International Journal of Education, Vocational and Social Science
https://e-journal.citakonsulindo.or.id/index.php/IJEVSS
E-ISSN : 2963 4954
IJEVSS

Assessing The Limits Of Algorithms: Reorienting Human Resource Management Policies For Mitigating Bias In AI-Based Performance Evaluation

Fajar Zainulloh
Universitas KH. Mukhtar Syafaat Banyuwangi
Siti Aimah
Universitas KH. Mukhtar Syafaat Banyuwangi
Wafi Bahrul Ilmi
Universitas KH. Mukhtar Syafaat Banyuwangi
Moh Ainul Yaqin
Universitas KH. Mukhtar Syafaat Banyuwangi

Corresponding Author: *fajarzainulloh95@gmail.com*

Abstract

This study aims to analyze the urgency of reorienting Human Resource Management (HRM) policies and updating operational procedures to mitigate the risks of bias generated by Artificial Intelligence (AI)-based performance evaluation systems. Employing a qualitative-descriptive approach, this study combines a literature review and field research through in-depth interviews to examine regulatory frameworks, organizational internal policies, and case studies concerning the implementation of performance evaluation algorithms in the digital industry sector. The findings reveal that the use of AI in performance evaluation systems has significant potential to generate historical data bias, algorithmic unfairness toward minority groups, and the erosion of human empathy in competency assessments. Therefore, the reorientation of HRM policies is imperative through the formulation of multilayered supervisory mechanisms (Human-in-the-Loop), periodic algorithm audits, and transparency in performance evaluation metrics for employees. Theoretically, this study contributes to enriching the body of knowledge in strategic human resource management, particularly in the area of technology governance in the era of digital disruption. Practically, it provides strategic guidance and concrete policy recommendations for HR practitioners and policymakers in designing performance evaluation standard operating procedures (SOPs) that are not only technologically efficient but also uphold the principles of

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

fairness, accountability, and organizational humanism.

Keywords: Human Resource Management Policy, Artificial Intelligence, Algorithmic Bias, AI-Based Performance Evaluation.

INTRODUCTION

The phenomenon of historical bias embedded in raw data is often unintentionally replicated by Artificial Intelligence (AI) systems in evaluating employee performance. The primary cause of this condition is that AI systems are trained using companies' historical data, which are frequently contaminated by subjective human prejudices related to gender, race, or particular backgrounds .(Wibowo & Kom, 2025). Empirical evidence can be observed across various global industries, where performance evaluation algorithms repeatedly assign lower scores to minority groups or women for leadership roles due to their limited historical representation. Consequently, the adoption of AI without ethical data curation merely prolongs structural discrimination in modern workplaces, thereby necessitating a rigorous reorientation of Human Resource Management (HRM) policies to eliminate such latent biases (Nasution et al., 2025)

Another emerging phenomenon is technostress and widespread anxiety among employees resulting from constant and inflexible algorithmic surveillance. The psychological pressure underlying this phenomenon stems from AI's tendency to measure productivity solely through quantitative indicators, such as mouse-click duration or other digital metrics, without understanding the qualitative dynamics of human labor(Paredes et al., 2026; Quiel et al., 2026). Empirical evidence indicates that many employees in the technology and customer service sectors experience severe mental exhaustion (*burnout*) because they feel continuously monitored by a "robot boss" that lacks empathy. This reality suggests that excessively mechanistic AI-based performance assessment undermines employees' mental well-being, thereby compelling Human Resource Management to establish clear limitations on the scope of surveillance in order to safeguard employee privacy and psychological health (Lubis & Nurhayati, 2026)

This issue is also associated with the devaluation or diminishing significance of interpersonal skills (*soft skills*) within automated performance assessment ecosystems. The primary reason lies in the inherent limitations of algorithms in measuring crucial yet abstract contributions, such as collaboration, empathy, conflict resolution, and emotional intelligence. For instance, many employees recognized as key contributors to team cohesion or exceptional mentors receive poor performance evaluations from AI systems merely because their individual

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

quantitative targets fall slightly below algorithmic standards. Consequently, excessive dependence on AI generates evaluative biases that undermine collaborative workplace cultures, making it imperative for HRM policies to be reoriented toward restoring human-centered approaches to performance assessment (Febriana, 2025).

The implementation of AI-driven performance assessment systems has also led to collective resistance and a substantial decline in employees' trust in the objectivity of the institutions in which they work. Such social resistance originates from the opaque nature of algorithms, which frequently operate as "black boxes," rendering their decision-making processes and performance evaluations inaccessible and indisputable to employees(Sadowski & Bryda, 2025; George, 2026). Empirical evidence demonstrates waves of protests, declining employee loyalty, and even legal actions initiated by labor unions affected by inaccurate AI-based decisions concerning layoffs or bonus withholding. This crisis of trust highlights the indispensable role of algorithmic transparency and underscores the necessity for HRM policies that guarantee human-managed appeal mechanisms to preserve organizational justice.

Furthermore, the adoption of AI technologies has created new forms of digital and competency gaps among HR managers themselves in operating these evaluation technologies(El Maknouzi, 2022; Dutta & Srinivasan, 2026). This disparity exists because not all HR practitioners possess adequate data literacy to interrogate, configure, or challenge the performance assessment outcomes generated by AI. Sociological evidence reveals that many middle managers unquestioningly accept AI-generated recommendations (*automation bias*) due to their limited technical understanding, despite the possibility that such recommendations may be inaccurate. Therefore, AI bias mitigation efforts will remain ineffective without substantial improvements in technological competence and digital literacy among HR professionals as part of broader policy reorientation aimed at ensuring that human control remains the ultimate authority(Archana, 2025; Kuzmanov, 2025; Marchena-Sekli & Godo, 2026).

The literature on bias in artificial intelligence systems demonstrates a broad academic consensus that algorithms are not inherently neutral; rather, they reflect the values, preferences, and prejudices embedded within their training data and designers' decisions. Various sociological studies of technology affirm that when AI is implemented in human resource management, it risks perpetuating historical injustices through what is known as "bias laundering," whereby hidden discrimination becomes legitimized under the guise of mathematical objectivity. Contemporary HRM literature further indicates that excessive reliance on quantitative metrics without critical human oversight frequently overlooks the qualitative contexts of performance, ultimately

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

contributing to the degradation of organizational ethics and complicating relationships between employees and institutional authorities.

Conversely, HRM policy frameworks within the literature on AI governance emphasize the urgency of transitioning toward the *Human-in-the-Loop* model to mitigate accountability risks inherent in black-box systems. Empirical studies reveal that the adoption of transparent policies that allow human review mechanisms for algorithmic performance evaluations significantly enhances employee trust while reducing the adverse consequences of systemic bias. Therefore, the literature suggests that HRM policy reorientation should not merely prioritize technical efficiency but must instead focus on integrating the principles of procedural justice and comprehensive digital literacy to ensure that organizations can balance machine autonomy with essential human moral judgment(Setiawan and Lely A., 2025; Suaib et al., 2026).

The novelty of this study lies not merely in acknowledging the existence of AI bias but in proposing a blueprint for HRM policy reorientation that integrates participatory algorithmic auditing with the principles of procedural justice through the *Human-in-the-Loop* approach. Unlike previous studies that predominantly emphasize technical improvements in programming codes, this analysis conceptualizes bias mitigation as a restructuring of socio-organizational governance aimed at balancing machine efficiency with the protection of employees' psychological rights. Accordingly, this study offers a novel adaptive framework for HR practitioners in determining the appropriate boundaries of algorithmic intervention while simultaneously redefining the ethical limits of artificial intelligence utilization in human performance assessment.

This study is particularly significant because the widespread adoption of AI in performance assessment systems frequently compromises organizational justice and employee mental well-being due to uncontrolled algorithmic bias. Without an immediate reorientation of HRM policies, organizations risk facing internal crises of trust, substantial declines in productivity, and legal challenges arising from technology-legitimized systemic discrimination.

RESEARCH METHOD

This study employed a qualitative-descriptive approach by combining a literature review and field research through in-depth interview techniques. It examined various regulatory frameworks, organizational internal policies, and case studies concerning the implementation of performance evaluation algorithms in the digital industry sector. Previous findings have indicated that the use of AI in performance evaluation systems has significant potential to generate historical data bias, algorithmic unfairness toward minority groups, and the erosion of human

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

empathy in competency assessments(Khotib et al., 2026; Setiawan, et al., 2025). Therefore, the reorientation of Human Resource Management (HRM) policies is imperative through the formulation of multilayered supervisory mechanisms (Human-in-the-Loop), periodic algorithm audits, and transparency in performance evaluation metrics for employees. Pondok Pesantren Minhajut Thullab was selected as the research site because the integration of its administrative technology poses the risk of reducing the qualitative value of ustadz service and dedication into quantitative metrics, thereby necessitating the reorientation of HRM policies to mitigate such automated evaluation biases (RITA, 2026) .

The literature review was conducted at the initial stage to collect secondary data from reputable journal articles, management textbooks, and AI governance regulations in order to establish a robust theoretical foundation concerning algorithmic bias and the mitigation of ethical risks. Meanwhile, in-depth interviews were employed as the primary instrument for collecting primary data at Pondok Pesantren Minhajut Thullab, South Konawe. The interviews were conducted directly with purposively selected informants, including the leaders of the Islamic boarding school, HR managers, and representatives of the teaching staff, to capture the sociological realities and cultural dynamics associated with the implementation of digital performance evaluation systems in practice.

To obtain in-depth, objective, and comprehensive data regarding the limitations of automation systems and the direction of educational workforce governance policies within the institution, the research involved key informants who possess relevant experiences, responsibilities, and insights related to policy formulation, digital system management, and the implementation of performance evaluation practices (Arifianto et al., 2025; Azhari et al., 2025). These informants were selected based on their roles and direct involvement in the institutional context to provide a comprehensive understanding of the sociocultural dynamics surrounding the performance evaluation system.

Tabel.1 Informan Penelitian

No.	Research Informants	Number	Rationale for Informant Selection
1	Pesantren Caretaker/Leader	1 Person	Holds the highest authority in determining the direction of Human Resource Management (HRM) policy reorientation and ensuring that AI systems do not undermine the noble values of the pesantren.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

2	Head of the Pesantren Human Resources Division	1 Person	Responsible for implementing performance evaluations and the party that most requires digital literacy to mitigate biases arising from algorithmic calculations.
3	Information System Operator / Pesantren IT Administrator	1 Person	The technical administrator who understands the operation, data inputs, and parameters of the evaluation algorithms, while also identifying potential historical data biases.
4	Head of Madrasah or School Principal	1 Person	The direct evaluator in the field who understands the discrepancies between AI's quantitative indicators and the qualitative competencies demonstrated by teachers.
5	Representative of Senior *Ustadz* or Educators	1 Person	The affected party who can provide critical perspectives regarding system transparency, *technostress*, and the devaluation of service and devotion (*khidmah*) resulting from automated evaluations.
6	*Santri* (Islamic Boarding School Students)	1 Person	A stakeholder group that closely interacts with technology and can provide insights into the level of acceptance or resistance toward the accuracy of AI-based performance evaluations.

Data were analyzed interactively by integrating conceptual findings derived from the literature with empirical data obtained from the interviews. All interview recordings were transcribed verbatim and subsequently reduced and classified into thematic clusters, including the devaluation of service and dedication (*khidmah*), *automation bias*, and *technostress*. The researcher employed source triangulation techniques by comparing findings from the global literature with the oral testimonies of pesantren stakeholders to examine the validity of the data. The final synthesis of this content analysis served to contextualize the theory of algorithmic limitations within the landscape of traditional Islamic education, ultimately generating adaptive Human Resource Management (HRM) policy recommendations grounded in the *Human-in-the-Loop* approach that are both relevant and equitable for the pesantren ecosystem.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

International Journal of Education, Vocational and Social Science
https://e-journal.citakonsulindo.or.id/index.php/IJEVSS
E-ISSN : 2963 4954
IJEVSS

RESULTS AND DISCUSSION

The Devaluation of Religious *Khidmah* Due to the Limitations of Quantitative Algorithmic Metrics

The implementation of Artificial Intelligence (AI)-based performance evaluation technology at Pondok Pesantren Minhajut Thullab, South Konawe Regency, has brought significant sociological and structural implications for the Islamic educational ecosystem. As an institution that has historically upheld traditional values, spirituality, and religious morality, the digitalization of its administrative system has created a new dialectical space that brings together the efficiency of modernity and the preservation of local wisdom. Rigid digital transformation often compels the pesantren's social institutions, which were originally fluid and family-oriented in nature, to adapt to mechanistic computerized bureaucratic systems. Sociological tensions emerge when AI instruments begin to assume managerial roles in measuring individual productivity, ultimately influencing the patterns of cultural relationships among institutional leaders, human resource managers, and *ustadz*. This dialectic requires institutions not merely to adopt technology uncritically but rather to critically assess the extent to which machine autonomy may intervene in human-centered evaluations that are deeply embedded with spiritual values(Aly, 2025).

The findings at Pondok Pesantren Minhajut Thullab reveal a sharp conceptual conflict between the operation of Artificial Intelligence, which relies on formal quantitative indicators, and the principle of *khidmah*, or the sacred devotion and service embodied by *ustadz*. AI systems rigidly record educators' performance through superficial digital variables, such as punctuality in application-based attendance log-ins, the completeness of classroom administrative documents, and the duration of formal face-to-face teaching sessions. In contrast, the essence of pesantren-based service, which is qualitative, sincere, and extends beyond official working hours, is entirely overlooked by the mathematical calculations of soulless machines. The reductionist characteristics of algorithms fail to translate the spiritual values of teaching into binary data, thereby generating evaluative biases that disadvantage dedicated educators. Consequently, *ustadz* who devote their lives to serving the pesantren beyond formal working hours risk receiving poor performance evaluations merely because of technical negligence when interacting with digital application interfaces.

The evaluative bias resulting from the limitations of algorithmic detection directly contributes to the devaluation of the religious and cultural activities that constitute the lifeblood of Pondok Pesantren Minhajut Thullab. Crucial activities, such as accompanying students in

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

congregational dawn prayers, supervising Qur'anic memorization sessions until late at night, providing personal spiritual counseling in the dormitories, and participating in religious educational activities within the South Konawe community, are regarded as nonexistent by the system because they leave no digital footprints. The algorithm excludes these informal contributions from the calculation of the Key Performance Indicators (KPIs), thereby creating artificial and inequitable standards of performance. This condition generates social anxiety among educators with strong cultural and spiritual commitments who are systematically marginalized. Conversely, the system rewards pragmatic and opportunistic educators who merely fulfill formal administrative obligations punctually through digital applications without demonstrating profound emotional and spiritual commitment to the moral development of the students residing within the pesantren.

This socio-technological conflict underscores the urgency of radically reorienting Human Resource Management policies to preserve the noble values of the pesantren from algorithmic dehumanization. The inability of Artificial Intelligence to capture human realities and the values of *khidmah* demonstrates that HR governance within religious institutions must not be entirely delegated to the absolute autonomy of evaluative machines. This institution in South Konawe requires policy restructuring through a hybrid governance approach that restores humans as the ultimate decision-makers (*Human-in-the-Loop*). Under this policy framework, the quantitative outputs generated by AI are no longer positioned as binding final decisions but merely serve as secondary data or instruments to enhance administrative efficiency. The authority to determine final evaluation policies must be returned to the deliberative process of the Board of Caretakers (*Dewan Pengasuh*), ensuring that moral considerations, humanity, service loyalty, and character-based evaluations remain the principal compass in safeguarding organizational justice while preserving the spiritual dignity of the pesantren.

This finding is reinforced by the statement of one senior educator during the interview session:

> *"Honestly, we feel that this new system is somewhat rigid. When the evaluation report was released yesterday, my score decreased simply because I was frequently late in clicking the attendance application at 7 a.m. In fact, I had been at the mosque since 4 a.m. accompanying students during their Qur'anic memorization sessions and then continued my dormitory cleanliness duty until after 6 a.m. The system does not recognize that we teach Qur'anic studies at night or deliver religious lectures for the people of South Konawe*

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

outside official working hours. The values of sincerity and service (khidmah) cultivated over decades in the pesantren seem to have been defeated by the rules imposed by machines."

The interview findings demonstrate that AI-based performance evaluation algorithms generate biases that disadvantage dedicated educators. The system disregards the essence of qualitative *khidmah*, such as dawn spiritual guidance and community service, because it exclusively measures rigid quantitative metrics, including digital attendance records. Consequently, it contributes to the devaluation of the values of sincerity and service that have long characterized the pesantren tradition.

The Phenomenon of Automation Bias and the Simplification of Verification at the Managerial Level

This study identifies the phenomenon of *automation bias* as both tangible and alarming among middle-level management within Pondok Pesantren Darussalam Minhajut Thullab. This psychological and sociological phenomenon particularly affects operational policymakers, especially those within the Human Resource Management division and the heads of madrasah who serve as direct evaluators in practice. Automation bias manifests itself in managers' tendency to blindly trust every output, decision, or recommendation generated by Artificial Intelligence systems. Middle-level management positions this technology as an infallible entity incapable of committing either technical or conceptual errors in its calculations. Consequently, passive acceptance has become widespread, whereby digital dashboards are no longer treated merely as instruments for administrative efficiency but have instead evolved into a new idol that exercises complete control over the professional destinies of *ustadz* (Hijazi, 2025).

The limited understanding of algorithmic mechanisms, programming logic, and the ways in which historical data are processed has trapped managers within an oversimplified and naïve perspective. They assume that recommendations generated by Artificial Intelligence algorithms represent absolute truths that are entirely objective, scientific, and free from human subjectivity. Management fails to recognize that Artificial Intelligence is fundamentally nothing more than a mathematical mirror that processes historical raw data, which remains highly susceptible to carrying inherent structural biases embedded by its designers. The technical inability to interrogate systems or identify data anomalies leaves managerial personnel feeling inferior before Artificial Intelligence, prompting them to adopt the expedient approach of unquestioningly accepting whatever numerical outputs are displayed on computer monitors (Rustiyana et al.,

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

2025).

The most destructive consequence of the dominance of automation bias is the complete paralysis of the social control and managerial functions that should inherently characterize effective leadership. Due to their excessive glorification of technological reliability, administrators tend to execute AI-generated performance scores immediately without undertaking any process of reconfirmation (*cross-checking*) regarding the actual circumstances experienced by teachers in the field (Raprap et al., 2025).



Figure 1. The Dominance of Automation Bias Erodes Human Leadership
Source: Processed by the Author, 2026

The figure illustrates the destructive impact of automation bias in undermining human leadership. Dependence on AI and blind compliance result in inadequate verification processes, the surrender of moral sovereignty, and the perpetuation of injustice, ultimately leading to the paralysis of management's social control functions.

Ultimately, this simplification and delegation of evaluative autonomy trigger a crisis of organizational justice that undermines the internal cohesion of Pondok Pesantren Minhajut Thullab. Personnel-related decisions based solely on rigid binary calculations without contextual field verification gradually create deep divisions and distrust between educators and administrators. The ustadz perceive that their professional rights and cultural service are being evaluated unfairly by a management system that shelters itself behind the façade of machine objectivity.

This blind dependence was openly acknowledged by the Head of the Human Resources Division of the pesantren, Mr. Priyadi:

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

"We in the Human Resources Division are actually in a difficult position. When the pesantren began implementing the AI-based automated evaluation system introduced by the central administration, we regarded it as highly sophisticated and unquestionably 100% objective, free from human favoritism. Therefore, whenever the system issued recommendations for poor performance reports or incentive reductions for particular teachers, we immediately approved and forwarded them to the leadership without conducting any further verification in the classroom. We felt that we lacked the technical foundation necessary to challenge the calculations produced by the computer system."

The interview findings confirm the occurrence of automation bias resulting from the limited digital literacy of HR management personnel. Administrators perceive AI-generated recommendations as absolutely objective, thereby paralyzing their capacity for human-centered verification. Consequently, crucial decisions, such as reductions in teachers' incentives, are executed unquestioningly without field verification because management personnel feel technically inferior and are reluctant to challenge the calculations generated by computer systems.

The Manifestation of Technostress and the Transformation of Cultural Relationships Among Traditional Educators

The psychological pressure to continuously adapt to inflexible algorithmic performance evaluation systems has triggered widespread manifestations of technostress among educators at Pondok Pesantren Minhajut Thullab. This phenomenon has significantly affected the emotional and psychological well-being of senior ustadz who have devoted decades of service through conventional educational approaches. Within this traditional Islamic educational institution, technostress manifests in the form of persistent digital anxiety, feelings of helplessness in responding to software updates, and mental exhaustion resulting from the rigid demands of fulfilling administrative performance targets .(Andryan & Suminar, 2024).

The transformation of supervisory models from a personal and familial approach grounded in exemplary leadership (uswah) to constant digital surveillance has generated acute psychological anxiety concerning the stability of educators' professional roles. Within the traditional pesantren ecosystem, the performance evaluation of educators has historically been conducted by the Board of Caretakers (Dewan Pengasuh) through direct observation characterized by empathy, wisdom, and familial dialogue. However, when such supervisory authority is replaced by real-time AI-driven

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

algorithmic monitoring that operates as a form of digital panopticon, ustadz increasingly perceive that their pedagogical privacy has been invaded. They are haunted by persistent anxiety that even minor errors in interacting with digital applications may be immediately interpreted as a decline in their professional performance. This radical transformation from a management model based on trust and cultural respect to one driven by suspicion and digital metrics has gradually eroded the sense of job security that they have long experienced.

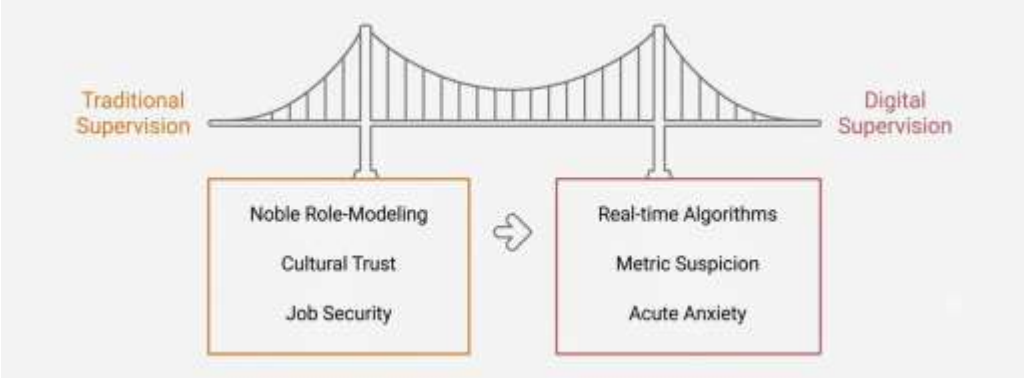


Figure 2. The Transformation of the Supervisory Model of Pesantren Educators
Sumber : Data olahan peneliti 2026

The figure illustrates the transformation of the supervisory model of pesantren educators. The transition from a traditional model grounded in exemplary leadership and cultural trust to digital surveillance characterized by real-time algorithms and metric-based suspicion has resulted in the emergence of acute psychological anxiety among educators.

The most concerning social consequence of the rigid requirement to interact continuously with digital interfaces is the erosion of the traditional emotional closeness between educators and students (santri). The time and energy that ustadz once devoted entirely to accompanying the spiritual development of santri beyond formal classroom hours are now increasingly consumed by the demands of completing administrative reports through digital devices. Noble cultural practices, such as engaging in heartfelt conversations on the dormitory terrace, providing personalized moral guidance, and demonstrating exemplary conduct through daily activities, have gradually diminished in their significance. These humanistic spaces of interaction, deeply embedded with the values of spiritual nurturing, are now being replaced by the mechanical routines of completing digital forms (Purnamasari, 2025).

Ultimately, the social estrangement resulting from technological constraints has extended

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

to erode the bonds of collective solidarity among fellow educators. The teachers' lounge, which once served as a warm space for exchanging ideas, discussing the moral development of santri, and strengthening ukhuwah islamiyah (Islamic brotherhood), has gradually become quieter as educators increasingly focus on their respective digital screens. The emphasis on digital competencies and individualized AI performance scores has inadvertently created a climate of artificial competition that undermines the values of togetherness that have long characterized Pondok Pesantren Minhajut Thullab.

The principal described these changes in workplace social dynamics as follows:

"The atmosphere in the teachers' room has changed somewhat since digital metric-based supervision became more stringent. Several senior ustadz, particularly the elderly ones, have shared with me that they feel stressed, pressured, and constantly monitored by a 'robot boss.' They experience tremendous anxiety whenever they encounter difficulties uploading their teaching administration documents or when the application encounters an error, fearing that the system will immediately regard them as incompetent. The familial relationships that once characterized this pesantren now feel increasingly mechanical and distant. Educators have become more preoccupied with staring at their digital screens and worrying about achieving their quantitative performance targets."

The interview findings highlight the manifestation of technostress, whereby digital metric-based surveillance generates psychological pressure among senior ustadz due to their fear of system errors and technological inadequacy. Socially, the traditional familial relationships within the pesantren have gradually shifted toward a more mechanical and impersonal environment as educators become increasingly oriented toward fulfilling quantitative performance targets through digital devices.

The Urgency of Restructuring Human Resource Management Governance Through Human-in-the-Loop Policies

The final findings emphasize that the most appropriate solution to mitigating biases in Artificial Intelligence (AI)-based performance evaluation systems is not to reject or entirely discontinue the adoption of technology but rather to undertake a radical reorientation of Human Resource Management (HRM) policies through the implementation of the Human-in-the-Loop governance model. Amid the rapid advancement of digitalization, Pondok Pesantren Darussalam Minhajut Thullab in South Konawe Regency faces an urgent need to restructure its evaluation

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

system to avoid becoming entrapped in rigid machine autonomy. The implementation of this hybrid model is designed to bridge the demands of modern managerial efficiency with the obligation to preserve the dignity (muruah) and local wisdom of the pesantren, which have been deeply rooted in Southeast Sulawesi for decades (Yazid, 2025).

In principle, this policy reorientation restores human authority represented within this sociocultural context by the Board of Caretakers (Dewan Pengasuh) and the pesantren leadership as the ultimate decision-makers within the organizational hierarchy. In traditional Islamic educational institutions, the figures of the kiai and the Board of Caretakers serve as spiritual pillars endowed with profound emotional sensitivity and moral intuition qualities that can never be possessed by lines of Artificial Intelligence programming code. By returning managerial sovereignty to the hands of the caretakers, the pesantren can ensure that personnel evaluation processes do not lose their humanistic essence or devolve into a cold, mechanistic, and binary bureaucracy that disregards the sociological dimensions of ustadz service and dedication (Samsu, 2025).

Within this new governance framework, the quantitative analyses and performance scores generated by AI are positioned solely as supporting instruments or secondary sources of data. Artificial Intelligence no longer possesses absolute authority to render final and procedurally binding decisions; instead, its function is redefined as a comparative dashboard for monitoring administrative trends in educators' teaching activities. This redefinition of technology's role is particularly crucial for mitigating the risks of automation bias, which has long undermined the verification functions of personnel administrators. Consequently, the system is prohibited from automatically imposing sanctions, reducing incentive-related benefits, or assigning poor performance evaluations without the approval of human supervisory authorities.

At the final stage of the evaluation process, every recommendation generated by Artificial Intelligence concerning the professional future of educators must undergo collective moral scrutiny, human-centered considerations, and character-based assessments. Strategic decisions, including promotions, awards, and disciplinary sanctions, are returned to deliberative mechanisms directly led by the Board of Caretakers. This pesantren-specific Human-in-the-Loop approach ensures that the cultural values of service, sincerity in professional commitment, and the spiritual loyalty (khidmah) of ustadz, which remain undetectable by digital applications, continue to receive fair recognition. Consequently, both organizational justice and the sacred dignity of Islamic educational institutions can be sustainably preserved.

The Head of the Pondok Pesantren emphasized the policy reorientation currently being

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

implemented by the institution as follows:

> *"Technology is an excellent instrument for achieving efficiency, but computer algorithms possess neither a soul nor empathy, nor do they understand what constitutes moral character (akhlak). Therefore, the reorientation of policies at Minhajut Thullab is absolutely imperative. I have instructed the Human Resources Division and IT operators that AI-generated evaluation results must never be considered final decisions. The system merely provides comparative and secondary data for our consideration. The final decisions concerning the welfare, professional future, or disciplinary measures of our ustadz must always be determined through deliberation by the Board of Caretakers. We must uphold the pesantren-specific principle of Human-in-the-Loop."*

The interview findings confirm the institution's rejection of granting absolute autonomy to AI in pesantren-based performance evaluations. The pesantren leadership emphasized that algorithms neither possess empathy nor embody akhlak (moral character), making policy reorientation indispensable. Accordingly, the system is positioned solely as a provider of secondary data, while all final personnel-related decisions remain exclusively under human authority through deliberation by the Board of Caretakers in accordance with the Human-in-the-Loop principle.

RESULTS AND DISCUSSION

The synthesis of the field findings demonstrates that the implementation of performance evaluation algorithms cannot be detached from the unique sociocultural contexts in which they are applied. When Artificial Intelligence systems are imposed upon traditional Islamic educational ecosystems, value distortions frequently emerge as a consequence of epistemological conflicts. Digital technologies embody the characteristics of modern rationality that tend to be mechanistic, secular, and reductionist, which subtly challenge the traditional value systems of pesantren that are grounded in local wisdom. Therefore, the successful adoption of technological innovation within this religious educational institution in Southeast Sulawesi is highly dependent upon policymakers' ability to undertake sociological adaptations, ensuring that managerial digitalization does not compromise the institution's spiritual nobility (Apriadi et al., 2025)

Furthermore, the fundamental inability of Artificial Intelligence to capture the essence of *khidmah*, namely the noble service and dedication embodied by *ustadz*, substantiates the theoretical proposition that systems based on historical data processing are inherently vulnerable

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

to generating new forms of structural discrimination. Unless local qualitative indicators that embody moral values, sincerity, and spiritual dedication beyond formal working hours are integrated into the architecture of computerized systems, mathematical bias will remain unavoidable. Algorithms will continue to marginalize educators who demonstrate high levels of cultural integrity while disproportionately rewarding those whose performance is primarily pragmatic and administratively oriented. Such systemic discrimination constitutes a tangible threat that may gradually undermine organizational justice and diminish the profound meaning of sincerity in teaching religious knowledge within the pesantren (Aly, 2025).

As this Islamic educational institution in South Konawe Regency increasingly embraces the digitalization of Human Resource Management, the ethical boundaries governing the use of algorithms must be carefully defined and formulated. Internal regulations are indispensable to ensure that rigid technological autonomy does not extinguish the unique characteristics and spiritual dignity (*muruah*) of an institution that has been preserved across generations. Pesantren administrators must possess the ethical courage to establish clear limitations on machine-based decision-making, restore the sovereignty of moral judgment to the deliberative processes of the Board of Caretakers (*Dewan Pengasuh*), and reposition Artificial Intelligence solely as an instrument for enhancing operational efficiency. This strategic measure is essential to ensure that bureaucratic modernization can progress harmoniously without compromising the very spirit of Islamic education.

Table 2. Comparative Analysis of the Results and Discussion

No.	Evaluation Indicator	Pure AI-Based Evaluation System	Recommended Human-in-the-Loop Policy Model
1	Nature of Output	Conclusive, automated, with limited opportunities for clarification.	Recommendation-based; subject to mandatory human verification.
2	Measurement Basis	Quantitative metrics (digital attendance and administrative documents).	Integration of digital metrics with the values of *khidmah* (service and devotion) and *akhlak* (moral character).
3	Final Authority	Application dashboard algorithm.	The Board of Caretakers (*Dewan Pengasuh*) and the leadership of the Pondok Pesantren.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

Through the comparative framework presented above, the reorientation of Human Resource Management (HRM) policies at Pondok Pesantren Minhajut Thullab is directed toward a hybrid governance model. This approach ensures that operational management efficiency remains optimized through the support of Artificial Intelligence while simultaneously safeguarding organizational justice and protecting the psychological well-being of educators through the moral intervention and empathetic policies exercised by the Board of Caretakers (Dewan Pengasuh) as the highest institutional authority.

CONCLUSION

This study demonstrates that the implementation of Artificial Intelligence (AI)-based performance evaluation technologies at Pondok Pesantren Darussalam Minhajut Thullab, South Konawe Regency, has generated significant sociocultural tensions. The digitalization of personnel administration, which embodies the characteristics of modern rationality, has proven incapable of capturing the essence of Khidmah the noble values of service and devotion that have long constituted the lifeblood of traditional Islamic educational institutions. The rigid quantitative metrics employed by algorithms are only capable of detecting formal activities that leave digital footprints, such as attendance records and classroom documentation, while the spiritual and cultural contributions of ustadz beyond official working hours are entirely overlooked. Consequently, this condition has contributed to the devaluation of sincerity in service and generated evaluative biases that disadvantage dedicated educators.

This study also identifies the existence of automation bias among middle-level management within the pesantren, including the Human Resources Division and the heads of madrasah. Due to limited data literacy concerning emerging technologies, administrators tend to hold the erroneous assumption that every recommendation generated by computer algorithms represents an absolute truth that is entirely objective and free from error. This condition has resulted in the paralysis of human social control and managerial functions, whereby management executes crucial decisions, such as reductions in incentive-related benefits, in a blind and mechanistic manner without conducting any process of reconfirmation (cross-checking) regarding the actual circumstances experienced by educators in the field.

The implementation of constant and inflexible digital surveillance systems has further triggered manifestations of technostress and acute psychological anxiety among educators, particularly senior ustadz. The radical transformation from a personal and familial leadership model grounded in exemplary conduct (uswah) to mechanistic supervision by Artificial Intelligence

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

has gradually eroded educators' sense of job security. Moreover, the requirement to continuously interact with time-consuming digital interfaces has created new forms of social distance within the pesantren, weakening the traditionally intimate emotional bonds between ustadz and santri, while simultaneously diminishing the collective solidarity among colleagues that has long served as a cornerstone of institutional cohesion.

As a strategic solution, the findings of this study emphasize the necessity of reorienting Human Resource Management policies through the implementation of a Human-in-the-Loop governance model adapted to the local wisdom of the pesantren. The appropriate response to the challenges of digitalization is not the rejection of technological advancement but rather the restructuring of institutional regulations to establish clear limitations on machine autonomy. Within this new governance framework, the quantitative outputs generated by AI are no longer positioned as final and binding decisions but are instead utilized solely as supporting instruments or comparative secondary data. Meanwhile, the sovereignty of ultimate decision-making authority must be fully restored to human actors through deliberative leadership mechanisms.

This study concludes that as Islamic educational institutions increasingly embrace digitalization, the ethical boundaries governing the utilization of teSchnology must be carefully and wisely defined to preserve their unique institutional characteristics. Evaluations concerning the future professional trajectories of ustadz must undergo moral scrutiny, human-centered considerations, and character-based assessments directly led by the Board of Caretakers (Dewan Pengasuh). Through this hybrid governance approach, Pondok Pesantren Minhajut Thullab in Southeast Sulawesi can continue to harmonize the demands of modern managerial efficiency with the preservation of spiritual nobility, organizational dignity (muruah), and the principles of justice for dedicated educators who have devoted themselves to the institution.

Bibliography

Aly, A. H. (2025). *Pesantren Digital: Masa Depan Pendidikan Islam di Era Kecerdasan Buatan*. Publica Indonesia Utama.

Andryan, A. A., & Suminar, D. R. (2024). Tinjauan Literatur Sistematis: Hubungan antara Technostress dengan Burnout. *Al-Kharaj J Ekon Keuang Bisnis Syariah*, *6*(12), 8079–8093.

Apriadi, E. A., Julianto, R., Dwiatmoko, F., Kom, S., Kom, M., Bisri, M., & Kom, M. (2025). *KECERDASAN BUATAN Teori, Implementasi, dan Aplikasi di Era Digital*. Eko Aziz Apriadi.

Archana, T. (2025). Artificial intelligence (AI) and digital competencies in the public sector. In *Digital competency development for public officials: Adapting new technologies in public*

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

services (pp. 95–120). IGI Global Scientific Publishing.

Arifianto, A., Prabowo, G., Setiawan, A., Sain, Z. H., Robbani, H. R. S., & Duerapho, Z. (2025). Management Strategies for Developing the PAI Learning Program at SMP NU Islamic Center Pesanggaran. *MUMTAZ: Jurnal Pendidikan Agama Islam*, *5*(1), 65–82.

Azhari, A. K., Khaudli, M. I., & Setiawan, A. (2025). Tantangan Kurikulum Inovatif di Sekolah Antara Tuntutan Modernisasi dan Otoritas Tradisional. *MUDABBIR Journal Research and Education Studies*, *5*(2), 596–603.

Dutta, D., & Srinivasan, V. (2026). Future-proofing HRM through technology-human conjoined agencies: emerging HR competencies for digital disruption. *The International Journal of Human Resource Management*, 1–43.

El Maknouzi, M. E. H. (2022). Human resource management in the era of artificial intelligence: future HR work practices, anticipated skill set, financial and legal implications. *Strategic Management Journal*, *21*(S1), 1–14.

Febriana, A. (2025). *Strategi Pengelolaan Sdm Di Era Industri 4.0 Menghadapi Isu-Isu Terkini*. Penerbit Widina.

George, B. (2026). The algorithmic university: rethinking learning, agency and knowledge in the age of AI. *Education Innovations: Systems and Future Learning?*, *1*(1), 130–148.

Hijazi, A. (2025). *Manajemen Strategik Modernisasi Pondok Pesantren Di Era Society 5.0 (Kajian Pondok Pesantren Khairul Ummah)*. Universitas islam negeri sultan syarif kasim riau.

Khotib, A., Khaudli, M. I., Setiawan, A., & Bahari, D. (2026). *Enhancing Public Trust Through E-WOM In Islamic Boarding Schools : a Cyber Public Relations Perspective*. *7*(1).

Kuzmanov, I. (2025). The AI generals: Mitigating human bias in artificial intelligence development. *Journal of Novel Research and Innovative Development*, *3*(1), a174–a187.

Lubis, D. S. W., & Nurhayati, S. E. (2026). *AI & Ethical Human Resource Management (HRM)*. Goresan Pena.

Marchena-Sekli, G., & Godo, A. (2026). Digital transformation and the future of human resources: emerging dimensions, implementation challenges and a strategic framework. *Journal of Organizational Effectiveness: People and Performance*, 1–21.

Nasution, U. H., Syauqi, T. M., Setyawibawa, R. A., Rusdi, R., & Rizky, M. C. (2025). *Manajemen di Era AI "Strategi, Etika, dan Adaptasi Organisasi."* Serasi Media Teknologi.

Paredes, P. E., Hakimi, S., Hong, M. K., Mora-Mendoza, M. A., Murakami, K., Bravo, N. S., Cogny, A., & Klenk, M. (2026). Unstuck-Ubiquitous Just-in-Time Interventions to Unleash Creativity and Sustain Wellbeing of Creative Professionals. *Proceedings of the ACM on Interactive,*

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

Mobile, Wearable and Ubiquitous Technologies, *10*(2), 1–50.

Purnamasari, N. (2025). *Dampak Media Sosial Terhadap Perilaku Sosial Peserta Didik Di Madrasah Aliyah Pondok Pesantren Ddi Pattojo Soppeng*. IAIN Parepare.

Quiel, J., Mix, K. A., Salmony-Wust, F., Kanbach, D. K., & Kraus, S. (2026). Ideation by mouse click–human creativity and generative artificial intelligence. *European Journal of Innovation Management*, 1–21.

Raprap, W. P., Ngabalin, M., Camerling, L. Y., Wahono, T. R., Subekti, P. A., Zega, T. F., Sahureka, Z., Simanungkalit, E. M., Manurung, M., & Nugroho, W. S. (2025). *Kepemimpinan pendidikan 5.0: Mengelola sekolah di era disrupsi*. Star Digital Publishing.

RITA, O. (2026). *Strategi Kepemimpinan Transformasional Dalam Optimalisasi Penyelenggaraan Pelatihan (Studi Kasus Pada Latihan Dasar CPNS tahun 2023 di Badan Pengembangan Sumber Daya Manusia Provinsi Lampung)*.

Rustiyana, R., Judijanto, L., Mahendra, G. S., Kamil, Z. A., Purba, D. N., Sutoyo, M. N., Hendrayana, I. G., Pasrun, Y. P., & Prayudani, S. (2025). *Data Mining: Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia.

Sadowski, D., & Bryda, G. (2025). Stakeholders' Visions of AI's Future—Analyzing Sociotechnological Narratives. *World Conference on Qualitative Research*, 3–37.

Samsu, S. A. (2025). *Analisis kebijakan dalam manajemen pendidikan Islam*. Zabags Qu Publish.

Setiawan, A., Ana, L., & Ekaningsih, F. (2025). *Education Financing Management in the Digital Era : An Analysis of the Role of Virtual Accounts Tuition Payments in Islamic Boarding School*. *January*, 1–15.

Setiawan, A., Azhari, A. K., & Rofiq, A. (2025). *TADBIR : Jurnal Manajemen Pendidikan Islam TADBIR : Jurnal Manajemen Pendidikan Islam*. *13*(01), 99–114. https://journal.iaingorontalo.ac.id/index.php/tjmpi/article/download/5772/2694

Suaib, H., Kurdi, S., Iskandar, I., & Sos, S. (2026). *TATA KELOLA AI DALAM MANAJEMEN SDM: Etika, Keadilan, dan Masa Depan Pengambilan Keputusan*. Goresan Pena.

Wibowo, I. A., & Kom, M. (2025). *KECERDASAN BISNIS (BI) DIDUKUNG AI: Meningkatkan prediksi dan pembuatan keputusan dengan ML (Machine Learning)*. yayasan penerbit.

Yazid, A. (2025). *SDM 5.0: Strategi Manusia dalam Era AI dan Otomatisasi*. CV Eureka Media Aksara.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License