

Association for Information Systems

**AIS Electronic Library (AISeL)**

---

Wirtschaftsinformatik 2021 Proceedings

Track 10: Design, management and impact of  
AI-based systems

---

## The evaluation of the black box problem for AI-based recommendations: An interview-based study

Jessica Ochmann

*Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany*

Sandra Zilker

*Friedrich-Alexander-Universität Erlangen-Nürnberg*

Sven Laumer

*Friedrich-Alexander-Universität Erlangen-Nürnberg*

Follow this and additional works at: <https://aisel.aisnet.org/wi2021>

---

Ochmann, Jessica; Zilker, Sandra; and Laumer, Sven, "The evaluation of the black box problem for AI-based recommendations: An interview-based study" (2021). *Wirtschaftsinformatik 2021 Proceedings*. 6. <https://aisel.aisnet.org/wi2021/QDesign/Track10/6>

This material is brought to you by the Wirtschaftsinformatik at AIS Electronic Library (AISeL). It has been accepted for inclusion in Wirtschaftsinformatik 2021 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# The evaluation of the black box problem for AI-based recommendations: An interview-based study

Jessica Ochmann<sup>1</sup>, Sandra Zilker<sup>2</sup>, Sven Laumer<sup>1</sup>

<sup>1</sup> Friedrich-Alexander-University Erlangen-Nuremberg, Schöller Endowed Professorship for Information Systems (Digitalization in Business and Society), Nuremberg, Germany  
jessica.ochmann@fau.de, sven.laumer@fau.de

<sup>2</sup> Friedrich-Alexander-University Erlangen-Nuremberg, Chair of Digital Industrial Service Systems, Nuremberg, Germany  
sandra.zilker@fau.de

**Abstract.** Organizations are increasingly adopting artificial intelligence (AI) for business processes. AI-based recommendations aim at supporting users in decision-making, e.g., by pre-filtering options. However, users can often hardly understand how these recommendations are developed. This issue is called “black box problem”. In the context of Human Resources Management, this leads to new questions regarding the acceptance of AI-based recommendations in the recruiting process. Therefore, we develop a model based on the theory of planned behavior explaining the relation between the user’s perception of the black box problem and the attitude toward AI-based recommendations distinguishing between a mandatory and voluntary use context. We conducted 21 interviews with experts from recruiting and AI. Our results show that the perception of the black box problem conceptualized by the awareness and the evaluated relevance relates to the user’s attitude toward AI-based recommendations. Further, we show that the use context has a moderating effect on that relation.

**Keywords:** *Black box, AI-based recommendations, Human Resources Management*

## 1 Introduction

The increasing demand for information technologies within the organizational context changes the way companies handle their business-related processes. Organizations have early recognized that the implementation of new technologies is beneficial for the company’s success. Therefore, companies are ambitious to expand the use of information systems further in order to design their processes more efficiently [1]. In this regard, artificial intelligence (AI) is increasingly addressed [2]. One possible application of AI are recommender systems. In general, recommender systems are software tools and techniques, providing suggestions for items that are of use to a user [3]. In the private context, AI-based recommendations are well-known. For example, if a person wants to watch a movie on Netflix, the platform suggests possible movies based on the user’s preferences, which is determined by personal data, like previously watched

movies or shows. In comparison, AI-based recommendations are also becoming more and more present in the business context in general and in the human resources management (HRM) in particular [4]. Focusing on HRM, AI-based recommendations offer new possibilities to realize various advantages from the organization's perspective [4]. Primarily the automated recruiting of candidates gains importance as organizations expect various benefits such as faster application handling and relief of recruiters [5]. In this regard, organizations make considerable efforts to push the application of AI-based recommendations in their recruiting process [5, 6]. One of the general advantages of these recommendations is that the information overload for the user is minimized. For example, by pre-filtering possible candidates with the use of AI, recruiters are provided with a pre-select list of job seekers that fit the vacancy. Therefore, information overload is prevented, and the recruiter's decision is supported. However, the application of AI-based systems in organizations also encourages controversial discussions as the prediction quality of the underlying technology of Machine Learning (ML) has continuously advanced by simultaneously increasing the complexity of the learning process. As a result, it is nearly impossible for humans to understand what the system is doing [7]. This so-called "black box problem" leads to challenges for organizations. For example, organizations have to develop a basic understanding of AI-based systems and clarify both potentials of this technology such as the production of accurate results and risks such as regulatory requirements [2]. In this regard, also the acceptance of AI-based technologies must be addressed as individuals might behave hesitantly when they are uncertain about the basis of decision-making and have difficulties to appraise the consequences [8].

Regarding AI-based recommendations in HR, the system undertakes the role of the recruiter. In this particular case, it is a debatable point whether the recruiter follows an AI-based recommendation on their own responsibility when neither the basis of decision-making (e.g., why a system makes a certain recommendation) nor the consequences of a certain decision (e.g., what happens if the system fails) are clear. Moreover, the extent to which the recruiter is willing to accept the competence-shift from the human to the system is questionable [9, 10]. Given these challenges and considering that "people factors" like user acceptance gain in importance for the successful adoption of AI-based recommendations [5], it is crucial to discuss the purport of black boxes for individuals in an organizational context. Prior research on the black box problem focuses primarily on the technical perspective and aims to identify methods that uncover how systems recognize patterns and make predictions [7, 11, 12]. However, the existing research lacks in covering the individual point of view in an organizational context. Nonetheless, it is essential to understand the consequences of non-interpretable AI-based recommendations for individuals. Against this backdrop, it has to be clarified how users (e.g., recruiters) of AI-based recommendation systems evaluate the consequences of missing transparency – provided that these users are aware of the black box accompanied by the application of AI-based systems. Based on this research gap, we intend to answer the following research question (RQ):

**RQ:** *How does the perception of the black box problem influence an individual's attitude toward AI-based recommendations?*

As the field of HRM is highly underrepresented in current research in regards to the black box problem of AI and it is of high relevance for individuals, we focus on this specific domain. To frame our research, we focus on the theoretical lens of attitude-based technology acceptance [13]. We also consider the use context of the AI-based recommendation following the approach presented by Krönung et al. [14]. In addition, we explain the terms AI-based recommendations and black box problem. From a methodology perspective, we conducted 21 explorative interviews. Based on the results extracted from these interviews, we discuss the relation between the perception of the black box problem and the user's attitude toward AI-based recommendations and we show that the use context moderates this relation. Finally, the paper consolidates the findings, shows limitations, and discusses research as well as practical implications.

## **2 Related work**

In this section, we will summarize related work on attitude-based technology acceptance research and the black box problem of AI-based recommendations to highlight the specific research gap that our approach is intended to fill.

### **2.1 Attitude-based technology acceptance research**

To examine a user's attitude toward AI-based recommendations in general and gain a basic understanding of the behavioral implications regarding the attitude toward the black box problem in particular, we build on attitude-based technology acceptance research. This research stream is based on the theory of planned behavior (TPB) that links users' behavioral beliefs to the attitude toward an object. In this context, attitude is defined as an individual's evaluation (i.e., positive or negative) of a certain object [15]. More precisely, TPB postulates that behavior is influenced by attitude, and attitude is influenced by beliefs [13]. As we strive to analyze how beliefs regarding the black box problem shape the attitude toward AI-based recommendations, TPB seems to be the most appropriate starting point to develop comprehensible models. In the context of our research, beliefs are defined as the perception of the black box problem. Complementary, we consider the attitude toward an object with the object being specified as AI-based recommendations. Besides, we consider the use context where a particular belief emerges as different perspectives might result in divergent attitudes. Behavioral theory suggests that the use context influences the attitude of users and thereby affecting their acceptance [16]. In other words, the actions users take when using an information system depend on the attitude they have in a particular situation and that the attitude formation process varies in different contexts. Therefore, we adopted an approach proposed by Krönung et al. [14]. The authors show that the use context influences the attitude formation process in the respective context of using information systems. The use context can either be mandatory or voluntary. A mandatory use context is characterized by the obligatory use of information systems. For example, an organization requires its employees to use a certain technology and deny them to handle business-

related tasks without the applied technology. In comparison, in a voluntary use context, the organization offers information systems as a supportive tool. However, the use is not required to fulfill the business-related task, and employees can decide themselves if they want to apply a certain system. In sum, this paper considers the perception of the black box problem as a belief that shapes the attitude toward an object, which are in our approach AI-based recommendations. In addition, we investigate the use context, thereby distinguishing between a voluntary and a mandatory technology use as an individual's attitude toward AI-based recommendations might differ in relation to the use context.

## **2.2 Recommender systems in human resources management**

Recommender systems describe information systems that process user data to provide personalized recommendations for products or services that match user preferences [3]. By sorting and filtering alternatives, recommender systems aim to reduce possible information overload for users and support them in making reasonable decisions. In the last decade, recommender systems have been increasingly implemented in a variety of application domains with a strong focus on e-commerce and media in the consumer context [17] to reduce a consumer's effort to find relevant products or services. In contrast, in an HR context, two types of recommender systems are discussed depending on their aim to support either the organization or the job seeker. The first ones are job recommender systems that support job seekers in their job search by matching vacancies with their job preferences. The second group is CV recommender systems that are implemented by organizations and support recruiters by pre-selecting suitable candidates for a certain vacancy.

For realizing a data-driven recommender system, there are different kinds of approaches like collaborative or content-based filtering [18]. Besides these more traditional statistical metrics, the methods behind those approaches are nowadays increasingly based on AI models that lead to more user-centric AI-based recommendations [19].

Given the ubiquity of recommender systems, scholarly research put much effort to investigate factors that influence the acceptance of both the recommender systems and the underlying recommendations in a business-to-consumer context. Prior research shows, for example, that personalization and familiarity positively influence the acceptance of recommender systems [18], while system biases might decrease the adoption of such systems [20]. However, it remains unclear under which circumstances these results can be adapted to a business-related context, which is characterized by higher stakes of the decisions. While the consequences of a failed product recommendation are comparability low, a wrong business decision might lead to more serious economic consequences. In an HR context, for example, the recruitment of an unskilled employee due to an insufficient recommendation might damage the company's reputation if this new employee is not able to perform adequately. Besides, the application of such recommender systems in HR can be either mandatory or voluntary, depending on the organization's strategy [14]. A review of the literature shows that academic research evaluates automation in HRM predominantly positive [21–23]. However, little is known

about its consequences for organizational users in general and recruiters in particular [24]. Accordingly, the understanding of how organizations can proactively influence the attitude toward these AI-based recommendations remains limited.

Therefore, it is essential to look at recommendations in the business context in general and in the HR context in particular as individuals might show different acceptance behaviors toward the use of recommender systems in organizations.

### **2.3 Black box problem**

Recent AI-based systems based on deep learning achieve higher predictive quality while at the same time gaining in complexity. Therefore, individuals find it increasingly difficult to understand the underlying reasoning for a certain recommendation. This phenomenon is subsumed under the term “black box problem”. Black boxes miss providing further information about how predictions are generated [25] as they cover multiple levels of abstraction that are barely interpretable [26]. Prior research has highlighted that uncertainty about the reliability of AI predictions might hurt trust in AI-based systems and reduce acceptance of an AI-based recommendation [12]. Hence, transparency and explainability of data-based decisions are crucial for most business processes (e.g., recruiting of candidates) and medical (e.g., interpretation of radiographic images) or safety-critical (e.g., autonomous driving) as well as commercial (e.g., lending decisions) applications [7, 11, 12]. Individuals want to understand the reasons for a certain recommendation to appraise the consequences and – especially when they are responsible for a certain task – to verify a decision. In comparison, black boxes might be acceptable in most business-to-consumer contexts as referred services (e.g., translation services, consumer entertainment, e-commerce) are merely uncritical, and the consequences of failed AI-based recommendations are unspectacular and manageable [12]. Thus, transparency requirements for AI-based systems might be higher for Business-to-Business than for Business-to-Consumer contexts [25].

To address this issue, regulatory authorities and other initiatives have lately focused on the transparency of algorithms and started to promote the concept of explainable AI (XAI). According to Zanni-Merk [27], the overall goal of XAI is to build an explainable model. This is, understandable by humans, yet providing justification for predictions or decisions. Previous research focuses on the black box problem from a technical perspective (e.g., underlying explanatory structures of an algorithm) [7, 11, 12, 26, 28], in the medical domain [29–32], or in the financial sector [25, 33, 34]. However, the impact of the black box problem in an organizational context in general and in HRM in particular is mostly unaddressed. Especially in HRM, it is crucial to understand how the output of an AI-based system is generated. First, the consequences of failed recommendations might be business-critical (e.g., algorithms with discriminatory tendencies, which harms a company’s reputation). Second, from an organization’s perspective, the individual’s attitude toward an AI-based recommendation must be investigated to encourage the usage of such systems. We aim to address this research gap by conducting an exploratory interview analysis to unveil the user’s perception of the black box problem to gain insights regarding the attitude toward AI-based recommendations.

### 3 Research method

The overall aim of this paper is to explain attitude-based technology acceptance of AI-based recommendations against the backdrop of the black box problem. To reveal the user's attitude toward AI-based recommendations in HR, we applied a two-stage approach. First, we conducted interviews to get further insights regarding the black box problem as a determinant for the user's attitude toward AI-based recommendations. Second, we examined the influence of the use context on the relation between the perception of the black box problem and the attitude toward AI-based recommendations. As already mentioned before, our research focuses on applications in the business context, more specifically on the application in HRM. Furthermore, we distinguish between a mandatory and a voluntary use context.

#### 3.1 Study design

Overall, 21 semi-structured interviews with experts from various fields (HRM, AI and ML, law, data protection) were conducted to gain a broad understanding of fundamental beliefs and concerns regarding the acceptance of AI-based recommendations in HRM [35]. All interviewed experts (I1 – I21) have experience with recruiting of candidates and can assess the black box problem as they deal with the potentials of innovative AI-technologies in their daily work. Our interviewees provide both a perspective from an organizational and from an individual point of view as the experts indicate that the black box problem affects the organization in general and the user of AI-based systems in particular. The identification of potential experts took place by screening their business-related networking sites in order to assess their involvement in the topic. In sum, 56 potential participants were contacted via E-Mail or the professional network platform LinkedIn. Twenty-eight contacted persons did not respond, and seven persons stated they could not take part due to time issues or because they are not highly involved in recruiting processes. In the end, we interviewed 21 experts exclusively in person or by phone. Table 1 illustrates the demographic characteristics.

**Table 1.** Overview on the study design

<i>Characteristics</i>	<i>Attribute</i>	<i>Share</i>
Gender	Male	12 (56,1%)
	Female	9 (42,9%)
Age	24 – 29	7 (33,3%)
	30 – 39	2 (9,5%)
	40 – 49	4 (19,1%)
	50 – 59	4 (19,1%)
	>59	4 (19,1%)
Background	Recruiting	9 (42,9%)
	AI and ML	7 (33,3%)
	Law/ Data Protection	5 (23,8%)

We conducted an interview guideline considering questions regarding the *awareness and relevance* of the black box problem, the evaluation of lack in transparency and explainability of AI-based recommendations (i.e., *positive vs. negative*), as well as the importance of an application's use context (i.e., *mandatory vs. voluntary*). The interview guideline follows a semi-structured format. This approach gives respondents sufficient freedom to describe their overall attitude toward AI-based recommendations. Moreover, it allows for consistency across the interviews and enables the interviewer to explain specific and new insights [36, 37]. To reduce possible response bias and encounter social desirability, the authors assured them to treat all answers anonymous and strictly confidential.

### **3.2 Interview-based research**

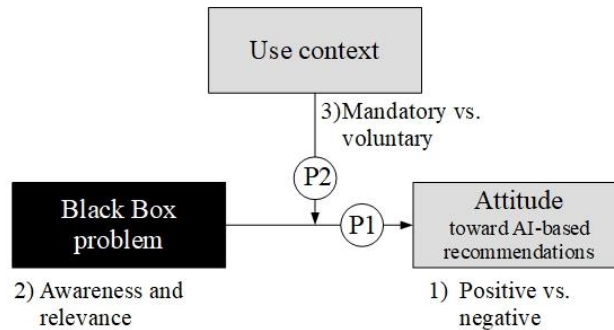
The transcription in preparation for the data analysis was accomplished after each interview using MAXQDA. This approach ensured that no thematic aspect is missing in the analysis. The systematic analysis and categorization of the insights from the interviews followed the method of qualitative content analysis [38]. In order to generate insights regarding the perception of the black box problem from the user's perspective in an organizational context, we followed an exploratory, inductive approach to code the interviews [38, 39]. Coding the interviews inductively allows the consideration of alternative solutions as we searched for statements that reflect alternative adoption behaviors as well as determinants for the attitude toward AI-based recommendations, thereby avoiding a bias toward the black box problem [35, 40].

## **4 Results**

Based on the interviews, we could derive two propositions, P1 and P2. A synthesis of determinants for the attitude toward AI-based recommendations results in the research model shown in Figure 1.

In our analysis, we identified the black box problem (i.e., awareness and relevance) as a determinant to explain the user's attitude toward AI-based recommendations (P1). Additionally, we found evidence that the use context (i.e., voluntary vs. mandatory) moderates the relation between the black box problem and user's attitude toward AI-based recommendations (P2). In the following subchapters, we will describe the resulting implications in more detail.





**Figure 1.** Relation between black box problem and attitude toward AI-based recommendations.

#### 4.1 Black box and attitude toward AI-based recommendations

To explain the user’s attitude toward AI-based recommendations, we examine the black box problem from an individual’s perspective in an organizational HRM context. Thereby, we identified users’ awareness regarding the lack of transparency and explainability of AI-based recommendations. Based on our analysis, we highlight users’ evaluation of the black box problem by scrutinizing the relevance of a black box when users rely on AI-based recommendations. In general, the interviews show that users are aware of the black box problem of AI-based recommendations and assess this issue as highly relevant for HR. Black boxes seem to be the major challenge when an organization considers the application of AI-based recommendations in HR-related processes, as the following statements show:

*“At the moment, black boxes are the greatest challenge for AI systems, because we do not know how they come to their decisions.” (I1)*

*“[...] there is no transparency, the accountability is lost, and these systems are not traceable.” (I13)*

More precisely, respondents criticize the incomprehensibility of AI-based recommendations resulting from obscure data processing. They are not able to interpret the recommendation’s underlying reasoning and would prefer a higher level of transparency:

*“The problem when applying AI-based recommendations is that data is inserted and then an answer emerges. It is unclear, however, how this answer has been produced. [...]. It is incomprehensible.” (I14)*

HR-related processes are linked to a certain degree to responsibility regarding the consequences of a decision. Therefore, users find it difficult to trust non-transparent recommendations and to assume responsibility for the consequences, as the following interviewee stated:

*“I imagine working together with an AI to be complicated, as I would need to raise a lot of trust, if not understanding its inner workings [...] because I like to understand causes and procedures.” (I21)*

To increase the attitude toward AI-based recommendation in HR, organizations have to clarify the level of responsibility for the users with regard to the acceptance of AI-based recommendations.

Especially in an organizational HRM context, users want AI-based systems to provide additional information regarding the underlying processes. Respondents see it critical when AI-based systems miss to exemplify how recommendations are generated and believe that AI-based recommendations should offer some detailed explanation for the suggested solution:

*“The user of AI systems should not only be told: ‘This is the result now and you trust it.’ Rather, requests should be allowed. For example: ‘When was a candidate filtered out?’ And then the user himself also thinks about what the reasons are. However, explainable artificial intelligence is still very much in development.” (I3)*

Regarding “soft skills” (e.g., empathy and intuition), humans are currently superior to systems [9]. Anyhow, respondents see potentials in the application of AI-based systems, in particular when it comes to processing a huge data amount:

*“In feeling, in intuition, in perception – generally in all ‘human competencies’, AI systems are inferior to humans. An AI system has no chance in this respect. On the other hand, where man has no chance is when it comes to the evaluation of very many numbers. If thousands of people apply for a job in a large corporation, then people have no chance to match their skills as quickly as AI.” (I2)*

AI-based recommendations might also be an appropriate instrument to prevent discriminatory tendencies in HRM as unbiased, and correctly programmed algorithms are directly linked to consistent and impartial decisions [41].

*“When it comes to recruiting, AI has the biggest potential here. It will help human beings to overcome their conclusive biases when making a decision. But again, AI or any intelligent system is only as good as the data you are feeding it with.” (I7)*

However, there are also more critical voices regarding the underlying algorithms of AI-based recommendations programmed by humans as the black box contains – strictly speaking – also the opinion of the programmer and represents his beliefs:

*“An algorithm is not objective. An algorithm does not have a moral. The machine does what humans tell it to do. In the end, in particular, it depends on humans becoming aware when making the decision to use an algorithm or AI system that ultimately, it is error-prone and, furthermore, where the boundaries and possibilities of the system lie.” (I6)*

In summary, the interviews show that users are aware of the black box problem in an organizational context and evaluate it as a relevant issue that influences their attitude toward AI-based recommendations. Particularly in HRM, where personal data is processed, transparent decisions are inevitable. For example, insufficient explanations in the process of applicant selection can harm an organization’s reputation due to discrimination accusations and even induce legal consequences [10]. Therefore, it is important that users of AI-based recommendations are aware of the consequences resulting from a lack of transparency and explainability.

Hence, we argue that there is a relation between the black box problem and the user’s attitude toward AI-based recommendations. This applies when the user is aware of the

lack of transparency and explainability. Simultaneously, the user has to evaluate the relevancy of the given challenge as crucial. Provided that both aspects are given, the user's attitude toward AI-based recommendations is more negatively connoted. Therefore, we conclude that the perception of the black box problem, conceptualized by awareness and relevance, influence the attitude toward AI-based recommendations, such that we assume:

**P1:** *The higher (lower) a user's perception of the black box problem in terms of awareness and relevance, the lower (higher) is the user's attitude toward AI-based recommendations.*

## 4.2 The influence of the use context

In our interview-based study, we examined the influence of the use context on the relation between the black box problem and the user's attitude toward AI-based recommendations. Organizations might pursue the mandatory use of AI-based systems to realize advantages from the organization's perspective (e.g., increase labor productivity). In a mandatory use context, the organization constrains its employees to use a specific AI-based system. Users have no other choice and must accept the recommendation created by the AI. For example, an organization implements a mandatory CV parser. Users have to accept the recommendations from the CV parser instead of screening CVs independently – even if they are not able to understand a certain recommendation and are eventually responsible for the consequences.

In contrast, organizations can implement AI-based systems as a supportive tool that can be used voluntarily by employees to support them in their daily work routine.

In our interviews, we found evidence that users' attitude toward AI-based recommendations depends on the level of voluntariness provided by an organizational AI-based system. In general, the interviews indicate that users hesitate to trust a mandatory tool to make business-critical decisions. Respondents evaluate it critically if AI-based systems force users to select a certain option without giving them any guidance regarding the consequences, even when the objective of an AI-based recommendation is relieving the user (e.g., recruiter), illustrated by the following statement:

*“But I don't think the AI should decide whether I want to hire someone or not, it should just simplify what the HR person does for them.” (I12)*

More precisely, the interviews show that users want to make the final decision (e.g., selection of candidates in the recruiting process) and are critical of AI-based systems that act autonomously:

*“That [who is responsible for AI-based decisions] is already a discussion. But AI should not be the ultimate decision-maker. The decision maker is the one who uses this tool.” (I1)*

Furthermore, respondents stated that users might feel uncomfortable to follow such a non-transparent recommendation when they are personally responsible for the decision. Moreover, interviewees find it difficult to verify a decision they are not able to interpret. Especially when they cope with reconstructing the reasoning of an AI-based recommendation:

*“The users receive some recommendations that should lead to a decision respectively contribute to a decision making the process. But in the end, the human decision maker, e.g., the recruiter, is held responsible for the decision.” (I1)*

In comparison, a voluntary use context allows the user to consider AI-based systems as a supportive tool. The recommendations provided are a kind of guidance for the user. For example, recruiters receive AI-based recommendations for a job platform regarding appropriate candidates. The recruiter can accept these recommendations and contact the proposed candidates or ignore them. In this particular case, AI-based systems support the user in self-assessment:

*“So, if the recommendation of the AI is not understandable, then I feel uncomfortable with it. If this is something where you would have come to the same conclusion or you can check the recommendation (e.g., by running an assessment center independently of the AI recommendation), then perhaps AI will help you to rethink your assessments.” (I3)*

Simultaneously, in a voluntary use context, the user receives enough freedom to make their own decisions. In this case, the black box problem is less critical. AI-based recommendations are merely options and support the user in the daily work – but the final decision resides with the user:

*“So I’m always a friend of any kind of support, but the last decision must be made by the human being.” (I9)*

In summary, the interviews indicate that the relation between the black box problem and user’s attitude is moderated by the use context. A mandatory use context strengthens the negative relation between the perception of the black box problem and the attitude toward an AI-based recommendation, whereas a voluntary use context weakens this effect. Especially when users are responsible for a decision, they want to understand the underlying reasoning of a recommendation and evaluate the use of mandatory AI-based systems more negatively. In contrast, AI-based recommendations are evaluated positively when they are considered as a supportive tool, and the user maintains the decisional power. Therefore, we propose:

**P2:** *The use context moderates the relation between the black box problem and user’s attitude toward AI-based recommendations such that the relation is stronger (weaker) in a mandatory (voluntary) use context.*

## **5 Discussion**

We propose that the attitude toward AI-based recommendations is based on the perception of the black box problem in terms of awareness and relevance and that the influence of this perception depends on the context of AI-based recommendations, whereas we assume that the effect is stronger in mandatory use contexts. This proposed theoretical model has implications for theory and practice, as we will discuss in the following. Prior studies have generated insights by outlining the technical perspective of the black box problem, thereby focusing on methods to increase explainability of the underlying

algorithms [7, 11, 12, 28]. Further, approaches that deal with the black box problem consider in particular the medical domain [29–32] as well as the financial sector [25, 33, 34] with a strong focus on the technical perspective, thereby unveiling *how* an algorithm works to derive recommendations.

The issue of *why* perceptions on the black box problem influence an individual's attitude toward AI-based recommendations in HR is largely unaddressed, especially in regard to the use context (i.e., mandatory vs. voluntary). Therefore, our theoretical model contributes to the outlined XAI discussion in the literature by providing an individual focused perspective that explains how AI users evaluate the black box problem in HR and if the black box problem shapes their attitude toward AI-based recommendations. This perspective enriches the rather technical perspective in the literature as it highlights the need for a comprehensive model to explain the relation of the black box problem of AI-based recommendations and the role of the use context. Our proposed model addresses this challenge by providing valuable insights regarding the black box perceptions on the attitude toward AI-based recommendations. The overall results show that the attitude toward AI-based recommendations is more negative when user awareness and user evaluation as determinants of the black box problem are high. Furthermore, we contribute by considering the organizational use context as a moderator on the relation between AI-based recommendations and the perception of the black box problem. We show that a mandatory use context, which is characterized by the obligated use of AI-based systems, strengthens the relation between the black box problem and user's attitude toward AI-based recommendations. In contrast, users evaluate the voluntary use of AI-based systems as beneficial because the given recommendations are experienced as suggestions and support them in their business-related tasks.

Regarding research dealing with AI-based systems in HRM, we followed the call for an examination of attitude-based technology acceptance research [42]. So far, much research has focused on the application of AI to enhance the strategic role of HR [43]. AI-based systems are implemented to increase value creation for the company and therefore realize economic advantages. Furthermore, research has focused on using AI to improve employees' performance from an organizational point of view [44]. In this context, research has tackled the challenges of discrimination and unfairness and discussed AI as an appropriate instrument to prevent both discriminatory tendencies and unfairness [41]. Besides these research efforts linking AI and HRM and to focus on general challenges such as discrimination, less is known about how HR employees value the use of AI in their job. Hence, this paper contributes to this stream of research by focusing on the user's attitude toward such AI-based systems in HR. Our study contributes to this discussion in the literature as it shows that HR employees are aware of the black box problem. They evaluate the lack of transparency and explainability as critical, especially when they are responsible for a decision, which they are not able to verify (mandatory use context). However, HR employees see potential in AI-based systems as recommendations can support them in non-critical business-related decision making (voluntary use context). Our findings also give practical implications. When implementing AI-based systems, companies should be aware of the fact that user acceptance depends on the transparency and understandability of such systems. In

addition, AI-based recommendations are not suitable for every business-related task. When users are of high responsibility for the possible consequences, they wanted to be more involved in the decision-making process and hesitate to trust AI-based recommendations unconditionally.

Besides these contributions, the presented paper underlies several limitations. First, the generalizability of the findings is restricted as only 21 German experts with experience in recruiting or AI affinity were interviewed. As the majority of respondents deal with innovative technologies and AI in their daily business, the sample might be biased in regard to awareness of the black box problem. Based on our findings, we aim at expanding future research in regard to the implementation of XAI in HRM. In this context, it is not only important to understand the perception of the black box problem within an organization, but also to focus on the black box problem for AI-based recommendations that are provided to external individuals, such as in an HR context potential employees. Therefore, future research should expand our study to black box problems of AI-based recommendations offered for candidates. This might enable to study additional use contexts that are relevant in the general AI-based recommendation context. In addition, the identified effects can be further validated by using a quantitative study that enables tests of additional context factors such as user personality or experience with AI. These factors have not been tested in the development of the proposed model so far, but they are expected to better explain which users in which contexts especially perceive the black box problem such that it influences their attitude toward AI-based recommendations.

## **6 Conclusion**

This paper contributes to both academia and practice. We provide insights into the application of XAI in HRM. We highlight the importance of perception of the black box problem conceptualized by the awareness and the relevance in regard to AI-based recommendation systems. Adding to the existing technical-oriented literature on the black box problem, this work recognizes the individual perspective and examines an attitude-based view on this strategically relevant topic. According to use of these findings in practice, we highlight the need for awareness regarding the trade-off between realization of an organization's efficiency and the user's responsibility for their decisions.

## **Acknowledgements**

This project is funded by the Adecco Stiftung "New Ways for Work and Social Life".

## **References**

1. Melville, N., Kraemer, K., Gurbaxani, V.: Information Technology and Organizational Performance: An Integrative Model of IT Business Value. *MIS Quarterly* 28, 283–322 (2004)

2. Ransbotham, S., Kiron, D., Gerbert, P., Reeves, M.: Reshaping Business With Artificial Intelligence: Closing the Gap Between Ambition and Action. *MIT Sloan Management Review* 59, 1–17 (2017)
3. *Recommender Systems Handbook*. Springer US, Boston, MA (2011)
4. Eckhardt, A., Laumer, S., Maier, C., Weitzel, T.: The transformation of people, processes, and IT in e-recruiting. *Employee Relations* 36, 415–431 (2014)
5. van Esch, P., Black, J.S., Ferolie, J.: Marketing AI recruitment: The next phase in job application and selection. *Computers in Human Behavior* 90, 215–222 (2019)
6. Strohmeier, S., Piazza, F.: Artificial Intelligence Techniques in Human Resource Management—A Conceptual Exploration. In: *Intelligent Techniques in Engineering Management*, 87, pp. 149–172. Springer International Publishing, Cham (2015)
7. Castelveccchi, D.: Can we open the black box of AI? *Nature* 538, 20–23 (2016)
8. Nugent, C., Cunningham, P.: A Case-Based Explanation System for Black-Box Systems. *Artificial Intelligence Review* 24, 163–178 (2005)
9. Jarrahi, M.H.: Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons* 61, 577–586 (2018)
10. Orwat, C.: Diskriminierungsrisiken durch Verwendung von Algorithmen. Eine Studie erstellt mit einer Zuwendung der Antidiskriminierungsstelle des Bundes. *Nomos; Antidiskriminierungsstelle des Bundes, Baden-Baden* (2019)
11. Ghahramani, Z.: Probabilistic machine learning and artificial intelligence. *Nature* 521, 452–459 (2015)
12. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Springer International Publishing, Cham (2019)
13. Ajzen, I.: The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes* 50, 179–211 (1991)
14. Kroenung, J., Eckhardt, A.: The attitude cube: A three-dimensional model of situational factors in IS adoption and their impact on the attitude–behavior relationship. *Information & Management* 52, 611–627 (2015)
15. Ajzen, I., Fishbein, M.: Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin* 84, 888–918 (1977)
16. Mallat, N., Rossi, M., Tuunainen, V.K., Öörni, A.: The impact of use context on mobile services acceptance: The case of mobile ticketing. *Information & Management* 46, 190–195 (2009)
17. Lu, J., Wu, D., Mao, M., Wang, W., Zhang, G.: Recommender system application developments: A survey. *Decision Support Systems* 74, 12–32 (2015)
18. Komiak, Benbasat: The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents. *MIS Quarterly* 30, 941 (2006)
19. Dietvorst, B.J., Simmons, J.P., Massey, C.: Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science* 64, 1155–1170 (2018)
20. Adomavicius, G., Bockstedt, J.C., Curley, S.P., Zhang, J.: Do Recommender Systems Manipulate Consumer Preferences? A Study of Anchoring Effects. *Information Systems Research* 24, 956–975 (2013)
21. Bondarouk, T.V., Ruël, H.J.M.: Electronic Human Resource Management: challenges in the digital era. *The International Journal of Human Resource Management* 20, 505–514 (2009)

22. Bondarouk, T., Harms, R., Lepak, D.: Does e-HRM lead to better HRM service? *The International Journal of Human Resource Management* 28, 1332–1362 (2017)
23. Ruta, C.D.: HR portal alignment for the creation and development of intellectual capital. *The International Journal of Human Resource Management* 20, 562–577 (2009)
24. Wirtky, T., Laumer, S., Eckhardt, A., Weitzel, T.: On the Untapped Value of e-HRM – a Literature Review. *Communications of the Association for Information Systems* 38, 20–83 (2016)
25. Samek, W., Müller, K.R.: Towards Explainable Artificial Intelligence. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11700 LNCS, 5–22 (2019)
26. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* 521, 436–444 (2015)
27. Zanni-Merk, C.: On the Need of an Explainable Artificial Intelligence. In: *Information Systems Architecture and Technology: Proceedings of 40th Anniversary International Conference on Information Systems Architecture and Technology – ISAT 2019*, 1050, p. 3. Springer International Publishing, Cham (2020)
28. Zhu, J., Liapis, A., Risi, S., Bidarra, R., Youngblood, M.: Explainable AI for Designers: A Human-Centered Perspective on Mixed-Initiative Co-Creation. *Proceedings of the 2018 IEEE Conference on Computational Intelligence and Games (CIG'18)* (2018)
29. Holzinger, A., Biemann, C., Pattichis, C.S., Kell, D.B.: What do we need to build explainable AI systems for the medical domain? *CoRR abs/1712.09923* (2017)
30. Holzinger, A., Langs, G., Denk, H., Zatloukal, K., Müller, H.: Causability and explainability of artificial intelligence in medicine. *Wiley interdisciplinary reviews. Data mining and knowledge discovery* 9, e1312 (2019)
31. Lamy, J.-B., Sekar, B., Guezennec, G., Bouaud, J., Séroussi, B.: Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach. *Artificial intelligence in medicine* 94, 42–53 (2019)
32. Meldo, A.A., Utkin, L.V.: A new approach to differential lung diagnosis with CT scans based on the Siamese neural network. *J. Phys.: Conf. Ser.* 1236, 12058 (2019)
33. Weller, A.: Transparency: Motivations and Challenges. In: *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, 11700, pp. 23–40. Springer International Publishing, Cham (2019)
34. Wang, D., Yang, Q., Abdul, A., Lim, B.Y.: Designing Theory-Driven User-Centric Explainable AI. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, pp. 1–15. ACM Press, New York, New York, USA (2019)
35. Gioia, D.A., Corley, K.G., Hamilton, A.L.: Seeking Qualitative Rigor in Inductive Research. *Organizational Research Methods* 16, 15–31 (2013)
36. Myers, M.D.: *Qualitative research in business & management*. Sage, Los Angeles (2010)
37. Bryman, A.: *Social research methods*. Oxford University Press, Oxford (2016)
38. Mayring, P.: *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. Klagenfurt (2014)
39. Schreier, M.: *Qualitative content analysis in practice*. Sage, Los Angeles, London, New Delhi, Singapore, Washington DC (2012)
40. Yin, R.K.: *Case study research and applications. Design and methods*. Sage, Los Angeles, London, New Dehli, Singapore, Washington DC, Melbourne (2018)



41. Xu, P., Barbosa, D.: Matching Résumés to Job Descriptions with Stacked Models. In: *Advances in Artificial Intelligence*, 10832, pp. 304–309. Springer International Publishing, Cham (2018)
42. Burton-Jones, A., Stein, M., Mishra, A.: IS Use. *MIS Quarterly Research Curations* (2017)
43. Marler, J.H., Parry, E.: Human resource management, strategic involvement and e-HRM technology. *The International Journal of Human Resource Management* 27, 2233–2253 (2016)
44. Tursunbayeva, A., Di Lauro, S., Pagliari, C.: People analytics—A scoping review of conceptual boundaries and value propositions. *International Journal of Information Management* 43, 224–247 (2018)