CHAPTER 4

MOTIVATED TRUST IN AI

An Integrative Model Considering Multiple Stakeholder Views in HRM

Guido Hertel
University of Münster

Sandra L. Fisher
Münster University of Applied Sciences

Jenna A. Van Fossen
Münster University of Applied Sciences

ABSTRACT

Artificial intelligence (AI) is increasingly used across human resource management (HRM) functions, yet successful integration is contingent upon these tools being trusted within the organizations. Whereas longstanding cognitive models of trust have been extended to technology and AI as references of trust, motivational influences on trust in technologies have been neglected thus far. However, the consideration of motivational drivers is particularly important to understand and predict different stakeholder views on trust in AI.

The Future of Human Resource Management, pp. 123–161

We integrate cognitive frameworks of trust in AI with more recent approaches to trust motivation stemming from interpersonal trust research. Based on this new integrative model, we specify motivational drivers and cognitive processes for trust in AI for four separate HRM stakeholder perspectives: the employer, decision makers, decision targets, and HR professionals. Particularly, the perspective of HR professionals has been neglected so far, despite the fact that this group is often closely involved with the implementation of AI systems and may see considerable changes to their job tasks after AI adoption. We discuss the theoretical implications of the different stakeholder perspectives for future research and outline specific practical implications of our integrative model.

INTRODUCTION

The use of artificial intelligence (AI) technologies in human resource management (HRM) has increased significantly in recent years. AI-based tools are used, for example, in recruiting by automatically analyzing job application documents or evaluating applicant performance in interviews (Langer et al., 2019; Lukacik et al., 2022), or to support leadership by predicting which employees might be likely to turnover (Chowdhury et al., 2022). However, the research to understand how, when, and why such technologies should be implemented has lagged behind (Tippins et al., 2021). Although a compelling case can be made that AI-based technologies offer practical advantages in HRM, such as high prediction rates or exploitation of large and dynamic data sets, we still lack fundamental knowledge of how these technologies can be effectively implemented into organizational processes. A central precondition of successful implementation is that people across different organizational roles develop and maintain *trust* in AI applications (e.g., Ferrario et al., 2020; Gillespie et al., 2021; Glikson & Woolley, 2020; Huang et al., 2021; Saßmannshausen et al., 2021).

Whereas trust in conventional computer technology and information systems has been addressed in a considerable number of studies and conceptual work (e.g., McKnight et al., 2011; Meeßen et al., 2019; Thielsch et al., 2018), we are just starting to understand trust in AI-based technologies. Conventional computer technology in HRM, such as the use of algorithms for decision-making (e.g., Grove & Meehl, 1996; Hertel et al., 2019; Highhouse, 2008), online testing (e.g., Lievens & Burke, 2011; Tippins, 2015), and internet-based recruiting (e.g., Chapman & Gödöllei, 2017), follow deductive decision rules which have been developed and implemented by humans. These decision rules are characterized by predictable and repeatable analyses that allow managers to process large amounts of data in a rapid, accurate fashion. Their processes and outcomes are controlled and usually well understood (Glikson & Woolley, 2020). AI-based information systems follow *inductive* decision rules developed by machine learning algorithms independent of human insight. This "black box" nature, or opacity (e.g., Fisher & Howardson, 2022; Langer & König, 2021) of AI-based applications in HRM makes trust development much more difficult. Moreover, AI-based applications can have self-optimizing features that lead to constantly changing decisions, creating a situation in which "the workings of machine learning algorithms can escape full understanding and interpretation by humans, even for those with specialized training, even for computer scientists" (Burrell, 2016, p. 10).

The existence of quite diverse definitions and operationalizations of AI makes the concept of AI also somewhat fuzzy, which might further contribute to people's hesitation to place trust in these technologies (Glikson & Woolley, 2020; Pan & Froese, 2023). For example, Solberg et al. (2022) focused on "computer programs that use AI to generate decision alternatives or recommended courses of actions to achieve a specific objective" (p. 2). Turnover prediction tools and interview assessment applications are examples of this category, as they offer recommendations to decision makers but do not make decisions autonomously. Other scholars focus on intelligent machines more broadly, including robots in addition to AI and algorithms (Tang et al., 2022), with autonomous decision making as the core construct. The general concept behind AI is to simulate human intelligence or even exceed it in certain circumstances, although the present status of AI is characterized as "narrow" or "weak" AI that can perform relatively limited functions in a specific field (Glikson & Woolley, 2020; Strohmeier, 2022), often depending on the availability of sufficient and reliable data to develop ("train") the AI algorithm.

The recent empirical record suggests that people often have less trust and are leerier of AI in comparison to humans when decisions are being made (Braganza et al., 2021; Höddinghaus et al., 2021; Langer et al., 2022; Yeomans et al., 2019). Yet, the type of task that AI performs (e.g., tasks requiring mechanical or human skills) as well as the physical representation of AI (e.g., robot, avatar, none/embedded) also matters for trust (Glikson & Woolley, 2020; Lee, 2018), and findings on trust in AI within organizations are mixed. For example, Oracle and Future Workplace (2019) surveyed employees, managers, and HR professionals across 10 countries and found that the majority (64%, ranging from 56% in France to 89% in India) agreed that they would trust AI more than they would trust their managers. Suseno et al. (2022) surveyed HR professionals in China and found that they held fairly positive views and generally lower anxiety about AI. Therefore, in addition to concerns about a lack of trust in AI, there is also the potential for people to have undiscerning "overtrust" in AI (Chugunova & Sele, 2020; Glikson & Woolley, 2020; Keding & Meissner, 2021).

In this chapter, we develop an integrative model of trust in AI in HRM as a framework for current and future research on this topic. In so doing, we understand AI as highly complex computer systems that can automatically analyze large datasets, make predictions, and recommend or even make decisions. Within the field of HRM, these systems are usually based on machine learning (ML) focused on acquisition of knowledge through automated detection of patterns in data (Strohmeier, 2022) to develop rules that can be applied in future decisions. ML can be conducted using a variety of existing algorithms, can be supervised (e.g., with a specific prediction target in mind), unsupervised (e.g., clustering employees according to their behavior to detect training needs), or reinforced (e.g., comparing strategies to optimize specific indicators, such as performance management, job ad campaigns, or sustainability of employee behavior). Moreover, ML can include continuous self-learning and related changes over time (Pan & Froese, 2023), although few ML applications currently in use already incorporate such autonomous and dynamic learning. ML is also used in other types of AI applications that could be applied in HRM, such as robotics, natural language processing or prediction of human preferences (see Strohmeier, 2022, for a recent overview). For example, interview assessment tools could use several different AI functions, such as natural language processing to convert the spoken word to written text and text analysis to evaluate the content of what was said relative to the requirements of the position (Fan et al., 2023; Lukacik et al., 2022). Moreover, ML has been used to predict aesthetic appeals of company websites (Eisbach et al., 2022), supporting a more efficient design of websites for recruiting or e-learning. Finally, generative AI systems, such as ChatGPT (Sackett, 2023), will have considerable influence in multiple HRM fields, for example, as tools supporting the creation of job ads (employer) but also job applications (employees). While the potential of such generative AI systems is quite impressive, overconfidence can be a problem, as these systems can also create fictitious information (e.g., CV components) without reliability indicators (Alkaissi & McFarlane, 2023).

Our integrative approach builds on recent cognitive models of trust in AI (e.g., Solberg et al., 2022), but also considers the different perspectives and motives of people engaging with AI, that is, those persons who might place trust in the technology. HRM processes affect a wide range of stakeholders with differing needs and motivations, which might also affect their trust in AI (Lockey et al., 2021). For example, AI applications might come with separate risks and opportunities for employees as compared to managers (Chowdhury et al., 2022, 2023). Existing research on fairness and trust in AI has already addressed such different perspectives of key stakeholders, for example, the perspectives of decision makers in organizations and the people about whom decisions are made (e.g., Langer & Landers, 2021; Lockey et al., 2021). However, less attention has been paid to the viewpoint of HR professionals (Laurim et al., 2021; Radonjić et al., 2022) who nevertheless occupy core positions when it comes to the implementation and promotion of AI systems in HRM. On the one hand, HR professionals should highly benefit from such technological support, helping them to make high-quality decisions, provide better support to internal customers, and reduce their workload and cognitive strain (e.g., during recruiting and selection). On the other hand, HR professionals might fear losing influence and status and even being replaced by AI systems (Del Giudice et al., 2021; Lockey et al., 2021). These fears may create reluctance to trust and to support the implementation of AI systems, particularly if HR professionals are serving a gatekeeper role in organizations regarding the use of AI technologies for HR. Therefore, we specifically address the perspective of HR professionals with respect to developing trust in AI in organizations.

The chapter is organized as follows. We first review existing models of trust more generally, including both cognitive and motivational perspectives on trust. Whereas cognitive approaches to trust have already been generalized from humans to technology as trust reference, motivational perspectives are so far limited to humans as trust reference. We propose a new integrative model of trust in AI, connecting cognitive and motivational perspectives on trust with respect to technology and AI, and apply this model to the field of HRM. Based on this model, we derive specific propositions that illustrate how different motivations in different organizational roles might affect trustworthiness assessments of AI in HRM, as well as the motivation to develop and maintain trust in such AI systems. We conclude by examining boundary conditions of the model and proposing future research directions in the field of HRM and beyond.

Cognitive Perspectives on Trust in AI

Stemming from the importance of trust for many (if not most) organizational processes (e.g., Colquitt et al., 2007; Schoorman et al., 2007), there is a sizeable interdisciplinary literature on trust and trust development. Most of the extant models so far understand trust as an experienced state following rational considerations of a *trustor* (the subject who trusts). Trust is thus presented as a decision based on more or less elaborate cognitive processes. For instance, in Mayer and colleagues' (1995) model of organizational trust, one of the most foundational frameworks in this literature, trust is defined as the willingness of a trustor to engage in risk-taking and to assume vulnerability to a *trustee* (the reference object or target who is trusted), for instance, by not monitoring the trustee's actions even though these actions affect the trustor. In addition to the dispositional characteristics of the trustor (i.e., their trust propensity), Mayer et al. positioned the trustworthiness of the trustee, including their perceived ability, benevolence, and integrity (ABI), as a key determinant of trust. Mayer et al. (1995)

also proposed that perceived contextual factors affect the consequences of trust, leading to a stronger connection between trust (willingness to assume risks) and trusting behaviors (actually assuming risks) when contextual factors suggest high rather than low risk. Notably, the ABI model does not explicitly specify humans as targets of trust and may thus also be applied to nonhuman targets, including technologies (Schoorman et al., 2007).

Indeed, extant models of trust in technologies follow similar principles to those established by Mayer et al. (1995), assuming that both trustors' dispositions and trustworthiness assessments are precursors of rational (i.e., cognitive) trust decisions. For instance, McKnight et al. (2011) proposed that trust in technology occurs similarly to trust in human targets. The authors addressed dispositional propensities to trust in technology and translated Mayer et al.'s dimensions of ability, benevolence, and integrity into the field of information technology, proposing perceived functionality, helpfulness, and reliability as the main dimensions for trustors' assessments of the trustworthiness of information technologies. Functionality describes whether a technology has sufficient capability to complete a required task, helpfulness addresses whether the technology has adequate and responsive help functions, and reliability is the extent to which a technology works consistently and predictably. As McKnight et al. referred to traditional information technologies (e.g., office applications), helpfulness is constrained to straightforward features, such as whether a help function is installed, but does not consider more complex and advanced aspects such as the implicit goals of the information system. However, the reliability dimension covers many trustworthiness aspects that have also been mentioned in other research on trust in automation or technology acceptance, such as accuracy or system quality (e.g., Delone & McLean, 2003; Thielsch et al., 2018; see Glikson & Woolley, 2020, and Kaplan et al., 2021).

Extending trust in technology to trust in AI more specifically, Solberg et al. (2022) more recently introduced a model that likewise follows the classic ABI structure (Mayer et al., 1995). Integrating extant work on trust in automation (Lee & Moray, 1992), Solberg et al. delineate performance-, process-, and purpose-based trust dimensions. Performance-based trust refers to the perceived capability (e.g., computing speed) of an AI technology, speaking to the perceived ability of an AI system. Interestingly, Solberg et al. considered the reliability of an AI system as part of the performance-based trust dimension, somewhat different from McKnight et al. (2011). Purpose-based trust refers to the assessment of whether the inherent goals of an AI system support its users with respect to their interests (e.g., achieving higher job performance), similar to benevolence in the ABI model. In contrast to the adaptation of McKnight et al. purpose-based trust also includes the implicit goals of an information system. Finally, process-based trust is considered following from assessments of an AI system's transparency and adherence to normative values, corresponding to the integrity dimension. Solberg et al. concentrate here on implicit values of the AI system rather than its reliability and predictability. The latter makes sense in light of the far greater complexity of AI systems, which make assessments of reliability or predictability more difficult and sometimes impossible for a human trustor (e.g., Fisher & Howardson, 2022; Langer & König, 2021). However, besides these minor conceptual differences, Solberg et al.'s more recent model of trust in AI also adheres to a general ABI structure as a general framework.

This brief (and selective) review shows that a focus on cognitive mechanisms underlying the formation of trust is a continued trend from classic approaches of trust in organizations more generally (e.g., Mayer et al. 1995) to specific models with AI systems as the target of trust (Solberg et al., 2022). Trustors are depicted as appraising a target's characteristics in terms of relevant trustworthiness information and determining their trust on the basis of their evaluations. In this line of theorizing, scholars have provided valuable conceptual development into understanding the cognitive mechanisms of trust in AI.

However, far less attention has been paid to understanding how perceptions of trustworthiness are generated and may be influenced over time, despite abundant evidence coming from research on decision making and motivated cognition that such evaluations are often influenced by perceivers' motivation (e.g., Blanchette & Richards, 2010; Higgins & Molden, 2003). At the same time, although there has been much attention placed on understanding how different characteristics of a target matter for trust, there is less insight into how differences between trustors may impact their trustworthiness assessments, and ultimately, their trust. This particularly holds for research on trust in technology.

Motivational Perspectives on Trust Development

Although less often addressed than cognitive perspectives, the question of whether and how people may be motivated to trust is nevertheless longstanding in the literature of trust. Indeed, the definition of trust as "*willingness*" to make oneself vulnerable (e.g., Mayer et al., 1995; Rousseau et al., 1998) reflects its motivational reference. For instance, Luhmann (1968) positioned trust as a means of reducing complexity and uncertainty, underlining the functional impact of trust. As such, people may be motivated to trust to obtain a sense of psychological security given situational uncertainty and risk. In a similar way, Weber et al. (2004) argued that, when dependent upon another, a focal actor will be motivated to reduce the anxiety that stems from this dependence by perceiving the target as trustworthy.

More recently, van der Werff et al. (2019) incorporated motivation theories directly into trust research, applying self-determination theory (Ryan & Deci, 2017) and control theory (Carver & Scheier, 2012) to conceptualize trust as a desirable goal state. Drawing from principles of control theory, van der Werff and colleagues conceive of trust as a dynamic process during which trustors actively monitor their progress towards the trust goal. Momentary trust levels and related cognitions, affect, and behaviors are assumed to influence consecutive perceptions of risk and trustworthiness, which again informs the goal state of trust towards which a trustor is striving. Drawing from self-determination theory, the authors also differentiate between different types of trust motivation. In the case that the relationship with a trustee satisfies trustors' basic psychological needs (for autonomy, competence, and relatedness), individuals are assumed to be intrinsically motivated to develop trust (for instance, in a supportive colleague). Alternately, trustors may also be motivated to develop trust when such a need fulfillment is lacking. Extrinsically motivated trust might either arise due to trustors' perceptions that their goals and values align with those of the trustee (autonomous motivation to trust), or trustors might be motivated to maintain trust because they expect to gain specific rewards or avoid punishment (controlled motivation to trust).

In line with self-determination theory, van der Werff et al. (2019) assume that intrinsically motivated trust has stronger effects and is longer-lasting than extrinsically motivated trust. However, in their theoretical work, van der Werff et al. focused on interpersonal trust and remained silent on technology as a referent of trust. We assume extrinsic trust motivation to be more relevant for trust in technologies at work and HRM, as such technologies are unlikely to fulfill basic psychological needs for most persons. Moreover, many workers have only limited control over whether or not to use a technology, along with performing many other job requirements.

Baer et al. (2022) also assume motivational processes to be central for understanding trust. Connecting motivated reasoning theory (Kunda, 1990) with trust research, these authors argue that trustors' personal motives, defined as "conscious desires for particular outcomes" (Baer et al., 2022, p. 1561), affect their risk propensity, which in turn feed into their trust experience and related behavior. Notably, Baer et al. assume that trustors' personal motives vary considerably over time, leading to within-person variance in trust and trusting behavior. In an experience sampling study with matched trustors and trustees, the authors indeed found systematic correlations between trustors' daily motives (i.e., achievement, stimulation, affiliation, and security), their momentary risk propensity and trust, and related assessments of trustors' behavior collected from their trustees. Thus, the data demonstrate that trustors' momentary motives can influence their momentary decision to make themselves vulnerable to a trustee on a particular day.

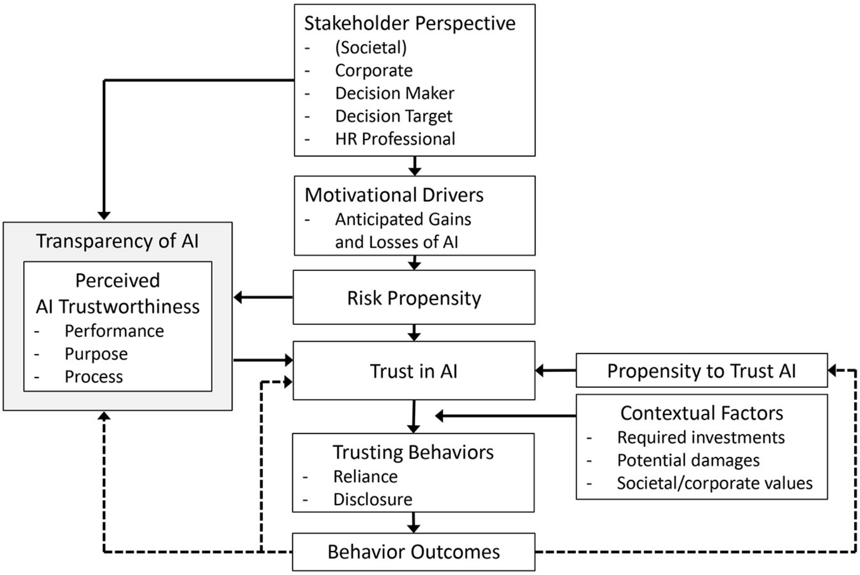
As with van der Werff et al. (2019), Baer et al. (2022) focused on interpersonal trust and remained silent on technology as a reference of trust. However, we assume that similar principles might hold for trust in technology, and for trust in AI in particular. However, we go beyond the approaches of van der Werff et al. and Baer et al. by arguing that extrinsic (as compared to intrinsic) motivation might play a more prominent role for trust in AI, and by focusing on more stable (as compared to daily changing) motives as predictors of risk propensity and trust. Indeed, the articulation of how trustors' motivation affects trust in AI and trusting behavior opens the door to theorizing how individuals occupying different roles within an organization may be more or less likely to trust in AI. Within the HRM context, the use of an AI decision aid holds different implications for need fulfillment for employers, managers, employees and job applicants, as well as HR professionals. We will consider these implications when deriving specific assumptions for different roles in HRM. Before we do so, we describe our integrated model of trust in AI that we developed based on the various precursors described earlier.

INTEGRATIVE MODEL OF TRUST IN AI

In order to conceptualize AI-related trust in HRM, we introduce an integrative model that connects both cognitive and motivational streams of research in this field (see Figure 4.1). While cognitive models of trust have already been adapted from human to technical references of trust (e.g., McKnight et al., 2011; Solberg et al., 2022), motivational trust models have only focused on human trustees so far (Baer et al., 2022; van der Werff et al., 2019). We believe that trust in technology, and trust in AI more particularly, includes both cognitive and motivational processes, particularly in more complex settings, such as HRM contexts with multiple influences on trust experience and decisions. Below, we first describe the cognitive processes of our model (lower part of Figure 4.1), considering established models on trust (Mayer et al., 1995) as well as more recent models on trust in AI (Solberg et al., 2022). Then, we describe the motivational processes of our model (upper part of Figure 4.1), which are newly developed based on extant literature on interpersonal trust and offer propositions on the connections. Finally, we apply the integrative model to core questions in the field of HRM, illustrating the potential value of the model.

Figure 4.1

Integrative Model on Trust in AI in HRM



Following the established trust literature, we consider *trust* as a dynamic experience state of the trustor connected with the willingness to assume vulnerability to a trustee, which can also be an AI system. This state of trust can be followed by specific *trusting behaviors*, during which the trustor makes her/himself vulnerable to the trustee. Extant research on interpersonal trust has suggested *reliance* and *disclosure* as major types of trusting behaviors (Breuer et al., 2020; Gillespie, 2003, as cited in Baer et al., 2022). Reliance behavior can also be observed in human interaction with AI in HRM, for instance, when HR professionals or supervisors decide to rely on an AI recommendation system for personnel selection. While saving valuable personal resources, such reliance might nevertheless be connected with certain risks for the trustor when the recommendations of the AI system are biased or wrong, leading to poor selection decisions and inferior performance of new hires, early turnover, and need for replacements. Apart from monetary losses, such outcomes would also be detrimental for the reputation of the responsible HR professional or supervisor. Disclosure behavior as a second type of trusting behavior (i.e., sharing sensitive personal information) can also come with risks in the interaction with AI in HRM. For example, employees may consent (sometimes under pressure) to share detailed and continuous data about their work processes. For instance, GPS data of truck drivers can be used by an AI system that optimizes work efficiency. One outcome of this process might be that existing time periods with lower work pressure are identified and changed by the AI system, leading to higher workload and strain for the employee. Employees subject to such performance monitoring and optimization systems might make attempts to block transmission of data, thus reducing disclosure. We consider both reliance and disclosure as broader types of trusting behaviors with respect to AI in HRM (see Figure 4.1), noting that both reliance and disclosure imply a certain level of acceptance of the AI system. When individuals are willing to rely on an AI system for decision making or disclose their data to such a system, they are using the system, and such behavior is generally considered an indicator of technology acceptance (Solberg et al., 2022). Consequently, we propose:

> **Proposition 1:** Trust in an AI system is positively related to trusting behaviors, such as reliance on the AI system or disclosure of information to the AI system.

Trusting behaviors are usually a central precondition for the successful implementation of an AI system. However, the connection between trust and trusting behaviors is not automatic but assumed to be qualified by *contextual factors* that determine the level of trust needed for trusting behaviors to occur (e.g., Mayer et al., 1995; Solberg et al., 2022). For example, if the required financial investments for an AI-based selection system are high, more trust is needed and trusting behaviors more strongly depend on trust as compared to situations when required investments are low. That is, when required investments are high, trust (e.g., on the part of top managers) must be high to lead to trusting behaviors (e.g., implementation of this AI-based system). If investment costs are low, trusting behaviors might also occur when trust is low because the financial risks are low. Other contextual factors relevant for the relation between trust in AI and trusting behaviors are the potential damages due to failures of an AI system, for instance, the costs of poor selection decisions, or the potential discrimination of job applicants after their disclosure of personal information (see Solberg et al., 2022, for other examples). The higher these potential damages are, the higher trust must be to lead to trusting behaviors.

Moreover, general attitudes towards AI-based technologies (e.g., in the organization or society) should also moderate the connection between trust and trusting behaviors. In this way, processes at higher levels of analysis (e.g., organization or society) can influence trust at lower levels (e.g., the trust—trusting behavior link for managers or employees), allowing for the examination of multi-level processes within our model. We assume that the prevailing attitudes within a worker's environment exert a downward pressure to shape and guide behavior, and this includes influencing individual use of AI (Schneider et al., 2013). For instance, societal or corporate values (such as a tendency towards tradition and conserving the status quo, versus being oriented towards change, Schwartz et al., 2012) may engender more skeptical or more positive attitudes towards AI generally held within the workforce. Whereas workers located in societies, industries, organizations, departments, and teams where attitudes about AI are optimistic may require less trust to engage in trusting behaviors, workers in environments where AI is viewed pessimistically will need to hold more trust, and trust will be more strongly associated with their trusting behaviors. More generally, we propose:

> **Proposition 2:** The relation between trust in an AI system and trusting behaviors is qualified by contextual factors, such that the relationship is stronger when contextual factors are less conducive for the AI system.

Proposition 2 differs from a related proposition made by Solberg et al. (2022) on contextual moderation, which they describe as "situational risk" and define as "person's beliefs about the extent to which relying on an AI decision aid … will result in undesirable outcomes, outside of considerations about the trustworthiness of the AI decision aid." The authors propose that the relation between trust in AI and reliance behavior "will be less positive when perceived situational risk is high" (p. 20) because, under these conditions, it is possible that persons do not use an AI system even if they trust the system. We agree that less conducive contexts for implementation of AI (e.g., high implementation costs) generally reduce the likelihood that a person adopts an AI system as compared to more conducive contexts (e.g., low implementation costs). However, less conducive contexts (e.g., high implementation costs) also require *more* trust of persons to show trusting behavior (e.g., reliance) as compared to more conducive contexts (e.g., low implementation costs), leading to the predicted higher correlation between trust and trusting behaviors in our Proposition 2.

Trusting behaviors, such as reliance on AI or disclosure of information to an AI system, should lead to specific *behavior outcomes*, such as the quality of recruitment decisions based on AI or performance indicators after applying an AI-based logistic system, which further *feedback* to the trust experience of the trustor either directly or via adjusted trustworthiness assessments and trust propensity, thereby enabling dynamic cycles and adjustments over time. Given that we assume these feedback loops (dotted arrows in Figure 4.1) to be directional and *follow* trusting behavior outcomes, our assumption can be tested in longitudinal study designs. Thus, we propose:

Proposition 3: The outcomes of trusting behaviors affect subsequent trust in AI as well as subsequent trustor's perceptions of AI trustworthiness and propensity to trust AI.

However, what are the factors that determine AI-related trust experience in the first place? In line with established trust models (e.g., Mayer et al., 1995), we assume both dispositional and perceptional influences. First, a trustor's *propensity to trust AI* should play a role with respect to AI in HRM. Propensity to trust in AI describes a very general and stable disposition to make oneself vulnerable to AI, regardless of the specific AI system or ML application, or the context (e.g., HRM). Propensity to trust AI reflects general attitudes towards technology (e.g., algorithm aversion, Dietvorst et al., 2015), which are influenced by persons' personality (e.g., Tang et al., 2022) but also reflect common stereotypes and the public discourse about AI systems in society (Fisher & Howardson, 2022; Lanz et al., 2023). Second, trust in an AI system will also be affected by the trustor's perceptions about that specific AI system. We assume that the *perceived trustworthiness of an AI system* follows similar core dimensions as described for technology more broadly (e.g., McKnight et al., 2011) or human trustees (e.g., Mayer et al., 1995). Borrowing from Solberg et al.'s (2022) model, we expect that a trustor's perceptions of an AI system's performance features (e.g., computing speed and capacity, reliability), purpose-related aspects (e.g., perceived goals of the system with respect to the interests of the trustor), and process-related aspects (e.g., adherence to general fairness norms) determine their willingness to trust the system. Together, we propose:

Proposition 4: Trust in an AI system is determined by a trustor's general propensity to trust AI and the currently perceived trustworthiness of the AI system (performance-, purpose-, and process-related aspects).

However, while Solberg et al. (2022) consider the transparency of an AI system as part of process-related trust, we argue that *transparency* is an independent feature of AI trustworthiness assessment. This is in line with earlier research considering "openness" as an additional predictor of trustworthiness assessments of human trustees (see, e.g., Mayer et al., 1995, for a discussion). More importantly, recent empirical research has provided evidence for transparency being a central trustworthiness dimension, particularly with respect to technology applications (e.g., Breuer et al., 2020; Glikson & Woolley, 2020; Vorm & Combs, 2022). We consider transparency to be particularly important for the assessment of AI as a trust object because the high complexity and opaque nature of AI systems often provides little information for trustworthiness assessments. The current discussion about the opacity of AI and potential ways to make AI understandable (XAI; e.g., Arrieta et al., 2020; Langer & König, 2021) illustrates the importance of this dimension for AI assessments.

Notably, with respect to the interrelation between transparency and the other trustworthiness dimensions of AI, we consider transparency as a qualifying factor of the other trustworthiness dimensions (performance-, purpose-, and process-based trust) because available and comprehensible information is a precondition for the trustor to assess these other trustworthiness dimensions. Thus, we argue that a person's trust in an AI system is determined by their assessment of the AI system's performance, process, and purpose, which in turn are qualified by the perceived transparency of the AI system. Notably, one implication of this conceptualization is that dispositional factors (a person's general trust propensity) become particularly strong when the transparency of an AI system is low rather than high.

> **Proposition 5:** The effect of AI trustworthiness assessments on a person's trust is qualified by AI transparency, so that effects of trustworthiness assessments on trust are higher when AI transparency is high as compared to low.

With respect to the interrelations of performance-, purpose- and process-related components of trustworthiness perceptions, specific predictions (additive, multiplicative, etc.) are difficult in light of the scarce research on interactions among ABI factors in general (Dirks & de Jong, 2022). It is plausible that purpose-related assessments (benevolence) have main effects on trust experience, given that this component relates directly to the trustor's own interests. Performance- and process-related assessments (e.g., how capable the AI system is to accomplish its goals, and how systematically the AI system follows these goals) might rather qualify and moderate the main effect of purpose-related trust assessments. For instance, a job applicant's trust in an AI-based interview assessment tool should be more strongly affected by the perceived purpose of the system, rather than its performance. In fact, if the purpose of an AI system is perceived to be in conflict with the interests of the job applicant (e.g., if the job applicant perceives the AI system as unfair), the final trust might be even lower the more capable the AI system is perceived to be. Further research on the interrelations of the three trustworthiness factors as well as their interaction with transparency is highly desirable. However, a conclusive discussion and more specific propositions are beyond the scope of this chapter.

After describing the (predominantly) cognitive processes, we now turn to the motivational aspects of our integrated model (see the upper part of Figure 4.1) and the construct of *risk propensity* that links the two parts of the model. Risk propensity is defined as a trustor's tendency to *take* risks

(Baer et al., 2022), and contributes to a more complete conceptualization of trust emergence. Whereas propensity to trust as well as trustworthiness assessments focus on the perception of a trust object (e.g., does an HR professional generally trust AI systems and perceive a specific AI system as trustworthy), risk propensity focuses on the trustor, and why they might take risks in interaction with the trust object (e.g., does an HR professional follow the recommendation of an AI-based recruiting system). Both aspects are important for a complete understanding of trust and can explain, for instance, why persons sometimes show trusting behaviors even though they do not perceive the trustee to be very trustworthy (e.g., an HR professional following recommendations of an AI-based recruiting system even though she does not find the system very reliable).

In general, risk propensity reflects a trustor's current interests and motives in relation to the trust target. While Baer et al. (2022) introduced risk propensity as a rather volatile construct, reflecting daily changes in within-person states ("motives") of a trustor, we consider risk propensity as a more stable construct that might also reflect more enduring interests of different stakeholders in a technology. A high risk propensity should stimulate more optimistic trustworthiness assessments as well as higher trust due to motivated cognition (e.g., biased memory or confirmation strategies, see Kunda, 1990). For instance, if a CEO has a high propensity to trust an AI system that identifies potential high performers, that CEO will focus more strongly on positive as compared to critical information about the AI system, with positive effects on the trustworthiness assessment and experience of trust. More formally, we postulate:

Proposition 6: The perception of AI trustworthiness and the experience of trust in an AI system are affected by a trustor's propensity to take risks in the interaction with the AI system.

Risk propensity in the context of AI applications at work is considered to be determined by *motivational drivers* of the trustor, that is, anticipated gains and losses of using the AI that reflect both interests and conscious motives of the trustor as well as external pressures within the work contexts (see Figure 4.1). Stakeholders of an AI system will have varying arrays of perceived gains and losses associated with the use of the technology that impact their motivation to trust the system. Perceived gains can be associated with the fulfillment of basic needs, such as feelings of competence or interest in the technology. However, such intrinsic motivation seems less likely to occur with trust in AI because stakeholders would need considerable computer literacy to actively participate in the development of an AI system, and to date, most HR professionals do not possess such skills (Malik et al., 2022). Instead, we argue that extrinsic motivation is much more likely to affect the drivers of trust AI in HRM at this point in time.

For instance, users might be motivated to take risks when perceiving that the goals and implicit values of an AI system align with their own goals and values. For instance, HR professionals or line managers might be motivated to develop trust in an AI system for recruiting when they perceive that this system pursues similar goals, for example, enhancing gender equality and fairness in candidate selection. By contrast, if HR professionals or line managers perceive the goals and values of the AI system to conflict with their own, for instance, when using the AI system results in discrimination of certain groups or is in conflict with accepted competency models, motivation to trust the AI system should be low. Moreover, motivation to trust AI should occur when using an AI system is instrumental for specific rewards, such as saving working time or higher efficiency of work processes. For instance, employees should be more strongly motivated to trust AI-based task management tools if these tools allow better alignment with their individual preferences.

In general, a trustor's risk propensity to trust an AI in HRM should be a consequence of the currently *anticipated gains and losses* connected with the implementation of an AI system. This rather broad conceptualization of anticipated gains and losses allows us to aggregate different types of motives in a holistic process. For instance, HR professionals might perceive that an AI-based selection system implicitly pursues similar goals as themselves (e.g., high fairness in selection), but at the same time anticipate high additional workload to advertise and implement the system. Individualized weightings of these different motives could lead to an overall low risk propensity toward this system if the workload is a stronger motivator than the similar goals. Thus, our conceptualization of anticipated gains and losses enables an examination of different motivational drivers to help understand how people approach trust of an AI system. Therefore, we suggest:

> **Proposition 7:** A person's risk propensity to trust an AI system is affected by their motivational drivers, defined as anticipated gains and losses experienced through the AI system.

Moreover, we postulate that the anticipated gains and losses of an AI system vary systematically between different groups of involved persons. Specifically, we assume that a major part of expected gains and losses connected with AI in HRM systematically differ between different stakeholder groups in an organization (e.g., employer, supervisor, employee, job applicant, HR professional). For instance, whereas corporate boards and managers should particularly expect monetary gains and efficiency from an AI-based personnel selection system, job applicants might instead be concerned with whether they will be treated fairly. HR professionals, as a third stakeholder group, might expect support from the AI system in their daily work but also have concerns about being replaced by the AI system in the long-term. Thus, we postulate:

Proposition 8: Stakeholders differ in their expected gains and losses when interacting with an AI system, leading to differences in their risk propensity to trust an AI system.

In the next section, we elaborate on the different stakeholder perspectives on anticipated gains and losses of AI adoption in HRM in more detail. In doing so, we also discuss how these motivational drivers might affect trustors' risk propensity, as well as subsequent trust and trustworthiness assessment with respect to AI in HRM.

Stakeholder Perspectives and Trust in AI in HRM

Our integrated trust model addresses the different perspectives that trustors might have toward the use of AI, which should affect both cognitive and motivational processes (i.e., trustworthiness assessments and motivation to trust). Applied to HRM, stakeholders will receive and process different information about an AI-system because they have different access to and interaction with the system. Research has consistently demonstrated that different stakeholder groups have different experiences with HR technology and, consequently, a multi-stakeholder approach is recommended (Bondarouk et al., 2017). For example, managers and HR professionals receive different details about the reliability of an AI system than employees or job applicants. These different information backgrounds should directly influence trustworthiness assessments according to our model (see Figure 4.1). Moreover, stakeholders should also vary with respect to expected or perceived gains and losses connected with an AI-based system, leading to different motivations to develop trust in the AI. These motivational processes should also influence trustworthiness assessments as well as trust according to our model.

Within the field of HRM, we distinguish four core stakeholder perspectives: Employer, decision maker, decision target, and HR professional. For simplicity, we restrict our model to the individual level. Each stakeholder perspective refers to the perspective that a given individual *within this particular role* in HRM is likely to hold.

First, the employer perspective reflects the management or organizational view on the implementation of AI systems in HRM, in particular expected strategic gains such as higher decision accuracy and speed, but also potential savings of workforce cost. This perspective represents the viewpoint of individuals within upper-level management positions who are more likely to be concerned with these broader considerations for AI adoption. The second perspective, decision makers, represent individuals who are responsible for making HR decisions in organizations, such as hiring, promoting, or coaching, but are not within the HR profession themselves. These may be supervisors, team leaders, line managers, or mid-level managers, all of whom are charged with making decisions about utilizing and developing talent. The third perspective is the decision target, which includes both current and potential new employees (i.e., job applicants). This perspective reflects potential benefits (e.g., higher fairness and procedural justice) but also risks due to lacking transparency (opacity) and "cold" automatic decision making. Finally, the fourth perspective reflects the HR professional viewpoint, representing experts in staff positions in various HRM fields such as recruiting, selection, and training but also in strategic HRM. HR professionals' trust in AI systems is a major determinant of a successful implementation process and needs to be examined in more detail outside of the more general user role posed by Langer and König (2021) and Lockey et al. (2021). HR professionals are likely to have complex and potentially conflicting perspectives on AI systems in HRM as they provide both benefits and challenges.

While the managerial and employee perspectives have already been addressed in respect to effects of opacity (Langer & König, 2021), the perspective of HR professionals has received far less attention. There is overlap between the roles of managerial decision makers and HR professionals in how they use AI systems. However, HR professionals as a user group are likely to have both more breadth and depth in their interactions with such AI systems than other types of managers. For example, HR professionals are likely to play a central role in the implementation process of AI systems as their work can include the selection of the AI system, the adjustment of the AI system to the specifics of the organization (e.g., access to recruiting data and job descriptions), promoting the new system within the workforce, and validating the effectiveness of the system. These perspectives are explained in more detail below and summarized in Table 4.1.

Employer Perspective

This stakeholder perspective reflects upper-level management, representing the strategic interests of a corporation or employer. In this sense, the employer perspective shares features with the deployer group described by Langer and König (2021), as they are ultimately responsible for making choices about which technology systems to implement. Employers

Table 4.1

Perceived Gains and Losses and Implications for Trust for Stakeholders of AI Tools in HRM

	Stakeholders of AI Tools in HRM			
	Employer	**Decision Makers**	**Decision Target**	**HR Professionals**
Perceived Gains	• Increased competitiveness in hiring • Decision-making speed • Employer branding benefits	• Reduced responsibility for HR decisions • More time for other tasks	• Clear rules and high consistency in HR decisions • Avoidance of decision biases (attractiveness, liking, etc.) • Voice • Adaptions of individual working conditions	• Relief from routine tasks, more time for other tasks • Support for complex tasks and better decision-making • Advancing own skills • Reduced responsibility for some HR decisions
Perceived Losses	• Monetary investments • Non-financial costs • Potential poor or biased decisions • Potential lawsuits	• Potential displacement by technologies • Reduced control over decisions	• Opaque and potential unfair decision making, algorithmic reductionism • Loss of privacy and control over personal data	• Loss of professional identity • Learning costs • Change management processes • Potential displacement by technologies
Risk Propensity Trend	Positive	Positive	Negative	Mixed, many moderators

are particularly likely to view strategic benefits of implementing AI systems in HRM, such as staying competitive in hiring new employees, increasing speed in making HR decisions, and bolstering employer branding by presenting the firm as modern and updated in its use of technology. On the negative side (i.e., potential losses) are the investment costs, financial and otherwise, of acquiring and using AI technology for HR functions. Use of new technology in HRM requires time to implement, often causes disruptions in routines and processes, and places additional demands on employees to learn the new tools. Upper-level management will be concerned with the resources and capacity to use technology in the company, expected productivity gains, and government regulations for making the decision about whether to adopt AI (Malik et al., 2022; Pan et al., 2022).

We expect that employers weigh these different gains and losses for a specific AI system. Depending on the net result, employers should then proceed with a positive or negative risk propensity toward the AI system, which in turn should influence the following trustworthiness assessments and experienced trust. For instance, if a corporate board committee focusses particularly on the positive prospects of an AI system in HRM, a resulting high-risk propensity might stimulate motivated reasoning through which individual committee members pay less attention to or even ignore negative information about the system. Further, once the organization has devoted substantial resources to acquiring and integrating an AI system, employer stakeholders may take a "sunk costs" perspective (Arkes & Blumer, 1985) and continue trusting the system over time, even if direct experience with the system suggests that trust should be reduced as part of the trust regulation feedback loop. Thus, employer stakeholders might trust the system too much, or continue trusting it for too long, if the sunk costs are high. This type of overreliance on AI systems may lead to negative outcomes (Glikson & Woolley, 2021) if the system is not performing effectively.

However, if an employer focusses more on the implementation costs or potential risks that the organization becomes the target of legal action (for making unexplainable employment decisions or for violating emerging laws regulating the use of AI decision-making tools), the resulting low risk propensity should lead to greater attention on critical information about the system. It should be noted, though, that we consider the implementation costs of an AI system not only as a predictor of trustor's risk propensity with respect to the AI system itself (among other losses and gains), but also as a moderating context factor of the link between trust and trusting behavior later in the process (see Figure 4.1). For instance, when the financial cost of an AI system is high, employers should not only be hesitant to trust the AI system (low risk propensity), for example, by focusing on negative information, but also should need a high level of trust in the AI system to express reliance through a purchase decision (moderating context factor).

Inexpensive solutions, or systems acquired through software-as-a-service contracts, may not require as much trust from the corporate stakeholder as the overall risk is lower. Thus, the high costs of an AI system should complicate its implementation not only by making a positive trust motivation and risk propensity more difficult, but also by raising the bar of required trust from an employer. While these twofold effects of implementation costs are unique for the employer perspective (decision makers and HR professionals are usually not directly responsible for such investment decisions), the ability to tease apart different motivational and contextual factors is one of the contributions of our multi-stakeholder model.

Decision Maker Perspective

The decision maker stakeholder group represents managers and supervisors at a variety of levels in an organization who are responsible for using AI systems for making HR decisions, such as hiring new workers, deciding who to promote, or deciding who should participate in training or development. These decision makers may need to figure out how to incorporate AI recommendations into their own choices, which can be difficult when there is a lack of transparency (Chowdhury et al., 2022). Another potential loss for members of this decision maker group is being replaced by the technology in this aspect of their jobs. Apart from job loss in the worst case, being replaced by technology means loss of power and control, such as control over HRM decisions and the ability of managers to hire or promote employees for other reasons than regular organizational policies (for instance, difficult family situation of the employee, strategic or political developments within the department, e.g. König et al., 2010). If AI systems can make relatively autonomous decisions in these matters, human decision makers have reduced responsibility in this area. However, this loss of control is not always negative for decision makers but could be also perceived as a gain, for instance, when the AI system releases decision makers from routine tasks. Indeed, Hertel et al. (2019) have shown that users of a management information system not only save cognitive resources when trustfully relying on the information system ("directed forgetting"), but also report higher well-being as compared to users who did not trust the information system. Similar effects might be expected with AI-based information systems. Little research has been conducted on this specific question, with some initial evidence suggesting that the type of AI system will affect how satisfied decision makers are with their role in making hiring decisions (Langer et al., 2021).

We argue that decision makers might actually prefer being replaced in some contexts. If they can rely on AI systems to make hard decisions, the managers do not have to be responsible for the decision outcomes, can make external attributions for poor quality or unpopular decisions (i.e., blame the technology), and may be able to refocus their efforts on other parts of the job. Therefore, we argue that the managerial decision makers will have a relatively high-risk propensity toward the use of AI systems in HRM that reduces their own decision-making responsibility. This risk propensity facilitates a motivation to trust in these systems. In addition, the technological affinity of the organization or its developmental orientation might also influence whether decision makers perceive being partly replaced by an AI system as a gain or loss.

Decision Target Perspective

The next stakeholder group is the individuals who are the targets of decisions being made in the HR context, such as employees and job applicants. This group has been studied extensively from the perspective of understanding how they are likely to react to AI technologies and when they perceive such technologies as fair (e.g., Langer & Landers, 2021; Newman et al., 2020; Höddinghaus et al., 2021). The decision targets are likely to experience many more potential losses than gains. There is the risk of not understanding how decisions are made, and therefore the perceived risk of not being hired or promoted due to an unfair or discriminatory system. There is a risk of losing privacy and of losing ownership of personal data and how it is used (Chowdhury et al., 2023). Perceived gains could include a more consistent hiring process and potential for reduced bias and discrimination. Research supports that applicants appreciate the consistency element of AI, in which all applicants are treated the same, but show concerns about the extent to which AI can consider them as a unique human ("algorithmic reductionism"; Newman et al., 2020) and perceive AI-based decision making as less fair than human decision making (Acikgoz et al., 2020). In addition, research has identified some potential gains for employees that are different from those of applicants. For example, employees may experience greater engagement with the use of AI chatbots, as chatbots can grant them more opportunities for voice, fostering the organizational climate for trust (Dutta et al., 2022). The use of AI also allows for HR processes to be individually tailored to employees, which may positively relate to their job satisfaction and commitment, and negatively to turnover intentions (Malik et al., 2022).

In sum, employees and applicants might be more likely to have an explicit, controlled motivational pattern characterized by a low-risk propensity. The use of AI decision-making tools offers few direct gains as perceived by this decision target stakeholder group. While job applicants, for example, may benefit directly from some technologies that employ AI, such as asynchronous video interviews (Basch & Melchers, 2019), these benefits largely arise from the use of asynchronous communication technologies rather than from the AI analysis and decision support features. Yet because applicants and employees have lower control regarding these technologies, the decision targets are nevertheless likely to cooperate by relying on them and disclosing information to them, even with relatively low levels of trust.

These interests of protecting privacy and being able to understand how the technology is used are even more strongly expressed by representatives of employees, such as unions and works councils. We observe that these representative bodies are often fighting against the implementation of AI systems (e.g., Quackenboss & Meisburg, 2020; Trades Union Congress, 2021) in order to protect the individuals. Applicants, on the other hand, must rely on society more broadly, a stakeholder group discussed by Lockey et al. (2021), to protect their rights and interests regarding the use of AI systems in the hiring processes. Unions and works councils may also have an impact on the employer stakeholder perspective given their role in technology adoption decisions affecting the workforce, or even in a more regulatory role (Langer & König, 2021), but their interests are more clearly aligned with the decision target stakeholders.

HR Professional Perspective

HR professionals are a complex stakeholder group, as they inhabit multiple roles in relation to the use of AI systems. They are in some respects decision makers, using the software to make decisions about other people. They may serve as domain experts (Lockey et al., 2021) on the processes for which AI systems are being implemented, change management agents in the technology deployment role (Langer & König, 2021), or even in a regulatory role from an internal viewpoint, helping to ensure that organizations are using AI systems consistent with existing laws and regulations. Complicating matters, expertise in AI is relatively low among HR professionals (Malik et al., 2022). These varied roles result in conflicting viewpoints and complex motivation patterns that make it difficult to predict resulting levels of risk propensity and trust in AI. Therefore, we choose to examine the unique trust motivation concerns of the HR professional group directly.

On the side of potential gains due to AI systems, HR professionals may welcome these tools as providing support and relief in routine or complex tasks (e.g., Laurim et al., 2021; World Economic Forum, 2021), thus providing HR professionals leeway for more creative and strategic activities. Some HR professionals want to learn more about the technology and how to use it, advancing their own skills in the profession. There is also potential for gains related to outcomes aligned with their own goals and values. Similar to the strategic stakeholders, they may see gains in shorter time to fill positions, higher accuracy and fairness of selection decisions, more diverse new hires, and a modern organizational image that would encourage them to be open to taking more risks with AI systems. Moreover, HR professionals might perceive AI systems as technical support that releases them from routine tasks and provides more time for qualitative social interactions with employees or job applicants.

However, there is also substantial potential for perceived losses for HR professionals in the use of AI systems that could reduce motivation to trust. There are risks of being replaced by the technology in this aspect of their jobs, although these risks seem low given the historical context of other HR technologies that were accompanied by unrealized predictions that these technologies would replace HR professionals. More direct is a potential loss of professional identity that HR professionals may experience (Lockey et al., 2021). For example, HR professionals who have been heavily involved in recruiting may have defined their role as representing the organization through direct contact with applicants. When such interactions are instead handled by AI-based chatbots that can respond to applicant questions and conduct initial assessments, HR professionals may experience a loss of identity. The risk of reduced autonomy in some aspects of decision making could occur, particularly if AI systems are more autonomous rather than supporting. There are also reputational risks (Lockey et al., 2021) involved in adoption of AI systems, and these risks potentially conflict. An HR professional who tries to avoid using AI may experience a reputation risk of not being a modern professional, while one who embraces AI is taking on risks of using technology that results in poor decisions. Other risks are associated with change management, such that target users (such as the decision maker and decision target groups) reject the technology and are unwilling to use it. Supporting tasks may increase and be perceived as another loss for HR professionals, as users and decision targets have more questions about processes and results and require more explanation.

With the HR professionals, we see a complex mix of motives. There are some similarities between the corporate/employer and the decision maker groups that could result in positive risk propensities, particularly when HR professionals serve as decision makers themselves. However, there is also a set of risk factors that should result in a more hesitant risk propensity. The final direction might be a function of moderating factors in how HR professionals view gains and losses regarding AI. Attitudes and skills related to technology in general, and AI in particular, are likely to affect motivational drivers for HR professionals, such that those HR professionals with more experience and higher self-efficacy for using complex technologies may view more gains, even perceiving the opportunity to learn how to use AI technologies as positive and interesting. Because HR professionals as a group are likely required to dedicate more time and personal resources to learning how to use AI systems, age and time horizon within an organization should also have an impact. Meta-analytic evidence suggests that chronological age is negatively related to acceptance of new technologies, primarily through a negative relationship with ease of use (Hauk et al., 2018). This suggests that older HR professionals may need more training in order to develop trust with AI systems. Further, and related to age, is the concept of future time perspective (e.g., Zacher & Frese, 2009), or the perception of how much time is remaining for an individual relative to a particular pursuit. If an individual's future time perspective for working is low, the costs of needing to learn complex new AI systems might be more salient. While these moderators should be particularly relevant for HR professionals, they might also be relevant for other stakeholder groups to a lesser degree.

Societal Perspective

In addition to these four core stakeholder perspectives, we have noted several times in this discussion the relevance of the societal perspective and how this may play a role in individuals' trust in AI. For instance, values related to risk propensity, transparency, and tolerance of risk from technology vary also at the country level and can affect trust decisions about AI tools. Concerns about privacy losses are higher in many European countries than in the United States or China, and values of trust, openness, and transparency are particularly high in Nordic countries (Robinson, 2020), leading to differences in governmental policies. Some societies are motivated to accept greater individual losses from AI in exchange for increased business opportunities or perceived safety and security (e.g., Kostka et al., 2021). Notably, these differences not only affect the adoption of AI systems but—perhaps even more importantly—the access to learning data to develop AI systems in the first place. Legislation on development and use of AI may signal potential losses, but also help assure trustors that the technology is being regulated and, therefore, more trustworthy.

DISCUSSION

Artificial intelligence (AI) applications are increasingly used in HRM and have the potential to change this field dramatically. However, one core precondition for successful usage and integration of AI-based technologies is that the involved persons place sufficient trust in these new technologies, for instance, by following recommendations of AI-systems, allowing AI-technologies to use their personal data, or deciding to implement an AI system in HRM in the first place.

We introduce a new model of trust in AI as a potential framework of current and future research on this topic. In this model, we integrate various research lines that have been partly developed in isolation and/or only in the field of interpersonal trust but not trust in technology. First, our model includes classic *dispositional factors* (propensity to trust AI) and *cognitive processes* of trustworthiness assessments with respect to the target of trust, building on seminal general trust models (Mayer et al., 1995) as well as more recent models on trust in AI (Solberg et al., 2022). In addition, our model also considers context factors that moderate the connection between trust experience and trusting behaviors, acknowledging that trust in AI is not always followed by implementation or usage of an AI system, and also that implementation and usage of an AI system is not always accompanied by high trust, for instance, when implementation costs are very low.

Second, we also consider the *motivational processes* of trust (Baer et al., 2022; van der Werff et al., 2019) that have been developed for interpersonal relations but not yet outlined with respect to trust in technology or trust in AI systems. In this respect, our model provides a more inclusive framework for our understanding of trust in AI and provides the first outline of stable trust motives (as compared to Baer et al. 2022, who addressed daily changing motives). Third, our model provides a more differentiated perspective on trust in AI by considering *stakeholder perspectives* with respect to AI usage. While such perspectives have been already discussed in extant work (e.g., Bach et al., 2022; Lockey et al., 2021; Langer & König, 2022), our integrative model allows a more precise analysis of these perspectives. Specifically, our model allows examination of cognitive and motivational processes based on the stakeholders' different access to information (e.g., on trustworthiness or risks of an AI system) and different expectations of gains and losses connected to AI usage. Finally, our model is complemented by various *feedback mechanisms* that cover dynamic processes and changes over time and allows for *cross-level processes* by which factors at higher levels (e.g., societal or corporate values) affect trust in AI at the lower level of the individual, via contextual risk factors. By applying this new integrative model of trust in AI to the field of HRM, we illustrate the analytic depth and potential of this model. Based on our model, we suggest some plausible general trust tendencies of stakeholder groups resulting from different general motivation processes. However, due to the complexity of the motivated trust processes in this context, we suggest explorative research as a first step to better establish how the stakeholder differences play out in practice.

Another unique feature of our model is that we do not assume direct linkages between system features and trust motivation. Rather, we argue that the impact of system features depends on the gains and losses (and their combination) that stakeholders expect as a consequence of AI implementation and usage. The psychological processes and perceptions are the focus of the model rather than the features of the technology (see also Bach et al., 2022). The consideration of different gains and losses may also help advance understanding of how trustors develop overreliance (i.e., acceptance of incorrect recommendations) and underreliance (i.e., neglect of correct recommendations) with respect to AI systems (e.g., Buçinca et al., 2021). A focus on gains only would likely lead to overreliance, or trusting too much, while a focus on losses would result in underreliance, or trusting too little. Our model suggests that stakeholders coming from the employer and decision maker perspectives may be more susceptible to overreliance, whereas stakeholders coming from the decision target perspective may be more susceptible to underreliance, in contrast to suggestions that all stakeholder groups may be vulnerable to overreliance on AI (Lockey et al., 2021).

Finally, while prior models of trust considered risk perceptions either as a context factor moderating the link between experienced trust and trusting behaviors (Mayer et al., 1995), or as a direct influence on trust motivation (Baer et al., 2022), we consider risk perceptions at two steps in our model, thereby integrating prior approaches. First, risk perceptions can refer to the trust target itself (an AI system) reflecting a trustor's risk propensity with respect to the trustee. While low risk propensity and related trust motivation should hamper trust development even for high trustworthy AI systems, trustors can place (and experience) trust even in an unreliable AI system if their risk propensity and trust motivation towards an AI system is high. Second, risk perceptions can also refer to assessments of the conduciveness of context factors (organizational, societal, etc.) for trusting behaviors. If the context is perceived as not conducive, trusting behaviors, such as implementing an AI system or relying on recommendations from the AI system, might not be shown even when trust in the AI system is high. Thus, high trust does not automatically lead to trusting behaviors if contextual risks, such as required financial investments, are perceived as being too high. At the same time, low trust in an AI-system can still be followed by trusting behaviors if contextual risks are perceived as very low. However, the latter is not likely in HRM, where AI applications affect working people.

We note that trust theories are not the only approach for understanding and predicting use of AI systems in HRM. The technology acceptance model (TAM) from the information systems literature also offers useful perspectives on the use of these tools (e.g., Del Giudice et al., 2023; Vorm &

Combs, 2022), suggesting that perceived usefulness and perceived ease of use of technology predict attitude toward the technology, which then leads to usage behavior. Our model addresses the main components of TAM in a different configuration. We address two constructs that are similar to perceived usefulness. The gains that stakeholders expect from using an AI tool could be construed as perceived usefulness, particularly those gains that are related to enhanced job functioning (e.g., HR professionals can more efficiently identify and communicate with top job applicants). Moreover, the purpose-based component of trustworthiness deals with the goals of the system and stakeholders could determine how useful those are. Researchers have examined alternative relationships between ease of use and trust, with some arguing they are relatively unrelated predictors of attitude (Dickson et al., 2021) and others arguing that ease of use could be a predictor of trust (Beldad & Hegner, 2018). In our framework, the potential costs of using AI could be represented by ease of use, such that a system that is very difficult to use presents a loss because of extra effort or added difficulty. Overall, we argue that the active management of trust and AI acceptance in a motivation-based self-regulatory perspective offered by our framework can be more informative than a strictly TAM-based perspective that tends to be more static.

Open Questions and Implications for Future Research

The introduced new model of trust in AI and its application in the field of HRM provides a roadmap for further research on this topic. First and foremost, many of the proposed trust processes still await empirical investigation. In such research, studies on existing AI systems (e.g., Fan et al., 2023) would be preferable to studies using only simulated (vignette) applications of AI in HRM contexts. Moreover, our theorizing about trust in AI systems is based on the understanding that AI in HRM is still relatively constrained, such that most algorithms do not engage in autonomous self-improvement. However, trust in AI systems could change when self-improving algorithms become more available. On one hand, there is evidence that algorithms with a self-learning capability are viewed as more human (Kim & Duhachek, 2020) and therefore more trustworthy. On the other hand, if an algorithm can continue to learn and change without human direction, the algorithm may be viewed as even more opaque and less trustworthy, as a user would not know when changes are made, nor what the changes are. Similarly, we have not addressed the complexity inherent in hybrid human-AI systems where the human and the AI work together (see Kares et al., 2023; Lanz et al., 2023). When it is unclear which actions are taken or recommendations are given by the AI compared to the human, it will be more difficult for a trustor to assess the perceived trustworthiness of a system, complicating the motivational and cognitive processes in the model.

In order to manage the complexity of our model, we have not explicitly included within-person changes in affective states or motivation. Indeed, empirical work has shown that a trustor's emotions can influence trust (see Dirks & de Jong, 2022, for a review). The seminal approach to trust by McAllister (1995) even explicitly distinguishes cognitive and affective precursors of trust (see also Glikson & Woolley, 2020, for an application of this conceptualization to trust in technologies). While affective cues can be integrated in cognitive models as cues for a "mood as information" heuristic (see Dirks & de Jong, 2022), daily affective states nevertheless can "color" assessment processes of trustees as well as context factors, and thus moderate trust processes.

In addition to changes in risk propensity, trustors' affective or motivational states can also affect the degree of heuristic or systematic information processing (e.g., Shin, 2021) with respect to deciding whether to trust an AI or not. For instance, positive affective states might signal that the current situation is safe, allowing heuristic decision making (e.g., following AI recommendations because others are doing so) whereas negative affective states might lead to systematic information acquisition about the reliability and benevolence of the AI system. In a similar way, daily changes in the salience of trustors' motives (Baer et al., 2022) can affect their momentary risk propensity and assessment of risks. We did not elaborate such daily influences in our model for space constraints, however, momentary affective states and the current salience of specific motives should be a valuable avenue for future research on trust in AI in HRM.

Stable individual differences between trustors could have major influences on trust in AI for HRM. Tang et al. (2022) found that individuals with high levels of conscientiousness were more likely to have negative reactions (i.e., higher role ambiguity, lower role breadth self-efficacy) to working with AI systems than people low on this trait. They argued that highly conscientious employees are motivated to control work processes and may be less likely to see information coming from AI systems to be credible. In our model, high conscientiousness should lead stakeholders to perceive AI usage as a loss, thus reducing motivation to trust AI tools. However, empirical tests are still needed.

Finally, within our model, we propose that higher-order influences, such as societal and corporate values, will inform the relation between trust in AI and trusting behaviors. There are further opportunities to conceptualize how, exactly, levels, such as team- and organization-level factors (e.g., unit-level motivation and trustworthiness perceptions, climate for innovation) inform individual stakeholders' motivation and trust in AI. Indeed, social pressure has been theorized to drive perceptions of the usefulness of technology and acceptance (Del Giudice et al., 2021). From a policy perspective, the availability of programs to help people whose jobs are replaced or radically changed by the implementation of AI systems could positively impact trust by minimizing the perceived losses of adopting AI.

Implications for HRM Practice

Even though many of the proposed processes of our model still await empirical tests, we nevertheless can derive initial practical implications as guidelines for designing, implementing, and managing AI systems in HRM. Below, we describe such implications, particularly for HR professionals, following the logic of our integrative model of trust in AI. In so doing, we focus on developing high trust in AI, however, it should be noted that overreliance on AI technology can be as dangerous as underreliance (Buçinca et al., 2021), thus, appropriate trust in AI (or "trust calibration," e.g., Glikson & Woolley, 2020) would be the ultimate goal.

AI Trustworthiness Assessment

A future-oriented perspective on HR suggests that stakeholders, in particular HR professionals, suffering from a lack of knowledge about AI (opacity due to illiteracy) should be encouraged to learn more about how AI systems are developed and how they function. Developing this expertise is likely to allow them to make more accurate judgments about the performance, purpose, and processes involved in an AI system. HR professionals with higher levels of expertise may also increase their ability to identify AI tools that are a good match for their own goals within the organization, increasing perceived gains of AI adoption, and, as a result, motivation to trust in this AI. Based on our arguments about the importance of transparency for a sound assessment of AI systems, other aspects of opacity also should be addressed. Intentional opacity (Langer & König, 2021) is a more difficult practical challenge to overcome, as vendors or developers obscure information about how their tools work in an effort to maintain competitive advantage. This could rely on societal level solutions, such as requirements for algorithmic auditing, or pressure from managerial and corporate stakeholders to demand such information prior to making purchase decisions.

Motivational Drivers and Risk Propensity

For effective change management, given the adoption of AI tools, organizations can draw from our theorizing to promote gains and account for perceived losses connected with AI usage separately for each relevant stakeholder group (cf. Table 4.1). In this way, the salient opportunities and concerns held by different stakeholders can be accounted for, so that AI may be accepted and effectively introduced. Openly recognizing potential losses as well as embedded transparency tools (XAI) may help guard against overreliance. Organizations may also seek to provide leeway and flexibility for individuals to craft working conditions connected with AI systems. For example, HR professionals could help decision makers identify where an AI decision support tool may be most aligned with their values, such as increasing equal treatment in making decisions about training opportunities and allowing them to start by using only that function. This greater goal alignment can allow decision makers to perceive greater gains from using AI, thus making them more likely to develop trust in the tool.

Context Factors

The incorporation of context factors in our model also provides an indication of the societal as well as organizational contexts that may be conducive to AI implementation. Even if stakeholders trust an AI system, they will be more likely to engage in trusting behaviors of reliance and disclosure if the context factors are supportive. For example, work environments characterized by high psychological safety (Newman et al., 2017) may be beneficial to encourage use of AI systems, making it safer to take risks related to the trusting of AI. High levels of psychological safety could help mitigate fear about potential losses related to low decision quality during early phases of AI implementation and encourage stakeholders to identify errors, ultimately leading to improvements in the use of such systems that could actually increase trust over time. We also believe that our model is not restricted to HRM but should be also relevant for other fields in which trust of AI is a relevant concern, such as customer relationship management or managerial decision making (e.g., Keding & Meissner, 2021).

CONCLUSION

Tools based on artificial intelligence are an increasing part of the HRM landscape and if current projections are correct, will continue to become a more important part of organizational HRM processes and procedures. The integrated trust model presented in this chapter makes several contributions to the emerging literature in this area. We integrate cognitive and motivational trust processes, providing a more inclusive framework for our understanding of trust in AI. We also examine the impact of AI on a broad range of HR stakeholders, thus providing a framework for better understanding how different stakeholders of HRM in organizations develop trust in these systems and decide how to behave when faced with such interactions. This new model provides directions for future research in this rapidly evolving field.

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*(4), 399–416. https://doi.org/10.1111/ijsa.12306

Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus, 15*(2). https://doi.org/10.7759/cureus.35179

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, *35*(1), 124–140.

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/https://doi.org/10.1016/j.inffus.2019.12.012

Bach, T. A., Khan, A., Hallock, H., Beltrão, G., & Sousa, S. (2022). A systematic literature review of user trust in AI-enabled systems: An HCI perspective. *International Journal of Human-Computer Interaction*. https://doi.org/10.1080/10447318.2022.2138826

Baer, M. D., Sessions, H., Welsh, D. T., & Matta, F. K. (2022). Motivated to "roll the dice" on trust: The relationships between employees' daily motives, risk propensity, and trust. *Journal of Applied Psychology*, *107*(9), 1561–1578. https://doi.org/10.1037/apl0000959

Basch, J. M., & Melchers, K. G. (2019). Fair and flexible?! Explanations can improve applicant reactions toward asynchronous video interviews. *Personnel Assessment and Decisions*, *5*(3), 2. https://doi.org/10.25035/pad.2019.03.002

Beldad, A. D., & Hegner, S. M. (2018). Expanding the technology acceptance model with the inclusion of trust, social influence, and health valuation to determine the predictors of German users' willingness to continue using a fitness app: A structural equation modeling approach. *International Journal of Human–Computer Interaction*, *34*(9), 882–893. https://doi.org/10.1080/10447318.2017.1403220

Blanchette, I., & Richards, A. (2010). The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Cognition & Emotion*, *24*, 561–595. https://doi.org/10.1080/02699930903132496

Bondarouk, T., Parry, E., & Furtmueller, E. (2017). Electronic HRM: four decades of research on adoption and consequences. *The International Journal of Human Resource Management, 28*(1), 98–131. https://doi.org/10.1080/09585192.2016.1245672

Braganza, A., Chen, W., Canhoto, A., & Sap, S. (2021). Productive employment and decent work: The impact of AI adoption on psychological contracts, job engagement and employee trust. *Journal of Business Research*, *131*, 485–494. https://doi.org/10.1016/j.jbusres.2020.08.018

Breuer, C., Hüffmeier, J., Hibben, F., & Hertel, G. (2020). Trust in teams: A taxonomy of perceived trustworthiness factors and risk-taking behaviors in face-to-face and virtual teams. *Human Relations*, *73*(1), 3–34. https://doi.org/10.1177/0018726718818721

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–21. https://doi.org/10.1145/3449287

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

Carver, C. S., & Scheier, M. F. (2012). *Attention and self-regulation: A control-theory approach to human behavior*. Springer Science & Business Media.

Chapman, D. S., & Gödöllei, A. F. (2017). E-recruiting: Using technology to attract job applicants. *The Wiley Blackwell handbook of the psychology of the Internet at work*, *7696*, 213–230. https://doi.org/10.1002/9781119256151.ch11

Chowdhury, S., Dey, P., Joel-Edgar, S., Bhattacharya, S., Rodriguez-Espindola, O., Abadie, A., & Truong, L. (2023). Unlocking the value of artificial intelligence in human resource management through AI capability framework. *Human Resource Management Review*, *33*(1). https://doi.org/10.1016/j.hrmr.2022.100899

Chowdhury, S., Joel-Edgar, S., Dey, P. K., Bhattacharya, S., & Kharlamov, A. (2022). Embedding transparency in artificial intelligence machine learning models: Managerial implications on predicting and explaining employee turnover. *The International Journal of Human Resource Management*, 1–32. https://doi.org/10.1080/09585192.2022.2066981

Colquitt, J. A., Scott, B. A., & LePine, J. A. (2007). Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance. *Journal of Applied Psychology*, *92*(4), 909–927. https://doi.org/10.1037/0021-9010.92.4.909

Chugunova, M., & Sele, D. (2020). We and it: An interdisciplinary review of the experimental evidence on human-machine interaction. *Center for Law & Economics Working Paper Series*, *12*. https://doi.org/10.3929/ethz-b-000442053

Del Giudice, M., Scuotto, V., Orlando, B., & Mustilli, M. (2023). Toward the human–centered approach. A revised model of individual acceptance of AI. *Human Resource Management Review*, *33*(1). https://doi.org/10.1016/j.hrmr.2021.100856

DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems*, *19*(4), 9–30. https://doi.org/10.1080/07421222.2003.11045748

Dickson B. U., Oby B. O., Samuel N. N., & Udoka S. O. (2021). Integrating trust into Technology Acceptance Model (TAM), the conceptual framework for E-Payment platform acceptance. *British Journal of Management and Marketing Studies 4*(4), 34–56. https://doi.org/10.52589/BJMMS-TB3XTKPI

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dirks, K. T., & de Jong, B. (2022). Trust within the workplace: A review of two waves of research and a glimpse of the third. *Annual Review of Organizational Psychology and Organizational Behavior*, *9*, 247–276. https://doi.org/10.1146/annurev-orgpsych-012420-083025

Dutta, D., Mishra, S. K., & Tyagi, D. (2022). Augmented employee voice and employee engagement using artificial intelligence-enabled chatbots: A field study. *The International Journal of Human Resource Management*, 1–30. https://doi.org/10.1080/09585192.2022.2085525

Eisbach, S., Daugs, F., Thielsch, M. T., Böhmer, M., & Hertel, G. (2022). Predicting rating distributions of website aesthetics with deep learning for AI-based research. *ACM Transactions on Computer-Human Interaction (TOCHI)*. https://doi.org/10.1145/356988

Fan, J., Sun, T., Liu, J., Zhao, T., Zhang, B., Chen, Z., Glorioso, M., & Hack, E. (2023). How well can an AI chatbot infer personality? Examining psychometric properties of machine-inferred personality scores. *Journal of Applied Psychology*. Advance online publication. https://doi.org/10.1037/apl0001082

Ferrario, A., Loi, M., & Viganò, E. (2020). In AI we trust incrementally: A multilayer model of trust to analyze human-artificial intelligence interactions. *Philosophy & Technology*, *33*, 523–539. https://doi.org/10.1007/s13347-019-00378-3

Fisher, S. L. & Howardson, G. (2022). Fairness of AI in HR: Held to a Higher Standard? In S. Strohmeier (Ed.), *Handbook of Research on Human Resource Management and Artificial Intelligence* (pp. 303–322). Edward Elgar.

Gillespie, N. (2003). *Measuring trust in work relationships: The behavioral trust inventory* [Paper presentation]. Annual meeting of the Academy of Management, Seattle, WA.

Gillespie, N., Lockey, S., & Curtis, C. (2021). Trust in artificial intelligence: A five country study. (2021). *The University of Queensland and KPMG Australia*. https://doi.org/10.14264/e34bfa3

Glikson, E., & Woolley, A. W. (2020). Human trust in Artificial Intelligence: Review of empirical research. *Academy of Management Annals, 14(2),* 627–660. https://doi.org/10.5465/annals.2018.0057.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, *2*(2), 293–323.

Hauk, N., Hüffmeier, J., & Krumm, S. (2018). Ready to be a silver surfer? A meta-analysis on the relationship between chronological age and technology acceptance. *Computers in Human Behavior*, *84*, 304–319. https://doi.org/10.1016/j.chb.2018.01.020

Hertel, G., Meeßen, S. M., Riehle, D. M., Thielsch, M. T., Nohe, C., & Becker, J. (2019). Directed forgetting in organisations: The positive effects of decision support systems on mental resources and well-being. *Ergonomics, 62*(5), 597–611. https://doi.org/10.1080/00140139.2019.1574361

Higgins, E. T., & Molden, D. C. (2003). How strategies for making judgments and decisions affect cognition: Motivated cognition revisited. In G. V. Bodenhausen & A. J. Lambert (Eds.), *Foundations of social cognition: A festschrift in honor of Robert S. Wyer, Jr.* (pp. 211–236). Erlbaum.

Highhouse, S. (2008), Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology, 1*(3), 333–342. https://doi.org/10.1111/j.1754-9434.2008.00058.x

Höddinghaus, M., Sondern, D., & Hertel, G. (2021). The automation of leadership functions: Would people trust decision algorithms? *Computers in Human Behavior, 116*. https://doi.org/10.1016/j.chb.2020.106635

Huang, L., Cooke, N. J., Gutzwiller, R. S., Berman, S., Chiou, E. K., Demir, M., & Zhang, W. (2021). Distributed dynamic team trust in human, artificial intelligence, and robot teaming. In *Trust in human-robot interaction* (pp. 301–319). Academic Press. https://doi.org/10.1016/B978-0-12-819472-0.00013-7

Kares, F., König, C. J., Bergs, R., Protzel, C., & Langer, M. (2023). Trust in hybrid human-automated decision-support. *International Journal of Selection and Assessment*. Online first edition. https://doi.org/10.1111/ijsa.12423

Kaplan, A. D., Kessler, T. T., Brill, J. C., & Hancock, P. A. (2021). Trust in artificialintelligence: Meta-analytic findings. *Human Factors*. https://doi.org/10.1177/00187208211013988

Keding, C., & Meissner, P. (2021). Managerial overreliance on AI-augmented decision-making processes: How the use of AI-based advisory systems shapes choice behavior in R&D investment decisions. *Technological Forecasting and Social Change, 171*. https://doi.org/10.1016/j.techfore.2021.120970

Kim, T. W., & Duhachek, A. (2020). Artificial intelligence and persuasion: A construal-level account. *Psychological Science*, *31*(4), 363–380. https://doi.org/10.1177/0956797620904985

König, C. J., Klehe, U. C., Berchtold, M., & Kleinmann, M. (2010). Reasons for being selective when choosing personnel selection procedures. *International Journal of Selection and Assessment*, *18*(1), 17–27. https://doi.org/10.1111/j.1468-2389.2010.00485.x

Kostka, G., Steinacker, L., & Meckel, M. (2021). Between security and convenience: Facial recognition technology in the eyes of citizens in China, Germany, the United Kingdom, and the United States. *Public Understanding of Science*, *30*(6), 671–690. https://doi.org/10.1177/09636625211001555

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Lanz, L., Briker, R. & Gerpott, F. H. (2023). Employees adhere more to unethical instructions from human than AI supervisors: Complementing experimental evidence with machine learning. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-023-05393-1

Lockey, S., Gillespie, N., Holm, D., & Someh, I. A. (2021). A review of trust in artificial intelligence: Challenges, vulnerabilities and future directions. In *Proceedings of the 54th Hawaii international conference on system sciences* (pp. 5463–5472).

Langer, M., König, C. J., & Papathanasiou, M. (2019). Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment*, *27*(3), 217–234. https://doi.org/10.1111/ijsa.12246

Langer, M., & König, C. J. (2021). Introducing a multi-stakeholder perspective on opacity, transparency and strategies to reduce opacity in algorithm-based human resource management. *Human Resource Management Review*. Aadvanced online publication. https://doi.org/10.1016/j.hrmr.2021.100881

Langer, M., König, C. J., Back, C., & Hemsing, V. (2022). Trust in Artificial Intelligence: Comparing trust processes between human and automated trustees in light of unfair bias. *Journal of Business and Psychology*, 1–16. https://doi.org/10.1007/s10869-022-09829-9

Langer, M., König, C. J., & Busch, V. (2021). Changing the means of managerial work: Effects of automated decision support systems on personnel selection tasks. *Journal of Business and Psychology*. https://doi.org/10.1007/s10869-020-09711-6

Langer, M., & Landers, R. N. (2021). The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2021.106878

Lanz, L., Briker, R., & Gerpott, F.H. (2023). Employees adhere more to unethical instructions from human than AI supervisors: Complementing experimental evidence with machine learning. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-023-05393-1

Laurim, V., Arpaci, S., Promegger, B., & Krcmar, H. (2021). Computer, whom should I hire?—Acceptance criteria for Artificial Intelligence in the recruitment process. In *Proceedings of the 54th Hawaii International Conference on System Sciences,* 5495–5504.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1). https://doi.org/10.1177/2053951718756684

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*(10), 1243–1270. https://doi.org/10.1080/00140139208967392

Lievens, F., & Burke, E. (2011). Dealing with the threats inherent in unproctored Internet. testing of cognitive ability: Results from a large-scale operational test program. *Journal of Occupational and Organizational Psychology, 84*, 817–824. https://doi.org/10.1348/096317910X522672

Luhmann, N. (1968). *Trust and power: Two works by Niklas Luhmann*. Wiley.

Lukacik, E. R., Bourdage, J. S., & Roulin, N. (2022). Into the void: A conceptual model and research agenda for the design and use of asynchronous video interviews. *Human Resource Management Review*, *32*(1). https://doi.org/10.1016/j.hrmr.2020.100789

Malik, A., Budhwar, P., Patel, C., & Srikanth, N. R. (2022). May the bots be with you! Delivering HR cost-effectiveness and individualised employee experiences in an MNE. *The International Journal of Human Resource Management*, *33*(6), 1148–1178. https://doi.org/10.1080/09585192.2020.1859582

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

McAllister, D. J. (1995). Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, *38*(1), 24–59. https://doi.org/10.5465/256727

McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems (TMIS)*, *2*(2), 1–25. https://doi.org/10.1145/1985347.1985353

Meeßen, S. M., Thielsch, M. T., & Hertel, G. (2019). Trust in management information systems (MIS). *Zeitschrift für Arbeits-und Organisationspsychologie A&O*. https://doi.org/10.1026/0932-4089/a000306

Newman, A., Donohue, R., & Eva, N. (2017). Psychological safety: A systematic review of the literature. *Human Resource Management Review, 27(3*), 521–535. https://doi.org/10.1016/j.hrmr.2017.01.001

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

Oracle. (2019). *New study: 64% of people trust a robot more than their manager*. Retrieved February 1, 2023, from https://www.oracle.com/corporate/pressrelease/robots-at-work-101519.html

Pan, Y., & Froese, F. J. (2023). An interdisciplinary review of AI and HRM: Challenges and future directions. *Human Resource Management Review*. https://doi.org/10.1016/j.hrmr.2022.100924

Pan, Y., Froese, F., Liu, N., Hu, Y., & Ye, M. (2022). The adoption of artificial intelligence in employee recruitment: The influence of contextual factors. *The International Journal of Human Resource Management*, *33*(6), 1125–1147. https://doi.org/10.1080/09585192.2021.1879206

Quackenboss, R. T., & Meisburg, R. (2020*). Union strategies to confront automation in the workplace*. Society for Human Resource Management.

Radonjić, A., Duarte, H., & Pereira, N. (2022). Artificial intelligence and HRM: HR managers' perspective on decisiveness and challenges. *European Management Journal*. https://doi.org/10.1016/j.emj.2022.07.001

Robinson, S. C. (2020). Trust, transparency, and openness: How inclusion of cultural values shapes Nordic national public policy strategies for artificial intelligence (AI). *Technology in Society*, *63*. https://doi.org/10.1016/j.techsoc.2020.101421

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404. https://doi.org/10.5465/amr.1998.926617

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford.

Sackett, T. (2023). *ChatGPT 101 for HR Pros. Making HR Tech Easy*. Society for Human Resource Management.

Saßmannshausen, T., Burggräf, P., Wagner, J., Hassenzahl, M., Heupel, T., & Steinberg, F. (2021). Trust in artificial intelligence within production management—An exploration

of antecedents. *Ergonomics*, *64*(10), 1333–1350. https://doi.org/10.1080/00140139.2021.1909755

Schneider, B., Ehrhart, M. G., & Macey, W. H. (2013). Organizational climate and culture. *Annual Review of Psychology*, *64*, 361–388. https://doi.org/10.1146/annurev-psych-113011-143809

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. *Academy of Management Review*, *32*(2), 344–354. https://doi.org/10.5465/amr.2007.24348410

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., Ramos, A., Verkasalo, M., Lönnqvist, J.-E., Demirutku, K., Dirilen-Gumus, O., & Konty, M. (2012). Refining the theory of basic individual values. *Journal of Personality and Social Psychology*, *103*(4), 663–688. https://doi.org/10.1037/a0029393

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies, 146*. https://doi.org/10.1016/j.ijhcs.2020.102551

Solberg, E., Kaarstad, M., Eitrheim, M. H. R., Bisio, R., Reegård, K., & Bloch, M. (2022). A conceptual model of trust, perceived risk, and reliance on AI decision aids. *Group & Organization Management*, *47*(2), 187–222. https://doi.org/10.1177/10596011221081238

Strohmeier, S. (2022). HR machine learning-An introduction. In S. Strohmeier (Ed.), *Handbook of research on artificial intelligence in human resource management* (pp. 25–45). Edward Elgar.

Suseno, Y., Chang, C., Hudik, M., & Fang, E. S. (2022). Beliefs, anxiety and change readiness for artificial intelligence adoption among human resource managers: The moderating role of high-performance work systems. *The International Journal of Human Resource Management*, *33*(6), 1209–1236. https://doi.org/10.1080/09585192.2021.1931408

Tang, P. M., Koopman, J., McClean, S. T., Zhang, J. H., Li, C. H., De Cremer, D., Yizhen Lu & Ng, C. T. S. (2022). When conscientious employees meet intelligent machines: An integrative approach inspired by complementarity theory and role theory. *Academy of Management Journal*, *65*(3), 1019–1054. https://doi.org/10.5465/amj.2020.1516

Thielsch, M. T., Meeßen, S. M., & Hertel, G. (2018). Trust and distrust in information systems at the workplace. *PeerJ,* 6. https://doi.org/10.7717/peerj.5483

Tippins, N. T. (2015). Technology and assessment in selection. *Annual Review of Organizational Psychology and Organizational Behaviour*, *2*, 551–582. https://doi.org/10.1146/annurev-orgpsych-031413-091317

Tippins, N. T., Oswald, F. L., & McPhail, S. M. (2021). Scientific, legal, and ethical concerns about AI-based personnel selection tools: A call to action. *Personnel Assessment and Decisions*, *7*(2), 1. https://doi.org/10.25035/pad.2021.02.001

Trades Union Congress (2021). *When AI is the boss*.

van der Werff, L., Legood, A., Buckley, F, Weibel, A., & De Cremer, D. (2019). Trust motivation: The self-regulatory processes underlying trust decisions. *Organizational Psychology Review. 9*(2-3), 99–123. https://doi.org/10.1177/2041386619873616

Vorm, E. S. & Combs, D. J. Y. (2022). Integrating transparency, trust, and acceptance: The intelligent systems technology model (ISTAM). *International Journal of Human–Computer Interaction, 38*, 18–20. https://doi.org/10.1080/10447318.2022.2070107

Weber, J. M., Malhotra, D., & Murnighan, J. K. (2004). Normal acts of irrational trust: Motivated attributions and the trust development process. *Research in Organizational Behavior*, *26*, 75–101. https://doi.org/10.1016/S0191-3085(04)26003-8

World Economic Forum (2021). *Human-centred artificial intelligence for human resources: A toolkit for human resources professionals*. https://www3.weforum.org/docs/WEF_Human_Centred_Artificial_Intelligence_for_Human_Resources_2021.pdf

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*, *32*(4), 403–414. https://doi.org/10.1002/bdm.2118

Zacher, H., & Frese, M. (2009). Remaining time and opportunities at work: Relationships between age, work characteristics, and occupational future time perspective. *Psychology and Aging*, *24*(2), 487–493. https://doi.org/10.1037/a0015425