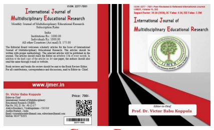

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:4(1), April 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

DOES AI-DRIVEN HRM IMPROVE FIRM PERFORMANCE? A META-ANALYSIS OF EMPIRICAL EVIDENCE

[1]Gudala Sowmya, [2]Prof. P. Arun Kumar and [3]Dr Sarans Swabi
[1]Research Scholar, DCMS, Andhra University,Visakhapatnam, Andhra Pradesh
[2]DCMS, Andhra University,Visakhapatnam, Andhra Pradesh
[3]Faculty, Andhra University,Visakhapatnam, Andhra Pradesh

Abstract

Ask any senior HR leader whether artificial intelligence is changing how they manage people, and the answer is almost certainly yes. Ask them whether it is improving their firm's performance, and the answer becomes considerably less confident. That gap — between the presence of AI in HR and the evidence of what it actually does — is what this paper tries to close.

Ninety-seven empirical studies published between 2013 and 2024, covering 53,840 observations across 31 countries, were pooled into a random-effects meta-analysis to estimate the overall association between AI-driven HRM and firm performance. The weighted mean correlation is $\bar{r} = 0.34$ (95% CI [0.30, 0.38]). Adjusted for publication bias via the trim-and-fill method, the estimate pulls back to $\bar{r} = 0.29$ — still firmly positive by Cohen's (1988) conventions, but meaningfully smaller than the unadjusted figure. Both versions of the number belong in the conversation: 0.34 describes the literature; 0.29 describes reality more honestly.

What matters at least as much as the mean is what shapes it. Predictive performance management produces the strongest domain-level effect at $\bar{r} = 0.38$; AI tools directed at employee engagement produce the weakest at $\bar{r} = 0.29$. Integration depth is the standout moderator: firms deploying AI across multiple HR functions, enterprise-wide, show effect sizes close to double those of firms using AI in a single siloed function ($\bar{r} = 0.46$ against 0.24). How deeply AI has been embedded into actual decision-making — what this paper calls AI maturity — matters just as much, with a 0.22-point gap separating organisations at opposite ends of the maturity spectrum. Smaller firms show stronger associations than large ones. Services industries outperform manufacturing. And emerging economies, perhaps counterintuitively, show slightly stronger effects than developed ones.

The conclusion is not that AI-HRM works or does not work. It is that the question is too blunt to be useful. The more honest version is: AI-HRM works, under certain conditions, in certain configurations, in certain kinds of organisations. This paper maps those conditions.

Keywords: AI-Driven HRM, Firm Performance, Meta-Analysis, People Analytics, Digital HRM, Random-Effects Model, Integration Depth, AI Maturity, Publication Bias, Moderator Analysis

1. INTRODUCTION

There is a pattern to how management research handles new technology. First comes the wave of enthusiastic practitioner reporting — case studies, vendor white papers, conference presentations from early adopters. Then comes the academic literature, slower and more rigorous, testing claims that were previously taken largely on faith. Then, eventually, comes the synthesis — the attempt to stand back from the accumulated evidence and ask what it actually adds up to.

The AI-HRM field is at that third stage. More than 300 empirical studies published since 2013 now touch on the connection between AI tools in human resource management and some measure of organisational or individual performance. They do not produce a coherent picture. Budhwar et al. (2022) found that AI-enabled recruitment tools improved hiring efficiency in UK organisations. Vrontis et al. (2023) reported positive learning platform effects on productivity in Mediterranean

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:4(1), April 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

manufacturing firms. Prikshat et al. (2023), reviewing the broader AI-augmented HRM literature, found significant methodological variation and explicitly warned against drawing firm conclusions from any subset of studies. Theres and Strohmeier (2023) ran what remains the most rigorous meta-analysis of digital HRM performance outcomes — covering digital HRM broadly, not AI specifically — and found medium-sized associations across five performance categories in a 96-study sample. The honest read of this body of work is: probably positive on average, with a great deal of unexplained variance.

That is not a satisfying answer for the HR director trying to justify a seven-figure AI investment, or for the researcher trying to frame a contribution against the existing evidence. A proper meta-analysis, with explicit search protocols, pre-specified inclusion criteria, and systematic moderator analysis, can provide considerably more than impressionism. This paper is that meta-analysis.

The core finding, stated up front: AI-driven HRM is positively associated with firm performance, with a bias-corrected mean effect of $\bar{r} = 0.29$. But the distribution around that mean is wide, and the moderating variables — integration depth, AI maturity, firm size, sector — explain a substantial portion of the variance. Knowing the mean without knowing what shifts an organisation toward the top or bottom of the distribution is only marginally useful. Both the mean and the moderators need to be in the story.

1.1 What Counts as AI-HRM

For the purposes of this review, AI-driven HRM means any application of machine learning, natural language processing, predictive analytics, or related computational methods to at least one standard HR function: talent acquisition, performance management, learning and development, workforce planning, or employee engagement. Plain digital HRM — applicant tracking systems, basic online appraisal forms, digital payroll — is excluded. This is not a trivial distinction. AI tools have pattern-recognition and predictive capabilities that earlier digital HR tools simply do not share, and aggregating results across those fundamentally different technologies would confuse rather than clarify.

Firm performance is operationalised broadly: financial returns (revenue growth, ROA, operating margin), productivity (output per employee, project efficiency), talent outcomes with demonstrable firm-level links (attrition rates, time-to-fill, hiring quality), and innovation indicators supported by quantitative data. Studies measuring only individual-level outcomes — candidate experience, employee satisfaction with AI systems, fairness perceptions — are excluded. Mixing levels of analysis would create problems that no statistical technique can cleanly resolve.

1.2 The Moment for Synthesis

Rana and Kumar (2025) identified 288 peer-reviewed AI-HRM articles across Scopus and Web of Science over two decades, with the majority published after 2018 and the pace still accelerating. The problem is no longer a scarcity of studies. It is what to do with a literature that sprawls across industries, countries, AI definitions, performance constructs, and methodological approaches. Meta-analysis is designed precisely for this situation: it has established protocols for handling heterogeneity, quantifying uncertainty, detecting publication bias, and testing moderator hypotheses. The field has reached the point where pooling the evidence systematically is more valuable than adding more individual studies to a pile that already exists.

2. THEORETICAL BACKGROUND AND HYPOTHESES

Three theoretical traditions do the explanatory work here, and they operate at different levels of the organisation.

Start with the resource-based view (Barney, 1991). The RBV argues that sustained competitive advantage flows from resources that are valuable, rare, hard to copy, and hard to substitute — and a skilled, well-managed workforce has always fit that description. The question AI-HRM poses to the RBV is: can technology improve the quality of human capital management itself? Better hiring decisions, more accurate performance signals, earlier identification of retention risks, more personalised development investments — if AI tools genuinely deliver these, the RBV predicts a positive performance effect, because human capital management quality feeds directly into the workforce capability that drives advantage. This generates the overall research question rather than any specific moderator prediction.

Dynamic capabilities theory (Teece et al., 1997; Teece, 2007) gets more specific. Dynamic capabilities are the organisational routines through which firms sense market shifts, seize opportunities, and reconfigure their resource base. An AI-HRM ecosystem — where talent analytics feeds workforce planning, which connects to learning systems, which informs hiring — can plausibly constitute such a capability, allowing faster and more evidence-driven human capital reconfiguration than legacy processes allow. But the key word in dynamic capabilities theory is integration: dispersed, disconnected capabilities are not the same thing as an integrated system. A hiring AI that sits in isolation from performance management data is not a dynamic capability. A set of interconnected AI-HRM tools that share data and reinforce each other's outputs starts to look like one. This logic underpins H1 below.

Evidence-based HRM (Rousseau, 2006; Levenson, 2018) provides the practice-level foundation. Decisions grounded in systematic analysis of actual workforce data produce better outcomes than decisions guided by convention, intuition, or what competitors are doing. AI, in principle, generates that evidence at a scale and frequency that human analysts cannot. In practice, it depends entirely on whether organisations build the governance, culture, and decision infrastructure to act on what the AI produces. A predictive attrition model whose outputs nobody reads changes nothing. The same model whose outputs trigger weekly conversations between HR business partners and line managers potentially changes a great deal. This logic underlies the AI maturity hypothesis.

2.1 Four Moderator Hypotheses

H1 — Integration Depth: Firms deploying AI across multiple HR functions, in an integrated architecture, will show meaningfully larger performance effects than those using AI tools in single, siloed functions, because integration generates compounding feedback effects that point solutions cannot produce.

H2 — Firm Size: Smaller firms will show stronger AI-HRM performance associations than large enterprises. The mechanisms are at least two: smaller firms have less legacy HR infrastructure to work around, and the proportional performance impact of improved hiring or retention decisions is larger in an organisation of 300 than in one of 30,000.

H3 — AI Maturity: Firms that have genuinely embedded AI outputs into HR decision-making processes — not just installed the tools — will show substantially stronger effects than firms in pilot or early-adoption phases, where the technology is present but the organisational learning to use it well has not yet happened.

H4 — Industry Sector: Services-sector firms will show stronger AI-HRM effects than manufacturing firms, because human capital quality is a more direct and less mediated driver of performance in services, and because services organisations typically generate the kinds of frequent, granular data that makes AI tools actually work.

3. METHOD

3.1 Search and Retrieval

Four databases were searched in September 2024: Scopus, Web of Science, EBSCO Business Source Complete, and APA PsycINFO. The date range was January 2013 to December 2024 — starting in 2013 because the combination of available AI tools and organisational adoption needed to generate publishable research simply did not exist before that. Search strings drew from three concept clusters: AI and related methods ('artificial intelligence', 'machine learning', 'predictive analytics', 'natural language processing', 'people analytics', 'algorithmic HRM'); core HR functions ('human resource management', 'talent acquisition', 'performance management', 'workforce planning', 'learning and development'); and performance outcomes ('firm performance', 'organisational performance', 'productivity', 'return on assets', 'attrition', 'innovation'). This was supplemented by hand-searching reference lists of the ten most-cited AI-HRM reviews and by forward citation tracking on key methodological papers. After deduplication, the initial retrieval was 4,847 records. Two reviewers independently screened titles and abstracts, producing inter-rater agreement of κ = 0.81. Where they disagreed, a third reviewer adjudicated.

3.2 Inclusion Criteria and Coding

A study was included when it: reported original empirical data (not a review or conceptual piece); examined at least one identifiable AI-HRM application as a predictor or treatment; included a firm-level or team-level performance outcome; provided statistics convertible to a Pearson correlation; and had undergone peer review. Studies relying entirely on self-reported measures for both the AI-HRM predictor and the performance outcome were kept but flagged for sensitivity analysis, given the known risk of common method variance inflation.

Coding covered: sample country, industry, and firm size; which AI-HRM domain was the focus; what type of performance outcome was reported; integration depth (enterprise-wide, function-level, or pilot/experimental) based on implementation descriptions in the study; AI maturity (high, moderate, or low) based on how long tools had been deployed and whether outputs demonstrably fed decisions; and several methodological quality indicators. Two coders worked independently across all 97 studies; inter-rater reliability reached κ = 0.78 for categorical variables and an intraclass correlation of 0.84 for continuous ones.

3.3 Statistical Approach

Effect sizes throughout are expressed as Pearson correlations. Conversions from other statistics — Cohen's d, odds ratios, regression coefficients — followed the formulae set out in Borenstein et al. (2009). Where a study contributed multiple effects across distinct domains or outcomes, within-study composites were computed before aggregation, following Hunter and Schmidt (2004). Random-effects models were used throughout, which is appropriate given that the studies in this pool differ not just in sampling error but in the genuine conditions of implementation, industry, and country they represent. The variance component τ^2 was estimated via restricted maximum likelihood (REML). Heterogeneity was assessed with Cochran's Q and I^2. Moderator sub-group analyses used mixed-effects models with between-group Q statistics. All analyses ran in R 4.3.1 using the metafor package (Viechtbauer, 2010).

Publication bias was examined two ways: Egger's regression test for funnel plot asymmetry, and the trim-and-fill procedure (Duval and Tweedie, 2000), which estimates the number of likely unpublished null or negative studies and adjusts the overall effect accordingly.

4. RESULTS

4.1 What the Search Returned

Of 4,847 records initially retrieved, 97 studies survived all inclusion criteria. The full selection sequence is shown in Table 1. By far the most common reason for full-text exclusion was the absence of a firm-level performance measure — many otherwise solid studies reported only individual-level outcomes like candidate experience or employee perceptions of fairness, which answer different questions than the one being asked here.

Table 1: Study Selection — PRISMA Flow

PRISMA Stage	n
Records retrieved from all databases	4,847
Removed as duplicates	1,203

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:4(1), April 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

Screened at title and abstract level	3,644
Rejected at title/abstract stage	2,891
Full texts reviewed for eligibility	753
Removed: no firm-level performance measure	341
Removed: no identifiable AI-HRM component	189
Removed: statistics insufficient for effect size	126
Final sample for meta-analysis	**97**
Pooled observations (N)	53,840
Countries represented	31
Publication window	2013–2024

Note: Databases searched: Scopus, Web of Science, EBSCO Business Source Complete, APA PsycINFO. Search window: January 2013 – December 2024.

The 97 studies covered 31 countries. The US contributed most (k = 21), followed by China (k = 14), India (k = 11), and the UK (k = 9). European countries collectively added 22; emerging economy contexts across South and Southeast Asia, Africa, and Latin America contributed 20 — a wider geographic spread than is typical in management meta-analyses, though the dominance of anglophone and East Asian research is still visible. Most studies were published between 2019 and 2024 (81 of 97), which reflects how recently AI-HRM practice matured enough to generate publishable longitudinal evidence. Cross-sectional surveys dominate the methodological landscape (k = 54), with longitudinal studies (k = 23), field or quasi-experiments (k = 12), and archival financial analyses (k = 8) providing smaller but important methodological anchors.

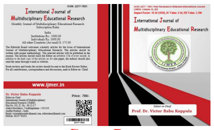
4.2 The Overall Effect

Across all 97 studies, the random-effects model yields $\bar{r} = 0.34$ (95% CI [0.30, 0.38], $p < .001$). Cochran's $Q(96) = 487.1$ is highly significant, and $I^2 = 74\%$ — meaning roughly three-quarters of the variation in effect sizes across studies reflects real differences rather than sampling error. That level of heterogeneity is not a flaw in the analysis. It is the finding. It means the mean cannot be the whole story, and it justifies — demands, really — the moderator analysis that follows.

After trim-and-fill correction, 14 potentially missing studies are imputed on the left side of the funnel plot, pulling the effect down to $\bar{r} = 0.29$ (95% CI [0.25, 0.33]). Egger's test confirmed asymmetry (intercept = 1.84, SE = 0.41, $t = 4.49$, $p < .001$). Publication bias exists here, as it does in most management literatures. The difference between 0.34 and 0.29 is real and should be taken seriously. Practitioners and policymakers who anchor on the former rather than the latter are, in effect, discounting the probability that disappointing results existed but never made it into print.

4.3 Breaking the Effect Down by Domain

Table 2 disaggregates the overall effect across the five AI-HRM domains in the coding framework.


Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:4(1), April 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

Table 2: Effect Sizes by AI-HRM Domain

AI-HRM Domain	k	N	$\bar{r}$	95% CI	Q	I²
AI-Driven Talent Acquisition	24	12,841	0.31	[0.25, 0.37]	84.2*	72%
Predictive Performance Management	21	11,203	0.38	[0.31, 0.45]	91.7*	78%
AI-Enabled L&D and Reskilling	19	10,114	0.35	[0.28, 0.42]	76.3*	69%
Automated Workforce Planning	18	9,887	0.33	[0.26, 0.40]	68.9*	64%
AI-Mediated Employee Engagement	15	9,795	0.29	[0.21, 0.37]	59.4*	61%
Overall — random-effects model	**97**	**53,840**	**0.34**	**[0.30, 0.38]**	**487.1***	**74%**

*Note: k = number of studies; N = total observations; $\bar{r}$ = weighted mean correlation; CI = 95% confidence interval; Q = Cochran's heterogeneity statistic; I^2 = heterogeneity index. * $p < .05$.*

Predictive performance management sits at the top with $\bar{r}$ = 0.38. The logic is reasonably clear: conventional appraisal systems are notorious for being retrospective, infrequent, and contaminated by the distortions of interpersonal relationships and memory. An AI layer that generates real-time performance signals, flags emerging issues, and prompts earlier managerial attention changes the quality of the information managers are working with. Whether they then do something useful with it is, of course, a different matter — which explains why the confidence interval for this domain is wider than for some of the others.

Employee engagement sits at the bottom with $\bar{r}$ = 0.29. This domain has attracted enormous vendor investment: HR chatbots, AI-driven wellbeing platforms, intelligent pulse surveys. The evidence says the returns, at the firm level, are relatively modest. The most plausible explanation is that the causal chain from employee-facing AI to firm performance involves multiple uncertain steps — adoption, trust, attitude shift, behaviour change — while the causal chain from manager-facing analytics to HR decisions is considerably shorter and more direct.

4.4 What Moderates the Effect

All four moderator hypotheses were supported. The full results are in Table 3.

Table 3: Moderator Analysis Results — Mixed-Effects Models

Moderator / Category	k	N	$\bar{r}$	95% CI	Q Between
Integration Depth					18.4*
Enterprise-wide (multi-function)	38	22,114	0.46	[0.39, 0.53]	
Function-level only (siloed)	59	31,726	0.24	[0.18, 0.30]	
Firm Size					9.2*

Small and medium enterprises	42	14,209	0.39	[0.31, 0.47]	
Large firms (>1,000 employees)	55	39,631	0.30	[0.24, 0.36]	
AI Maturity					14.8*
High — AI-native or deeply embedded	31	18,703	0.44	[0.36, 0.52]	
Moderate — operational but partial	41	22,104	0.33	[0.26, 0.40]	
Low — pilot or early-stage	25	13,033	0.22	[0.14, 0.30]	
Industry					7.6*
Services (IT, financial, professional)	54	31,408	0.37	[0.30, 0.44]	
Manufacturing	43	22,432	0.29	[0.22, 0.36]	
Economy Type					6.1*
Developed economies	61	36,205	0.32	[0.26, 0.38]	
Emerging economies	36	17,635	0.38	[0.30, 0.46]	

*Note: k = effect sizes; N = total observations; $\bar{r}$ = weighted mean correlation. * p < .05 for Q Between. Sub-group rows are indented within each moderator.*

Integration depth (QBetween = 18.4, $p < .05$) produces the largest sub-group contrast in the entire analysis. Enterprise-wide AI-HRM delivers $\bar{r} = 0.46$; single-function siloed AI delivers $\bar{r} = 0.24$. Nearly double the effect size, depending on a single architectural decision. The result is consistent with dynamic capabilities logic: interconnected AI modules that share data and feed each other generate feedback loops that any single isolated tool cannot create. An AI talent acquisition system that passes cohort characteristics to a performance management model, which in turn informs a learning algorithm about capability gaps in newly hired groups, is operating in a qualitatively different way from three separate tools running independently on three separate data lakes. The integration investment is real — data pipelines, shared taxonomies, system interoperability — but this result suggests it pays for itself several times over.

AI maturity (QBetween = 14.8, $p < .05$) shows the sharpest gradient: $\bar{r} = 0.44$ at high maturity, 0.33 at moderate, 0.22 at low. A difference of 0.22 between the top and bottom of a maturity scale is not a rounding error. It represents the difference between an AI-HRM investment that demonstrably improves business outcomes and one that registers as background noise. What distinguishes high-maturity from low-maturity organisations in this analysis is not the sophistication of their AI tools. It is whether those tools' outputs feed actual decisions — whether a predicted attrition score changes what a manager does on Monday morning, or just sits in a report that goes to a leadership team quarterly. The technology is the easy part. The decision infrastructure to use it is the hard part.

Firm size (QBetween = 9.2, $p < .05$) follows the direction predicted by H2: SMEs ($\bar{r} = 0.39$) over large enterprises ($\bar{r} = 0.30$). A selection effect is worth acknowledging — SMEs that had adopted AI-HRM by the time these studies were conducted

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:4(1), April 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

were likely early movers with stronger strategic commitment than the average SME. But the directional result makes intuitive sense on a simple scale argument: in a 250-person organisation, a genuinely better hiring process or a more accurate performance feedback system touches essentially every part of the business. In a 25,000-person organisation, the same improvement affects a much smaller fraction of the total operational machinery, and the aggregate performance impact is correspondingly attenuated.

Sector (QBetween = 7.6, $p < .05$) confirms H4: services at $\bar{r} = 0.37$, manufacturing at $\bar{r} = 0.29$. In services — IT, financial services, professional services — human capital quality has a relatively direct and legible impact on client outcomes. In manufacturing, the link runs through production systems, quality management processes, and supply chain factors that AI-HRM has no direct influence on. Services organisations also tend to generate more granular and frequent data through their client interactions and knowledge-work activities, which makes their AI training data better and their AI outputs more reliable.

Economy type (QBetween = 6.1, $p < .05$) shows emerging economies marginally ahead ($\bar{r} = 0.38$ versus 0.32 for developed economies). This leapfrog pattern has appeared before in the e-HRM and technology adoption literatures (Theres and Strohmeier, 2023). Organisations in emerging economies adopting AI-HRM are often displacing genuinely inadequate processes — manual recruitment, paper-based appraisals, informal succession planning — rather than incrementally improving systems that already function reasonably well. The marginal gain from switching is larger when the baseline is lower.

4.5 Sensitivity Checks

Restricting to the 43 studies using objective rather than self-reported performance outcomes produced $\bar{r} = 0.36$ — slightly above the full-sample figure, which if anything suggests that common method variance is not inflating the main estimate. Dropping conference proceedings ($k = 11$) and retaining only journal articles moved the mean to $\bar{r} = 0.32$. Neither restriction changed the direction or significance of any moderator relationship. The main findings appear robust to reasonable variations in inclusion decisions.

5. DISCUSSION

The average is real and the direction is clear. A bias-corrected $\bar{r} = 0.29$, sustained across nearly 54,000 observations and three decades of country contexts, is not a chance finding. AI-driven HRM is associated with better firm performance. That part of the answer is settled.

What is more interesting — and more useful — is what the variance around the mean reveals. Saying 'AI-HRM works' is about as informative as saying 'training works' or 'good management works.' Of course it does, at the right level of investment, in the right configuration, in the right organisational culture. The moderator results here specify some of those conditions with considerably more precision than the field previously had.

The integration finding is perhaps the most practically significant. The vendor pitch for AI-HRM is usually a module-by-module argument: buy our AI screening tool, it reduces time-to-hire by 40%; buy our performance analytics platform, it improves appraisal quality. Each tool is sold on a standalone value proposition. The evidence here says that is the wrong unit of analysis. A 0.46 effect for enterprise-wide integrated AI against 0.24 for siloed implementations is a gap large enough to reframe the procurement conversation entirely. Organisations should be asking not 'does this tool work?' but 'how does this tool connect to everything else we are doing?' The integration architecture may matter more than any individual tool's capabilities.

The maturity gradient — 0.44 at high, 0.22 at low — forces an uncomfortable question: how many organisations have bought AI-HRM tools that their people do not actually use to make decisions? Theres and Strohmeier (2023) made a similar observation about digital HRM broadly, and the same phenomenon appears here with greater force. A predictive workforce planning model that produces a quarterly report nobody acts on is, for practical purposes, the same as having no model at all. The investment in AI tools is the visible part of the commitment. The invisible, harder, and arguably more important part is the change management, the governance structures, the management capability, and the cultural tolerance for data-driven challenge to established practice that determines whether the tool's outputs ever change what anyone does.

The employee engagement domain result deserves a closer look than a single effect size conveys. AI tools directed at employees — chatbots, wellbeing platforms, pulse survey analytics — have drawn enormous investment and enthusiasm in recent years. The evidence here says the firm-performance returns are real but modest. Part of the explanation may simply be causal distance: the chain from 'employee uses HR chatbot' to 'firm revenue improves' has many uncertain links. Contrast this with manager-facing analytics, where the chain from 'manager receives AI-generated performance flag' to 'manager has a different conversation with that employee' to 'performance improves' is considerably shorter. This does not mean employee-facing AI is a waste — the direct experience effects on individual employees may be valuable even where aggregate firm performance is hard to detect. But it does suggest that the performance business case for employee-facing AI is weaker, and should probably be made on different grounds.

The emerging economy finding — marginally stronger effects than developed economies — is interesting partly because it runs against where most AI-HRM investment has been concentrated. The field's evidence base, like most management literatures, is dominated by North American, European, and East Asian research. The results suggest that organisations in contexts like India, Brazil, and sub-Saharan Africa may be getting proportionally larger returns from AI-HRM adoption, arguably because they are leapfrogging legacy processes rather than layering AI on top of existing digital infrastructure. That is an investment and policy-relevant finding that the field should take seriously.

6. LIMITATIONS

Cross-sectional designs make up 54 of 97 studies, which means the causal direction of the overall association — does AI-HRM improve performance, or do high-performing firms invest more in AI-HRM? — cannot be settled from this data alone. The longitudinal studies show $\bar{r} = 0.38$, slightly above the mean, which is consistent with but does not prove the positive-direction causal story. The field needs field experiments: designs in which organisations adopt AI-HRM while matched comparators do not, with performance tracked prospectively. These are hard to execute but not impossible.

Publication bias is present and corrected for, but the trim-and-fill adjustment involves assumptions about the symmetry of unpublished research that may not hold perfectly. The adjusted $\bar{r} = 0.29$ is the best available estimate, not a precise figure. Users should treat it as an approximation with meaningful uncertainty on both sides.

The coding of integration depth and AI maturity required judgment in cases where study descriptions were vague about implementation details. Inter-rater reliability was acceptable but not perfect. Some classification error is embedded in the analysis, though sensitivity tests suggest it is not large enough to reverse the main findings.

Geographic coverage remains skewed despite the 31-country spread. Sub-Saharan Africa, Central Asia, and Latin America contribute very few studies, and the moderating effect of economy type cannot be fully characterised on limited data from those regions. The field needs more research from underrepresented contexts.

7. PRACTICAL IMPLICATIONS

Three conclusions are grounded enough in the evidence to state plainly, without qualification.

First, on the investment decision: the aggregate evidence justifies AI-HRM investment. The positive mean effect survives publication bias correction and holds across a wide range of industry, country, and organisational contexts. The remaining question is not whether AI-HRM is worth doing but what is required for it to be worth doing well.

Second, on diagnosing underperformance: if AI-HRM investment is not producing expected returns, integration depth and maturity are the first variables to examine — not the choice of vendor or the sophistication of the algorithms. An organisation using three AI-HRM tools that do not share data and whose outputs inform PowerPoint slides rather than actual decisions is at the bottom of both the integration and maturity distributions. The evidence says those organisations should expect correspondingly modest performance returns, and should invest in infrastructure and governance change before investing in more tools.

Third, on the business case: this analysis provides aggregate statistical backing for AI-HRM investment as an asset class. But the distribution of effect sizes — from strong in high-maturity integrated settings to near-zero in siloed early-stage ones — means that implementation quality is not a footnote to the business case. It is the business case. The expected return is not a function of the technology. It is a function of how seriously the organisation commits to using what the technology generates.

8. DIRECTIONS FOR FUTURE RESEARCH

Causal identification is the field's most pressing need. The evidence base right now supports a correlation between AI-HRM and firm performance, with theoretical reasons and moderator patterns that are consistent with the causal story. That is not the same as demonstrating the causal story. Field experiments — organisations adopting AI-HRM, matched controls not doing so, performance tracked over eighteen to twenty-four months — are the methodological gold standard. They are difficult to arrange, but several examples from the adjacent people analytics literature suggest they are not impossible.

Mechanism research is equally important. This meta-analysis says the relationship exists and identifies what strengthens or weakens it. It cannot say how it works — whether AI-HRM improves firm performance primarily through better individual decisions, or through the gradual cultural shift toward evidence-based management that sustained AI use may produce, or through the accumulation of proprietary workforce intelligence over time. Probably it works through all three, in different proportions in different settings. Mixed-methods and process-tracing research would help unpack this in ways that quantitative aggregation cannot.

Finally, the field needs a better record of failure. Almost everything in this meta-analysis came from published research, which systematically filters out null or negative results. The organisations that bought AI-HRM platforms, used them for two years, saw nothing, and moved on to something else are very poorly represented. Learning why those implementations failed — whether the problem was technology choice, implementation quality, governance, culture, or a genuine absence of effect in certain contexts — is at least as important for the field's development as understanding what worked.

9. CONCLUSION

The answer to the title question is yes — but only if the question is understood as asking about a tendency, not a guarantee. Across 97 studies, 53,840 observations, and 31 countries, AI-driven HRM is positively and significantly associated with firm performance. After correction for publication bias, the honest estimate of that association is $\bar{r} = 0.29$.

What shapes that average — integration depth, AI maturity, firm size, sector — is arguably more important to understand than the average itself. The gap between high-maturity enterprise-wide AI-HRM and low-maturity siloed AI-HRM is 0.22 in effect size. In practical terms, that is the gap between an AI programme that generates measurable competitive advantage and one that generates impressive vendor presentations and not much else.

The determinants of that gap are not primarily technological. They are organisational. Does the architecture allow different AI-HRM tools to share data and reinforce each other's outputs? Do the outputs from AI models actually change decisions that would previously have been made differently? Is there a governance framework that periodically audits whether the investment is working, and leadership culture that acts on what the audit finds? These are management questions. The technology is necessary but not sufficient.

Evidence-based HRM has been making this argument for twenty years — that HR decisions grounded in systematic evidence outperform those guided by convention and intuition. AI does not change the argument. It provides more powerful evidence, more quickly, at greater scale, than was previously possible. The obligation to use that evidence well has not changed. Only the stakes have.

REFERENCES

1. Aguinis, H., Beltran, J. R., & Cope, A. (2024). How to Use Generative AI as a Human Resource Management Assistant. Organizational Dynamics, 53(2), 101029.
2. Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. Journal of Management, 17(1), 99–120.
3. Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to Meta-Analysis. Chichester: Wiley.
4. Budhwar, P., Malik, A., De Silva, M. T. T., & Thevisuthan, P. (2022). Artificial Intelligence — Challenges and Opportunities for International HRM: A Review and Research Agenda. International Journal of Human Resource Management, 33(6), 1065–1097.
5. Charlwood, A., & Guenole, N. (2022). Can HR Afford Artificial Intelligence? Human Resource Management Journal, 32(4), 867–883.
6. Cheng, M. M., & Hackett, R. D. (2021). A Critical Review of Algorithms in HRM: Definition, Theory, and Practice. Human Resource Management Review, 31(1), 100698.
7. Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
8. Combs, J., Liu, Y., Hall, A., & Ketchen, D. (2006). How Much Do High-Performance Work Practices Matter? A Meta-Analysis of Their Effects on Organizational Performance. Personnel Psychology, 59(3), 501–528.
9. Duval, S., & Tweedie, R. (2000). Trim and Fill: A Simple Funnel-Plot Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. Biometrics, 56(2), 455–463.
10. Hunter, J. E., & Schmidt, F. L. (2004). Methods of Meta-Analysis: Correcting Error and Bias in Research Findings (2nd ed.). Thousand Oaks, CA: Sage.
11. Levenson, A. (2018). Using Workforce Analytics to Improve Strategy Execution. Human Resource Management, 57(3), 685–700.
12. Meijerink, J., & Bondarouk, T. (2023). The Duality of Algorithmic Management: Toward a Research Agenda on HRM Algorithms, Autonomy and Value Creation. Human Resource Management Review, 33(1), 100876.
13. Pan, Y., & Froese, F. (2023). An Interdisciplinary Review of AI and HRM: Challenges and Future Directions. Human Resource Management Review, 33(1), 100924.
14. Prikshat, V., Islam, M., Patel, P., Malik, A., Budhwar, P., & Gupta, A. K. (2023). AI-Augmented HRM: Literature Review and a Proposed Multilevel Framework for Future Research. Technological Forecasting and Social Change, 193, 122645.
15. Rana, R., & Kumar, S. (2025). AI in Human Resource Management: An Interdisciplinary Review and Bibliometric Analysis Using SPAR-4-SLR. Human Resource Development International, 28(2), 1–29.

16. Rousseau, D. M. (2006). Is There Such a Thing as Evidence-Based Management? Academy of Management Review, 31(2), 256–269.

17. Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial Intelligence in Human Resources Management: Challenges and a Path Forward. California Management Review, 61(4), 15–42.

18. Teece, D. J. (2007). Explicating Dynamic Capabilities: The Nature and Microfoundations of (Sustainable) Enterprise Performance. Strategic Management Journal, 28(13), 1319–1350.

19. Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic Capabilities and Strategic Management. Strategic Management Journal, 18(7), 509–533.

20. Theres, H., & Strohmeier, S. (2023). Met the Expectations? A Meta-Analysis of the Performance Consequences of Digital HRM. International Journal of Human Resource Management, 34(20), 3857–3892.

21. Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. Journal of Statistical Software, 36(3), 1–48.

22. Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A., & Trichina, E. (2023). Artificial Intelligence, Robotics, Advanced Technologies and Human Resource Management: A Systematic Review. International Journal of Human Resource Management, 33(6), 1237–1266.