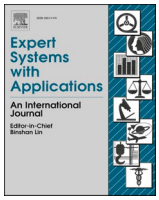
Expert Systems With Applications 296 (2026) 128954

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa



Beyond aversion – principles of appropriate algorithmic decision-making in human resource management

Stefan Strohmeier *, Mathias Becker, Ellen Scheer-Weller

Saarland University, Chair of Management Information Systems, Saarbrücken, Germany

ARTICLE INFO

Keywords:
Algorithmic Decision-Making
AI Principles
Human Resource Management
Machine Learning
Ethical AI
Prescriptive HR Analytics

ABSTRACT

As algorithmic decision-making (ADM) becomes increasingly embedded in human resource management (HRM), concerns such as a lack of fairness and accountability raise urgent questions about its appropriateness. This study addresses the need for ADM evaluation by developing a coherent framework of principles grounded in the task-technology fit approach. It elaborates a balanced triad of nine indispensable ADM principles—methodical (veracity, accuracy, validity), managerial (relevancy, quality, efficiency), and ethical (fairness, accountability, transparency)—and validates them through a systematic literature review of 126 ADM artifacts in HRM. The analysis reveals a troubling lack of attention to ethical and managerial dimensions, while even methodical aspects are often neglected—with the notable exception of accuracy. Building on these findings, the study outlines a forward-looking agenda to operationalize, calibrate, implement, evaluate, and codify ADM principles, ultimately promoting responsible, appropriate ADM in HRM that reflects an evaluative stance beyond mere aversion.

1. Introduction – rationale and relevance

The rapid advancement of machine learning, fueled by increasing data availability and enhanced computing power, has positioned it as a core field within Artificial Intelligence (AI) (e.g., Delipetrev et al., 2020; Oliveira & Figueiredo, 2024). Among its versatile applications (e.g., Jordan & Mitchell, 2015; Mienye & Swart, 2024), machine learning plays a central role in augmenting and automating human decision-making, commonly referred to as *algorithmic decision-making* (*ADM*). ADM follows a two-phase process: first, machine learning algorithms process training data to generate decision models; second, these models are applied to produce actionable recommendations (Hüllermeier, 2021; Mariscal et al., 2010). ADM is now widely adopted in domains such as credit scoring (Wilson Drakes, 2021), healthcare (Tilala et al., 2024), and criminal justice (Završnik, 2021). Increasingly, Human Resource Management (HRM) is also integrating ADM (Duggan et al., 2020; Meijerink & Bondarouk, 2021; Sienkiewicz, 2024; Strohmeier, 2020). While earlier algorithms for decision support were restricted to structured, quantitative decisions such as scheduling (Lin et al., 2020), machine learning has massively expanded this to include *all types* of decisions, including subjective, judgment-based decisions such as selection (Mollay et al., 2024), compensation (Jafari et al., 2020), performance evaluation (Yang & Tang, 2023), and training decisions (Yel, 2025).

Accompanying its growing relevance, however, ADM raises serious concerns in HRM. These include, for instance, discrimination, as ADM may reinforce biases present in training data; opacity, as ADM often lacks transparency and understandability; and decision errors, as ADM risks relying on incorrect data, leading to flawed outcomes (e.g., Bryce et al., 2022; Cachat-Rosset & Klarsfeld, 2023; Gal et al., 2020; Giermindl et al., 2021; Hamilton & Davison, 2022; Hunkenschroer & Luetge, 2022; Köchling & Wehner, 2020; Pessach & Shmueli, 2021; Simbeck, 2019). Moreover, real-world cases show that such concerns are not merely theoretical. For instance, methodical, managerial, and ethical issues contributed to the early termination of a promising ADM system in recruitment—despite its potential benefits—due to the high level of scrutiny and caution required for its responsible deployment (Dastin, 2022).

Given the significance of these concerns, they require serious consideration in both research and practice. However, rather than uncritically accepting all concerns and dismissing ADM as inherently inappropriate, its suitability should be determined through systematic evaluation. Only a rigorous assessment ensures that ADM in HRM is rejected when demonstrably inappropriate while being recognized as

* Corresponding author.
E-mail addresses: s.strohmeier@mis.uni-saarland.de (S. Strohmeier), m.becker@mis.uni-saarland.de (M. Becker), e.weller@mis.uni-saarland.de (E. Scheer-Weller).

https://doi.org/10.1016/j.eswa.2025.128954
Received 23 December 2024; Received in revised form 18 June 2025; Accepted 6 July 2025
Available online 10 July 2025

valuable where appropriate.

This need for systematic evaluation raises the critical question of suitable evaluation criteria—general *principles* that determine ADM (in) appropriateness. Such principles are essential for assessing both existing and future ADM applications, making them highly relevant for research and practice. *Principles* can be defined as *fundamental requirements that appropriate ADM must meet*. For example, *fairness* mandates non-discriminatory treatment of individuals (e.g., Fisher & Howardson, 2022; Jui & Rivas, 2024), while *veracity* ensures the correctness and completeness of training data (e.g., Reimer & Madigan, 2019; Surur et al., 2025). Being fundamental requirements, the principles are generally valid regardless of how ADM is implemented—whether as fully automated (human-out-of-the-loop), semi-automated (human-in-the-loop), or even as participatory, collaborative decision-making involving human stakeholders. However, as fundamental requirements, principles are not directly actionable and must be operationalized to guide practical ADM (e.g., Canca, 2020). For instance, while *fairness* establishes the necessity of non-discrimination, it does not specify which aspects of discrimination, such as age, gender, or ethnicity, should be considered or how this should be ensured in practice. Consequently, operationalization follows elaboration as the second step in making principles actionable (e.g., Canca, 2020). Despite their lack of direct actionability, principles provide crucial insights and overarching guidance on appropriate ADM, forming a necessary foundation for both research and practice (e.g., Morley et al., 2021).

Given the critical importance of appropriate ADM in HRM, this study develops a framework of evaluation principles in four steps. The methodology reviews existing research and identifies four key challenges in current ADM evaluations: unfoundedness, incompleteness, proliferation, and ambiguity. While the task-technology fit (TTF) approach specifically addresses the problem of unfoundedness, the remaining issues are tackled by broadening the focus to include methodical and managerial principles, defining a parsimonious set of minimal requirements, and applying integrative yet selective conceptualizations. The *elaboration* derives nine core principles of appropriate ADM—three methodical, three managerial, and three ethical—rooted in this framework. *Validation* consists of a systematic review of 126 ADM artifacts in HRM, revealing a concerningly narrow focus on accuracy, while other principles are largely overlooked. The *discussion* synthesizes the findings and outlines concrete directions for future research and practice, emphasizing the operationalization, calibration, implementation, and codification of ADM principles in HRM.

2. Methodology – challenges and approach

Given the growing relevance of ADM, there is fortunately an expanding body of research proposing potential principles—often also referred to as challenges, guidelines, recommendations, or requirements. These contributions can be categorized along several dimensions. First, by *scope*: some studies focus on a single principle (Arrieta et al., 2020; Suk & Han, 2024; Zweig & Raudonat, 2022), while others present comprehensive sets (Fjeld et al., 2019; Hagendorff, 2020). Second, by *domain specificity*: some address ADM in general contexts (Lima et al., 2022; Mittelstadt, 2019), whereas others focus specifically on HRM (Bankins, 2021; Hamilton & Davison, 2022; Waymond et al., 2023). Third, by *topical focus*: some contributions concentrate directly on ADM (Krafft et al., 2020; Zerilli et al., 2019), while others engage with related fields such as AI (Greene et al., 2019; OECD, 2019), machine learning (Larus et al., 2018; Lepri et al., 2018), data science (Egger et al., 2022; Saltz & Dewar, 2019), or analytics (Simbeck, 2019; Tursunbayeva et al., 2021).

Despite the field's momentum, current research on ADM principles faces four interrelated challenges: *unfoundedness*, *incompleteness*, *proliferation*, and *ambiguity*. These challenges result in complex—and at times conflicting—requirements that are difficult to reconcile in future research. The current section thus briefly elaborates on each of these challenges and develops a reasoned and pragmatic approach to addressing them, recognizing that ideal or universally accepted solutions are difficult to achieve.

2.1. Foundation of principles

The first challenge lies in the lack of a *theoretical foundation* for ADM principles. Although such principles are broadly discussed in the literature, they are often derived from normative assumptions or plausibility considerations rather than grounded in systematic theoretical frameworks (e.g., Gal et al., 2022; Hunkenschroer & Luetge, 2022). Only a few contributions explicitly draw on justice theory (Acikgoz et al., 2020; Newman et al., 2020), while others rely on ethical frameworks such as utilitarianism or deontology to justify specific, particularly ethical, principles (Gal et al., 2022; Hagendorff, 2020; Hunkenschroer & Luetge, 2022). Some scholars have also proposed TTF as a foundation (Bankins, 2021; Sturm & Peters, 2020). Yet, an integrative theory of ADM requirements—one that systematically conceptualizes and justifies the heterogeneous set of ADM principles—remains absent. The first challenge, therefore, concerns this lack of a solid theoretical foundation.

One possible approach is to draw on a range of theories from different disciplines to underpin specific ADM principles. For instance, the principle of *accountability* might be grounded in virtue ethics (Hursthouse & Pettigrove, 2018), *accuracy* in statistical learning theory (Vapnik, 1998), and *efficiency* in scientific management theory (Taylor, 1911). This eclectic strategy allows for a broad and diverse theoretical base by leveraging well-established theories across disciplines, delivering suitable justifications for individual principles. However, it presents significant challenges. Beyond the considerable effort required to engage with multiple theories, their origins in different domains—often with heterogeneous and partially incompatible assumptions—make a coherent and integrated derivation of principles difficult.

To address this issue, this study adopts the TTF framework, as recommended in the literature (Bankins, 2021; Sturm & Peters, 2020). TTF is a well-established approach for explaining technology utilization and performance (Goodhue & Thompson, 1995) and has been widely applied across various digital technologies and organizational contexts (Cane & McCarthy, 2009; Chavarnakul et al., 2024; Furneaux, 2012; Spies et al., 2020). By focusing on the "fit" between task and technology characteristics, TTF posits that a stronger fit leads to higher utilization and better performance (Goodhue & Thompson, 1995). Unlike the fragmented multi-theoretical approach, TTF offers a coherent and unified framework for elaborating and evaluating ADM principles. Its openness to defining task and technology characteristics contextually makes it suitable for grounding principles across multiple dimensions—including methodological, managerial, and ethical aspects. For instance, Bankins (2021) has already applied TTF to conceptualize ethical principles such as fairness, transparency, and accountability as integral characteristics of HR decision tasks, thus linking normative expectations directly to the logic of task-technology fit.

Building on this understanding, TTF links principles to the characteristics of tasks and technologies (Bankins, 2021; Sturm & Peters, 2020) by acknowledging that HR decisions represent a distinct class of tasks, while algorithmic decision models constitute a distinct class of technologies. Its core premise is that ADM will be effectively utilized and perform well in HRM when the characteristics of algorithmic decision models align with those of HR decision tasks. As such, ADM principles can be understood as manifestations of task-technology fit, reinforcing their relevance and applicability in HRM settings.

However, TTF has its limitations. While it explains why and how adherence to principles ensures "fit"—which in turn drives ADM utilization and success—it does not directly specify what those principles are. Their identification remains a matter of conceptual derivation or empirical investigation. Previous applications of TTF have consequently determined principles either conceptually (Bankins, 2021) or empirically (Sturm & Peters, 2020). This study adopts the conceptual

approach, deriving principles conceptually from the characteristics of HR tasks and ADM technologies. Although TTF does not prescribe specific ADM principles, it offers a more systematic and adaptable framework for their development and evaluation than a fragmented, multi-theoretical approach.

2.2. Completion of principles

A second challenge lies in the *incompleteness* of current ADM principles. Existing research largely focuses on ethical aspects, repeatedly emphasizing principles such as *fairness*, *accountability*, and *transparency* (e.g., Decker et al., 2025; Grimmelikhuijsen, 2022; Hagendorff, 2020; Hunkenschroer & Luetge, 2022; Lim & Kwon, 2021). While this emphasis is understandable—given that core concerns about ADM are indeed ethical—ethical considerations alone are insufficient. They provide moral guidance but fail to capture the full range of factors that determine whether ADM is appropriate. The challenge, therefore, is to identify additional categories of principles that extend beyond "ethical fit" and contribute to a more complete understanding of what constitutes a good fit in ADM. Following the TTF approach, these additional principles must be grounded in the specific characteristics of both ADM technologies and the decision tasks, while ensuring conceptual parsimony and practical manageability to avoid an unstructured proliferation of principles.

A central technical characteristic is the methodological demandingness of machine learning—the key technology underpinning ADM. Machine learning is neither simple nor deterministic; it involves complex, iterative processes that require careful tuning, validation, and scrutiny (e.g., Domingos, 2012; Sammut & Webb, 2017). For instance, since machine learning learns from historical data, ensuring the *veracity* of training data remains a persistent challenge (e.g., Reimer & Madigan, 2019; Rubin & Lukoianova, 2013; Surur et al., 2025). Moreover, decision models inherently yield probabilistic rather than perfect outcomes, making *accuracy* control essential (e.g., Ting, 2017a; Webb, 2017). These factors highlight the need for a *methodical fit*—ensuring sufficient methodological rigor when applying machine learning to decision-making.

Equally important is a central characteristic of the task itself: the managerial demands inherent in HR decision-making. Such decisions are typically complex, requiring the integration of multiple, often conflicting, criteria. As a result, ADM risks oversimplifying this decision space, thereby compromising domain relevance (e.g., Cao, 2010; Liu et al., 2023; Strohmeier & Piazza, 2013). Furthermore, since ADM systems serve as alternatives or complements to human judgment, their decision quality must at least match—and ideally exceed—that of human decision-makers (e.g., Früh et al., 2019). These factors underscore the necessity of a *managerial fit*, meaning the systematic alignment of ADM with the substantive demands of the HR domain—the very rationale for adopting ADM in the first place.

Taken together, the methodical, ethical, and managerial dimensions of fit provide a coherent and comprehensive foundation for evaluating the appropriateness of ADM. Each dimension derives directly from the defining characteristics of ADM: its technological complexity (methodical fit), its domain-specific demands (managerial fit), and its normative implications (ethical fit). Combined, these dimensions capture the essential requirements for ADM systems to be considered suitable and justifiable. Introducing additional dimensions would increase complexity without substantive benefit. For reasons of conceptual parsimony and practical usability, the proposed triad should be regarded as a framework that is both necessary and—while not always exhaustive—generally sufficient.

2.3. Limitation of principles

A third challenge concerns the *proliferation* of ADM principles. As the literature continues to expand, so too does the number of proposed principles (Floridi & Cowls, 2019). For instance, a review of general AI guidelines identified no fewer than *22 distinct ethical principles* (Hagendorff, 2020). While an expanding set of principles helps ensure that all facets of "fit" are considered, it also increases the complexity and effort required for their application—making their practical use more demanding and, as their number continues to grow, potentially unmanageable. This leads to a central dilemma for both research and practice: whether to prioritize a manageable yet inevitably incomplete set of principles or to pursue a comprehensive but increasingly impractical one.

While this dilemma is not easily resolved, this study follows a suggestion from the literature to develop a parsimonious set of core principles representing the "*minimal requirements*" for ADM applications (Hagendorff, 2020). These core principles define the *minimum threshold of fit* that must be met. In other words, if one or more of these principles is violated, the ADM application should be considered unsuitable and must be dismissed. Following the TTF approach, the selection of these core principles must be grounded in arguments that demonstrate their *indispensable relevance* to achieving fit. At the same time, identifying minimal requirements should not preclude the inclusion of *complementary principles* where specific task or technology characteristics warrant them. Accordingly, this study supports future research and practice by proposing a *universal baseline* of mandatory principles, while explicitly leaving open the task of evaluating and determining *additional principles* appropriate to specific ADM contexts.

2.4. Clarification of principles

A fourth challenge concerns the *ambiguity* of ADM principles. Amidst their proliferation, two related problems emerge: first, there is no widely accepted conceptualization of core principles; second, many principles appear to overlap significantly in both content and intent. These issues can be illustrated with the example of the ethical principle of *transparency*. While the literature offers various explicit definitions, they reflect *divergent interpretations*—ranging from technical transparency (e.g., the visibility of data and algorithms) to managerial transparency (e.g., disclosure of stakeholders and institutional interests) (e.g., Andrada et al., 2023). Additionally, *transparency* is surrounded by a cluster of related and often overlapping principles, including *auditability* (Bracci, 2023), *explainability* (Arrieta et al., 2020), *explicability* (Mittelstadt, 2019), *inspectability* (Simbeck, 2019), *interpretability* (Shrestha et al., 2019), *legibility* (Pilling et al., 2020), *reviewability* (Cobbe et al., 2021), *traceability* (Larus et al., 2018), and *understandability* (Arrieta et al., 2020). Yet no widely accepted distinctions exist that would allow for systematic comparison, consolidation, or prioritization.

To address this, we propose an approach based on *integrative yet selective conceptualizations*. *Integrative conceptualization* means defining each principle in a way that captures the *semantic nuances* of its various uses and closely related concepts. For example, an integrative conceptualization of *transparency* might combine technical and managerial transparency while incorporating elements of *explainability*, such as causal reasoning behind decisions. *Selective conceptualization*, by contrast, involves clearly distinguishing these broader *principles as separate constructs*. For example, *fairness* and *transparency* must be treated as distinct, since an ADM system can be highly fair yet completely opaque, or transparent but unfair. Selectivity does not preclude *relationships* between principles; for instance, the methodological principle of *veracity* (of training data) supports the principle of *accuracy* (of model output), even though both remain distinct in scope and definition.

3. Elaboration – derivation and conceptualization

In the following, three core principles are respectively derived based on their indispensable importance for the methodical, managerial, and ethical fit of ADM. These principles are conceptualized in a way that

integrates overlapping principles while remaining clearly differentiated from one another. The discussion uses applicant selection as an illustrative decision-making example, although the principles apply to *all* algorithmic HRM decisions. For initial orientation, the derivation of extensible core principles of ADM in HRM is summarized in Table 1.

3.1. Methodical principles

Methodical principles outline procedural requirements to ensure appropriate algorithmic decisions, as established in general machine learning research (Sammut & Webb, 2017). These principles form the foundation for both managerial and ethical principles, as methodologically flawed ADM can be neither managerially nor ethically appropriate. Therefore, the derivation of principles begins with methodical ones. Although less frequently discussed than ethical principles, three methodical requirements are indispensable for the methodical fit of ADM:

The first indispensable methodical requirement concerns the *veracity* of training data used to generate ADM decision models (Garcia-Arroyo & Osca, 2019; Reimer & Madigan, 2019; Rubin & Lukoianova, 2013). For example, since applicant selection models use historical data to predict whether a candidate is suitable, the quality of these data directly affects outcomes such as interview invitations. Deficient (inaccurate, outdated, or incomplete) training data undermine decision quality during preselection. This issue is exacerbated by the frequent use of secondary ("big") data, often repurposed from external sources of unknown quality rather than collected explicitly for ADM (Garcia-Arroyo & Osca, 2019)—for instance, scraping applicant data from social media instead of collecting it directly. These challenges underscore the critical need for high-quality training data, making *veracity* indispensable. The first core methodical principle of ADM is therefore *veracity*, stating that "ADM is based on correct data." This principle includes the concepts of *data accuracy* (e.g., Mohammed et al., 2022), *data integrity* (e.g., Oladoyinbo et al., 2024), and *data quality* (e.g., Strohmeier, 2020).

The second indispensable methodical requirement is the *accuracy* of decision models (Ting, 2017a; Tolan, 2018; Webb, 2017). In machine learning, decision models must not only be generated but also evaluated to ensure compliance with methodical standards (Webb, 2017). Typically, training data are split into training and test sets: the training data generates the model, and the test data evaluates its performance using quality measures such as confusion matrices (Ting, 2017a). For example, in applicant selection, a two-by-two confusion matrix categorizes predictions as true positives (invited suitable applicants), false positives (invited unsuitable applicants), false negatives (rejected suitable applicants), and true negatives (rejected unsuitable applicants), revealing the model's error rate. Since machine learning models rarely produce error-free predictions (Ting, 2017a), ensuring model accuracy is indispensable for ADM. The second core methodical principle is therefore *accuracy*, asserting that "ADM demonstrates methodical goodness." This principle includes related concepts such as *model quality* (Webb, 2017), *methodical goodness* (Tolan, 2017a), and quality sub-concepts such as *precision* and *recall* (Ting, 2017b).

The third indispensable methodological requirement is the *validity* of decision models (Calude & Longo, 2017; Giermindl et al., 2021; Simbeck, 2019). Machine learning's empirical-inductive approach can detect statistical regularities—yet some of these may be spurious and misleading (Calude & Longo, 2017). For example, a selection model might associate alcohol consumption with poor job performance. This relationship could reflect a true causal link if alcohol impairs occupational functioning. However, it could also be spurious—for instance, if a third factor, such as managerial mistreatment, causes both increased alcohol consumption and reduced performance. Even if such a spurious regularity remains statistically stable over time, it is unsuitable for prediction without a valid causal explanation (Giermindl et al., 2021; Simbeck, 2019). The *validity* of decision models is therefore essential for appropriate algorithmic decision-making (ADM). The third core methodological principle is thus *validity*, which asserts that *ADM relies on explainable regularities*. This principle incorporates the concepts of *causality* (Hünermund et al., 2021), *causal inference* (Simbeck, 2019), and *generalizability* (Nay & Strandburg, 2019).

Table 1
Core ADM Principles.

	PRINCIPLE	DEFINITION	INCLUDED PRINCIPLES
METHODICAL PRINCIPLES	Veracity	ADM is based on correct and reliable data.	Data Accuracy, Data Integrity, Data Quality
	Accuracy	ADM demonstrates methodical goodness.	Model Quality, Methodical Goodness, Precision, Recall, …
	Validity	ADM is based on explainable and generalizable regularities.	Causal Inference, Causality, Generalizability
MANAGERIAL PRINCIPLES	Relevancy	ADM addresses an actual decision-making problem.	Actionability, Interestingness, Operability
	Quality	ADM delivers a satisficing decision outcome.	Decision Effectiveness
	Efficiency	ADM balances necessary inputs with usable outputs.	Decision Rationalization, Decision Speed
ETHICAL PRINCIPLES	Fairness	ADM is non-discriminatory.	Discrimination Awareness, Diversity, Equality (of Treatment), Inclusion, Justice, Unbiasedness
	Transparency	ADM is understandable.	Auditability, Comprehensibility, Explainability, Explicability, Intelligibility, Interpretability, Inspectability, Legibility, Reviewability, Traceability, Understandability
	Accountability	ADM is attributable, justifiable, and correctable.	Contestability, Justifiability, Responsibility

3.2. Managerial principles

Managerial principles delineate the domain-specific requirements essential for ensuring appropriate algorithmic decisions, as established in general HR decision research (Strohmeier, 2020; Vaiman et al., 2012). These principles underscore the fundamental rationale for implementing ADM systems, emphasizing that such systems are not ends in themselves but tools to enhance managerial effectiveness. Consequently, managerial principles build upon methodical principles, extending them toward practical utility. Although, again, less frequently discussed than ethical principles, three managerial principles are indispensable for achieving the managerial fit of ADM:

The first managerial principle is the *relevancy* of decision models (Cao, 2010; Liu et al., 2023; Strohmeier & Piazza, 2013). Ensuring that ADM addresses actual and significant decision problems is paramount in HRM. Despite its apparent self-evidence, *relevancy* is often overlooked, leading to the development of models that address irrelevant or overly simplistic issues. This oversight is frequently attributed to the limited HR domain expertise of developers from technical disciplines (Strohmeier & Piazza, 2013). For instance, some ADM tools attempt to predict the "big five" personality traits from applicant videos and

suggest using these for selection, despite ongoing debates about the utility of personality testing in recruitment. Meta-analyses indicate that only *conscientiousness* moderately predicts employee performance (Diekmann & König, 2018). Such applications exemplify a lack of relevancy. The prevalence of these gaps has even led to the emergence of a subfield within AI, known as *domain-driven machine learning*, which emphasizes the centrality of domain relevance and actionability (e.g., Cao, 2010; Liu et al., 2023). Consequently, relevancy is indispensable for the managerial fit of ADM. The managerial principle of *relevancy* thus asserts that "ADM tackles an actual decision problem" and incorporates the concepts of *actionability*, *operability*, and *interestingness*, all emphasizing the necessity of providing pertinent, actionable insights (e.g., Cao, 2010; Liu et al., 2023).

A second managerial requirement is the *quality* of ADM decision outputs (Früh et al., 2019; Nayak & Dhanaraj, 2020; Strohmeier, 2020). Decision quality pertains to the soundness of suggested actions in achieving managerial objectives, irrespective of the decisions' ultimate outcomes (Vaiman et al., 2012). For example, in applicant preselection, high-quality decisions involve accurately identifying and inviting the most suitable candidates while excluding those less fit for the role. However, concerns have been raised regarding ADM's ability to account for complex socio-psychological factors such as motivation and commitment, giving rise to the metaphorical question of whether employees can be "reduced to numbers" (Giermindl et al., 2021; Tambe et al., 2019). If such critical aspects are overlooked, resulting decisions may fail to meet acceptable quality standards. Therefore, ADM systems must achieve decision quality at least equivalent to, and ideally surpassing, that of human judgment (Früh et al., 2019). Consequently, *quality* emerges as an indispensable managerial principle, asserting that "ADM delivers satisfactory decision outcomes." This principle encompasses related concepts such as *decision effectiveness* (Green & Chen, 2019), emphasizing that ADM systems must produce decisions that are not only methodologically sound but also practically effective in achieving managerial objectives.

A third managerial requirement is the *efficiency* of the ADM process (Cao, 2010; Hickok, 2021; Nayak & Dhanaraj, 2020). Efficiency underscores the necessity for ADM processes to optimize resource utilization, ensuring that the efforts and costs involved are commensurate with the benefits derived. While ADM promises enhanced decision-making capabilities, its development and operationalization can be resource-intensive endeavors. For instance, consider the creation of an ADM application designed to predict personality traits from applicant videos. Such a system necessitates extensive primary data collection, involving the recording of numerous applicant videos and subsequent assessment of personality traits using validated instruments. This process demands significant time and financial investment, which can only be justified if the ADM application is scalable and applicable across a broad spectrum of decisions. The efficiency of ADM is also influenced by the degree of human involvement in the decision-making process. A full automation approach ("human-out-of-the-loop") can maximize efficiency by reducing the time and resources required for each decision; however, this frequently encounters legal constraints and managerial intentions to retain control, thereby favoring an augmentation approach ("human-in-the-loop"). However, the "human-in-the-loop" approach incurs additional time and resource expenditures, negatively impacting efficiency. In light of these considerations, *efficiency* emerges as a third indispensable managerial principle, asserting that "ADM balances necessary input and usable output." This principle includes related concepts such as *decision rationalization*, which focuses on enhancing the input–output ratio of decisions, and *decision speed*, which emphasizes accelerating decision-making processes without compromising quality.

3.3. Ethical principles

Ethical principles delineate the moral imperatives essential for ensuring appropriate ADM, as extensively discussed in AI ethics research (Floridi & Cowls, 2019; Hagendorff, 2020; Hunkenschroer & Luetge, 2022; Jobin et al., 2019). These principles acknowledge that HR decisions significantly impact individuals, often with lasting positive or negative consequences, thus necessitating morally justifiable procedures. Building upon methodical principles, ethical principles extend them toward the moral legitimacy of ADM. Despite the proliferation of AI ethical guidelines, three principles emerge as indispensable for achieving the ethical fit of ADM:

The most prominent and widely discussed ethical requirement is *fairness* in ADM decisions (Decker et al., 2025; Fisher & Howardson, 2022; Jui & Rivas, 2024; Köchling & Wehner, 2020; Robert et al., 2020). Fairness entails avoiding discrimination based on irrelevant characteristics such as age, gender, or ethnicity. Discrimination can arise from data-driven or model-driven causes (Barocas et al., 2023; Barocas & Selbst, 2016; Tolan, 2018). *Data-driven discrimination* results from issues in training data. For instance, *historical bias* occurs when data reflect past discriminatory practices: if, historically, females were underrepresented in selection decisions, algorithms trained on such data replicate this bias ("bias in, bias out"). Similarly, *sampling bias* arises when relevant groups are over- or underrepresented in training data. For instance, if qualified migrants—whose educational and vocational backgrounds systematically differ from those of non-migrants—are absent from training data, the model will neglect them in decision-making. *Model-driven discrimination*, in contrast, results from flaws in the decision model itself (Barocas et al., 2023; Barocas & Selbst, 2016; Tolan, 2018). *Model design bias* occurs when developers include features that encode bias, such as secondary schools attended, which may correlate with ethnicity due to residential segregation. Beyond unconscious forms of discrimination, concerns also persist that ADM's opacity might be misused to conceal deliberate discrimination (Giermindl et al., 2021). *Fairness*, therefore, is an evidently indispensable principle for the ethical fit of ADM. It asserts that "ADM is non-discriminatory" and encompasses related concepts such as *discrimination awareness* (Cardoso et al., 2019), *diversity* (Jobin et al., 2019), *equal opportunity* (Lepri et al., 2018), *equality* (Amani, 2021), *inclusion* (Jobin et al., 2019), *justice* (Hamilton & Davison, 2022), and *unbiasedness* (Tolan, 2018).

A second common ethical concern is *accountability* among the actors involved in developing and applying ADM (de Laat, 2017; Wieringa, 2020; Zweig & Raudonat, 2022). ADM involves multiple interacting actors, including decision-makers (whose past choices inform training data), developers (who select data, algorithms, and construct the decision model), senior managers (who approve ADM implementation), and HRM end-users (who apply ADM tools in practice) (Cobbe et al., 2021). The opacity of ADM systems can hinder the clear attribution of responsibilities to these actors, creating what has been described as an *accountability gap* (Martin, 2019). For example, a recruiter using a commercial ADM tool might blame poor hiring outcomes on senior managers who approved the system; those managers might shift responsibility to the vendor, who might, in turn, hold developers accountable for flawed model validation. Such deflections have raised concerns that ADM could *obscure human responsibility* for consequential decisions (Barocas & Selbst, 2016). *Accountability* complements *transparency*, which merely discloses system performance, by identifying responsible actors, justifying their choices, and enabling corrective action when necessary (Wieringa, 2020). For instance, while transparency might reveal a system's error rate, accountability clarifies *who* deemed that rate acceptable and *who* can be approached to explain or contest that choice (Cobbe et al., 2021; Lepri et al., 2018). In sum, *accountability* constitutes a second indispensable ethical principle, affirming that "ADM is attributable, justifiable, and correctable." It encompasses related principles such as *responsibility* (Lima et al., 2022), *justifiability* (Almada, 2019), and *contestability* (Henin & Le Métayer, 2021).

A third frequent ethical requirement is the *transparency* of the ADM process (Arrieta et al., 2020; Grimmelikhuijsen, 2022; Mueller et al., 2019; Zerilli et al., 2019). A core problem of ADM concerns the opacity of underlying machine learning procedures, which can be attributed to

three types of opacity (Burrell, 2016): Intrinsic opacity arises from the technical and methodological intricacy of some (though not all) machine learning algorithms. Intentional opacity results from providers withholding information about their machine learning procedures to protect their business interests. Illiterate opacity stems from the lack of digital skills and literacy required to understand ADM processes. Together, these forms of opacity often render ADM a "black box" for stakeholders, hindering both basic comprehension and, thus, the acceptance of ADM by diverse audiences (Arrieta et al., 2020; Burrell, 2016). At first glance, achieving "full" transparency—disclosing the entire ADM process in detail, including all actors, training data, algorithms, and decision models—might seem appropriate. However, while this may suit highly skilled data scientists, it is often unsuitable for stakeholders without technical expertise. Thus, stakeholder-specific transparency measures, such as providing illustrative examples or simplified decision models, are necessary to enhance understanding (Arrieta et al., 2020). *Transparency* therefore constitutes a third indispensable ethical principle, claiming that "ADM is understandable." As delineated above, numerous related principles exist and are subsumed within this overarching transparency concept.

4. Validation – demonstration and assessment

Since these principles aim to support future research and practice in evaluating the appropriateness of ADM, we demonstrate their applicability and utility through an application example. Specifically, we applied the principles to assess existing scholarly work on ADM artifacts. Our focus was on ADM within the context of HRM—regardless of specific HR functions or tasks—as the principles are designed to apply broadly across personnel decision-making scenarios.

To identify relevant literature, we conducted a systematic review in accordance with the PRISMA statement (Moher et al., 2009; Page et al., 2021).

We operationalized "ADM artifacts" as (a) concepts (e.g., elaborated use cases), (b) models (e.g., trained decision models), or (c) applications (e.g., realized prototypes). Our search strategy involved querying the databases EBSCOhost, Web of Science, ScienceDirect, ProQuest, and Google Scholar using a comprehensive set of search terms combining HRM-, machine learning algorithm-, and decision-related keywords (see Appendix A). Additionally, systematic backward and forward citation searches were conducted. The review process concluded at the end of Q3 2024.

This approach initially yielded 17,030 contributions. After removing duplicates, 11,767 unique entries remained. We then applied a two-step

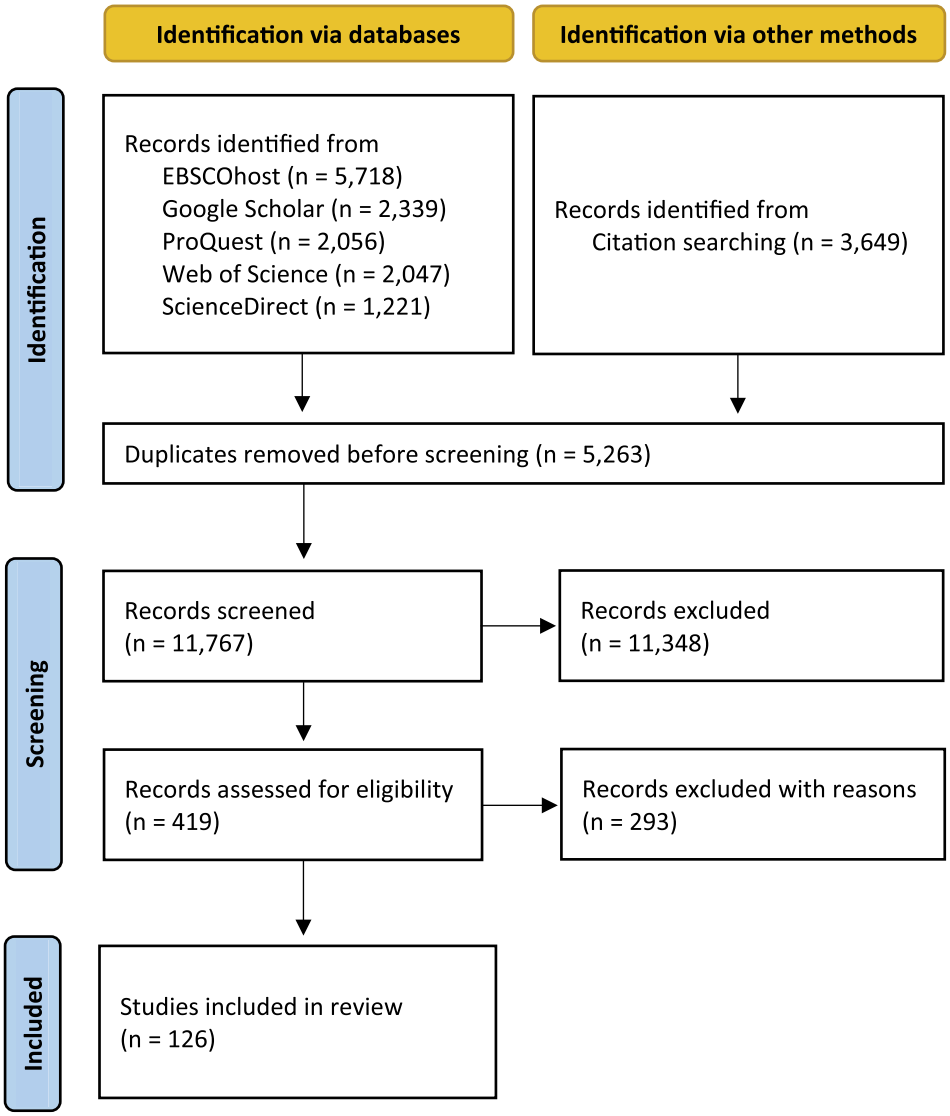


Fig. 1. Search flow diagram based on Page et al. (2021).

screening process: first reviewing titles and abstracts, followed by a full-text analysis of the remaining 419 contributions. Contributions were excluded if they: (a) did not address an HRM decision, (b) did not present an ADM artifact, (c) lacked explicit ("prescriptive") decision support, (d) were not based on machine learning, (e) were published before 2015 (to avoid evaluating older ADM work against principles that did not yet exist), (f) were not published in English, or (g) were student theses. This process yielded 126 eligible contributions (see Fig. 1 and Appendix B). Of these, 58.7 % (n = 74) were journal articles and 41.3 % (n = 52) conference papers; 92.9 % (n = 117) were peer-reviewed. Regarding artifact type, 38.1 % (n = 48) were concepts, 52.4 % (n = 66) models, and 9.5 % (n = 12) applications.

To assess whether—and to what extent—contributions evaluate the compliance of artifacts with principles of appropriate ADM, we developed a five-level scale grounded in established design science artifact evaluation (Peffers et al., 2012; Sonnenberg & vom Brocke, 2011; Venable et al., 2012). The scale differentiates between varying degrees of conceptual and empirical evaluation—two central approaches in design science evaluation—and enables a systematic assessment of the methodological depth of principle-related evaluations.

The five levels are defined as follows:

- *Level 0 – Neglection*: Contributions do not evaluate the respective principle at all.
- *Level 1 – Assertion*: Contributions claim compliance with a principle but provide no substantiation, demonstration, or empirical support.
- *Level 2 – Argumentation*: Contributions offer theoretical or plausibility-based reasoning to support compliance with the principle.
- *Level 3 – Demonstration*: Contributions empirically demonstrate compliance through a single case.
- *Level 4 – Confirmation*: Contributions empirically confirm compliance based on a representative sample of cases.

Based on these descriptions, we developed rules and examples for assigning artifacts to evaluation levels, recording them in a coding book. This coding book served as the basis for assigning suggested artifacts to levels. Coding was performed by two coders. To ensure consistency, one-third of the contributions (n = 42) were randomly selected, independently coded by both coders, and intercoder reliability was calculated (O'Connor & Joffe, 2020). Intercoder reliability was substantial across all coded principles, with an average Krippendorff's α of 0.93 and a weighted Cohen's κ of 0.88 (Cohen, 1968; Krippendorff, 1970). The lowest—but still fully satisfactory—Krippendorff's α values were observed for quality (α = 0.83), efficiency (α = 0.88), validity (α = 0.89), veracity (α = 0.90), and relevancy (α = 0.92). In contrast, the highest scores were recorded for fairness (α = 0.96), transparency (α = 0.97), accuracy (α = 0.98), and accountability (α = 1.0).

For analyzing results, we employed descriptive statistics and cluster analysis (k-means, using silhouette scores as a quality measure; MacQueen, 1967; Rousseeuw, 1987). This yielded valuable insights into the compliance of contributions with principles of appropriate ADM (see Table 2).

First, *methodical principles* exhibit polarized compliance: while *veracity* and *validity* show low levels of consideration (modal value: *neglection*), *accuracy* stands in stark contrast with consistently high values (modal value: *confirmation*). Despite their clear relevance, *veracity* and *validity* are often ignored or merely asserted, suggesting that these principles may be erroneously taken for granted by many scholarly ADM developers. The strong emphasis on *accuracy*—by far the most frequently addressed principle—can be attributed to disciplinary conventions within machine learning (Webb, 2017). Nevertheless, it is noteworthy that even for *accuracy*, one-third of the contributions fail to meet this standard.

Second, *managerial principles* are considered at a low to moderate level. *Relevancy* emerges as the most frequently addressed managerial principle (modal value: *argumentation*). Nearly half of the contributions support relevancy through theoretical reasoning, while more than a third demonstrate it empirically using a single case. In contrast, *quality* and *efficiency* receive somewhat less attention, with most contributions offering only assertions or theoretical arguments (modal values: *assertion* and *argumentation*). Overall, although addressed only moderately, managerial criteria represent the most well-considered category of principles.

Third, *ethical principles* constitute by far the least considered category (modal value: *neglection*). This is most apparent in *accountability*, which is not addressed at all. *Transparency* and *fairness* also exhibit low levels of compliance. The impression of widespread insensitivity to ethical concerns is further reinforced by the fact that over one-fifth of the contributions (21.4 % [n = 27]) present explicit ethical issues. These include the use of directly discriminatory variables—such as age, gender, nationality, or religion—as well as indirectly discriminatory proxies, such as military service (indicating gender), in decision-making tasks such as applicant selection.

Employing cluster analysis to detect potential differences in compliance patterns yielded two clusters (Cluster 1 with 70.6 % [n = 89] and Cluster 2 with 29.4 % [n = 37] of contributions), situated at the lower limit of acceptable cluster quality (silhouette s = 0.3). Broadly, both clusters reflect the modal value patterns discussed above (see Table 2). Their modal values align on accuracy, relevancy, fairness, transparency, and accountability, while showing slight differences in the remaining principles. Cluster 2 demonstrates somewhat stronger compliance with veracity, validity, quality, and efficiency. However, no systematic differences emerged between the clusters—such as consistently high or low compliance levels across all principles.

Overall, our evaluation of recent scholarly ADM artifacts for HRM purposes reveals a concerningly low level of sensitivity to, and consideration of, appropriate ADM principles. *Methodical principles* show polarized attention, *managerial principles* are addressed at a low to moderate level, and *ethical principles* are frequently neglected—with some even openly violated. Particularly notable is the rarity of empirical evaluations based on a single case (*Level 3*) and, even more so, those based on a representative sample (*Level 4*), despite such evaluations being an established standard in artifact assessment (e.g., Peffers et al., 2012). To date, ADM contributions appear largely confined to the narrow evaluation logic of machine learning, focusing primarily on methodical *accuracy* (Webb, 2017). As the above analysis reveals this focus to be insufficient, recent scholarly proposals—such as using ADM in applicant selection—cannot be considered suitable for practical application unless compliance with the broader set of appropriate ADM principles has been explicitly demonstrated.

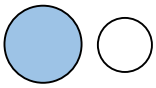
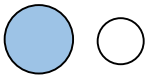
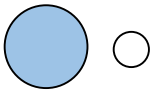
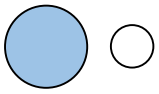
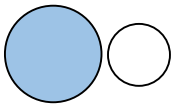
5. Discussion – limitations and implications

The key limitations of our study point directly to its core implications. Each limitation represents a conceptual or practical gap that future research should address.

First, the lack of concrete metrics highlights a limitation in *operationalization*—how can abstract principles be made measurable and actionable? Second, the absence of concrete thresholds reveals a need for *calibration*—how much fulfillment of a principle is sufficient? Third, the lack of practical guidance exposes a limitation in *implementation*—how can principles be effectively put into practice? Fourth, the absence of empirical testing indicates a need for *evaluation*—do these principles actually improve the appropriateness of ADM? Fifth, the lack of legal or institutional embedding underscores a gap in *codification*—how can principles be integrated into regulatory frameworks?

Taken together, these limitations define the components of a focused research agenda aimed at advancing ADM principles from conceptual elaboration to practical application.

Table 2
Descriptive Statistics and Clustering of ADM Contributions.

		COMPLIANCE LEVELS									
		0 Neglection		1 Assertion		2 Argumentation		3 Demonstration		4 Confirmation	
METHODICAL PRINCIPLES	Veracity										
		50.8%	64	20.6%	26	17.5%	22	11.1%	14	0%	0
	Accuracy										
		6.3%	8	10.3%	13	10.3%	13	5.6%	7	67.5%	85
	Validity										
		53.9%	68	23.8%	30	17.5%	22	4.8%	6	0%	0
MANAGERIAL PRINCIPLES	Relevancy										
		0.8%	1	11.1%	14	51.6%	65	36.5%	46	0%	0
	Quality										
		11.9%	15	38.1%	48	37.3%	47	12.7%	16	0%	0
	Efficiency										
		27.0%	34	31.7%	40	31.7%	40	9.6%	12	0%	0
ETHICAL PRINCIPLES	Fairness										
		61.1%	77	15.1%	19	15.9%	20	6.3%	8	1.6%	2
	Transparency										
		64.3%	81	14.3%	18	12.7%	16	8.7%	11	0%	0
	Accountability										
		100%	126	0%	0	0%	0	0%	0	0%	0

Modal Value Cluster 1 ○ Modal Value Cluster 2
(Circle size represents the number of cases)

5.1. Operationalization of principles

A first and evident implication concerns the *operationalization* of ADM principles—the translation of abstract normative requirements into forms that are both *measurable* and *actionable* (Canca, 2020; Mökander & Floridi, 2022; Morley et al., 2021; Morley et al., 2023). Operationalization serves two essential functions: first, it enables the *empirical assessment* of whether, and to what extent, a principle is fulfilled; second, it provides a foundation for *integrating* principles into technical development and organizational practice. Without operationalization, principles risk remaining vague, inconsistently interpreted, unevaluated, and unevenly—or not at all—applied (see Mittelstadt, 2019; Munn, 2022).

For example, while the principle of *fairness* is widely acknowledged as essential in ADM, its interpretation can vary significantly without a shared operational understanding. In the absence of concrete fairness metrics and implementation procedures, the concept risks being reduced to a rhetorical commitment—difficult to evaluate and impossible to enforce. In contrast, when fairness is operationalized—for instance, through measures of demographic parity or equal opportunity (Corbett-Davies et al., 2023)—it becomes possible to evaluate outcomes, identify disparities, and implement corrective mechanisms during model development and deployment.

Fortunately, some principles have already been well operationalized in the literature. *Accuracy* is addressed through diverse, well-established performance metrics (e.g., Naser & Alavi, 2020); *fairness*, through statistical and causal indicators (e.g., Corbett-Davies et al., 2023); *transparency*, through techniques from the field of Explainable Artificial Intelligence (XAI) (e.g., Arrieta et al., 2020); and *veracity*, through established data quality frameworks (e.g., Cichy & Rass, 2019; Ehrlinger & Wöß, 2022).

Other principles, however, remain only partially operationalized. *Accountability*, for instance, has been conceptually discussed—including roles and responsibilities (e.g., Novelli et al., 2024)—but lacks concrete models or indicators. *Quality* has been operationalized in the context of recommender systems (Pu et al., 2011), though broader applications to ADM in HRM remain limited.

Finally, some principles are operationalized only within narrow domains or not at all. *Validity* has been addressed in clinical contexts (Ryu et al., 2022) but requires adaptation for broader use. *Relevancy* is referenced in domain-driven design (Liu et al., 2023) yet lacks measurable indicators, while *efficiency* is typically framed in computational terms, with little attention to decision-making efficiency or cost-effectiveness in HRM (Dehghani et al., 2021).

In sum, while several ADM principles are already measurable and actionable, others lack the operational clarity necessary for meaningful evaluation and application. Advancing their operationalization—through conceptual refinement, domain-specific adaptation, and indicator development—remains a key task for future research to ensure these principles meaningfully inform ADM practice.

To support this effort, researchers can initially draw on general methodological procedures for operationalizing abstract concepts, including clarifying conceptual definitions, identifying measurable variables, selecting appropriate indicators, and designing reliable measurement tools—ideally through iterative testing and refinement (e.g., Goertz, 2006). Beyond these general strategies, emerging field-specific approaches to operationalizing (particularly ethical) principles, as illustrated by Morley et al. (2023), offer promising pathways for further development.

5.2. Calibration of principles

A related implication concerns the future *calibration* of ADM principles—that is, determining the degree of fulfillment required for each principle to be considered adequately met. At first glance, one might assume that maximum fulfillment—such as perfect accuracy or complete transparency—is necessary. In practice, however, such perfection is rarely achievable and, more importantly, not always required. For example, while accuracy is essential, an ADM system does not need to produce flawless predictions; its performance is sufficient if it surpasses that of the best available alternative, such as human decision-making. The appropriate benchmark, therefore, is not ideal performance but *comparative adequacy*.

Yet, much of the current critical discourse tends to hold ADM to unrealistic standards—demanding, for instance, perfect accuracy—without applying equivalent expectations to other forms of decision-making, particularly human judgment. This *double standard*, previously identified in discussions of transparency (Günther & Kasirzadeh, 2021; Zerilli et al., 2019), extends to other principles as well. A fair evaluation of ADM requires defining threshold levels that reflect the degree of principle fulfillment achieved by the best available alternative.

Future calibration efforts should prioritize establishing such benchmarks, striking a balance between normative ambition and practical feasibility. Researchers should build on existing operationalizations of principles to enable comparative assessments across different decision-making approaches—especially human judgment, which remains the dominant point of reference in HRM. The performance of the most effective available alternative should then define the threshold for ADM, ensuring that evaluations are both methodologically rigorous and substantively fair.

To advance these calibration efforts, researchers can draw on established methodological frameworks from comparative evaluation research. This includes defining context-sensitive criteria for what constitutes "adequate" fulfillment of a principle, selecting appropriate baseline comparisons (e.g., human expert performance), and empirically determining threshold levels through iterative testing. Depending on the principle, calibration may involve quantitative benchmarks—such as minimum accuracy rates or acceptable fairness disparities—or qualitative thresholds, such as stakeholder acceptability or procedural transparency. Ideally, these thresholds should be informed by both empirical evidence and stakeholder deliberation, ensuring that they are not only technically robust but also socially legitimate (see Scriven, 1991).

5.3. Implementation of principles

A fourth implication concerns the *implementation of ADM principles*—that is, putting principles into concrete practice. Two core implementation approaches can be distinguished. *Technical implementation* refers to the integration of principles into the development of ADM systems. *Organizational implementation*, by contrast, involves establishing institutional structures and processes to monitor, evaluate, and ensure adherence to these principles.

Technical implementation is increasingly pursued through the *by design* approach, which embeds principles directly into the development process of ADM systems (e.g., Canavese et al., 2024). Its core advantage lies in its anticipatory and mandatory nature: rather than treating principles as downstream compliance concerns, *by design* makes them integral from the outset, enforcing adherence whenever ADM is applied. The concept has primarily evolved under *ethics by design* (Kieslich et al., 2022) and has been actively applied to specific ethical principles such as *transparency by design* (Felzmann et al., 2020; Heidemann et al., 2024), *fairness by design* (Friedler et al., 2021; Lalor et al., 2024), and *accountability by design* (Horneber & Laumer, 2023; Lepri et al., 2018). These efforts have resulted in tangible system features such as explainability modules, bias mitigation tools, and traceability infrastructures. However, the approach should not remain limited to ethics. It holds equal promise for *methodical principles*—such as *veracity by design*, which ensures the quality of training data—and *managerial principles*, such as *relevancy by design*, which aligns ADM with practical decision contexts. Extending the *by design* approach across all principles toward general *compliance by design* (Canavese et al., 2024) would provide a coherent, proactive foundation for building responsible, context-sensitive systems.

Advancing this paradigm is a key task for future research and development.

Organizational implementation, in turn, is increasingly realized through *auditing*, which establishes structured processes to evaluate whether ADM systems comply with principles during development and application (e.g., Mökander & Floridi, 2022). The strength of auditing lies in its institutionalized enforcement: rather than relying on voluntary uptake, audits formally verify compliance and assign responsibility. To date, auditing has primarily focused on ethical principles and is gaining traction in regulatory frameworks. Tools such as bias assessments, documentation reviews, and third-party certifications have been applied to principles such as transparency, fairness, and accountability (Morley et al., 2021). Yet, the auditing logic applies equally to methodical principles—such as testing for accuracy or validity—and managerial principles, such as verifying whether ADM improves decision quality or operational efficiency. Future research must therefore develop standardized audit frameworks, clarify the interplay between internal and external audits, and ensure seamless integration into organizational practice without introducing excessive procedural burdens. Expanding the audit approach across all ADM principles is essential for establishing robust and sustainable governance.

When addressing the organizational and technical implementation of ADM principles, each principle can often be considered individually. However, their simultaneous application may lead to some interactions. For example, efforts to enhance fairness can at times compromise accuracy (e.g., Barocas & Selbst, 2016). Such interdependencies between goals are common in managerial practice and are manageable by deliberate balancing. Recognizing and addressing cross-principle interactions is therefore a critical element of responsible implementation.

5.4. Evaluation of principles

Another important implication concerns the *evaluation* of ADM principles—that is, empirically examining whether and how their application contributes to more appropriate ADM in practice. While operationalization, calibration, and implementation focus on making principles measurable, actionable, and enforceable, evaluation addresses the critical question of their real-world impact. It seeks to determine whether the adoption of these principles leads to better decision-making outcomes, greater organizational trust, or improved system acceptance.

This need is particularly urgent given the growing integration of ADM into high-stakes HRM contexts such as recruitment, performance assessment, and workforce planning—areas where biased or inappropriate decisions can have significant individual and institutional consequences. Without empirical validation, even well-intentioned principles risk remaining aspirational, failing to gain traction in practice, or producing unintended negative outcomes.

Such evaluation requires diverse methodological approaches. *Quantitative studies* can test whether ADM systems guided by principles such as fairness or accuracy outperform others in terms of decision quality, bias mitigation, or user trust. This may involve controlled experiments, comparative field studies, or longitudinal performance tracking. *Qualitative methods*—such as interviews and ethnographic studies—are equally essential for understanding how principles are interpreted, contested, or embedded in organizational practice. They can uncover contextual barriers, power dynamics, and unintended trade-offs. *Mixed-methods research* offers a particularly robust approach by linking measurable system outcomes with stakeholder perspectives and lived experiences. Together, these methods provide a multidimensional understanding of whether and how principles meaningfully support appropriate ADM.

In sum, systematic empirical evaluation is indispensable. It not only tests the real-world effectiveness of principles but also informs their ongoing refinement—ensuring they evolve in response to practical needs, empirical evidence, and ethical expectations.

5.5. Codification of principles

A final implication concerns the future *legal codification* of ADM principles—that is, embedding principles into regulatory frameworks to ensure their *mandatory* and *uniform* adoption across the territorial scope of the regulation. Initial steps in this direction have already been taken, most notably with the adoption of the *EU Artificial Intelligence Act (AIA)* (European Parliament and Council, 2024) and the ongoing preparation of the *U.S. Algorithmic Accountability Act (AAA)* (U.S. Congress, 2023).

A closer examination of the already adopted EU AIA, however, reveals that the principles discussed in this study are only partially addressed. While concepts such as *validity* are mentioned in the extensive but legally non-binding "recitals," the binding provisions remain limited. Legally enforceable articles address the *methodical principles* of *veracity* (framed as *data quality*, Art. 10) and *accuracy* (Art. 15), as well as the *ethical principles* of *fairness* (framed as *prevention of bias*, Art. 10), *transparency* (Art. 13 and 50), and, to some extent, *accountability* (via *human oversight*, Art. 14). However, these principles are largely presented without systematic operationalization and calibration. Without clear definitions of what constitutes adequate fulfillment and how it should be measured, both the practical implementation and effective monitoring of compliance remain difficult.

Regarding implementation, the EU AIA introduces mechanisms such as *conformity assessments* (Art. 43) and *post-market monitoring* (Art. 72). While the term *auditing* is not explicitly used, these instruments perform similar functions by systematically evaluating compliance and assigning accountability.

These observations yield three core implications for future codification efforts. First, indispensable principles—especially ethical and preceding methodical ones—must be explicitly and systematically codified. In contrast, the codification of managerial principles (e.g., *efficiency*) raises normative and practical concerns about their enforceability and desirability in legal form. Second, *all* codified principles must be supported by clear *operationalizations* and *calibrations* to unambiguously define expectations and compliance thresholds. Third, if technical and organizational *implementation measures*—such as conformity assessments or audits—are to be codified, these too require detailed specifications to ensure their effective realization and evaluation.

5.6. Application of principles

The preceding sections collectively outline not only a conceptual framework but also a set of practical pathways for translating principles into real-world ADM practice. *Operationalization* and *calibration* provide the empirical foundation for applying principles in a measurable and context-sensitive manner. *Implementation strategies*—both technical and organizational—demonstrate how principles can be embedded into systems and governance structures. The addition of a dedicated *evaluation* component ensures that the effectiveness of principles can be empirically assessed and iteratively refined. Finally, the section on *codification* links these developments to policy and regulatory action.

Together, these elements form a coherent and actionable research and realization agenda that balances scholarly rigor with practical feasibility. While some principles are already well supported by existing literature, others—particularly in the managerial domain—require further indicator development, contextual adaptation, and field testing. Addressing these gaps presents a concrete and timely opportunity for researchers and practitioners alike to help shape a more appropriate ADM in HRM.

6. Conclusion – contributions and outlook

This study addresses the urgent need for a structured approach to evaluating ADM in HRM—an area where practical application is accelerating, yet clear conceptual guidance remains limited. By proposing a coherent and theoretically grounded framework, the study makes a

substantial contribution to both academic research and organizational practice.

First, it advances theory by introducing the TTF framework into the ADM discourse. This reframing shifts the focus from general debates to a more precise understanding of *appropriateness* as the alignment between HR decision-making tasks and system functionalities—offering a transferable foundation for assessing ADM across contexts.

Second, the study delivers a methodological contribution through a triadic structure of nine key principles: methodical (veracity, accuracy, validity), managerial (relevancy, quality, efficiency), and ethical (fairness, accountability, transparency). This framework addresses gaps in the existing literature by broadening the perspective beyond ethics alone, reducing redundancy, and offering a clear, actionable structure.

Third, it offers empirical insight through a systematic review of recent ADM applications in HRM. The findings reveal a narrow focus on *accuracy*, limited attention to managerial concerns, and frequent neglect of ethical aspects—underscoring the need for more comprehensive and balanced evaluation principles.

Fourth, the study proposes a forward-looking research and realization agenda centered on five key tasks: operationalization to make principles measurable and actionable; calibration to establish meaningful benchmarks; implementation to integrate principles into systems and organizations; evaluation to assess the suitability of principles; and codification to enable enforceability within regulatory frameworks.

In sum, this study lays the foundation for a structured approach to ADM in HRM. It equips researchers, practitioners, and policymakers with the knowledge necessary to design and evaluate ADM systems that are not only methodically sound but also managerially useful and ethically responsible.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendices A & B. Supplementary Data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eswa.2025.128954.

Data availability

Data will be made available on request.

References

Acikgoz, Y., Davison, K. H., Compagnone, M., & Laske, M. (2020). Justice perceptions of artificial intelligence in selection. *International Journal of Selection and Assessment, 28*, 399–416. https://doi.org/10.1111/ijsa.12306

Almada, M. (2019). In *Human intervention in automated decision-making: Toward the construction of contestable systems* (pp. 2–11). Association for Computing Machinery, 10.1145/3322640.3326699.

Amani, B. (2021). AI and "equality by design". In F. Martin-Bariteau & T. Scassa (Eds.), *Artificial intelligence and the law in Canada*. LexisNexis Canada. https://dx.doi.org/10.2139/ssrn.3734665.

Andrada, G., Clowes, R. W., & Smart, P. R. (2023). Varieties of transparency: Exploring agency within AI systems. *AI & SOCIETY, 38*(4), 1321–1331. https://doi.org/10.1007/s00146-021-01326-6

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities, and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. MIT Press.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*(3), 671–732. https://doi.org/10.2139/ssrn.2477899

Bankins, S. (2021). The ethical use of artificial intelligence in human resource management: A decision-making framework. *Ethics and Information Technology, 23*(4), 841–854. https://doi.org/10.1007/s10676-021-09619-6

Bracci, E. (2023). The loopholes of algorithmic public services: An "intelligent" accountability research agenda. *Accounting, Auditing & Acountability Journal, 36*(2), 739–763. https://doi.org/10.1108/AAAJ-06-2022-5856

Bryce, V., Brooks, L., & Stahl, B. (2022). We need to talk about digital HR ethics! A review of the academic literature on ethical aspects of algorithmic Human Resource Management (HRM) technologies. https://dx.doi.org/10.13140/RG.2.2.26318.38726/1.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1). https://doi.org/10.1177/2053951715622512

Cachat-Rosset, G., & Klarsfeld, A. (2023). Diversity, equity, and inclusion in artificial intelligence: An evaluation of guidelines. *Applied Artificial Intelligence, 37*(1), Article 2176618. https://doi.org/10.1080/08839514.2023.2176618

Calude, C. S., & Longo, G. (2017). The deluge of spurious correlations in big data. *Foundations of Science, 22*, 595–612. https://doi.org/10.1007/s10699-016-9489-4

Canavese, D., Ferreira, A., Laborde, R., & Kandi, M. A. (2024). *Artificial Intelligence systems in the European Union: Guidelines and architectures for compliance-by-design* [Research report]. HAL Open Science. https://hal.science/hal-04794994.

Canca, C. (2020). Operationalizing AI ethics principles. *Communications of the ACM, 63*(12), 18–21. https://doi.org/10.1145/3430368

Cane, S., & McCarthy, R. (2009). Analyzing the factors that affect information systems use: A task-technology fit meta-analysis. *The Journal of Computer Information Systems, 50*(1), 108–123.

Cao, L. (2010). Domain-driven data mining: Challenges and prospects. *IEEE Transactions on Knowledge and Data Engineering, 22*(6), 755–769. https://doi.org/10.1109/TKDE.2010.32

Cardoso, L. R., Meira Jr, W., Almeida, V., & J. Zaki, M. (2019). A framework for benchmarking discrimination-aware models in machine learning. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 437–444. Association for Computing Machinery. https://doi.org/10.1145/3306618.3314262.

Chavarnakul, T., Lin, Y. C., Khan, A., & Chen, S. C. (2024). Exploring the determinants and consequences of task-technology fit: A meta-analytic structural equation modeling perspective. *Emerging Science Journal, 8*(1), 77–94. https://doi.org/10.28991/ESJ-2024-08-01-06

Cichy, C., & Rass, S. (2019). An overview of data quality frameworks. *IEEE Access, 7*, 24634–24648. https://doi.org/10.1109/ACCESS.2019.2899751

Cobbe, J., Lee, M. S. A., & Singh, J. (2021). Reviewable automated decision-making: A framework for accountable algorithmic systems. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 598–609. Association for Computing Machinery. https://doi.org/10.1145/3442188.3445921.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*(4), 213–220. https://doi.org/10.1037/h0026256

Corbett-Davies, S., Gaebler, J. D., Nilforoshan, H., Shroff, R., & Goel, S. (2023). The measure and mismeasure of fairness. *Journal of Machine Learning Research, 24*(312), 1–117.

Dastin, J. (2022). *Amazon scraps secret AI recruiting tool that showed bias against women*. In B. Mittelstadt, & L. Floridi (Eds.), *Ethics of data and analytics* (pp. 296–299). Auerbach Publications.

de Laat, P. B. (2017). Big data and algorithmic decision-making: Can transparency restore accountability? *ACM SIGCAS Computers and Society, 47*(3), 39–53. https://doi.org/10.1145/3144592.3144597

Decker, M. C., Wegner, L., & Leicht-Scholten, C. (2025). Procedural fairness in algorithmic decision-making: The role of public engagement. *Ethics and Information Technology, 27*(1), 1–16. https://doi.org/10.1007/s10676-024-09811-4

Dehghani, M., Arnab, A., Beyer, L., Vaswani, A., & Tay, Y. (2021). The efficiency misnomer [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2110.12894.

Delipetrev, B., Tsinaraki, C., & Kostic, U. (2020). *Historical evolution of artificial intelligence. Analysis of the three main paradigm shifts in AI*. Publications Office of the European Union, 10.2760/801580.

Diekmann, J., & König, C. J. (2018). Personality testing in personnel selection: Love it? leave it? Understand it! In I. Nikolaou, & J. Oostrom (Eds.), *Employee recruitment, selection, and assessment: Contemporary issues for theory and practice* (pp. 17–31). Psychology Press.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM, 55*(10), 78–87. https://doi.org/10.1145/2347736.2347755

Duggan, J., Sherman, U., Carbery, R., & McDonnell, A. (2020). Algorithmic management and app-work in the gig economy: A research agenda for employment relations and HRM. *Human Resource Management Journal, 30*(1), 114–132. https://doi.org/10.1111/1748-8583.12258

Egger, R., Neuburger, L., & Mattuzzi, M. (2022). Data science and ethical issues. In *Applied Data Science in Tourism* (pp. 51–66). Cham: Springer, 10.1007/978-3-030-88389-8_4.

Ehrlinger, L., & Wöß, W. (2022). A survey of data quality measurement and monitoring tools. *Frontiers in Big Data, 5*, Article 850611. https://doi.org/10.3389/fdata.2022.850611

European Parliament and Council. (2024). Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence and amending various

Regulations and Directives (Artificial Intelligence Act). Retrieved from https://eur-lex.europa.eu/eli/reg/2024/1689.
Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Towards transparency by design for artificial intelligence. *Science and Engineering Ethics, 26*, 3333–3361. https://doi.org/10.1007/s11948-020-00276-4
Fisher, S. L., & Howardson, G. N. (2022). Fairness of artificial intelligence in human resources-held to a higher standard?. In S. Strohmeier (Ed.), *Handbook of Research on Artificial Intelligence in Human Resource Management* (pp. 303-322). Edward Elgar Publishing. https://doi.org/10.4337/9781839107535.
Fjeld, J., Hilligoss, H., Achten, N., Daniel, M. L., Feldman, J., & Kagay, S. (2019). *Principled artificial intelligence: A map of ethical and rights-based approaches* [White paper]. Retrieved December 20, 2022, from https://ai-hr.cyber.harvard.edu/primp-viz.html.
Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1), 535–545. https://doi.org/10.1162/99608f92.8cd550d1
Friedler, S. A., Scheidegger, C., & Venkatasubramanian, S. (2021). The (im)possibility of fairness: Different value systems require different mechanisms for fair decision-making. *Communications of the ACM, 64*(4), 136–143. https://doi.org/10.1145/3433949
Früh, A., Thouvenin, F., Rudolph, T., Henseler, S., Bayamioglu, E., Braun Binder, N., Katzenbach, C., Kroll, C., Malik, M. & Spielkamp, M. (2019). *Towards principled regulation of automated decision-making – A work-shop report.* Zurich Open Repository and Archive, University of Zurich. https://doi.org/10.5167/uzh-183655.
Furneaux, B. (2012). Task-technology fit theory: A survey and synopsis of the literature. In Y. Dwivedi, M. Wade, & S. Schneberger (Eds.), *Information systems theory. Integrated series in information systems* (pp. 87–106). Springer. https://doi.org/10.1007/978-1-4419-6108-2_5.
Gal, U., Jensen, T. B., & Stein, M. K. (2020). Breaking the vicious cycle of algorithmic management: A virtue ethics approach to people analytics. *Information and Organization, 30*(2), Article 100301. https://doi.org/10.1016/j.infoandorg.2020.100301
Gal, U., Hansen, S., & Lee, A. S. (2022). Research perspectives: Toward theoretical rigor in ethical analysis: The case of algorithmic decision-making systems. *Journal of the Association for Information Systems, 23*(6), 1634–1661. https://doi.org/10.17705/1jais.00784
Garcia-Arroyo, J., & Osca, A. (2019). Big data contributions to human resource management: A systematic review. *International Journal of Human Resource Management, 32*(20), 4337–4362. https://doi.org/10.1080/09585192.2019.1674357
Giermindl, L. M., Strich, F., Christ, O., Leicht-Deobald, U., & Redzepi, A. (2021). The dark sides of people analytics: Reviewing the perils for organisations and employees. *European Journal of Information Systems, 1*(26), 410–435. https://doi.org/10.1080/0960085X.2021.1927213
Goertz, G. (2006). *Social science concepts: A user's guide.* Princeton University Press.
Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *Management Information Systems Quarterly, 19*(2), 213–236. https://doi.org/10.2307/249689
Green, B., & Chen, Y. (2019). The principles and limits of algorithm-in-the-loop decision making. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–24. https://doi.org/10.1145/3359152.
Greene, D., Hoffmann, A. L., & Stark, L. (2019). Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. In *Proceedings of the 52nd Hawaii International Conference on System Sciences* (pp. 2122–2131).
Grimmelikhuijsen, S. (2022). Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. *Public Administration Review, 83*(2), 241–262. https://doi.org/10.1111/puar.13483
Günther, M., & Kasirzadeh, A. (2021). Algorithmic and human decision making: For a double standard of transparency. *AI & SOCIETY, 1–7*. https://doi.org/10.1007/s00146-021-01200-5
Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines, 30*(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8
Hamilton, R. H., & Davison, H. K. (2022). Legal and ethical challenges for HR in machine learning. *Employee Responsibilities and Rights, 34*, 19–39. https://doi.org/10.1007/s10672-021-09377-z
Heidemann, A., Hülter, S. M., & Tekieli, M. (2024). Machine learning with real-world HR data: Mitigating the trade-off between predictive performance and transparency. *International Journal of Human Resource Management, 35*(14), 2343–2366. https://doi.org/10.1080/09585192.2024.2335515
Henin, C., & Le Métayer, D. (2021). Beyond explainability: Justifiability and contestability of algorithmic decision systems. *AI & SOCIETY*. https://doi.org/10.1007/s00146-021-01251-8
Hickok, M. (2021). Lessons learned from AI ethics principles for future actions. *AI and Ethics, 1*, 41–47. https://doi.org/10.1007/s43681-020-00008-1
Horneber, D., & Laumer, S. (2023). Algorithmic accountability. *Business and Information Systems Engineering, 65*(6), 723–730. https://doi.org/10.1007/s12599-023-00817-8
Hüllermeier, E. (2021). Prescriptive machine learning for automated decision making: Challenges and opportunities. arXiv. https://doi.org/10.48550/arXiv.2112.08268.
Hünermund, P., Kaminski, J., & Schmitt, C. (2021). *Causal machine learning and business decision making. SSRN*. https://doi.org/10.2139/ssrn.3867326
Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-enabled recruiting and selection: A review and research agenda. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-022-05049-6
Hursthouse, R., & Pettigrove, G. (2018). *Virtue ethics*. The Stanford Encyclopedia of Philosophy. Retrieved from https://plato.stanford.edu/entries/ethics-virtue/.
Jafari, A., Rouhanizadeh, B., Kermanshachi, S., & Murrieum, M. (2020). Predictive analytics approach to evaluate wage inequality in engineering organizations. *Journal of Management in Engineering, 36*(6), Article 04020072. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000841
Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2
Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415
Jui, T. D., & Rivas, P. (2024). Fairness issues, current approaches, and challenges in machine learning models. *International Journal of Machine Learning and Cybernetics, 15*, 3095–3125. https://doi.org/10.1007/s13042-023-02083-2
Kieslich, K., Keller, B., & Starke, C. (2022). Artificial Intelligence ethics by design. evaluating public perception on the importance of ethical design principles of Artificial Intelligence. *Big Data & Society, 9*(1), 1–15. https://doi.org/10.1177/20539517221092956
Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research, 13*, 795–848. https://doi.org/10.1007/s40685-020-00134-w
Krafft, T. D., Zweig, K. A., & König, P. D. (2020). How to regulate algorithmic decision-making: A framework of regulatory requirements for different applications. *Regulation & Governance*. https://doi.org/10.1111/rego.12369
Krippendorff, K. (1970). Bivariate agreement coefficients for reliability of data. *Sociological Methodology, 2*, 139–150. https://doi.org/10.2307/270787
Lalor, J. P., Abbasi, A., Oketch, K., Yang, Y., & Forsgren, N. (2024). Should fairness be a metric or a model? a model-based framework for assessing bias in machine learning pipelines. *ACM Transactions on Information Systems, 42*(4), 1–41. https://doi.org/10.1145/3641276
Larus, J., Hankin, C., Carson, S. G., Christen, M., Crafa, S., Grau, O., Kirchner, C., Knowles, B., McGettrick, A., Tamburri, D. A., & Werthner, H. (2018). *When computers decide: European recommendations on machine-learned automated decision making. Academy for Computing Machine*. https://doi.org/10.1145/3185595
Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology, 31*, 611–627. https://doi.org/10.1007/s13347-017-0279-x
Lim, J. H., & Kwon, H. Y. (2021). A study on the modeling of major factors for the principles of AI ethics. *The 22nd Annual International Conference on Digital Government Research*, 208–218. Academy for Computing Machinery. https://doi.org/10.1145/3463677.3463733.
Lima, G., Grgić-Hlača, N., Jeong, J. K., & Cha, M. (2022). In *The conflict between explainable and accountable decision-making algorithms* (pp. 123–134). Association for Computing Machinery, 10.1145/3531146.3534628.
Lin, Y., Wang, X., & Xu, R. (2020). Semi-supervised human resource scheduling based on deep presentation in the cloud. *EURASIP Journal on Wireless Communications and Networking, 2020*(1), 1–9. https://doi.org/10.1186/s13638-020-01677-6
Liu, C., Fakharizadi, E., Xu, T., & Yu, P. S. (2023). Recent advances in domain-driven data mining. *International Journal of Data Science and Analytics, 15*(1), 1–7. https://doi.org/10.1007/s41060-022-00378-1
MacQueen, J. (1967). In *Some methods for classification and analysis of multivariate observations* (pp. 281–297). University of California Press.
Mariscal, G., Marban, O., & Fernandez, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *Knowledge Engineering Review, 25*(2), 137–166. https://doi.org/10.1017/S0269888910000032
Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics, 160*, 835–850. https://doi.org/10.1007/s10551-018-3921-3
Meijerink, J., & Bondarouk, T. (2021). The duality of algorithmic management: Toward a research agenda on HRM algorithms, autonomy and value creation. *Human Resource Management Review, 33*(1), Article 100876. https://doi.org/10.1016/j.hrmr.2021.100876
Mienye, I. D., & Swart, T. G. (2024). A comprehensive review of deep learning: Architectures, recent advances, and applications. *Information, 15*(12), 755. https://doi.org/10.3390/info15120755
Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence, 1*, 501–507. https://doi.org/10.1038/s42256-019-0114-4
Mökander, J., & Floridi, L. (2022). Operationalising AI governance through ethics-based auditing: An industry case study. *AI and Ethics, 3*, 451–468. https://doi.org/10.1007/s43681-022-00171-7
Mohammed, S., Budach, L., Feuerpfeil, M., Ihde, N., Nathansen, A., Noack, N., Patzlaff, H., & Harmouch, H. (2022). *The effects of data quality on machine learning performance*. arXiv. https://arxiv.org/abs/2207.14529.
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), Article e1000097. https://doi.org/10.1371/journal.pmed.1000097
Mollay, M. H., Sharma, D. S., Kale, A. G., Anawade, P., & Gahane, S. (2024). In *Smart hiring: Leveraging AI for effective employee selection* (pp. 1–6). IEEE, 10.1109/IDICAIEI61867.2024.10842754.
Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). Ethics as a service: A pragmatic operationalisation of AI ethics. *Minds and Machines, 31*(2), 239–256. https://doi.org/10.1007/s11023-021-09563-w
Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., & Floridi, L. (2023). Operationalising AI ethics: Barriers, enablers and next steps. *AI and Society, 38*(1), 411–423. https://doi.org/10.1007/s00146-021-01308-8
Mueller, S. T., Hoffman, R. R., Clancey, W., Emrey, A., & Klein, G. (2019). *Explanation in human-AI systems: A literature meta-review, synopsis of key ideas and publications, and bibliography for explainable AI*. arXiv. https://doi.org/10.48550/arXiv.1902.01876.

Munn, L. (2022). The uselessness of AI ethics. *AI and Ethics, 1–9*. https://doi.org/10.1007/s43681-022-00209-w

Naser, M. Z., & Alavi, A. (2020). Insights into performance fitness and error metrics for machine learning [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2006.00887.

Nay, J., & Strandburg, K. J. (2019). Generalizability: Machine learning and humans-in-the-loop. In R. Vogl (Ed.), *Research handbook on big data law* (pp. 20–27). Edward Elgar Publishing. https://doi.org/10.2139/ssrn.3417436.

Nayak, D., & Dhanaraj, C. (2020). Strategic decision-making with artificial intelligence. Implications for decision speed and quality. *Academy of Management Proceedings, 2020*(1), 21603. https://doi.org/10.5465/AMBPP.2020.21603abstract

Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes, 160*, 149–167. https://doi.org/10.1016/j.obhdp.2020.03.008

Novelli, C., Taddeo, M., & Floridi, L. (2024). Accountability in artificial intelligence: What it is and how it works. *AI & SOCIETY, 39*(4), 1871–1882. https://doi.org/10.1007/s00146-023-01635-y

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods, 19*, 1–13. https://doi.org/10.1177/1609406919899220

OECD (2019). Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449. https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449.

Oladoyinbo, T. O., Olabanji, S. O., Olaniyi, O. O., Adebiyi, O. O., Okunleye, O. J., & Ismaila Alao, A. (2024). Exploring the challenges of artificial intelligence in data integrity and its influence on social dynamics. *Asian Journal of Advanced Research and Reports, 18*(2), 1–23.

Oliveira, A. L., & Figueiredo, M. A. T. (2024). Artificial intelligence: Historical context and state of the art. In H. S. Antunes, P. M. Freitas, A. L. Oliveira, C. M. Pereira, E. V. de Sequeira, & L. B. Xavier (Eds.), *Multidisciplinary perspectives on artificial intelligence and the law* (pp. 3–24). Springer, 10.1007/978-3-031-41264-6_1.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T., Mulrow, C., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal*. https://doi.org/10.1136/bmj.n71

Peffers, K., Rothenberger, M., Tuunanen, T., & Vaezi, R. (2012). Design science research evaluation. In K. Peffers, M. Rothenberger, & B. Kuechler (Eds.), *Design science research in information systems. Advances in theory and practice. DESRIST 2012. Lecture notes in computer science* (pp. 398–410). Springer. https://doi.org/10.1007/978-3-642-29863-9_29.

Pessach, D., & Shmueli, E. (2021). Improving fairness of artificial intelligence algorithms in privileged-group selection bias data settings. *Expert Systems with Applications, 185*, Article 115667. https://doi.org/10.1016/j.eswa.2021.115667

Pilling, F., Akmal, H., Coulton, P., & Lindley, J. (2020). The process of gaining an AI legibility mark. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–10, Association for Computing Machinery. https://doi.org/10.1145/3334480.3381820.

Pu, P., Chen, L., & Hu, R. (2011). A user-centric evaluation framework for recommender systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems* (pp. 157–164), 10.1145/2043932.2043962.

Reimer, A. P., & Madigan, E. A. (2019). Veracity in big data: How good is good enough? *Health Informatics Journal, 25*(4), 1290–1298. https://doi.org/10.1177/1460458217744369

Robert, L. P., Pierce, C., Marquis, L., Kim, S., & Alahmad, R. (2020). Designing fair AI for managing employees in organizations: A review, critique, and design agenda. *Human-Computer Interaction, 35*(5/6), 545–575. https://doi.org/10.1080/07370024.2020.1735391

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Rubin, V., & Lukoianova, T. (2013). Veracity roadmap: Is big data objective, truthful and credible? *Advances in Classification Research Online, 24*(1), 4–15. https://doi.org/10.7152/acro.v24i1.14671

Ryu, A. J., Romero-Brufau, S., Qian, R., Heaton, H. A., Nestler, D. M., Ayanian, S., & Kingsley, T. C. (2022). Assessing the generalizability of a clinical machine learning model across multiple emergency departments. *Mayo Clinic Proceedings: Innovations, Quality & Outcomes, 6*(3), 193–199. https://doi.org/10.1016/j.mayocpiqo.2022.03.003

Saltz, J. S., & Dewar, N. (2019). Data science ethical considerations: A systematic literature review and proposed project framework. *Ethics and Information Technology, 21*, 197–208. https://doi.org/10.1007/s10676-019-09502-5

Sammut, C., & Webb, G. I. (2017). *Encyclopedia of machine learning and data mining*. Springer Publishing Company, Incorporated, 10.1007/978-1-4899-7687-1.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). SAGE Publications.

Shrestha, Y. R., Ben-Menahem, S. M., & Von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review, 61*(4), 66–83. https://doi.org/10.1177/0008125619862257

Sienkiewicz, Ł. (2024). Algorithmic human resources management. In T. A. Adisa (Ed.), *HRM 5.0: Unpacking the digitalisation of human resource management* (pp. 57–85). Springer, 10.1007/978-3-031-58912-6_4.

Simbeck, K. (2019). HR analytics and ethics. *IBM Journal of Research and Development, 63*(4/5), 9. https://doi.org/10.1147/JRD.2019.2915067

Sonnenberg, C., & vom Brocke, J. (2011). Evaluation patterns for design science research artefacts. In M. Helfert, & B. Donnellan (Eds.), *Practical aspects of design science* (pp. 71–83). Springer, 10.1007/978-3-642-33681-2_7.

Spies, R., Grobbelaar, S., & Botha, A. (2020). A scoping review of the application of the task-technology fit theory. In M. Hattingh, M. Matthee, H. Smuts, I. Pappas, Y. K. Dwivedi, & M. Mäntymäki (Eds.), *Responsible design, implementation and use of information and communication technology* (pp. 397–408). Springer. https://doi.org/10.1007/978-3-030-44999-5_33.

Strohmeier, S. (2020). *Algorithmic decision making in HRM*. In T. Bondarouk, & S. Fisher (Eds.), *Encyclopedia of electronic HRM* (pp. 54–60). De Gruyter Oldenbourg. https://doi.org/10.1515/9783110633702-009.

Strohmeier, S., & Piazza, F. (2013). Domain driven data mining in human resource management: A review of current research. *Expert Systems with Applications, 40*(7), 2410–2420. https://doi.org/10.1016/j.eswa.2012.10.059

Sturm, T. & Peters, F. (2020). The impact of artificial intelligence on individual performance: Exploring the fit between task, data, and technology. *Proceedings of the 28th European Conference on Information Systems*. Association for Information Systems. https://aisel.aisnet.org/ecis2020_rp/200.

Suk, Y., & Han, K. T. (2024). A psychometric framework for evaluating fairness in algorithmic decision making: Differential algorithmic functioning. *Journal of Educational and Behavioral Statistics, 49*(2), 151–172. https://doi.org/10.3102/10769986231171711

Surur, F. M., Mamo, A. A., Gebresilassie, B. G., Mekonen, K. A., Golda, A., Behera, R. K., & Kumar, K. (2025). Unlocking the power of machine learning in big data: A scoping survey. *Data Science and Management, 6*, Article 100102. https://doi.org/10.1016/j.dsm.2025.02.004

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. https://doi.org/10.1177/0008125619867910

Taylor, F. W. (1911). *The principles of scientific management*. Harper & Brothers.

Tilala, M. H., Chenchala, P. K., Choppadandi, A., Kaur, J., Naguri, S., Saoji, R., & Devaguptapu, B. (2024). Ethical considerations in the use of artificial intelligence and machine learning in health care: A comprehensive review. *Cureus, 16*(6), Article e62443. https://doi.org/10.7759/cureus.62443

Ting, K. M. (2017a). Confusion matrix. In C. Sammut, & G. I. Webb (Eds.), *Encyclopedia of machine learning and data mining* (p. 260). Springer. https://doi.org/10.1007/978-1-4899-7687-1_50.

Ting, K. M. (2017b). Precision and recall. In C. Sammut, & G. I. Webb (Eds.), *Encyclopedia of machine learning and data mining* (pp. 990–991). Springer. https://doi.org/10.1007/978-1-4899-7687-1_659.

Tolan, S. (2018). *Fair and unbiased algorithmic decision making: Current state and future challenges*. arXiv. https://doi.org/10.48550/arXiv.1901.04730.

Tursunbayeva, A., Pagliari, C., Di Lauro, S., & Antonelli, G. (2021). The ethics of people analytics: Risks, opportunities and recommendations. *Personnel Review, 51*(3), 900–921. https://doi.org/10.1108/PR-12-2019-0680

U.S. Congress. (2023). *Algorithmic Accountability Act of 2023, S. 2892, 118th Congress.* Retrieved from https://www.congress.gov/bill/118th-congress/senate-bill/2892/text.

Vaiman, V., Scullion, H., & Collings, D. (2012). Talent management decision making. *Management Decision, 50*(5), 925–941. https://doi.org/10.1108/00251741211227663

Vapnik, V. (1998). Statistical learning theory now plays a more active role: After the general analysis of learning processes, the research in the area of synthesis of optimal algorithms was started (Doctoral dissertation).

Venable, J., Pries-Heje, J., & Baskerville, R. (2012). A comprehensive framework for evaluation in design science research. In K. Peffers, M. Rothenberger, & B. Kuechler (Eds.), *Design science research in information systems. Advances in theory and practice. DESRIST 2012. Lecture notes in computer science* (pp. 423–438). Springer. https://doi.org/10.1007/978-3-642-29863-9_31.

Waymond, R., Murray, J. M., Stefanidis, A., Degbey, W. Y., & Tarba, S. Y. (2023). An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes. *Human Resource Management Review, 33*(1), Article 100925. https://doi.org/10.1016/j.hrmr.2022.100925

Webb, G. I. (2017). Model evaluation. In G. Sammut, & G. I. Webb (Eds.), *Encyclopedia of machine learning and data mining* (pp. 844–845). Springer. https://doi.org/10.1007/978-1-4899-7687-1_555.

Wieringa, M. (2020). What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 1–18. Association for Computing Machinery. https://doi.org/10.1145/3351095.3372833.

Wilson Drakes, C.-A. (2021). Algorithmic decision-making systems: Boon or bane to credit risk assessment? *Proceedings of the 29th European Conference on Information Systems 2021*. Association for Information Systems. https://aisel.aisnet.org/ecis2021_rip/46.

Yang, Q., & Tang, Y. (2023). Big Data-based human resource Performance evaluation model using Bayesian network of deep learning. *Applied Artificial Intelligence, 37*(1), Article 2198897. https://doi.org/10.1080/08839514.2023.2198897

Yel, M. B. (2025). A machine learning approach to identify training needs and skill gaps in human resource development. *The Journal of Academic Science, 2*(4), 1126–1136.

Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019). Transparency in algorithmic and human decision-making: Is there a double standard? *Philosophy & Technology, 32*(4), 661–683. https://doi.org/10.1007/s13347-018-0330-6

Završnik, A. (2021). Algorithmic justice: Algorithms and big data in criminal justice settings. *European Journal of Criminology, 18*(5), 623–642. https://doi.org/10.1177/1477370819876762

Zweig, K. A., & Raudonat, F. (2022). Accountability of artificial intelligence in human resources. In S. Strohmeier (Ed.), Handbook *of Research on Artificial Intelligence in Human Resource Management* (pp. 323-336). Edward Elgar Publishing. https://doi.org/10.4337/9781839107535.