AI & SOCIETY
https://doi.org/10.1007/s00146-023-01699-w

OPEN FORUM

Application of artificial intelligence: risk perception and trust in the work context with different impact levels and task types

Uwe Klein[1] · **Jana Depping**[1] · **Laura Wohlfahrt**[1] · **Pantaleon Fassbender**[2]

Received: 24 November 2022 / Accepted: 22 May 2023

Abstract
Following the studies of Araujo et al. (AI Soc 35:611–623, 2020) and Lee (Big Data Soc 5:1–16, 2018), this empirical study uses two scenario-based online experiments. The sample consists of 221 subjects from Germany, differing in both age and gender. The original studies are not replicated one-to-one. New scenarios are constructed as realistically as possible and focused on everyday work situations. They are based on the AI acceptance model of Scheuer (Grundlagen intelligenter KI-Assistenten und deren vertrauensvolle Nutzung. Springer, Wiesbaden, 2020) and are extended by individual descriptive elements of AI systems in comparison to the original studies. The first online experiment examines decisions made by artificial intelligence with varying degrees of impact. In the high-impact scenario, applicants are automatically selected for a job and immediately received an employment contract. In the low-impact scenario, three applicants are automatically invited for another interview. In addition, the relationship between age and risk perception is investigated. The second online experiment tests subjects' perceived trust in decisions made by artificial intelligence, either semi-automatically through the assistance of human experts or fully automatically in comparison. Two task types are distinguished. The task type that requires "human skills"—represented as a performance evaluation situation—and the task type that requires "mechanical skills"—represented as a work distribution situation. In addition, the extent of negative emotions in automated decisions is investigated. The results are related to the findings of Araujo et al. (AI Soc 35:611–623, 2020) and Lee (Big Data Soc 5:1–16, 2018). Implications for further research activities and practical relevance are discussed.

Keywords Artificial intelligence · Automated decision-making · Trust · Risk perception · Workplace

1 Introduction

Artificial intelligence is not a short-term trend that will flatten out in the next few years. It is the future. Companies have realized that they can operate more cost-effectively, efficiently, and productively using artificial intelligence (Kirkpatrick 2019). In this context, decision-making processes are also increasingly being automated by artificial intelligence systems. Automated decision processes are being used in a variety of application contexts (e.g., Bickmore et al. 2016; Carey and Smith 2016; Dressel and Farid 2018, Graefe et al. 2018; Hodson 2014). Therefore, it is interesting to investigate how people perceive the use of AI in the enterprise, as AI will massively change working life. Araujo et al. (2020) explored this question in their study "In AI we trust? Perceptions about automated decision-making by artificial intelligence". Among other things, they investigated people's risk perceptions about automated decision-making by artificial intelligence. The authors distinguished between different application areas: Media, Health, and Medicine. Two scenarios were developed for each of the three application areas. One for a high-impact level—high-impact scenario—and one for a low-impact level—low-impact scenario—of automated decision-making by artificial intelligence. The subjects rated the different scenarios in terms of their perception of associated risks. Within the high- and low-impact scenarios, the authors compared the risk perception results for the case where human experts made the decisions and for the case where the decisions were made by artificial intelligence. Based on a Dutch sample, the authors reached the following conclusions, among others: The analysis of the

✉ Uwe Klein
uwe.klein@hs-fresenius.de

[1] Hochschule Fresenius University of Applied Science Wiesbaden, Moritzstraße 17 A, 65185 Wiesbaden, Germany

[2] Twisters Management Consulting LLC, 16751 NE 5th Street, Williston, FL 32696, USA

Published online: 02 June 2023

combined scenarios revealed no significant differences in risk perception for automated decisions made by artificial intelligence compared to decisions made by human experts in all three application domains. In all scenarios with a high-impact level, automated decisions by artificial intelligence were perceived as less risky. No differences were found for low-impact scenarios. Furthermore, a positive correlation was found between the age of the subjects and the degree of perceived risk. The present study investigated people's risk perception of automated decisions by artificial intelligence in a high-impact scenario compared to a low-impact scenario. The distinctive feature of this study was the presentation of scenarios in the form of two concrete case descriptions from the work context, through which the two impact level characteristics were operationalized. It was investigated whether the results described by Araujo et al. (2020) also apply to newly constructed scenarios.

In studies of artificial intelligence designed to support decision-making processes, contradictory results have been obtained regarding human acceptance. Some studies have shown that people trust decisions made by artificial intelligence more than decisions made by humans (Madhavan and Weigmann 2007). It has been shown that people trust their own opinions more than algorithmic decisions when they know that the algorithm has already made mistakes. The assumption here is that people discount the possibility that the algorithm is trainable in the first place and deny the algorithm the ability to learn (Dietvorst et al. 2015). According to Lee (2018), the reason for these different study results lies in the different contexts or tasks that require different capabilities. Basically, a distinction is made between "human" and "mechanical" skills.

The theoretical background of Lee's (2018) experiment is based on the literature about human choices in the algorithm development process (Barocas and Selbst 2016; Sweeney 2013), potential disparities between mathematical, computational definitions of fairness and social definitions of fairness (Lee and Baykal 2017), and the mental models people use to comprehend algorithms in social media (Rader and Gray 2015). The credibility and quality of information can be influenced by attitudes toward the information source (Sundar and Nass 2001). Lee also draws on Waytz and Norton's (2014) findings that suggest humans perceive computers and robots as having less emotional capability than humans. Based on these factors, Lee predicts that people will differentiate between tasks that require more "human" skills versus those that necessitate more "mechanical" skills. Lee (2018) found that there are differences in how artificial intelligence evaluates automated decision-making processes for management decisions that require "human" or "mechanical" skills. An online experiment examined perceptions about algorithmic decisions in a management context. Subjects were presented with descriptions of management decisions that were related to real examples from the work context. These decisions were made either by a person or by an algorithm and were based on human (e.g., a new hire) or mechanical skills (e.g., a planning task). Subjects' perceptions of fairness, trust, and emotional response were examined. A distinction was made between whether the decision was made by artificial intelligence, an algorithm, or a human (Lee 2018). Lee's (2018) research found that decisions made by artificial intelligence were perceived as more unfair and less trustworthy, as well as perceived more negatively when the task was "human" (e.g., a new hire). "Mechanical" tasks (e.g., a planning task), on the other hand, were perceived as equally fair or equally trustworthy, regardless of whether the decisions in this context were made by algorithms or humans. This was also found by examining the emotional response to "mechanical" tasks.

The present paper's contribution, in contrast to Lee's work, is the innovative development of scenarios utilizing the AI acceptance model proposed by Scheuer (2020). The model concentrates on AI-related enhancements that aim to foster trust in the system, such as the intelligence level of the system, the reliability of its results, and its perceived transparency.

The primary purpose of the study is to test new scenarios with a similar experimental design compared to Lee (2018) and Araujo et al. (2020). The scenarios were newly constructed and adapted to the context of "Artificial Intelligence in the Workplace". As we are primarily interested in the I/O-psychological impact of AI, adaptation and further development in one integrated step seem fitting due to the limited resources for this research project.

Hypotheses Following the studies of Araujo et al. (2020) and Lee (2018), a methodologically focused replication study in combination with a change of domain was conducted. For this purpose, new scenarios were constructed to test whether individual results of the two original studies could be replicated in other contexts or for other scenario descriptions. The scenarios contained real situations from everyday work or the work context in companies.

Based on the study by Araujo et al. (2020), two hypotheses are tested in the specific context of an AI-based job application process.

H1: There is a difference in subjects' risk perception of automated AI decisions that have a high-impact level compared to automated AI decisions that have a low-impact level.

H2: There is a positive relationship between subjects' age and risk perception in automated decisions by artificial intelligence.

Based on Lee's (2018) study, three hypotheses are tested in the specific context of an AI-based performance

evaluation of employees and an AI-based assignment of work tasks.

H3: The level of trust in decisions made in a partially automated way with the support of human experts is higher when dealing with tasks that require "human" skills than in fully automated decisions made by artificial intelligence alone.

H4: The level of negative emotions does not differ toward decisions that are partially automated by the assistance of human experts, compared to fully automated decisions made by artificial intelligence, in accomplishing tasks that require "mechanical" skills.

H5: The degree of trust in decisions that are partially automated with the support of human experts differs when accomplishing tasks that require mechanical skills compared to fully automated decisions made by artificial intelligence alone.

2 Methods

The study included two online experiments. Actual scenarios were used because studies show that recorded human behavior in experiments based on scenarios can be equated with the behavior of real people (Woods et al. 2006). The scenarios are constructed as realistically as possible and contain facts from everyday work. Scheuer's (2020) AI acceptance model was used to develop the scenarios. The model considers two possible selection options for the perception of artificial intelligence. Artificial intelligence can be perceived either as a person or from the perspective of emotional involvement. For the former, theories from psychology that include interpersonal acceptance (e.g., IPAT theory) are used. Since AI models are innovative technologies, it is believed that users will react emotionally. In this case, the IPAT theory would apply again. If the emotional involvement in the AI system is low, it is assumed that theories of technology acceptance models apply. However, according to Scheuer (2020), the focus is primarily on AI-specific extensions, as it is assumed that they control the acceptance of artificial intelligence regardless of the personality or technology perception of the AI system. The following influencing factors, which affect trust in AI systems, fall into this category: perceived intelligence level, result reliability, and transparency. The strongest impact can be measured in transparency. Result reliability also has a demonstrable influence on trust in AI systems. The perceived intelligence level ensures that basic trust in the functioning of AI systems is developed when a certain threshold is exceeded.

Based on the model, the scenarios used were expanded to include individual descriptive elements of AI systems compared to the original studies. These descriptive elements influence trust in artificial intelligence systems (Scheuer 2020) and include the described level of intelligence of the system, reliability of results, and perceived transparency. Subjects were told in the scenarios that the system was intelligent, reliable, as well as transparent. The scenarios were deliberately formulated in the third person to eliminate or reduce the effect of social desirability in order to capture honest assessments by the subjects (Nisbett et al. 1973).

The scenarios were based on the studies by Lee (2018) as well as Araujo (2020) and were set in concrete working contexts through detailed descriptions. In selecting the scenarios, care was taken to formulate examples that were as close to everyday life as possible so that subjects could empathize with them.

The first online experiment examines people's perception of risk in relation to automated decisions by artificial intelligence in a high-impact situation—high-impact scenario—(experimental group) compared to a low-impact situation—low-impact scenario (control group) (Tables 1, 2). In addition to Araujo et al. (2020), the two scenarios were not placed in a context with automated decisions on the application fields of media, health, and justice, but in relation to situations in the everyday work or work context of a company. In both scenarios, the identical application process of a fictitious company was described, which advertised jobs online in a recruiting portal due to restructuring. Subjects were instructed to think of themselves as applicants in this context. In both scenarios, automated decisions were made by artificial intelligence. In the high-impact scenario, the impact of the automated decision was to immediately send a job contract ready to be signed to suitable applicants. No human resources were used in this process. In the low-impact scenario, the impact of the automated decision was to immediately invite suitable applicants for an interview. Again, no human resources were used.

The second online experiment examined used two sub-experiments to investigate the perceived trust in decisions made by artificial intelligence systems in two

Table 1 Overview of experiment 1

AI-based application process	
Experimental Group—scenario high impact	Control Group—scenario low impact
The most suitable applicant will immediately receive a completed employment contract	The three most suitable applicants will be invited directly for an interview after receipt of the applications

Table 2 Experiment 1—scenarios presented (*variations in italics*)

Experimental Group scenario high-impact	Control Group scenario low-impact
Read the following scenario carefully. You will be asked questions about it below	Read the following scenario carefully. You will be asked questions about it below
Schneider and Co. KG has a number of positions to fill as part of a corporate restructuring. These have been advertised online on a recruiting portal. A number of applications have already been received. Now the company is checking which applicants are suitable for the respective positions. The selection is made with the help of new software that compares the job requirements with the applicants' profiles and evaluates them. *In order not to lose any time, an employment contract is automatically created for the suitable applicant and sent to him or her by e-mail*	Schneider and Co. KG has a number of positions to fill as part of a corporate restructuring. These have been advertised online on a recruiting portal. A number of applications have already been received. Now the company is checking which applicants are suitable for the respective positions. The selection is made with the help of new software that compares the job requirements with the applicants' profiles and evaluates them. *Based on this data, the three most suitable applicants are invited for personal interviews, which form the basis for the decision to hire them*
The system was tested before being deployed. It has been shown that the system works intelligently, provides reliable results and is transparent. The selected applicants are now informed online *that they have been hired.* They receive all important information as well as a detailed breakdown of why they were selected for the position. The relevant selection criteria can be viewed and understood by the applicants	The system was tested before being deployed. It has been shown that the system works intelligently, provides reliable results and is transparent. The selected applicants are now informed online *about the invitation to the interview*. They receive all important information as well as a detailed list of why they were invited to the interview. The relevant selection criteria can be viewed and understood by the applicants
Imagine you are an applicant at this company, and *you have been informed about your recruitment through the new system*	Imagine you are an applicant at this company, and *you have been invited to a personal interview* via *the new system*
On the following page, you will be asked questions about this situation. Please keep this scenario in mind	On the following page, you will be asked questions about this situation. Please keep this scenario in mind

sub-experiments, each depending on the nature of the task, which requires two different types of skill: "human skill" vs. "mechanical skill," and depending on the extent to which decisions are automated: partially automated with the assistance of human experts vs. fully automated. In the scenarios described, decisions have a direct impact on employees. Experiment 2 in the current study, which distinguishes between "human skills" and "mechanical skills," is consistent with Lee's (2018) operationalization. Lee employed a scheduling scenario in which an algorithm determined the work schedules of cafe baristas based on the projected number of customers ("mechanical skills"). In our study, we also used a scheduling scenario to assign employees to a new project, in which the algorithm determined which customer service employees had available capacity. To operationalize "human skills," Lee utilized a work evaluation scenario featuring an algorithm that assessed the performance of call center employees. Similarly, our study incorporated a work evaluation scenario in which the algorithm evaluated employee performance.

The first sub-experiment focused on the type of task that requires human skills. The "human skills" scenario presented describes a performance evaluation situation in a work context. In this scenario, the performance evaluation of the employees of the last quarter is to be performed using a new software introduced in the company. For this purpose, data on the amount of work performed, the pace of work, compliance with quality specifications, adherence to deadlines, etc. are used. In the experimental group, performance evaluation is partially automated by artificial intelligence with control or intervention by a human expert if corresponding wrong decisions are evident. In the control group, performance evaluation is fully automated without intervention or control by a human expert (Tables 3, 4).

The second sub-experiment focused on the type of tasks that require mechanical skills. The "mechanical skills" scenario presented describes a work assignment situation in a work context. In this scenario, the company wants to increase its sales and is planning a new sales campaign. Employees are still needed for this new project and are to be assigned internally for this purpose. A new system software

Table 3 Overview of experiment 2/sub-experiment 1

Performance Assessment Scenario—human skills	
Experimental Group—partially automatic	Control Group fully—automatic
Performance evaluation by artificial intelligence with intervention or control by a human expert (artificial intelligence + human expert)	Performance evaluation without intervention or control by a human expert (artificial intelligence only)

Table 4 Experiment 2/sub-experiment 1—scenarios presented (*variations in italics*)

Experimental Group—partially automatic	Control Group—fully automatic
Read the following scenario carefully. You will be asked questions about it below	Read the following scenario carefully. You will be asked questions about it below
In the company "Business Styles AG", a new system software was introduced 6 months ago that uses artificial intelligence to run work steps *semi-automatically*. Employees' performance is to be evaluated for the last quarter using the new system, which makes appropriate decisions semi-automatically. These assessments *will be cross-checked again by an experienced manager and corrected or adjusted if necessary*. The new system uses data on the amount of work done, the pace of work, compliance with quality specifications, adherence to deadlines, etc., for employee performance appraisals. The system was tested before it was implemented. It has been shown to work intelligently, provide reliable results, and be transparent. The relevant assessments can be viewed and understood by employees. The evaluation criteria used are known to the employees	In the company "Business Styles AG", a new system software was introduced 6 months ago that uses artificial intelligence to run work steps *fully automated*. Employees' performance is to be evaluated for the last quarter using the new system, which makes appropriate decisions semi-automatically. These *assessments will be adopted 1:1 and not cross-checked by a manager or corrected or adjusted if necessary.* The new system uses data on the amount of work done, the pace of work, compliance with quality specifications, adherence to deadlines, etc., for employee performance appraisals. The system was tested before it was implemented. It has been shown to work intelligently, provide reliable results, and be transparent. The relevant assessments can be viewed and understood by employees. The evaluation criteria used are known to the employees
On the following page, you will be asked questions about this situation. Please keep this scenario in mind	On the following page, you will be asked questions about this situation. Please keep this scenario in mind

is to check which employees of the customer service department have free capacities and can take over this task. In the experimental group, work assignment was fully automated by artificial intelligence, without any human expert intervention or control. In the control group, work allocation was partially automated with the control or intervention of a human expert (Tables 5, 6).

2.1 Sample

After cleaning the data, a total of 221 subjects from Germany participated in the study. 140 subjects are female, 80 subjects are male, and one subject reports his gender as diverse. The average age is 32 years, with the youngest subject being 18 years old and the oldest subject being 77 years old. When we categorize the age of the subjects, 40.3% of our sample is in the 18–25 age group. In the age group of 26–35 years, the percentage is 29.4%, in the age group of 36–45 years, 10.4%, and between 46 and 55 years, 13.6%. The last category includes all other age groups 56 years and older, representing 6.3% of our sample. People with different highest educational qualifications took part in the survey: "Hauptschulabschluss" 9%, "Realschulabschluss/Mittlere Reife" 8.1%, "Abitur/Fachabitur" 30.8%, "abgeschlossene Berufsausbildung" 20.4%, "Bachelor" 24.9%, "Master" 7.7% and "Promotion" 0.9%.

2.2 Procedure

In both online experiments, subjects were first presented with a definition of artificial intelligence (AI) and a definition of algorithm-based decision-making (ADM) before each scenario was presented to create a common understanding of this concept. The following definitions were used:

Definition of Artificial Intelligence (AI): Artificial intelligence is the generic term for applications in which machines perform human-like intelligence tasks such as learning, judgment, and problem-solving Machine learning technology (ML)—a subset of artificial intelligence—teaches computers to learn from data and experience and to perform tasks ever more effectively. Sophisticated algorithms can recognize patterns in unstructured data sets such as images, texts, or spoken language and make decisions independently based on these (source: news.sap.com/germany/2018/03/was-ist-kuenstliche-intelligenz).

Definition of algorithm-based decision processes (ADM): Algorithm-based decision processes (ADM) are self-learning algorithms that control processes and make increasingly automated decisions (source: civey.com/pro/unsere-arbeit/

Table 5 Overview experiment 2/sub-experiment 2

Work Assignment Scenario—mechanical skills	
Experimental Group—fully automatic	Control Group—partially automatic
Work assignment (planning of sales actions) without intervention or control of a human expert (artificial intelligence only)	Work assignment (planning of sales actions) by artificial intelligence with intervention or control by a human expert (artificial intelligence + human expert)

Table 6 Experiment 2/sub-experiment 2—scenarios presented (*variations in italics*)

Experimental Group—fully automatic	Control group—partially automatic
Read the following scenario carefully. You will be asked questions about it below	Read the following scenario carefully. You will be asked questions about it below
The company "Domicile Enterprise" wants to increase its sales. The goal is to mobilize regular customers and attract new ones. A sales campaign is to be planned and implemented. For about half a year, the company has been using new system software, which, among other things, can *fully automatically* assign tasks to available and competent employees. The software checks which employees in the customer service department currently have free capacity to take on these tasks. *These decisions are adopted 1:1 and are not cross-checked or manually approved by a company expert.* This ensures reliable results. The system was tested before it was deployed. It has been shown that the system works intelligently, delivers reliable results and is transparent. The selected employees are informed about the project online. They receive all the important information as well as a detailed breakdown of why they were selected for this task	The company "Domicile Enterprise" wants to increase its sales. The goal is to mobilize regular customers and attract new ones. A sales campaign is to be planned and implemented. For about half a year, the company has been using new system software, which, among other things, can semi-automatically assign tasks to available and competent employees. The software checks which employees in the customer service department currently have free capacity to take on these tasks. *These decisions are cross-checked by a company expert and manually approved*. This ensures reliable results. The system was tested before it was deployed. It has been shown that the system works intelligently, delivers reliable results and is transparent. The selected employees are informed about the project online. They receive all the important information as well as a detailed breakdown of why they were selected for this task
Imagine you are an employee of this company and you have been selected by the system for a task without prior personal consultation	Imagine you are an employee of this company and you have been selected by the system for a task without prior personal consultation
However, you are still busy in your current project with some activities that need to be completed in time	However, you are still busy in your current project with some activities that need to be completed in time
On the following page, you will be asked questions about this situation. Please keep this scenario in mind	On the following page, you will be asked questions about this situation. Please keep this scenario in mind

case-study/konsumgueter/verbraucherstudie-zu-automatisierten-entscheidungsprozessen).

Subsequently, subjects were randomly assigned to the scenarios described above using the SoSci Survey program. For the two sub-experiments of the second online experiment, after the initial assignment to experimental and control groups in the "mechanical skills" scenario, subjects were again randomly assigned to the experimental or control group in the "human skills" scenario. The different scenarios were considered independently.

After the presentation of each scenario in both experiments, the respective independent variables were recorded. Finally, all participants were asked about their age, gender, and highest educational attainment.

While we opted to balance the need for a well-structured research approach with the constraints of staying close enough to existing empirical data, in hindsight, we might have been better off by controlling some of the variables from the very beginning instead of the post-hoc-approach that is present in our analyses.

2.3 Measures

2.3.1 Independent variables

In the first experiment, the degree of impact of the automated decision by artificial intelligence formed the independent variable in the form of the two scenarios "high impact" vs. "low impact."

In our study, the direct hiring of an applicant—creation and sending of an employment contract—by the algorithm was operationalized as high impact. The automated invitation to interview 3 people, as a suggestion for possible new employees by the algorithm, was operationalized as low impact. This approach is consistent with Araujo et al. (2020), as the authors used far-reaching direct decisions by the algorithm as high impact and recommendations by the algorithm as low-impact scenarios.

For the second hypotheses, the age of the subjects was used as an additional independent variable.

In the second experiment, the level of automation was the independent variable in both sub-experiments in the two levels "partially automated with the support of human experts" vs. "fully automated".

2.3.2 Dependent variables

The dependent variable of the first experiment is the subjects' perceived risk, which was assessed based on the scenario presented in each case. The Simplified Conjoint Expected Risk Model (SCER model) of Holtgrave and Weber (1993) was used to capture the perceived risk. The SCER model is based on the original CER model of Luce and Weber (1986), which uses objective information to evaluate financial gambles. The SCER model can also be used for other areas outside of financial risk perception assessment. Unlike the original CER model, in the simplified CER model individuals define the harms and benefits

of the activity and provide subjective assessments of the probabilities and expected magnitude of those harms and benefits. Survey participants were then presented with five items to rate the impact of the scenarios in terms of benefits, harms, neutrality, percentage rating (0–100), and magnitude of expected benefits and harms (scale 0 neutral to 100 absolute positive/ negative). The magnitude of perceived risk-R (X)-was measured according to the SCER model as an additive linear combination of these five items (Carlstrom et al. 2000).

The dependent variables in the second experiment represent the degree of trust as well as the degree of negative emotions in the decisions made.

The construct trust was measured using the "Questionnaire for Measuring Trust in Dealing with Automated Systems" by Poehler et al. (2016). This instrument was developed for German-speaking countries and can be used independently of the domain. The construct "trust" is measured with the help of six items. The scale name was adapted with regard to the research question and the scenario descriptions. The logic remained unchanged, only the context of the scenario was considered. To illustrate this by way of example, the first item on trust ("The system offers security") was replaced by the name "In my opinion, the new system software for creating performance appraisals offers employees security." Poehler et al. (2016) statistically demonstrated that this question instrument subjectively measures the construct of "trust" according to the criteria of objectivity, reliability as well as validity. Internal consistency of $\alpha = 0.86$ was demonstrated. Poehler et al. (2016) found predominantly a medium item difficulty in their instrument: the minimum is thus Pi = 0.48, and the maximum Pi = 0.83. For the trust scale, the discriminatory power values ranged from $r = 0.14$ (item: "I am confident in using the system") to $r = 0.85$ (item: "The system is trustworthy").

The extent of negative emotions was measured using the German version of the Modified Differential Emotions Scale (mDES) by Brandenburg and Backhaus (2015). With their instrument, Brandenburg et al. (2015) have made a successful contribution to the assessment of emotions in the relationship between humans and machines. The scale for capturing negative emotions has an internal consistency of $\alpha = 0.79$ in the first measurement period, and a value of $\alpha = 0.89$ was found in the second measurement. The cognitive component of emotions was measured by having subjects rate a situation (scenario) within the experiment. People react emotionally differently depending on whether an event was evaluated positively or negatively (Brosch et al. 2010). The questionnaire used consists of a total of 10 items to assess negative emotions. Participants are instructed to think about the past two hours and rate on a five-point scale how they felt during that time. The emotions are listed below. The scale ranges from "0 = not at all" to "4 = very much." This five-point response scale was adopted for the present study. Brandenburg and Backhaus (2015) combined several emotions into one scale. For example, "annoyed," "irritated," and "irritated" form one item. In the present study, the subjects were confronted with one emotion per item in order to exclude a possible misinterpretation on the part of the participants. The instructions to the subjects were also adapted to the scenarios. Subjects were asked how they would have felt as an employee of the company if they had been selected by the system for a task without prior personal consultation (experimental group) or with a prior personal consultation and review by an expert (control group), although they were still engaged in activities from a current project. Subjects were asked about the following negative emotions: "annoyed," "irritated," "suspicious," "depressed," "unhappy," "anxious," "intimidated," "stressed," and "overwhelmed."

3 Results

Regarding the first hypothesis, no difference in subjects' risk perception was found for automated decisions by artificial intelligence that have a high impact compared to automated decisions by artificial intelligence that have a low impact. The mean value of risk perception was $M = 48.57$ for the experimental group with $N = 110$ and $M = 70.21$ for the control group with $N = 111$. The Mann–Whitney U test showed no significant difference between the magnitude of risk perception between the experimental and control groups, $U = 5349.000$, $Z = -1.591$, $p = 0.112$. The distributions between the experimental and control groups are not different from each other (Kolmogorov–Smirnov $p > 0.05$) (Table 7).

With regard to the second hypothesis, no positive relationship was found between the age of the subjects and the perception of risk in relation to automated decision-making by artificial intelligence. The methodological approach of simple linear regression did not show linearity between the subjects' age and their risk perception. Because linearity

Table 7 Results Mann–Whitney U test risk perception

	Total risk perception (RQ1)
Test statistics[a]	
Mann–Whitney U test	5349.000
Wilcoxon-W	11,454.000
Z	−1.591
Asymp. Sig. (2-sided)	0.112

[a]Group variable: experimental or control group

Table 8 Results linear regression

ANOVA[a]						
Model		Sum of squares	df	Mean of squares	*F*	Sig
1	Regression	1.176	1	1.176	0.012	0.913[b]
	Unstandardized residuals	21,428.531	217	98.749		
	Total	21,429.707	218			

[a]Dependent variable: total risk

[b]Influencing variables: (constant), age

Table 9 Results Mann–Whitney U-Test trust

Test statistics[a]	
	Total trust (RQ3)
Mann–Whitney U test	5791.500
Wilcoxon-W	12,461.500
Z	− 0.640
Asymp. Sig. (2-sided)	0.522

[a]Group variable: experimental or control group

Table 10 Results Mann–Whitney U test negative emotions

Test statistics[a]	
	Total negative emotions (RQ4)
Mann–Whitney-*U* test	5590.000
Wilcoxon-W	11,806.000
Z	− 1.084
Asymp. Sig. (2-sided)	0.278
Exact Sig. (2-sided)	0.279

[a]Group variable: experimental or control group

could not be demonstrated, four different methods of non-linear regression were chosen: The use of linear estimated regression, a test for a quadratic effect, a combination of a linear term with a quadratic term, and a fourth test for a sinusoidal relationship. The linear estimated regression showed no significant effect (Table 8).

The other three analyses also showed no relationship between subject age and risk perception in relation to automated decision-making by artificial intelligence.

The analysis of the data collected with respect to the third hypothesis showed no difference in the accomplishment of tasks requiring "human" skills with respect to the level of trust in decisions made partially automated with the assistance of human experts (experimental group) compared to fully automated decisions made by artificial intelligence alone (control group).

The mean value of perceived trust was $M = 4.09$ for the experimental group with $N = 115$ and $M = 4.25$ for the control group with $N = 105$. The Mann–Whitney U test showed no significant difference between the perceived trust of the experimental and control groups, $U = 5791.500$, $Z = -0.640$, $p = 0.523$. The distributions between the experimental and control groups are not different from each other (Kolmogorov–Smirnov $p > 0.05$) (Table 9).

„The original CER model includes (individual difference) parameterization by which gains and losses are raised to some power before the expected values of benefits and losses are calculated. Power parameters estimated from empirical data are often close in value to unity" (Holtgrave & Weber 1993: 554). While the authors state that "the simpler SCER model assumes power parameters of unity which makes it linear" (ibidem 554), this assumption is not proven. Out of an abundance of caution, we applied non-parametric testing procedures. In contrast, Poehler et al. (2016) (155) report KMO-statistics (KMO = 0.78), indicating that the data comply with normal-distribution assumptions expected as a prerequisite to carry out Factor Analyses.

With respect to the fourth hypothesis, no difference was found in the level of negative emotions toward fully automated decisions made by artificial intelligence alone (experimental group) compared to semi-automated decisions made with the support of human experts (control group) when dealing with tasks requiring "mechanical" skills.

The mean value of perceived negative emotions was $M = 3.11$ for the experimental group with $N = 110$ and $M = 2.93$ for the control group with $N = 111$. The Mann–Whitney U test showed no significant difference between the magnitude of negative emotions between the experimental and control groups, $U = 5590.000$, $Z = -1.084$, $p = 0.278$. The distributions between the experimental and control groups are not different from each other (Kolmogorov–Smirnov $p > 0.05$) (Table 10).

The data analysis with respect to the fifth hypothesis shows no difference in the accomplishment of tasks requiring "mechanical" skills in terms of trust in decisions made fully automatically by artificial intelligence alone (experimental group) compared to decisions made semi-automatically with the assistance of human experts (control group). The mean value of perceived trust was $M = 3.89$ (SD = 1.09, $n = 110$) for the experimental group and $M = 4.10$ (SD = 1.09,

$n = 111$) for the control group. The t-test shows no significant difference between the two groups ($t\,(219) = 1.42$, $p = 0.158$) (Table 11).

4 Discussion

In the present study, we aim for a replication that closely mirrors the rationale and statistical approach by Lee (2018). After a critical assessment of his design, we opted for a new domain with a focus on I/O psychology. Finally, we introduced a methodologically more rigorous approach to scenario design and testing, as suggested by Araujo et al. (2020).

We are very well aware of the added complexity and the methodological criticism that might be directed at Lee (2018) for his selection of (non-)parametric testing procedures. Feeling stuck "between Scylla and Charybdis" of staying too close to his approach or possibly venturing too far out from Lee's procedures, we stayed with his testing which clearly opens up our study to design challenges that are up to be solved by future researchers.

In contrast to Araujo et al. (2020), no difference in subjects' risk perception was found for automated decisions by artificial intelligence that have a high impact compared to automated decisions by artificial intelligence that have a low impact. Also, in contrast to Araujo et al. (2020), no positive relationship was found between age and risk perception in automated decisions by artificial intelligence.

Compared to Lee's (2018) study, consistent results were obtained with respect to the task type that required mechanical skills. Subjects equally trusted artificial intelligence decisions that were semi-automatic (with human expert intervention) and fully automatic (without human expert intervention). Also analogous to Lee (2018), the same level of negative emotion perception was observed for these two decision types. In contrast to Lee's (2018) study, no difference was found in perceived trust in decisions made by artificial intelligence that were partially automated (with human expert intervention) compared to fully automated decisions (without human expert intervention) with respect to the task type requiring human skill.

Different approaches at different levels are conceivable for explaining the different results of the two original studies. Similarly, limiting factors can be formulated for the present study.

At the scenario construction level, the first two studies by Araujo et al. (2020) and Lee (2018) did not provide further information about the artificial intelligence system/algorithm that made decisions in different contexts. In the present study, subjects were provided with further information about the AI system based on Scheuer's (2020) AI acceptance model. Large to moderate influences on AI acceptance

Table 11 Results T-test trust

	Levene's test of equality of variance		t-test for equality of means						
	F	Sig	T	df	Significance		Mean difference	Difference for standard error	95% confidence interval of difference	
					One-sided p	Two-sided p			Lower value	Upper value
Total trust (RQ5)										
Variances equal to	0.012	0.913	1.416	219	0.079	0.158	0.20816	0.14696	−0.08148	0.49781
Variances not equal to			1.416	218.965	0.079	0.158	0.20816	0.14697	−0.08149	0.49781

or trust were demonstrated for the AI-specific extensions. AI was presented in the same comprehensive manner in all scenarios. Transparency, reliability of results, and intelligence level of the system were highlighted. According to Scheuer (2020), these aspects are essential elements related to the acceptance of artificial intelligence systems. The scenario descriptions in Araujo et al. (2020) present individual decisions singularly without providing further contextual information. The contextualization in our study could be an explanation for the different results in risk perception of the subjects. Another possible explanation would be the different scope of application. The study by Araujo et al. (2020) examined the media, health, and justice domains. In the present work, an AI-based application process was used as a scenario from the everyday work or work context of a company. The difference not found with Lee (2018) in terms of perceived trust in decisions made by artificial intelligence in relation to the task type that required human skills could also be due to the scenario construction of the present study. In Lee's (2018) study, the decisions are made either by a human or by artificial intelligence. In the present study, the decisions within the scenarios were either partially automated with intervention or control by a human expert or fully automated by artificial intelligence. For the subjects, even the inclusion of AI appears to have an impact on decisions that require human skill. This does not seem to be the case for task types requiring mechanical skills.

The different operationalization of the dependent variables represents another level.

Since our measures do not perfectly align with those of Araujo et al. (2020) and Lee (2018), and multiple constructs were measured using the same methods, there is a potential for method bias to influence our results.

To measure perceived risk, Araujo et al. (2020) used a total of five adapted items from Cox and Cox (2001). In the present study, the SCER model according to Holtgrave and Weber (1993) was used because it is particularly suitable for assessing technologies (Carlstrom et al. 2000). For further studies in the context of automated decision-making by artificial intelligence, it could be examined whether possibly also the psychometric model according to Slovic et al. (1986), or the hybrid model according to Holtgrave and Weber (1993) could be considered for this context.

Furthermore, the question remains whether the subjects perceived the trust in a generalized or situational way. Thus, no statement can be made about whether subjects included prior experience with AI technologies in the trust evaluation or whether they merely evaluated the scenarios at hand (performance evaluation or work order) based on the data at hand in the description (Neumaier 2010). Logg (2017) has shown experimentally that people are more likely to trust algorithms than humans in certain situations. Individuals with above-average mathematical skills preferred algorithmic advice when objective decisions had to be made. It was found that this preference can be mitigated by overestimating oneself, providing an expert, and one's own expertise.

The listed aspects lead to the conclusion that the research question about trust could not be conclusively clarified within the framework of this research design. It cannot be determined whether fully automated decisions (by artificial intelligence) or semi-automated decisions are trusted more. This is true regardless of the situation or context, i.e., for both the performance evaluation scenario ("human task") and the work evaluation scenario ("mechanical task").

The nature of the sample also represents a different level. The subjects of the study by Araujo et al. (2020) are Dutch. The sample of Lee (2018) consists of Americans. In the present study, individuals from Germany participated. Due to country-specific attitude differences, it cannot be excluded that different basic attitudes toward AI systems influence the results. Araujo et al. (2020) point out the different privacy policies of different countries and recommend comparative studies accordingly. Lee (2018) also points out the diversity of country-specific attitudes. The study by Neudert et al. (2020) provides a very good overview of country-specific differences in attitudes.

In both online experiments, subjects were first presented with a definition of artificial intelligence and a definition of algorithm-based decision processes prior to each scenario to establish a common understanding of these terms. However, no data were collected on the extent of individual knowledge and experience with artificial intelligence systems. There was no differentiation of subjects between laypersons and experts. Logg et al. (2019) have confirmed that laypeople are more likely to follow tips from algorithms than tips that come from a human person. For future studies, we recommend collecting data on subjects' knowledge and experience of artificial intelligence systems.

When conducting online experiments, a number of confounding factors that affect the results cannot be excluded. The study uses scenarios. In a single-subject experiment, bias (confounding factors) cannot be adequately tested because subjects' attitudes about a particular topic are collected at a particular time. It is not possible to measure which variables influenced each other and how (Eifler 2014). It would be interesting to conduct a test in the laboratory excluding possible confounding factors. Subjects could be directly confronted with the AI technology and thus include the design features that are important for the development of technology trust and, for example, the performance of the AI system (Hoff and Bashir 2015).

Finally, it should be mentioned that the representativeness of the results is severely limited due to the self-selection of respondents in online surveys. To improve this, it would be necessary to draw samples from a suitable data pool.

5 Conclusion

Following the work of Araujo et al. (2020) and Lee (2018), the aim of the study was to develop new real facts from everyday work or the work context in companies and to compare the results on risk perception, trust and negative emotion perception with the already published results.

A key aspect seems to be the information about the AI system that is presented in the newly developed scenarios. Explicitly presenting information about a successful test of the system, resulting in AI that works intelligently, produces reliable results, and is transparent, could influence risk perception. This possible insight could be used for the design of human–AI interfaces, the design of AI systems themselves, and for further interaction conditions in the collaboration with AI systems.

Another aspect concerns the aspect of human control in tasks that require human skills. In this task cluster, already the reference to the human control function seems to influence the trust in artificial intelligence.

Further research activities could include experiments with additional scenarios. Likewise, from our point of view, the development of a "scenario taxonomy" would be interesting. This suggestion also applies to the development of a "task taxonomy". There is also a need for further research on psychological constructs such as acceptance, fairness, etc. in connection with the use of artificial intelligence in everyday work or in the work context of companies.

Overall, the present study makes a further contribution with regard to the methodological research diversity as well as the everyday relevance of application scenarios in order to be able to describe, explain and shape the influence on people's behavior and experience when working with artificial intelligence.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by UK, JD, LW, and PF. The first draft of the manuscript was written by UK, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. The authors did not receive support from any organization for the submitted work.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Araujo T, Helberger N, Kruikemeier S, de Vreese CH (2020) In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI & Soc 35(3):611–623. https://doi.org/10.1007/s00146-019-00931-w

Barocas S, Selbst AD (2016) Big data's disparate impact. California Law Rev 104(3):671–732. https://doi.org/10.15779/Z38BG31

Bickmore T, Utami D, Matsuyama R, Paasche-Orlow MK (2016) Improving access to online health information with conversational agents: a randomized controlled experiment. J Med Internet Res. https://doi.org/10.2196/jmir.5239

Brandenburg S, Backhaus N (2015) Zur Entwicklung einer deutschen Version der modified Differential Emotions Scale (mDES). In: Wienrich C, Zander T, Gramann K (eds) 11 Berliner Werkstatt Mensch-Maschine-systeme: tagungsband. Universitätsverlag der TU Berlin, Berlin, pp 63–67

Brosch T, Pourtois G, Sander D (2010) The perception and categorization of emotional stimuli: a review. Psychology Press, pp 76–108. https://doi.org/10.1080/02699930902975754

Carey D and Smith M (2016) How companies are using simulations, competitions, and analytics to hire. Harvard business review. https://hbr.org/2016/04/how-companies-are-using-simulations-competitions-and-analytics-to-hire

Carlstrom LK, Woodward JA, Palmer CGS (2000) Evaluating the simplified conjoint expected risk model. Comparing the use of objective and subjective information. Risk Anal 20(3):385–392. https://doi.org/10.1111/0272-4332.203037

Cox D, Cox AD (2001) Communicating the consequences of early detection: the role of evidence and framing. J Market 65(3):91–103. https://doi.org/10.1509/jmkg.65.3.91.18336

Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: People erroneously avoid algorithms after seeing them err. J Exp Psychol 144(1):114–126. https://doi.org/10.1037/xge0000033

Dressel J, Farid H (2018) The accuracy, fairness, and limits of predicting recidivism. Sci Adv. https://doi.org/10.1126/sciadv.aao5580

Eifler S (2014) Experiment. In: Baur N, Blasius J (eds) Handbuch Methoden der empirischen Sozialforschung. Springer, Wiesbaden, pp 195–209. https://doi.org/10.1007/978-3-658-37985-8_13

Eigenstetter M (2018) Digitalisierung und CSR II: Akzeptanz und Vertrauen in neue Technologien sicherstellen. csr.impulse.papier no. 5.2. Mönchengladbach: Hochschule Niederrhein

Graefe A, Haim M, Haarmann B, Brosius H-B (2018) Readers' perception of computer-generated news: credibility, expertise, and readability. Journalism 19:595–610. https://doi.org/10.1177/1464884916641269

Hodson H (2014) The AI Boss that deploys Hong Kong's subway engineers. New Scientist. https://www.newscientist.com/article/mg22329764-000-the-ai-boss-that-deploys-hong-kongs-subway-engineers/

Hoff KA, Bashir M (2015) Trust in automation: integrating empirical evidence on factors that influence trust. Hum Factors 57(3):407–434. https://doi.org/10.1177/0018720814547570

Holtgrave RD, Weber UE (1993) Dimensions of risk perceptions for financial and health risks. Risk Anal 13(5):553–558. https://doi.org/10.1111/j.1539-6924.1993.tb00014.x

Kirkpatrick K (2019) Artificial intelligence for enterprise applications. https://omdia.tech.informa.com/OM011943/Artificial-Intelligence-for-Enterprise-Applications

Lee MK (2018) Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management. Big Data Soc 5:1–16. https://doi.org/10.1177/2053951718775668

Lee MK and Baykal S (2017) Algorithmic mediation in group decisions: fairness perceptions of algorithmically mediated vs. discussion-based social division. In: Proceedings of the ACM conference on computer supported cooperative work and social computing, Portland, USA, 25 February–1 March 2017, pp 1035–1048. https://doi.org/10.1145/2998181.2998230

Logg JM, Minson JA, Moore DA (2019) Algorithm appreciation: people prefer algorithmic to human judgment. Organ Behav Hum Decis Process 151:90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Logg JM (2017) Theory of machine: when do people rely on algorithms? Working Paper, 17–086, Cambrige Massachusetts: Harvard University. http://nrs.harvard.edu/urn-3:HUL.InstRepos:31677474

Luce RD, Weber EU (1986) An axiomatic theory of conjoint, expected risk. J Math Psychol 30(2):188–205. https://doi.org/10.1016/0022-2496(86)90013-1

Madhavan P, Wiegmann DA (2007) Effects of information source, pedigree, and reliability on operator interaction with decision support systems. Hum Factors 49(5):773–785. https://doi.org/10.1518/001872007X230154

Neudert LM, Knuutila A & Howard PN (2020) Global attitudes towards AI, machine learning & automated decision making. Working paper 2020.10, Oxford Commission on AI & Good Governance. https://oxcaigg.oii.ox.ac.uk/wp-content/uploads/sites/124/2020/10/GlobalAttitudesTowardsAIMachineLearning2020.pdf

Neumaier M (2010) Vertrauen im Entscheidungsprozess. Der Einfluss unbewusster Prozesse im Konsumentenverhalten. Springer/gabler, Wiesbaden. https://doi.org/10.1007/978-3-8349-8863-8

Nisbett RE, Caputo C, Legant P et al (1973) Behavior as seen by the actor and as seen by the observer. J Pers Soc Psychol 27(2):154–164. https://doi.org/10.1037/h0034779

Pöhler G, Heine T, Deml B (2016) Itemanalyse und Faktorstruktur eines Fragebogens zur Messung von Vertrauen im Umgang mit automatischen Systemen. Zeitschrift Für Arbeitswissenschaft 70:151–160. https://doi.org/10.1007/s41449-016-0024-9

Rader E & Gray R (2015) Understanding user beliefs about algorithmic curation in the Facebook news feed. In: Proceedings of the 33rd annual ACM conference on human factors in computing systems, Seoul, South Korea, 18–23 April 2015, pp 173–182. https://doi.org/10.1145/2702123.2702174

Scheuer D (2020) Akzeptanz von Künstlicher Intelligenz. Grundlagen intelligenter KI-Assistenten und deren vertrauensvolle Nutzung. Springer, Wiesbaden. https://doi.org/10.1007/978-3-658-29526-4

Shikano S (2010) Introduction to inference by the nonparametric bootstrap. In: Wolf C, Best H (eds) Handbook of social science data analysis. VS Verlag für Sozialwissenschaften, Wiesbaden, pp 191–204. https://doi.org/10.1007/978-3-531-92038-2_9

Slovic P, Fischhoff B and Lichtenstein S (1986) The psychometric study of risk perception. In: Risk Evaluation and Management, Springer US, 3–24. https://doi.org/10.1007/978-1-4613-2103-3_1

Sundar S, Nass C (2001) Conceptualizing sources in online news. J Commun 51(1):52–72. https://doi.org/10.1111/j.1460-2466.2001.tb02872.x

Sweeney L (2013) Discrimination in online ad delivery. Queue 11(3):1–19. https://doi.org/10.2139/ssrn.2208240

Waytz A, Norton MI (2014) Botsourcing and outsourcing: Robot, British, Chinese, and German workers are for thinking—not feeling—jobs. Emotion 14(2):434–444. https://doi.org/10.1037/a0036054

Woods S, Walters M, Koay K & Dautenhahn K (2006) Comparing human robot interaction scenarios using live and video based methods: Towards a novel methodological approach. In: Proceedings of 9th IEEE international workshop on advanced motion control, pp 750–755. https://doi.org/10.1109/AMC.2006.1631754