Transparent Artificial Intelligence and Human Resource Management: A Systematic Literature Review

Alexis Megan Votto
University of Texas at San Antonio
Alexis.votto@utsa.edu

Charles Zhechao Liu, Ph.D.
University of Texas at San Antonio
Charles.liu@utsa.edu

Abstract

As the technological expansion of Artificial Intelligence (AI) penetrates various industries, Human Resource Management has attempted to keep pace with the new capabilities and challenges these technologies have brought. When adopting AI, transparency within HRM decisions is an increasing demand to establish ethical, unbiased, and fair practices within a firm. To this end, explainable AI (XAI) methods have become vital in achieving transparency within HRM decision-making. Thus, there has been a growing interest in exploring successful XAI techniques, as evidenced by the systematic literature review (SLR) performed in this paper. Our SLR starts by revealing where AI exists within HRM. Following this, we review the literature on XAI and accuracy, XAI design, accountability, and data processing initiatives within HRM. The integrated framework we propose provides an avenue to bridge the gap between transparent HRM practices and Artificial Intelligence, providing the industrial and academic community with better insight into where XAI could exist within HRM processes.

Keywords: Explainable Artificial Intelligence; Human Resource Management; Transparency; Systematic Literature Review; Decision-Making

1. Introduction

With the development and evolution of Artificial Intelligence (AI), Human Resource Management (HRM) within organizations has entered a new era characterized by the constant adoption and implementation of sophisticated technologies (Ahmad et al., 2019). The impacts of these newer technologies are double-edged. On the one hand, many existing organizational standards, practices, services, and processes have been greatly improved by the learning capabilities of AI software (Khan and Turowski, 2016). On the other hand, these same technologies have challenged the long-standing expectations levied upon HRM to make transparent, ethical, and unbiased decisions (Bissola and Imperatori, 2014).

AI and its relationship with HRM have started to receive growing attention from the academic and industrial communities. We seek to expand upon the limited literature surrounding transparent AI practices within HRM. Furthermore, we seek to identify what explainable AI (XAI) techniques exist to foster enhanced transparent, interpretable, and accurate HRM practices (Zhao et al., 2021). Recent studies within AI-related HRM highlight augmentation initiatives (Burton, 2019; Clifton et al., 2020), triage capabilities (He et al., 2019), and enhancing business intelligence practices (Berhil et al., 2020; Burton, 2019). Furthermore, previous literature reviews involving HRM and AI have examined a variety of new practices ranging from employee-centric HR activities (e.g., training and performance management) to organizational-centric HR activities (e.g., decision-making and robot/AI collaboration) (Qamar et al., 2021; Vrontis et al., 2021).

The contribution of this paper is two-fold. First, this paper proposes an updated XAI framework for HR professionals to use when building and implementing AI within processes. Secondly, this research identifies XAI Theory by Design components most which are prevalent within HRM and AI literature.

The arrangement of this paper is as follows. We first provide insights into the primary rationale and background theories that serve as the basis of this research and an introduction to the proposed AI and HRM framework. Then we review the methodology used to collect and analyze literature from the designated databases. Next, we present the results of the analysis. Finally, we conclude the paper by discussing our findings and future research opportunities.

2. Background

Within this section, we provide an overview of the two theories driving our proposed framework. We then identify previous literature reviews and meta-

URI: https://hdl.handle.net/10125/102762
978-0-9981331-6-4

analyses to comprehensively understand the research that has been conducted regarding AI and HRM. Lastly, we provide insight into popular AI methods which exist within HRM.

2.1 Dimensional Human Resource Management (D-HRM)

Given the unique and complex nature of HRM, Jia et al. (2018) propose six basic interconnected elements to HRM, which form effective management systems called Dimensional Human Resource Management (D-HRM). The first dimension is Human Resource Planning, which is the starting point for any HRM department. This component is a guiding point for organizations to identify and predict personnel needs and establish goals that promote long-term success. The second dimension of this framework is Recruitment and Deployment. Within this element, Human Resource Planning is considered an organization's input or life-essence. This component allows HRM departments to posture established organizational goals and personnel needs to staffing and staff matching challenges to overcome unique organizational problems. The third dimension is Training and Development, which focuses on upskilling and developing employee education. The fourth dimension is Performance Management which is the true "core" of this framework as it evaluates the organization's performance and where their employee fits. The fifth dimension is employee Compensation Management to ensure employees remain motivated to solve problems within the company. Lastly, the sixth dimension highlights Employee Relationship Management, which includes corporate culture and labor relationships. This dimension of D-HRM serves as an organization's self-evaluation component, as it measures employee satisfaction, retention, and morale based on company culture and initiatives. If retention and satisfaction are low, this dimension feeds into HR planning to explore remedies and solutions to enhance an employee's experience within a company.

2.2 Transparency within HRM

Given the increasing demand for governance within HRM, measures have come into existence to bolster transparency and accountability within HRM practices (Suwanda and Suryana, 2021). An observed lack of accountability and transparency within practices has led to decreased employee confidence, bringing processes to question due to ethical concerns surrounding fairness and equitability (Hood and Heald, 2006; Langer and König, 2021).

Artificial Intelligence and algorithmic technologies seek to automate HRM activities, inspiring new innovative solutions to tedious tasks. However, opaque and obscure practices, especially AI ones, have undermined trust in system outputs, thus diminishing decision-making performance within organizations and HRM departments (Langer and König, 2021; Yeomans et al., 2019). System transparency has been at odds with system performance, especially if the developer favors performance and expediency over accuracy (Brock, 2018). Nevertheless, this has not discouraged HRM professionals from intervening to ensure ethical, transparent, and accurate practices are honored to secure confidence and trust throughout organizational stakeholders (Schnackenberg and Tomlinson, 2014).

2.3 Transparency by Design (TbD)

The demand for transparency in AI use has constantly been growing, placing a special responsibility upon HRM professionals who utilize them (Qamar et al., 2021). However, simply demanding transparent AI technologies is an ineffective method of addressing the problems which arise from AI's "black-box" characteristic (Felzmann et al., 2020; Zarskey, 2013). Hence, Felzmann et al. (2020) developed a "Transparency by Design" (TbD) framework, which embeds accountability, system design, and information analysis components into the process of developing transparent AI technologies within an organization.

Inspired by the Privacy by Design principles developed by Cavoukian (2009) and codified by the European General Data Protection Regulation, TbD provides a practical and systematic approach to design, implementation, and account for transparency within AI practices within decision-making processes. This framework consists of three phases: 1) designing AI systems, 2) data processing and analysis, and 3) accounting for these capabilities. The designing phase establishes the requirements to enhance transparency. The second phase of this framework takes the established requirements in phase 1 and investigates the processing and analysis of data. Lastly, the accountability phase is a stakeholder-oriented component that promotes inspectable and responsive measures to all AI-related routines. Essentially acting as an auditing component to account for and monitor AI performance within an organization.

TbD exists for transparent XAI practices; addressing two key phrases when discussing XAI is essential. The first is "interpretable," where the algorithm or model itself is inherently and intrinsically interpretable (Adadi and Berrada, 2018; Rudin, 2019).

An example of this would be Letham et al.'s (2015) Bayesian Rule Lists-based decision trees, where the authors declare that preliminary interpretable models can provide concise and convincing capabilities to gain the trust of domain experts. The second is "explainable," which refers to the ability to explain the outcome of a model via post-hoc analysis. For example, Ribeiro et al. (2016) discuss Local Interpretable Model-Agnostic Explanation (LIME), which can approximate the black-box model's decisions in the area of interest, thus providing more intuitive insight into importance variables that contribute to the AI system's output.

2.4 AI Human Resource Transparency Model

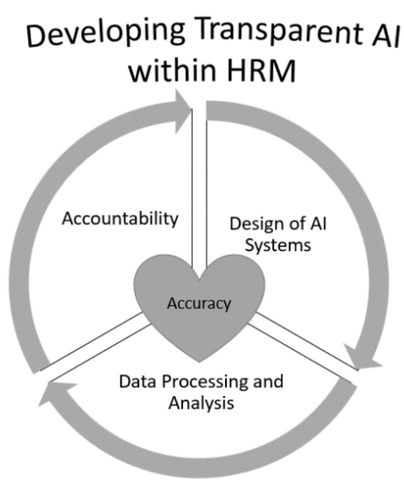

Figure 1. AI Human Resource Transparency Model

Given the demand for fair, accurate, and ethical HRM practices (Schnackenberg and Tomlinson, 2016), leveraging a framework that promotes the development of transparent AI systems becomes a necessary component within all elements of HRM. Thus, we extend the TbD model by adding an accuracy component to guide our literature review and ensure our research's scope maintains a clear focus. The extended model proposed in Figure 1 maintains the integrity of the cyclical nature and components of the TbD model (Felzmann et al., 2020). However, we enhance these respective considerations by adding an accuracy component within the model's heart, adhering to the demand for accurate and fair HRM practices (Lebovitz et al., 2021; Schnackenberg and Tomilson, 2016). This accuracy component directly relates to an algorithm's ability to correctly infer or predict an outcome based on historical data used during the development training and testing phase. When creating cross-disciplinary solutions to accelerate applications, AI algorithms delicately must balance transparent capabilities and accurate inferences (Gunning et al., 2019). Thus, Figure 1 details the extended TbD model to provide an avenue to explore accurate, systematically, and accountable XAI practices within HRM when developing transparent AI systems (see Figure 1).

2.5 Previous Literature Reviews and Meta-Analysis on AI and HRM

The recent explosion of academic and industry interest in AI and HRM has yielded a variety of literature reviews and meta-analysis that evaluate how AI fits into business practices. In an attempt to comprehensively review previous literature and convey the uniqueness of our contribution, we refer to four systematic literature reviews and three meta-analyses.

First, Budhwar et al. (2022) investigate a decade's worth of literature spanning the HRM and AI realm (2010-2020), uncovering emerging trends and future research direction for the international HRM agenda. Our research differs from the expanded scope of Budhwar et al. (2022). Our literature review explores the last 21 years of research, whereas Budhwar et al. (2022) focus on 12 years of research. Kaushal et al. (2021) propose an AI and HRM integration framework, clustering the embeddedness of various HRM functions through analyzing literature spanning from 1988 to 2020. Although their review covers a more extended research period, this research limited itself to Scopus as the only database. Our literature review leverages four databases within the information systems community to conduct our research to maximize search results.

Qamar et al.'s (2021) literature review focuses on AI and HRM research on or before July 2020. This research proposes a concept map illustrating where AI techniques aid HRM decision-making. Our research differs by investigating literature surrounding specific components of HRM and TbD. Qamar et al.'s (2021) paper convey that HRM decision-making uses various elements of AI (expert systems, fuzzy logic, data mining, neural networks, machine learning, and genetic algorithms). Our research differs by providing a framework to follow when developing the elements mentioned above of AI to be transparent. Vrontis et al. (2021) extensively review information management journals regarding AI and HRM capabilities, comprehensively reviewing challenges that exist for emerging technologies. We expand upon this by homing in on the ethical challenge of transparent AI processes by presenting a model for HRM and information systems professionals to use when developing AI tools.

Regarding the first meta-analysis, Anteby et al. (2021) explore eight years' worth of empirical literature involving deep learning AI capabilities. Within this meta-analysis, these authors review accuracy trends involving medical image processing systems to further understand an AI's capability in accurately identifying medical symptoms. The second meta-analysis focuses on pediatric obstructive sleep apnea (Gutiérrez-Tobal et al. (2021). This research evaluated accuracy thresholds involving test subjects and machine learning models, ultimately uncovering the high reliability of machine-learning methods to automatically diagnose severe pediatric patients. Lastly, Schwable and Finzel (2021) conduct a meta-analysis that investigates the terminology, motivations, and evaluation criteria across XAI. Their extensive analysis of seventy surveys across 2019 and 2021 shows accuracy and interpretability to serve as a grounded XAI metric. Although these meta-analyses veer toward the medical sector, empirical validate the importance of accuracy within industry practices. Thus, incorporating this component within our model was appropriate given the supported studies surrounding the importance of measuring accuracy within human-centric industry practices.

3. Methodology

This section explains how we implement an enhanced systematic literature review (SLR) process. Taking a unique two-step approach, we first focus on where AI exists within the components of D-HRM and then build on our findings to investigate further how the XAI literature can inform us on how to achieve transparency and accuracy within HRM.

3.1 Systematic Literature Review

Leveraging the SLR methodology generated by Votto et al. (2021), we generate copious amounts of literature to evaluate, analyze, and extract relevant pieces of academic literature relating to AI, D-HRM, XAI, and accuracy. This methodology is a 2-phase approach. The first phase involves evaluating the article demographics and establishing inclusion and exclusion criteria, while the second phase focuses on evaluating the content of each article.

3.1.1 Phase 1 Overview

Our search strategy focused on peer-reviewed papers identified through electronic searches within major academic databases. Specifically, Business Source Complete (EBSCO), Web of Science (WOS), Association of Information Systems e-Library (AIS), and 4) Scopus (Gittins and Fink, Idris et al., 2017; Koomson et al., 2020; Shela et al., 2021; Suhonen and Paasivaara, 2011; Votto et al., 2021). We searched through these four databases to extract articles relating to D-HRM, XAI, and TbD. After identifying the databases, we generated a standardized search string to query our databases. We leveraged previous literature reviews and meta-analysis keywords associated with those pieces of literature and the guiding theories mentioned in section two to generate our search string. Additionally, we used Boolean operators "OR" and "AND" to connect keywords and parentheses to compartmentalize the AI-specific keywords from the Human Resource-specific, XAI-specific, and TbD-specific keywords.

Table 1: Search String

Code	Search String
HRM	("Human Resource Management" OR "HRM")
D-HRM	("Human Resource Planning" OR "Recruitment" OR "Talent Management" OR "Deployment" OR "Training" OR "Development" OR "Performance Management" OR "Compensation Management" OR "Employee Relations")
XAI	("Explainable Artificial Intelligence" OR "XAI" OR "Interpret*" OR "Transparent" OR "Transparency")
TbD	("Accountability" OR "Accountable" OR "System Design" OR "Data Processing" OR "Data Analysis")
AI	("Artificial Intelligence" OR "AI" OR "Neural Networks" OR "Deep Learning" OR "Computer Vision" OR "Machine Learning" OR "Natural Language Processing" OR "NLP" OR "Robotic" OR "Bot" OR "Information Technology" OR "IT")

We then proceeded to generate inclusion and exclusion criteria to ensure each article was from a peer-reviewed journal, written in English, and published between 2000 – 2021 to ensure the currency of the research topics.

The last filtration component consisted of verifying the journal the article came from based on the Academic Journal Guide (AJG) of 2021. The Chartered Association of Business Schools publishes the renowned ABS journals list. The purpose of this guide is to assist researchers in making informed decisions on highly impactful outlets to publish their research. A journal's highest rankings can earn 4 or 4* (4* being the strongest). Therefore, we chose to limit our search to journals with this distinguished ranking within the information systems, human resources, and business community (Atewologun et al., 2017; Pereira et al., 2021; Vrontis and Christofi, 2021).

Overall, we considered 25 different journals when filtering each database's journals. A few notable journals included in this list are MIS: Quarterly, Journal of the Association of Information Systems (JAIS), Information Systems Research, and Human Resource Management Journal. The articles which did not belong to any of the 25 journals identified did not proceed to Phase 2.

3.1.2 Phase 2 Overview

We performed our content analysis, where the articles were meticulously screened and coded to produce our primary articles for our literature analysis. Phase 2 consists of three primary steps: 1) abstract, keyword, and title analysis, 2) context analysis (full read of the selected article), and 3) Operational Model Diagramming (Onwuegbuzie et al., 2016; Saldaña, 2016). For each of these steps, we leveraged Johnny Saldaña's (2016) qualitative coding manual to aid in appropriately profiling each of our articles heuristically (given the procedural and intuitive nature of coding) (Lincoln and Guba, 1985). Following the operational framework proposed in Figure 1, we sought to gain insight into where AI existed within the D-HRM components, XAI, and accuracy. We used a qualitative analysis software called MAXQDA, which automates our coding process. Utilizing this software during the initial coding processes limited human error as it automatically combed each article for embedded critical phrases associated with the search string that we may have otherwise overlooked (Gizzi and Rädiker, 2021; Ochmann et al., 2021; Onwuegbuzie et al., 2016; Saldaña, 2016). Continuing this iterative process (Onwuegbuzie et al., 2016), we proceeded to Step 2 of this content filtering process and conducted a full read of each article to understand the context. In this step, using the attribute coding scheme from Saldaña (2016), we identified whether an article is empirical (Emp), theoretical (Theo), literature-based (SLR), mixed method, or qualitative (Qual) in nature. Concurrently, we read through each article to verify that the subcodes identified were related to the scope of this research. For example, if we identified the word "train" within the article and initially marked it as a D-HRM component, we would then go through it and gain context to its use. If it did not refer to the training and development of employees, we eliminated it from our list as it did not fit the scope of our research interests.

The last step of this analysis is framing our codes into an operational model (see figure 1) to convey actionable relationships. To accomplish this, we quantified each of our subcodes to showcase how many published works discuss components within TbD, accuracy, and D-HRM. We then took those numbers and framed them in a graph to understand how they fit in our proposed model (see Results section).

4. Results

The findings of this literature analysis exist within this section. Based on these findings, we derive the status of AI literature and HRM components concerning D-HRM and TbD.

4.1 AI and D-HRM Overview

The first iteration of our integrated literature review focuses on D-HRM components, AI, and HRM. Phase 1 of this methodology yielded 18,268 total articles to sift through. Upon applying our comprehensive exclusion and inclusion criteria, we narrowed the extensive list to 129 articles from highly impactful journals, which serve as the basis for the primary analyses.

These 129 articles then entered the content filtration phase of our analysis. Upon coding these articles via MAXQDA, we eliminated 97 articles because the title, abstract, or keywords did not contain a keyword from the AI category and at least one other coding category (HRM or D-HRM). The remaining 32 articles were read in their entirety to verify they fit the scope of our interests. One article was eliminated from this list due to utilizing the acronym "AI" outside the scope of Artificial Intelligence. The rigorous filtering process resulted in having 31 primary research articles. We used Saldaña's (2016) attribute coding mechanism to distinguish empirical, qualitative, literature reviews, and theoretical papers. The results yielded 19 empirical articles, three qualitative articles, three literature reviews, five theoretical articles, and one mixed-method research article within the primary findings.

4.1.1 D-HRM, AI, and Accuracy

This iterative literature analysis supported all dimensions of D-HRM even though their representations vary. Although accuracy is not in the original TbD framework, we identified that this component was prevalent within the literature. For instance, Human Resource Planning appears in 15 articles. Topics surrounding Human Resource Planning and AI discussed procedures to build better processes for adopting AI and accurately tracking work production to achieve company goals. (Abraham et al., 2019).

Furthermore, five articles exist within the Recruitment and Deployment domain. As an example, one of the articles highlighted how the AI-interfaced YouTube platform has served as a tool to bolster recruitment operations (Ali-Hassan et al., 2015). Moreover, AI has been seen in research to accurately identify competitive candidates for a job via interview analysis and candidate selection (Michelotti et al., 2021).

There were 17 AI articles surrounding Training and Development. Specifically, Berente et al. (2021) express the importance of the HR expert's role in developing accurate AI Tools, as the performance of an AI relies upon its ground truth. Furthermore, articles addressed developing best practices to identify employees eligible for professional development and training (or retraining) opportunities related to intelligent technologies.

Performance Management is the next D-HRM component, represented by 13 articles. Specifically, literature on this topic highlighted the importance of how accurate AI-enhanced practices need to be, especially when monitoring organizational productivity and employee performance. Furthermore, these articles discussed and demonstrated how inaccuracy could lead to misinformed decision-making.

The next D-HRM component analyzed was Compensation Management. Comparatively, this D-HRM component was the least represented in the literature, with four articles addressing the relationship between performance and employee compensation regarding AI-related job positions. Specifically, these articles highlight how AI can improve and increase the accuracy of decisions on maintaining the competitiveness of employee compensation and benefits on a global scale (Whitaker et al., 2019).

Finally, the last D-HRM component examined in this analysis was Employee Relations Management. The 22 studies surrounding this component highlight strategic value in considering the social relationships developed between the organization and employees when deploying newer technologies. This literature highlighted the importance of establishing trust through accurate algorithms and explored how AI may positively or negatively impact the relationship between workers and employment standards. Given the broad nature of AI within HR, we identified more primary articles in this iteration of our analysis compared to the second iteration (discussed in later sections). Spanning over 21 articles, we identified that accuracy had a thematic impact on the literature, ultimately tying in the concept of trust and fidelity within an AI-enhanced process.

4.2 XAI, TbD and Accuracy in HRM Overview

The second part of the integrated literature review focuses on TbD, XAI, and accuracy to better understand where it fits in in AI-enhanced HRM practices. We identified a drastic decrease in the number of articles identified to a total of 10,051. After applying the exclusion and inclusion criteria, we narrowed the extensive list to 78 total articles from highly impactful journals.

The 78 articles then entered the content filtration phase of our methodology. We coded these articles via MAXQDA and eliminated 62 articles if the title, abstract, or keywords did not provide a keyword from the AI category and at least one other coding category (TbD, ACC, or XAI). The remaining 16 articles were then read to verify they fit the scope of our focuses. This rigorous filtration process resulted in 13 eligible research articles. We then attribute these primary findings to empirical, qualitative literature reviews and theoretical papers. The results yielded eight empirical articles, two qualitative articles, 0 literature reviews, three theoretical articles, and 0 mixed-method research articles within the primary findings.

4.2.1 TbD and Accuracy

Developing transparent and interpretable AI models, otherwise known as XAI, within AI systems and practices is the guiding intent behind Transparency by Design (TbD). We proposed an element of accuracy that should be accounted for, primarily when operating within HRM, where consequences may affect employees' livelihood within the various decision-making processes. Through this integrated SLR, we explore the literature to understand better the relationship between TbD and accuracy and how these elements fit into D-HRM. Seven articles represented the accountable component of TbD. Collectively, these articles posit a responsibility to justify the accuracy of decisions suggested by AI-capabilities to enhance the transparency of decision-making, enhancing Human Resource Planning goal-setting initiatives.

Understanding HRM decision-making relies upon ethics and accountability. The literature emphasized the importance of developing accurate AI capabilities to meet the demands of accountable expectations within HRM deployment, especially in recruitment (accurately selecting candidates based on skills). Regarding the System Design, four articles contributed, collectively posting the responsibility to develop AI capabilities that embed fairness, ethics, and safety within operations without compromising model accuracy.

Lastly, 11 articles were identified within the Data Processing and Analysis component. Within this component, insight into archival data potentially harming prediction engines came to light (Benbya et al., 2021; Benítez-Peña et al., 2021), as well as the importance of accurate long-term forecasts (Feuerriegel and Gordon, 2019) and the consequences of imbalanced data sets (Gunnarsson et al., 2021; Lebovitz et al., 2021). Furthermore, it provided insight into some of the benefits of utilizing AI interpretability to expediently process and forecast data (Kraus et al., 2020; Ma and Fildes, 2021; Shin et al., 2020; Zhu et al., 2021) and ineffective regression models which are not reliably accurate (Pedro Duarte Silva, 2017).

4.3 Integrated Operational Model

The integrated SLR identified 44 primary articles, which we meticulously analyzed and coded. Given some articles' nature, several HRM factors exist at once. Thus, an article that belongs to Human Resource Planning may represent two or more XAI factors, whereas other D-HRM components may be absent. To better convey where these XAI factors exist within current HRM literature, we graphically represent the findings that intuitively convey the most and least addressed factors. For instance, Berente et al. (2021) emphasize the importance of being able to scrutinize the interpretability and accuracy of AI systems concerning performance and understanding the relationship between AI and human behavior.

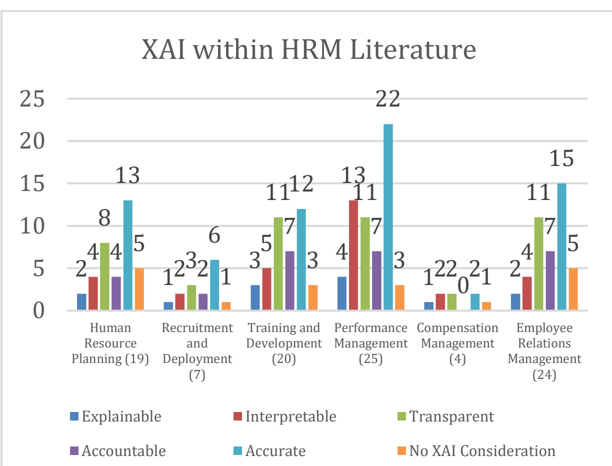

Figure 2. XAI within HRM Literature

In summary, Figure 2 shows the demographic breakdown of XAI factors concerning D-HRM components. For instance, within Human Resource Planning, two articles studied explainability, 4 studied interpretability, 8 studied transparency, 4 studied accountability, 13 studied accuracies, and five did not mention any XAI factor.

5. Discussion

The above findings broadly connect how XAI can be infused within the D-HRM framework, providing further insight into constructing transparent AI practices. This section summarizes the theoretical and practical considerations of this research. We also acknowledge that there are limitations within this study, which may offer directions toward new research opportunities.

5.1 Theoretical and Practical Implications

Our SLR complements existing literature regarding HRM and XAI within AI literature. Extant literature on these subjects focused on the implementation of AI and streamlining individual D-HRM components, so employee relations and performance received the most attention in the literature. Of the articles reviewed, none provided a framework for embedding XAI within HRM. Furthermore, incorporating accuracy within our proposed theoretical framework ensures the relevancy of our SLR, as it was one of the most general considerations discussed throughout the literature (see Figure 2). It also reinforced that our research model (shown in Figure 1) is an appropriate framework to guide HRM professionals and academics in understanding essential elements when incorporating transparent and accurate practices within D-HRM and conducting related research.

Among the various D-HRM components, we discovered that the performance management component received the most attention, with 25 articles directly relating to it. Within the TbD framework, accuracy and transparency were these studies' most ubiquitous TbD elements. Thus, this indicates a special relationship between accurate and transparent XAI practices within performance management functions, such as evaluating an employee's or team's productivity. Despite performance management having the most representation, we identified this as the only D-HRM component where the interpretability keyword was more represented throughout the literature than transparency. Contrarily, compensation management was the least represented D-HRM component. This discovery echoes the findings of Votto et al. (2021), as compensation management was the least explored HRM component across AI literature. All other D-HRM components share a consistent and common theme that accuracy is the most prevalent component. However, it was absent from the original model proposed by Feltzmann et al. (2020). Additionally, all but performance management share transparency, the second-most represented TbD component, indicating a connection between accuracy and transparency when applying AI to HRM processes.

Contrarily, explainability was the least represented TbD component, overshadowed by interpretability. This gap reinforces Cynthia Rudin's (2019) concern that there is a lack of research investigating the differences between interpretable and explainable AI. Furthermore, explainable AI consistently remained at the bottom. Contrarily, interpretability averaged the second-most frequently studied subject. These observations imply that research is expanding on XAI techniques that are inherently interpretable versus explaining outcomes via post-hoc analysis. Thus, expanding upon post-hoc decision-making analysis could benefit both the academic and industry communities in further developing ethical practices of XAI within HRM.

5.2 Future Research

By implementing our proposed theoretical model to conduct our integrated SLR, we identified three themes for future research: 1) There is a lack of research on employee compensation management and its relationship to XAI, 2) Accuracy is a principal component when evaluating XAI within a process, 3) The phrasing of “interpretability” is used more than “explainability” concerning XAI.

Moving forward, transparent and interpretable AI within HRM can be further explored, as we identified potential shortages in the literature surrounding AI within compensation management, literature reviews, and mixed-method research surrounding XAI. Given that employee compensation benefits are directly related to employee relations and retention (Kim et al., 2021), exploration of where AI could exist within this D-HRM component could bolster compensation processes such that suitable candidates are offered raises for performance or tenure. Additionally, research directly related to XAI and HRM is lacking, and where it is present, it is selective to the HR component (recruiting and other data-driven domains). Additional research must exist to understand the balance of transparency and accuracy regarding AI within HRM decision-making. Furthermore, the MAXQDA suite contains extensive tools for deeper text-frequency analysis. Future research interests could examine specific components of XAI and HRM to conduct term frequency.

We recognize that explainability and interpretability have been used interchangeably in AI literature. Based on our analysis, we have identified that interpretability is more frequent than explainability. We contend that using these two phrases interchangeably could confuse readers. This observation also implies that two potential research areas exist when exploring transparent models within D-HRM. One focuses on inherent model transparency, while the other provides post-hoc analysis to explain “black-box” outputs.

Although there is a vast amount of research surrounding general AI within D-HRM, there are growing opportunities to explore how to implement these newer technologies ethically and secure transparency and accuracy within decision-making. We discovered that accuracy is the most prevalent component when implementing XAI within HRM, despite not being initially included in the framework. These findings provide both industry and academic professionals an opportunity to educate themselves on the status of XAI and HRM, giving firm awareness to how transparency should be embedded in AI decision-making processes to promote accurate and fair processes.

5.3 Limitations

It is essential to acknowledge that there are limitations to this research. Although systematic, the methodology for data collection may have led us to miss articles, given our choice in databases and usage of keywords. Additionally, This study only considered peer-reviewed articles available in the selected databases and excluded other publication types such as conference proceedings and book chapters.

We also limited our journal section to those with the two highest ratings from the AJG (4, 4*). This excluded many other renowned academic journal sources from our search criteria. Through this limitation, we sought to highlight where HRM and XAI literature exists within top-tier journals and showcase potential research pathways to bolster research interests surrounding this phenomenon.

Lastly, we acknowledge the heuristic nature of literature reviews, as interpretations of literature may vary. Although we took a scientific approach designed with precautions to ensure the acceptable reliability of our work, we cannot entirely dismiss the possibility of misinterpretations or personal bias when analyzing the literature. However, utilizing the specified software to guide our coding and analysis helped mitigate possible bias from subjective interpretations.

6. Conclusion

This research systematically reviews the status of XAI, HRM, and TbD literature. Additionally, we explore the intersection of transparency regarding AI, accuracy, and HRM and extend the TbD framework to include an accuracy component when designing transparent XAI practices within HRM.

Through a two-stage analysis, this study examines the various components of D-HRM and TbD within AI literature to provide insight into what research exists concerning XAI practices within HRM. The results show that our integrated SLR provides a deeper insight into XAI's theoretical and practical application within HRM. Additionally, it showcases the demand for transparent and fair decision-making within organizations. We hope the proposed theoretical framework can guide future research interests in HRM and transparent AI practices.

7. References

A full list of the 86 references cited throughout this document will be provided by request. The references listed below reflect the articles chosen for our integrated analysis.

Abbasi, A., Li, J., Adjeroh, D., Abate, M., & Zheng, W. (2019). Don't Mention It? Analyzing User-Generated Content Signals for Early Adverse Event Warnings. Information Systems Research, 30(3), 1007–1028.

Abraham, M., Niessen, C., Schnabel, C., Lorek, K., Grimm, V., Möslein, K., & Wrede, M. (2019). Electronic monitoring at work: The role of attitudes, functions, and perceived control for the acceptance of tracking technologies. Human Resource Management Journal, 29(4), 657–675.

Ali-Hassan, H., Nevo, D., & Wade, M. (2015). Linking dimensions of social media use to job performance: The role of social capital. The Journal of Strategic Information Systems, 24(2), 65–89.

Asatiani, A., Malo, P., Aalto, Nagbøl, P. R, Penttinen, E., Rinta-Kahila, T., Salovaara, A. (2021). Sociotechnical Envelopment of Artificial Intelligence: An Approach to Organizational Deployment of Inscrutable Artificial Intelligence Systems. Journal of the Association for Information Systems, 22(2), 325–352.

Benbya, H., Pachidi, S., & Jarvenpaa, S. L. (2021). Special Issue Editorial: Artificial Intelligence in Organizations: Implications for Information Systems Research. Journal of the Association for Information Systems, 22(2), 281–303.

Benítez-Peña, S., Carrizosa, E., Guerrero, V., Jiménez-Gamero, M. D., Martín-Barragán, B., Molero-Río, C., Ramírez-Cobo, P., Romero Morales, D., & Sillero-Denamiel, M. R. (2021). On sparse ensemble methods: An application to short-term predictions of the evolution of COVID-19. European Journal of Operational Research, 295(2), 648–663.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). MANAGING ARTIFICIAL INTELLIGENCE. 18.

Coombs, C., Hislop, D., Taneva, S. K., & Barnard, S. (2020). The strategic impacts of Intelligent Automation for knowledge and service work: An interdisciplinary review. The Journal of Strategic Information Systems, 29(4), 101600.

Dejaeger, K., Goethals, F., Giangreco, A., Mola, L., & Baesens, B. (2012). Gaining insight into student satisfaction using comprehensible data mining techniques. European Journal of Operational Research, 218(2), 548–562.

Feuerriegel, S., & Gordon, J. (2019). News-based forecasts of macroeconomic indicators: A semantic path model for interpretable predictions. European Journal of Operational Research, 272(1), 162–175.

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Will Humans-in-the-Loop Become Borgs? Merits and Pitfalls of Working with AI. MIS Quarterly, 45(3), 1527–1556.

Glezer, C. (2003). A conceptual model of an interorganizational intelligent meeting-scheduler (IIMS). The Journal of Strategic Information Systems, 12(1), 47–70.

Gunnarsson, B. R., vanden Broucke, S., Baesens, B., Óskarsdóttir, M., & Lemahieu, W. (2021). Deep learning for credit scoring: Do or don't? European Journal of Operational Research, 295(1), 292–305.

Guo, F., Gallagher, C. M., Sun, T., Tavoosi, S., & Min, H. (2021). Smarter people analytics with organizational text data: Demonstrations using classic and advanced NLP models. Human Resource Management Journal.

Guo, X., Wei, Q., Chen, G., Zhang, J., & Qiao, D. (2017). Extracting representative information on intra-organizational blogging platforms. MIS Quarterly: Management Information Systems, 41(4), 1105–1127. Scopus.

Kane, G. C., Young, A. G., Majchrzak, A., & Ransbotham, S. (2021). Avoiding an Oppressive Future of Machine Learning: A Design Theory for Emancipatory Assistants. MIS Quarterly, 45(1), 371–396.

Kang, L., Jiang, Q., Chih-Hung, P., Sia, C. L., & Ting-Peng, L. (2020). Managing Change with the Support of Smart Technology: A Field Investigation of Ride-Hailing Services. Journal of the Association for Information Systems, 21(6), 4.

Kettinger, W. J., Ryoo, S. Y., & Marchand, D. A. (2021). We're engaged! Following the path to a successful information management capability. The Journal of Strategic Information Systems, 30(3), 101681.

Kim, S., Wang, Y., & Boon, C. (2021). Sixty years of research on technology and human resource management: Looking back and looking forward. Human Resource Management, 60(1), 229–247.

Kraus, M., Feuerriegel, S., & Oztekin, A. (2020). Deep learning in business analytics and operations research: Models, applications and managerial implications. European Journal of Operational Research, 281(3), 628–641.

Lebovitz, S., Levina, N., Lifshitz-Assa, H., & New York University. (2021). Is AI Ground Truth Really True?

The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What. MIS Quarterly, 45(3), 1501–1526.
Li, J., Li, M., Wang, X., & Thatcher, J. B. (2021). Strategic directions for AI: The role of cios and boards of directors. MIS Quarterly: Management Information Systems, 45(3), 1603–1643. Scopus.
Litwin, A. S. (2013). Not Featherbedding, but Feathering the Nest: Human Resource Management and Investments in Information Technology. Industrial Relations: A Journal of Economy and Society, 52(1), 22–52.
Litwin, A. S., & Tanious, S. M. (2021). Information Technology, Business Strategy and the Reassignment of Work from In-House Employees to Agency Temps. British Journal of Industrial Relations, 59(3), 816–847.
Liu, A. X., Li, Y., & Xu, S. X. (2021). Assessing the unacquainted: Inferred reviewer personality and review helpfulness. MIS Quarterly: Management Information Systems, 45(3), 1113–1148. Scopus.
Ma, S., & Fildes, R. (2021). Retail sales forecasting with meta-learning. European Journal of Operational Research, 288(1), 111–128.
Martens, D., Baesens, B., Van Gestel, T., & Vanthienen, J. (2007). Comprehensible credit scoring models using rule extraction from support vector machines. European Journal of Operational Research, 183(3), 1466–1476.
Martens, D., & Provost, F. (2014). Explaining Data-Driven Document Classifications. MIS Quarterly, 38(1), 73–100.
Michelotti, M., McColl, R., Puncheva-Michelotti, P., Clarke, R., & McNamara, T. (2019). The effects of medium and sequence on personality trait assessments in face-to-face and videoconference selection interviews: Implications for HR analytics. Human Resource Management Journal, n/a(n/a).
Osei-Bryson, K.-M., Dong, L., & Ngwenyama, O. (2008). Exploring managerial factors affecting ERP implementation: An investigation of the Klein-Sorra model using regression splines. Information Systems Journal, 18(5), 499–527.
Popovič, A., Hackney, R., Coelho, P. S., & Jaklič, J. (2014). How information-sharing values influence the use of information systems: An investigation in the business intelligence systems context. The Journal of Strategic Information Systems, 23(4), 270–283.
Preston, D. S., & Karahanna, E. (2009). Antecedents of IS Strategic Alignment: A Nomological Network. Information Systems Research, 20(2), 159–179.
Renaud, A., Walsh, I., & Kalika, M. (2016). Is SAM still alive? A bibliometric and interpretive mapping of the strategic alignment research field. The Journal of Strategic Information Systems, 25(2), 75–103.
Sadler-Smith, E., Akstinaite, V., & Akinci, C. (2021). Identifying the linguistic markers of intuition in human resource (HR) practice. Human Resource Management Journal, n/a(n/a).
Salge, T. O., Kohli, R., & Barrett, M. (2015). Investing in Information Systems: On the Behavioral and Institutional Search Mechanisms Underpinning Hospitals' Is Investment Decisions. MIS Quarterly, 39(1), 61–90.
Sarker, S., Ahuja, M., & Sarker, S. (2018). Work–Life Conflict of Globally Distributed Software Development Personnel: An Empirical Investigation Using Border Theory. Information Systems Research, 29(1), 103–126.
Schuetz, S., & Venkatesh, V. (2020). Research Perspectives: The Rise of Human Machines: How Cognitive Computing Systems Challenge Assumptions of User-System Interaction . Journal of the Association for Information Systems, 460–482.
Shin, D., He, S., Lee, G. M., Whinston, A. B., Cetintas, S., & Lee, K.-C. (2020). Enhancing Social Media Analysis with Visual Data Analytics: A Deep Learning Approach. MIS Quarterly, 44(4), 1459–1492.
Teodorescu, M., West Virginia University, Awwad, Y., Center for Complex Systems, King Abdulaziz City for Science & Technology, Kane, G., & Morse, L. (2021). Failures of Fairness in Automation Require a Deeper Understanding of Human-ML Augmentation. MIS Quarterly, 45(3), 1483–1500.
van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the machine meets the expert: An ethnography of developing ai for hiring. MIS Quarterly: Management Information Systems, 45(3), 1557–1580. Scopus.
Whitaker, J., Mithas, S., & Liu, C.-W. (2019). Beauty is in the eye of the beholder: Toward a contextual understanding of compensation of information technology professionals within and across geographies. Information Systems Research, 30(3), 892–911. Scopus.
Yang, Y., National Sun Yat-sen University, Taiwan, Ying, H., The Chinese University of Hong Kong, Hong Kong, Jin, Y., The Hong Kong Polytechnic University, Hong Kong, Cheng, H. K., University of Florida, USA, Liang, T.-P., & National Sun Yat-sen University, Taiwan. (2021). Special Issue Editorial: Information Systems Research in the Age of Smart Services. Journal of the Association for Information Systems, 22(3), 579–590.
Zhu, H., Samtani, S., Brown, R., & Chen, H. (2021). A Deep Learning Approach for Recognizing Activity of Daily Living (ADL) for Senior Care: Exploiting Interaction Dependency and Temporal Patterns. MIS Quarterly, 45(2), 859–896.