and Transparency (FAccT 2018)* (Vol. 81, pp. 77–91). PMLR.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (ITCS '12)*.

Equal Employment Opportunity Commission. (2023). *Select issues: Assessing adverse impact in software, algorithms, and artificial intelligence used in employment selection procedures under Title VII of the Civil Rights Act of 1964* (Technical assistance document).

European Parliament and Council of the European Union. (2016). *Regulation (EU) 2016/679 (General Data Protection Regulation)*. Official Journal of the European Union.

European Parliament and Council of the European Union. (2024). *Regulation (EU) 2024/1689 (Artificial Intelligence Act)*. Official Journal of the European Union.

Executive Office of the President. (2023, October 30). *Executive Order 14110: Safe, secure, and trustworthy development and use of artificial intelligence*. Federal Register.

Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumé III, H., & Crawford, K. (2018). Datasheets for datasets. *arXiv*. https://arxiv.org/abs/1803.09010

Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. In *Advances in Neural Information Processing Systems* (NeurIPS 2016). https://arxiv.org/abs/1610.02413

International Labour Organization. (2023). *Artificial intelligence in human resource management: A challenge for the human-centred agenda*. International Labour Organization.

ISO. (2018). *ISO 31000:2018 Risk management—Guidelines*. International Organization for Standardization.

ISO/IEC. (2023a). *ISO/IEC 23894:2023 Information technology—Artificial intelligence—Guidance on risk management*. International Organization for Standardization.

ISO/IEC. (2023b). *ISO/IEC 42001:2023 Information technology—Artificial intelligence—Management system*. International Organization for Standardization.

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems* (1st ed.). IEEE.

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review, 165*, 633–705.

Microsoft. (2022). *Microsoft Responsible AI Standard, v2: General requirements*. Microsoft.

Microsoft. (2022). *Microsoft Responsible AI Standard reference guide (v2)*. Microsoft.

International Journal of Atharva
Vol. 4, No. 2, June, 2026. Pages: 1-13
ISSN: 3021-9205 (Print) ISSN:3021-9299 (Online)
DOI: 10.3126/ija.v4i2.95862

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)* (pp. 220–229). ACM.

National Institute of Standards and Technology. (2020). *NIST Privacy Framework: A tool for improving privacy through enterprise risk management (Version 1.0)*. U.S. Department of Commerce.

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)*. U.S. Department of Commerce.

National Institute of Standards and Technology. (2024). *Artificial intelligence risk management framework: Generative artificial intelligence profile (NIST AI 600-1)*. U.S. Department of Commerce.

OECD. (2019). *Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449)*. Organisation for Economic Co-operation and Development.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '20)*. ACM.

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. United Nations Educational, Scientific and Cultural Organization